THÈSE

présentée par

Peter Franz STURM

pour obtenir le grade de **DOCTEUR**
de l'**INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE**
(Arrêté ministériel du 30 Mars 1992)
Spécialité : **Informatique**

VISION 3D NON CALIBRÉE : CONTRIBUTIONS À LA RECONSTRUCTION PROJECTIVE ET ÉTUDE DES MOUVEMENTS CRITIQUES POUR L'AUTO-CALIBRAGE

Date de soutenance : 17 Décembre 1997

Composition du jury :

Président :	**Roger MOHR**
Rapporteurs :	**Olivier FAUGERAS**
	Andrew ZISSERMAN
Examinateurs :	**Radu HORAUD**
	Stephen MAYBANK
	Long QUAN

Thèse préparée au sein du laboratoire GRAVIR - IMAG - INRIA Rhône - Alpes
sous la direction de Long QUAN

Meinen Eltern.

Remerciements

En lisant les à peu près 200 pages qui suivent, on rencontre des résultats scientifiques qui ne reflètent probablement pas très bien les sentiments qui ont accompagné la préparation de cette thèse. Je ne vais pas m'étendre trop sur cela, mais je tiens à dire dire que j'ai vécu des moments extraordinaires durant ces trois années, avec beaucoup de joie mais aussi du désespoir. Cette thèse a demandé beaucoup d'énergie et de passion. Mais j'ai aussi eu beaucoup de chance : de travailler dans un environnement agréable avec des gens sympas et avec l'esprit ouvert et de rencontrer des personnes qui m'ont aidé à trouver mon chemin.

Tout d'abord, je suis très obligé à Long Quan ; il a éveillé ma passion pour mon domaine de recherche et m'a guidé tout au long de cette thèse avec beaucoup d'enthousiasme et d'amitié. Je lui remercie pour sa disponibilité et les nombreuses discussions qu'on a eues, pour ses conseils pratiques et surtout pour sa chaleur humaine.

Je tiens à remercier particulièrement Roger Mohr, qui m'a accueilli dans son équipe, où il a su créer une atmosphère ouverte et sereine très profitable. Je pouvais toujours compter sur son soutien et c'est un honneur pour moi qu'il a présidé ma commission d'examen.

Je suis très reconnaissant à Olivier Faugeras et Andrew Zisserman d'avoir accepté de juger ce travail et d'en rédiger les rapports. Même à distance, leurs travaux ont eu un grand impact sur mes recherches. Je remercie également Radu Horaud et Steve Maybank pour l'intérêt qu'ils ont porté à ce travail en acceptant d'en être les examinateurs.

Richard Hartley a donné un impulse important à ma thèse, en m'écoutant, moi, petit étudiant en DEA à l'époque, et en m'offrant sa collaboration ! J'ai également beaucoup appris de Bill Triggs avec lequel j'ai partagé le bureau ; sa disponibilité et ses connaissances étaient inépuisables et c'était toujours un plaisir de discuter avec lui.

Je remercie tous les membres de l'équipe MOVI pour avoir passé une période agréable et riche parmi eux. Un grand merci à Cyrille Gauclin et Patrick Gros pour avoir accepté la lourde tâche de relire et de corriger mon manuscrit et à Jérôme Blanc pour son logiciel de texture mapping. Je suis également très obligé à Danièle Herzog pour sa disponibilité charmante à tout moment et au personnel du service administratif et technique de l'INRIA Rhône-Alpes pour leur support.

Je tiens à remercier Wu Chengke et tous mes amis en Chine où j'ai passé deux mois excitants juste avant de commencer la rédaction, ce qui m'a donné la distance nécessaire à mon travail avant de me plonger en plein dedans. « Mille grazie » à Paolo pour m'avoir hébergé durant le mois précédant la soutenance et pour sa cuisine exotique.

Si j'ai réussi à trouver la force pour ce travail pendant ces trois dernières années, ce n'est que grâce à Irina. Son amour et son humour étaient ma plus profonde source.

Table des matières

Table des figures

Liste des tableaux

1.

Introduction

L'œil humain nous permet de percevoir notre environnement et ainsi de pouvoir, par exemple, reconnaître des personnes ou des objets. Grâce à nos deux yeux, nous sommes en plus capables d'estimer des distances ou des vitesses. En vision par ordinateur, aussi appelée vision artificielle, le développement de méthodes et de nouvelles technologies permet d'améliorer ou de compléter la vision humaine dans de nombreuses applications voire même de la remplacer.

Parmi les nombreux exemples d'applications de la vision par ordinateur, on peut citer : la création de machines mobiles et autonomes, la surveillance, la création automatique ou assistée de modèles tridimensionnels d'objets existants, l'assistance pour des opérations chirurgicales ou bien la recherche d'informations dans des bases de données constituées d'images ou de séquences vidéo. Les raisons de l'utilisation de la vision par ordinateur sont liées aux atouts qu'offre l'informatique : rapidité du traitement de gros volumes d'informations, fiabilité et disponibilité. La recherche dans de grands ensembles d'images, pour l'identification de personnes ou pour la recherche d'images avec un contenu similaire à une image donnée, est un bon exemple où une grande quantité d'informations est présente et où le temps de réponse peut être un facteur critique.

Nous nous intéressons, dans cette thèse, surtout à l'utilisation de la vision par ordinateur pour la perception tridimensionnelle de l'environnement, aussi appelée vision 3D. Les données à traiter sont des images de l'environnement prises par des caméras. Un système de vision 3D a pour but d'extraire des informations de nature tridimensionnelle. Par exemple, un véhicule autonome conduisant sur un terrain inconnu doit être capable de détecter les obstacles et d'estimer leur distance afin de pouvoir les éviter. Le contournement d'obstacles et la navigation en général peuvent s'appuyer sur des informations provenant du système de vision, de même que la saisie d'objets par un bras articulé.

D'autres applications sont basées sur la création de modèles tridimensionnels d'objets ou d'environnements, éventuellement enrichis par une texture réaliste ou par des informations sur la réflectance de surfaces. De tels modèles rendent possible la navigation virtuelle dans ces environnements ; on peut par exemple penser à la visite virtuelle de musées ou de maisons à vendre ou, pourquoi pas, d'un trou de ver de terre qui a été modélisé à l'aide d'une caméra miniature. Ces applications se distinguent des techniques classiques de télévision 3D par la liberté du choix de point de vue : une expérience excitante serait par exemple de suivre un penalty de football de la position du gardien de but ou encore mieux, depuis le ballon ! Ce dernier exemple semble être futuriste, vu l'aspect dynamique de l'environnement à modéliser. Mais il existe déjà des logiciels qui permettent de créer des modèles 3D « photoréalistes » d'objets à partir de quelques images de ces objets. Si ceci demande encore un certain taux d'interaction de la part de l'utilisateur, la tâche de la vision par ordinateur dans ce domaine est d'augmenter à la fois le degré d'automatisme et de réalisme dans la création de modèles.

La détermination de mesures tridimensionnelles demande un certain calibrage du capteur, à savoir la caméra. Les caméras effectuent des projections de l'espace tridimensionnel dans un espace bidimensionnel, une surface photosensible. Cette projection peut être modélisée de plusieurs façons et le calibrage d'une caméra consiste en la détermination des valeurs des paramètres d'un modèle choisi. Le procédé classique pour des applications de vision 3D consiste en un calibrage hors ligne, suivi de l'application proprement dite. Le calibrage hors ligne est typiquement effectué avec un équipement spécial. Or, parmi les applications décrites ci-dessus, il y en a où cet équipement n'est pas disponible. C'est par exemple le cas pour des applications grand public pour lesquelles l'utilisation des caméras doit être la plus confortable possible. Beaucoup d'applications nécessitent une mise à jour fréquente du calibrage, due à des changements délibérés (zoom, mise au point) ou accidentelle (influences thermiques, vibrations) de paramètres de la caméra. La mise à jour par un nouveau calibrage hors ligne est pour le moins fastidieuse, et est souvent complètement intolérable si l'application en cours ne doit pas être interrompue.

À ce stade, on se pose naturellement quelques questions.

- ▷ D'une manière générale, est-ce qu'on peut s'affranchir du besoin de calibrer les caméras pour effectuer des mesures tridimensionnelles ?
- ▷ Si oui, quels types de mesures sont possibles avec des caméras non calibrées ?

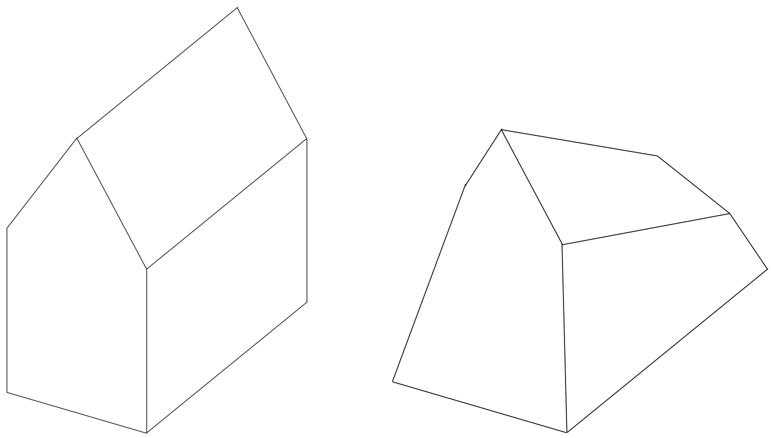
Avant de donner des réponses, nous considérons un autre type de capteur où la situation est similaire, mais plus simple. Que peut-on mesurer avec un thermomètre dont la graduation a été enlevée ? On ne peut plus qu'*ordonner* des températures. L'ajout d'une marque sur l'échelle permet de mesurer des températures relatives à cette marque et deux marques permettent finalement de mesurer des températures absolues.

Revenons aux caméras. Habituellement, on veut mesurer des informations tridimensionnelles qui sont conformes à notre perception du monde : on s'attend par exemple à ce que les dimensions des objets soient mesurées en unités métriques ou que les angles formés par les murs d'une maison soient reconnus comme étant des angles droits. Avec des caméras non calibrées, de telles informations métriques ne peuvent pas être déterminées. Tout de même, il est possible d'extraire des informations tridimensionnelles, mais qui sont moins riches. Le type de modèle 3D que l'on peut obtenir à partir d'images non calibrées d'un objet est présenté dans la figure 1.1 (b). Un tel modèle n'est certainement pas satisfaisant pour le regard, mais il permet tout de même de reconnaître l'objet comme étant une maison.

Nous pouvons alors donner une réponse positive à la première question ci-dessus. Nous allons répondre à la deuxième question dans la suite de la thèse, en termes mathématiques ; ici, nous nous contentons de la réponse qualitative donnée par la figure 1.1 (b). Deux nouvelles questions émergent directement.

- ▷ À quoi servent des informations tridimensionnelles non métriques ?
- ▷ Quelles sont les informations faciles à intégrer qui permettent d'obtenir des informations métriques ?

La vision 3D non calibrée est la partie de la vision par ordinateur qui donne des réponses à ces questions. Des modèles non métriques trouvent des applications dans les domaines de la synthèse de vues, de la reconnaissance d'objets et du contrôle de mouvements de robots manipulateurs. Ces modèles peuvent aussi former une première étape pour la construction de modèles métriques : l'information non métrique est souvent facile à obtenir et ceci de manière stable. Ensuite, l'introduction



(a) Un modèle métrique.

(b) Un modèle dit « projectif ». Les angles et les distances ne sont pas préservés, mais la coplanarité et la « topologie » de l'objet le sont.

FIG. 1.1: *Deux modèles tridimensionnels d'une maison.*

de quelques informations supplémentaires permet de trouver un modèle métrique. Ces informations supplémentaires peuvent être de plusieurs types. Nous considérons surtout le cas où il est connu que toutes les images d'une séquence à interpréter proviennent de la même caméra. D'autres possibilités intéressantes sont des mouvements appropriés de la caméra.

Organisation de la thèse. Dans cette introduction, nous avons donné des idées générales sur le sujet étudié dans cette thèse.

Le chapitre 3 présente une introduction étendue et plus spécifique. Les concepts importants en vision non calibrée y sont expliqués, tels que la reconstruction projective, l'auto-calibrage et la stratification de la reconstruction. Ce chapitre définit un cadre général dans lequel nous plaçons nos contributions, qui sont réparties sur les chapitres 4 à 7.

Dans le chapitre 2, nous décrivons des éléments de base qui seront utiles pour la lecture de la thèse et nous fixons les notations que nous allons utiliser.

Dans le chapitre 4, nous considérons le problème de la reconstruction tridimensionnelle à partir de deux images. Nous y expliquons pourquoi les méthodes classiques sont inadaptées au problème de la reconstruction projective. Une nouvelle méthode est proposée qui effectue une reconstruction projective optimale au sens d'un critère raisonnable.

Ensuite, dans le chapitre 5, nous proposons une méthode pour la reconstruction projective qui utilise plusieurs images simultanément, améliorant ainsi la stabilité et la précision de la reconstruction. La méthode est élégante et performante, mais nous discutons aussi ses inconvénients.

Le chapitre 6 est dédié à l'auto-calibrage. Ces dernières années, beaucoup de travaux sur le sujet ont émergé ; nous les résumons dans un état de l'art. Dans ce chapitre, nous proposons une méthode pour l'auto-calibrage d'une caméra linéaire, qui se présente relativement simplement par rapport au cas des caméras habituelles. Nous considérons aussi l'auto-calibrage de caméras avec un objectif à focale variable. Une méthode est proposée qui s'appuie sur un pré-calibrage d'une caméra.

Dans le chapitre 7, nous étudions les situations de dégénérescence de l'auto-calibrage, plus concrètement les mouvements de caméra qui causent des dégénérescences pathologiques. Ce chapitre contient une étude théorique du problème, qui mène à des conclusions importantes pour la pratique.

La thèse est close par le chapitre 8, où nous résumons nos conclusions et indiquons quelques perspectives pour la recherche sur les sujets traités.

Des détails et des preuves complétant le contenu des chapitres sont donnés dans les annexes A à D.

2.

Éléments de base

Dans ce chapitre, nous décrivons des éléments de base qui seront nécessaires pour la lecture de cette thèse. Tout d'abord, nous définissons les notations utilisées. Ensuite, les modèles de caméra employés dans cette thèse sont décrits, ainsi que la géométrie épipolaire, les quadriques et d'autres éléments géométriques.

Pour compléter le contenu du chapitre, nous renvoyons au livre de Faugeras [58] et aux livres de géométrie, par exemple [20, 171].

2.1 Conventions et notations

Tout au long de cette thèse, nous travaillons en *coordonnées homogènes* : à défaut de mention contraire, les points, les droites, les transformations etc. sont représentés par des vecteurs ou matrices, parfois des tenseurs, définis à un facteur multiplicatif près. L'égalité à un facteur multiplicatif près entre vecteurs ou matrices est désignée par le symbole $\sim$. Nous ne distinguons en général pas les entités géométriques et leurs représentations algébriques, par exemple nous ne faisons pas de différence entre un point et son vecteur de coordonnées ou entre une conique et sa représentation matricielle, donnée par rapport à un repère qui est déterminé par le contexte.

Les vecteurs sont représentés par des caractères gras, comme $\mathbf{q}$, et les matrices par des caractères sans-sérif, comme A. Les points 3D sont notés en majuscules, le plus souvent $\mathbf{Q}$ ou $\mathbf{Q}_p$. Un point image (2D) correspondant est désigné par le même caractère, en minuscule ; ici $\mathbf{q}$ ou $\mathbf{q}_p$. L'index p énumère les points et i les images ou vues. Ainsi, $\mathbf{q}_{ip}$ est la projection du p^{e} point dans la i^{e} vue. Les matrices de projection sont nommées P ou P_i. La droite reliant deux points $\mathbf{Q}_1$ et $\mathbf{Q}_2$ est désignée par $\langle \mathbf{Q}_1 \mathbf{Q}_2 \rangle$.

Les coniques, quadriques et plans sont désignés par des caractères grecs, minuscules pour les coniques, comme ϕ, et majuscules pour les autres, comme Λ pour un cône et Π pour un plan. Les coniques dans l'espace 3D (voir plus bas) sont notées en majuscule : Φ.

$\mathbf{q}$ représente un vecteur colonne. Un vecteur ligne est donné par $\mathbf{q}^{\mathsf{T}}$, où $^{\mathsf{T}}$ désigne la transposition, d'un vecteur ou d'une matrice. Par $\mathsf{A}^{-\mathsf{T}}$, nous abrégeons $(\mathsf{A}^{\mathsf{T}})^{-1}$ ou $(\mathsf{A}^{-1})^{\mathsf{T}}$. Une sous-matrice d'une matrice A est souvent notée $\bar{\mathsf{A}}$. Le signe $\wedge$ est l'opérateur du produit vectoriel et $[\mathbf{q}]_\wedge$ est la matrice anti-symétrique associée à un vecteur $\mathbf{q}$ de longueur 3, pour le produit vectoriel :

$$[\mathbf{q}]_\wedge = \begin{pmatrix} 0 & -q_3 & q_2 \\ q_3 & 0 & -q_1 \\ -q_2 & q_1 & 0 \end{pmatrix} .$$

Parfois, nous désignons la matrice adjointe à A par A^*, c'est-à-dire que :

$$\mathsf{A}^* = (\mathsf{A})^{-\mathsf{T}} .$$

Les vecteurs et matrices nuls sont représentés par $\mathbf{0}, \mathbf{0}_2, 0_3$ ou $0_{3\times 4}$, leurs dimensions étant connues du contexte ou indiquées par des indices. La matrice identité est désignée par $\mathtt{I}$, ou $\mathtt{I}_m$ pour indiquer ses dimensions.

Par abus de langage, nous appelons les similitudes souvent **transformations euclidiennes**.

Par $\mathcal{P}^n$ nous désignons l'espace projectif de dimension n.

Quand nous parlons de projections perspectives ou affines, sans préciser les dimensions des espaces concernés, nous sous-entendons des projection de $\mathcal{P}^3$ dans $\mathcal{P}^2$. Par des **bruits gaussiens** tout court, nous entendons des bruits de distribution normale, centrés, indépendants et de variance égale.

2.2 Des modèles de caméra

Dans cette section, nous décrivons les modèles de caméra utilisés dans cette thèse. Principalement, le modèle de **caméra perspective** est employé (voir 2.2.1 et 2.2.2). Dans 2.2.3, nous adoptons un modèle sténopé qui permet de traiter aisément son cas limite, la **caméra affine** et ses sous-classes, les projections orthographique, orthographique à l'échelle (perspective faible) et para-perspective. Les **caméras linéaires** sont brièvement discutées en 2.2.4.

2.2.1 La projection sténopé

Le modèle sténopé modélise une projection perspective avec un **centre de projection $\mathbf{C}$** et un **plan image** Π. Le centre de projection est souvent aussi appelé **centre optique** et le plan image, **rétine**. La projection d'un point $\mathbf{Q}$ de l'espace est définie comme étant l'intersection $\mathbf{q}$ du **rayon de projection** $\langle\mathbf{CQ}\rangle$ avec le plan image. Cette projection peut être représentée par une transformation projective P de $\mathcal{P}^3$ vers $\mathcal{P}^2$:

$$\mathbf{q}_3 \sim \mathsf{P}_{3\times 4}\mathbf{Q}_4 \quad ,$$

où les indices indiquent les dimensions des entités. La matrice P est appelée **matrice de projection**. Nous supposons dans la suite que la matrice de projection est toujours de plein rang, c'est-à-dire de rang 3. Le centre de projection est représenté (en coordonnées homogènes) par n'importe quel vecteur $\mathbf{C}$ du noyau de P. Ceci correspond bien au fait que la projection du centre de projection n'est pas définie : $\mathsf{P}\mathbf{C} = \mathbf{0}_3$ (nous rappelons que les vecteurs nuls ne sont pas admis comme vecteurs de coordonnées homogènes).

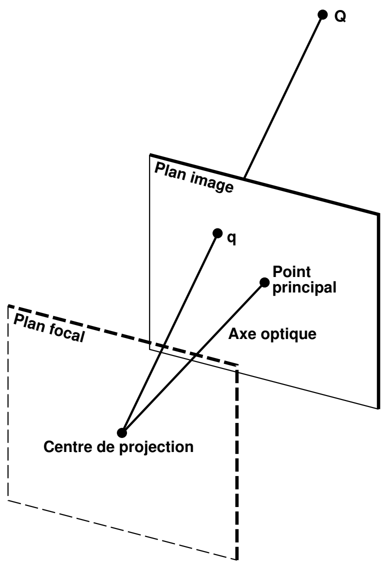


FIG. 2.1: *Le modèle sténopé.*

Nous définissons maintenant quelques autres notations associées au modèle sténopé : l'**axe optique** est la droite passant par le centre de projection et perpendiculaire au plan image. Son intersection avec le plan image est le **point principal**. Le **plan focal** est constitué des points de $\mathcal{P}^3$ qui se projettent sur des points à l'infini du plan image. Il s'agit donc du plan passant par le centre de projection qui est parallèle au plan image.

Nous distinguons maintenant les deux principaux modèles de caméra : une **caméra perspective** est une caméra effectuant une projection sténopé, dont le centre de projection est un point qui n'est pas à l'infini. Une telle projection est aussi appelée projection perspective parfaite. Une **caméra affine** [134] est une caméra dont la matrice de projection est une transformation affine, c'est-à-dire de la forme suivante :

$$\mathsf{P} \sim \begin{pmatrix} \bar{\mathsf{P}}_{2\times 3} & \bar{\mathbf{p}}_2 \\ \mathbf{0}_3^\mathsf{T} & 1 \end{pmatrix} \quad .$$

Une condition nécessaire et suffisante pour qu'une caméra soit affine est que son plan focal soit le plan à l'infini. Ceci implique que le centre de projection est à l'infini. *Toutes* les droites perpendiculaires au plan image passent par le centre de projection et donc les concepts d'axe optique et de point principal n'ont plus de sens pour la caméra affine.

Dans le prochain paragraphe, nous décrirons les paramètres intrinsèques et extrinsèques de la caméra perspective de manière classique. En 2.2.3 nous présentons une modélisation légèrement différente qui permettra d'unifier la formulation algébrique des caméras perspective et affine.

2.2.2 Paramètres intrinsèques et extrinsèques de la caméra perspective

La matrice de projection P ne révèle pas directement les grandeurs physiques qui caractérisent le processus de prise d'images avec une caméra réelle. Nous allons donc, dans la suite, décrire une façon de décomposer P en des paramètres reflétant des grandeurs physiques de la caméra (et du système d'acquisition). Cette caractérisation de la caméra ne capte pas les effets non modélisables par le modèle sténopé idéal, comme par exemple les distorsions optiques (voir 2.2.5). La plupart des chercheurs en vision par ordinateur se contentent pourtant du degré d'approximation de la réalité fourni par ce modèle (ce qui est différent dans le domaine de la photogrammétrie où la précision maximale des mesures est l'un des buts essentiels).

D'abord, nous introduisons quelques repères de coordonnées par rapport auxquels nous allons définir les transformations composant P.

- ▷ Un repère global par rapport auquel sont donnés les points tridimensionnels. Nous l'appelons le *repère monde*.
- ▷ Un repère 3D local attaché à la caméra, le *repère caméra*.
- ▷ Un repère 2D du plan image, le *repère image*.
- ▷ Un deuxième repère 2D du plan image, le *repère pixels*, associé à la grille des pixels.

Le repère caméra est souvent choisi de la façon suivante [58, 98] : le centre de projection est pris comme origine, l'axe optique coïncide avec l'axe des Z et le plan image est le plan $Z = f$, où f est la distance focale de la caméra. Le repère image a son origine au point principal et ses axes sont parallèles aux axes des X et Y du repère caméra. Quant au repère pixels, nous choisissons de mettre son origine au coin supérieur gauche de la matrice des pixels ; un des deux axes du repère est parallèle à un axe du repère image. Le deuxième axe est généralement perpendiculaire au premier, mais l'omission de cette contrainte permet de modéliser des caméras avec des pixels non rectangulaires, en forme de parallélogramme. Ces repères sont présentés dans la figure 2.2.

La projection perspective P peut maintenant être décomposée en une suite de transformations entre les repères définis ci-dessus :

$$\mathsf{P} \sim \mathsf{A} \; \mathsf{P}^{P} \; \mathsf{T} \;,$$

où nous avons :

- ▷ T est la transformation rigide entre le repère monde et le repère caméra. Elle est de la forme :

$$\mathsf{T}_{4\times 4} = \begin{pmatrix} \mathsf{R}_{3\times 3} & -\mathsf{R}\mathbf{t}_3 \\ \mathbf{0}_3^{\mathsf{T}} & 1 \end{pmatrix} \;,$$

  où R est une matrice de rotation (une matrice orthogonale). R indique l'orientation de la caméra par rapport au monde et $\mathbf{t}$ sa position.

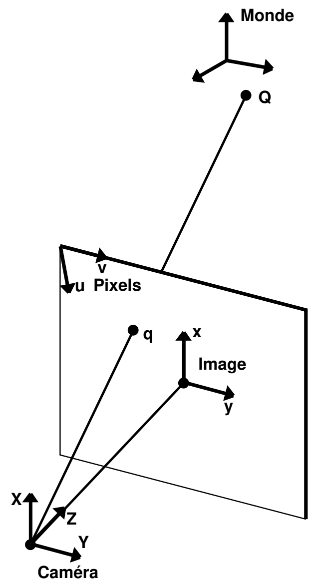


FIG. 2.2: *Les repères du modèle sténopé.*

▷ P^P est la projection perspective proprement dite. Un point $\mathbf{Q} \sim (X, Y, Z, 1)^\mathsf{T}$ dans le repère caméra se projette sur un point $\mathbf{q} \sim (x, y, 1)^\mathsf{T}$ avec :

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z} \quad .$$

En utilisant des entités homogènes, cette projection peut être représentée par la matrice :

$$\mathsf{P}^P_{3\times 4} \sim \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad . \tag{2.1}$$

▷ A est la transformation (affine) reliant les repères image et pixels. Sa forme générale est :

$$\mathsf{A} = \begin{pmatrix} k_u & -k_u \cot \Theta & u_0 \\ 0 & \frac{k_v}{\sin \Theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad .$$

Les paramètres utilisés sont les suivants : k_u et k_v sont les dimensions des pixels (la longueur inverse de leurs « côtés ») ; Θ, l'angle entre les axes (les côtés) des pixels ; $(u_0, v_0)^\mathsf{T}$, les coordonnées du point principal dans le repère pixels.

Puisque R et **t** définissent le positionnement de la caméra dans son environnement, ils sont souvent appelés **paramètres extrinsèques** de la caméra. Les autres grandeurs (k_u, k_v, f, Θ, u_0 et v_0) décrivent des propriétés physiques spécifiques de la caméra et de la carte d'acquisition et sont donc appelées **paramètres intrinsèques**.

Il est facile de voir que les paramètres k_u, k_v et f ne comptent que pour deux paramètres indépendants : le produit de A et P^P est :

$$\mathsf{A} \ \mathsf{P}^P = \begin{pmatrix} k_u f & -k_u f \cot \Theta & u_0 & 0 \\ 0 & \frac{k_v f}{\sin \Theta} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \mathsf{K}_{3\times 3} & \mathbf{0}_3 \end{pmatrix}$$

où k_u et k_v apparaissent toujours avec le facteur f. Nous restreignons donc l'ensemble des paramètres intrinsèques aux suivants : $\alpha_u = k_u f, \alpha_v = k_v f, \Theta, u_0, v_0$. Les paramètres α_u et α_v mesurent la distance focale, avec comme unité les dimensions des pixels.

Nous appelons K la **matrice des paramètres intrinsèques** :

$$\mathsf{K} = \begin{pmatrix} \alpha_u & -\alpha_u \cot\Theta & u_0 \\ 0 & \frac{\alpha_v}{\sin\Theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad .$$

Le rapport $\tau = \frac{\alpha_u \sin\Theta}{\alpha_v}$ est appelé **rapport d'échelle** de la caméra. Avec des pixels rectangulaires ($\Theta = 90°$), le rapport d'échelle équivaut le rapport des dimensions d'un pixel.

Étant donnée une matrice de projection P d'une caméra perspective, il est aisé d'en extraire les paramètres intrinsèques et extrinsèques. Soit $\mathsf{P} = \left(\; \overline{\mathsf{P}} \; | \overline{\mathbf{p}} \right)$. Si $\overline{\mathsf{P}}$ est normalisée de telle manière que sa troisième ligne soit de norme 1, la décomposition QL de $\overline{\mathsf{P}}^{\mathsf{T}}$ nous fournit les matrices des paramètres intrinsèques (triangulaire supérieure) K et d'orientation de la caméra (orthogonale) R : $\overline{\mathsf{P}} = \mathsf{KR}$. Le vecteur de translation est ensuite calculé par $\mathbf{t} = -\mathsf{R}^{\mathsf{T}}\mathsf{K}^{-1}\overline{\mathbf{p}}$.

Une décomposition courante et utile de la matrice de projection est la suivante :

$$\mathsf{P} = \mathsf{KR} \left(\; \mathtt{I}_3 \; \middle| -\mathbf{t} \; \right) \quad .$$

2.2.3 Une autre formulation – Unification des modèles perspectif et affine

Le choix du repère caméra fait dans la section précédente, et la transformation perspective P^P induite ne permettent pas d'en dériver la transformation de projection d'une caméra affine. La raison est que le centre de projection est fixé à l'origine et que la variation de la distance focale f fait « bouger » le *plan image*. Or, pour obtenir une caméra affine, il faut déplacer le *plan focal*, et avec lui le centre de projection, à l'infini. Nous adoptons donc un autre modèle, où l'origine du repère caméra est le point principal, au lieu du centre de projection. Ce dernier se trouve alors à la position $\mathbf{C} \sim (0, 0, -f, 1)^{\mathsf{T}}$ (cf. la figure 2.3). Ce modèle a été introduit par Szeliski et Kang [196, 197] (voir aussi une brève discussion dans [58]) et est utilisé pour la reconstruction multi-image par eux-mêmes et Azarbayejani et Pentland [9].

Avec cette modélisation, un point $\mathbf{Q} \sim (X, Y, Z, 1)^{\mathsf{T}}$ dans le repère caméra se projette sur un point $\mathbf{q} \sim (x, y, 1)^{\mathsf{T}}$ par :

$$x = \frac{X}{1 + Z\beta} \qquad y = \frac{Y}{1 + Z\beta} \quad ,$$

où $\beta = \frac{1}{f}$ est l'inverse de la distance focale. Cette projection peut être représentée par la matrice :

$$\mathsf{P}^P \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \beta & 1 \end{pmatrix} \sim \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & f \end{pmatrix}$$

qui est à comparer avec P^P pour l'autre choix de repères, dont la quatrième colonne est le vecteur nul (voir l'équation (2.1)). Changer β fait se déplacer le centre de projection, et notamment si l'on annule β, on obtient une projection affine : le plan focal est le plan à l'infini et le centre de projection est le point à l'infini dans la direction normale au plan image.

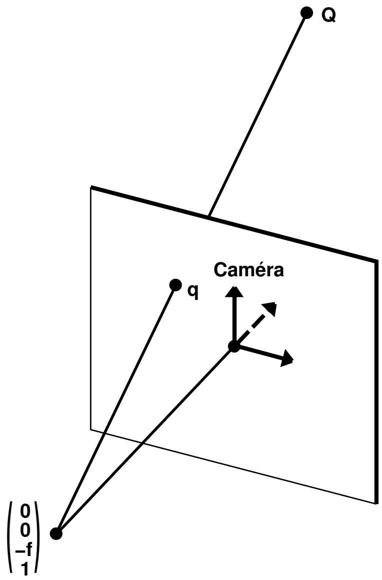


FIG. 2.3: *Le repère caméra attaché au plan image.*

2.2.3.1 Caméra affine

Nous considérons, dans la suite de ce paragraphe, uniquement le cas de la caméra affine, c'est-à-dire le cas où $\beta = 0$. Nous allons dériver les paramètres intrinsèques de la caméra affine directement des paramètres intrinsèques de la caméra perspective. Ceci est complémentaire à l'approche de Quan [164] qui les extrait de la matrice de projection affine réduite, par une décomposition QL.

Examinons la forme de la matrice de projection complète :

$$\begin{aligned}
\mathsf{P} &\sim \mathsf{A}\ \mathsf{P}^P\ \mathsf{T} \\
&= \begin{pmatrix} k_u & -k_u \cot\Theta & 0 & u_0 \\ 0 & \frac{k_v}{\sin\Theta} & 0 & v_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathsf{R} & -\mathsf{R}\mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \\
&= \begin{pmatrix} \overline{\mathsf{K}}_{2\times 2}\overline{\mathsf{R}}_{2\times 3} & -\overline{\mathsf{K}}\,\overline{\mathsf{R}}\mathbf{t} + \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} ,
\end{aligned}$$

où :

$$\overline{\mathsf{K}} = \begin{pmatrix} k_u & -k_u \cot\Theta \\ 0 & \frac{k_v}{\sin\Theta} \end{pmatrix}$$

et $\overline{\mathsf{R}}$ est une matrice de rotation « amputée », constituée des deux premières lignes de R. Elle représente pourtant entièrement la rotation R, puisque la troisième ligne peut en être reconstituée : elle est orthogonale aux deux autres et de norme 1. Il existe deux solutions pour la troisième ligne dont il faut choisir celle qui fait que le déterminant de R est égal à 1.

Le point principal n'est plus défini pour la caméra affine – tous les points du plan image peuvent être considérés comme la projection orthographique du centre de projection. On peut donc librement choisir un point de référence pour l'image. Par un choix adéquat, on peut annuler les deux premiers éléments de la dernière colonne de la matrice de projection, et ainsi absorber la translation de la caméra dans l'espace. Ceci correspond à choisir comme point de référence de l'image la projection du point de référence du repère monde. Avec cette matrice de projection simplifiée, l'équation de projection devient :

$$\mathbf{q} \sim \begin{pmatrix} \overline{\mathsf{K}}\,\overline{\mathsf{R}} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{Q} \ .$$

Si $\mathbf{q}$ et $\mathbf{Q}$ sont donnés en coordonnées non homogènes ($q_3 = 1$ et $Q_4 = 1$), cette équation est *exacte* ($=$ au lieu de $\sim$) et on peut alors utiliser une matrice de projection affine réduite, comme elle est considérée pour l'auto-calibrage par Quan [164] ou utilisée pour la reconstruction multi-image par factorisation par Tomasi, Poelman, Weinshall et Kanade [202, 149, 219] :

$$\mathsf{P}^A_{2\times 3} = \overline{\mathsf{K}} \ \overline{\mathsf{R}} \ .$$

Puisque nous travaillons maintenant en coordonnés non homogènes, cette matrice est définie de manière unique – il n'y a pas de liberté de facteur d'échelle.

Les matrices $\overline{\mathsf{K}}$ et $\overline{\mathsf{R}}$ jouent, pour la caméra affine, le même rôle que K et R pour la caméra perspective.

- ▷ $\overline{\mathsf{K}}$ est la **matrice des paramètres intrinsèques** de la caméra affine, contenant ses **3 paramètres intrinsèques** : k_u, k_v et Θ. Ces paramètres ont la même signification que α_u, α_v et Θ pour la caméra perspective : k_u et k_v représentent les échelles de la projection par rapport aux deux directions du plan image et nous retrouvons le même rapport d'échelle que pour la caméra perspective : $\tau = \frac{k_u \sin\Theta}{k_v} = \frac{\alpha_u \sin\Theta}{\alpha_v}$. Θ est l'angle entre les axes des pixels. Le point principal n'est plus défini.

- ▷ $\overline{\mathsf{R}}$ représente entièrement la position de la caméra : soit $\bar{\mathbf{r}}$ le vecteur complétant la matrice de rotation : $\mathsf{R} = \begin{pmatrix} \overline{\mathsf{R}} \\ \bar{\mathbf{r}}^\mathsf{T} \end{pmatrix}$. R représente l'*orientation* du plan image et la *position* du centre de projection est donnée par :

  $$\mathbf{C} = \begin{pmatrix} \bar{\mathbf{r}} \\ 0 \end{pmatrix} \ .$$

  Si nous décomposons R en trois matrices de rotation autour des axes des X, Y et Z, dans le bon ordre, nous pouvons attribuer des rôles individuels aux **3 paramètres extrinsèques**, c'est-à-dire les angles des trois rotations :

  $$\mathsf{R}_Z = \begin{pmatrix} \cos\rho_Z & -\sin\rho_Z & 0 \\ \sin\rho_Z & \cos\rho_Z & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
  $$\mathsf{R}_Y = \begin{pmatrix} \cos\rho_Y & 0 & \sin\rho_Y \\ 0 & 1 & 0 \\ -\sin\rho_Y & 0 & \cos\rho_Y \end{pmatrix} \quad \mathsf{R}_X = \begin{pmatrix} 1 & 0 & 1 \\ 0 & \cos\rho_X & -\sin\rho_X \\ 0 & \sin\rho_X & \cos\rho_X \end{pmatrix} \ .$$

  Avec $\mathsf{R} = \mathsf{R}_Z(\mathsf{R}_Y\mathsf{R}_X)$, les angles ρ_X et ρ_Y définissent uniquement la *position* du centre de projection qui, pour la caméra affine, définit aussi l'*orientation* du plan image dans le monde :

  $$\mathbf{C} = \begin{pmatrix} -\sin\rho_Y \\ \sin\rho_X\cos\rho_Y \\ \cos\rho_X\cos\rho_Y \\ 0 \end{pmatrix} \ .$$

Quant à l'angle ρ_Z, il décrit la rotation *dans* le plan image. S'il y avait un axe optique, on dirait que c'est la rotation autour de cet axe.

Cette distinction des paramètres extrinsèques nous sera très utile pour l'étude des séquences de mouvements qui sont critiques pour l'auto-calibrage de la caméra affine.

En faisant varier les paramètres intrinsèques nous pouvons maintenant reconstituer les sous-classes bien connues de la caméra affine – les caméras orthographique, orthographique à l'échelle et para-perspective [4, 94, 142]. Si nous supposons que les pixels sont rectangulaires, c'est-à-dire $\Theta = 90°$, nous obtenons la matrice des paramètres intrinsèques suivante :

$$\overline{\mathsf{K}} = \begin{pmatrix} k_u & 0 \\ 0 & k_v \end{pmatrix} \quad .$$

Quant à k_u et k_v, nous distinguons trois cas : des pixels carrés ($k_u = k_v$) donnent lieu à une projection **orthographique à l'échelle** (ou perspective faible). La restriction supplémentaire $k_u = k_v = 1$ mène à une projection **orthographique**. Quant au cas $k_u \neq k_v$, nous pourrions parler d'une projection **orthographique aux échelles an-isotropiques**.

Quant aux projections **para-perspectives**, il est connu qu'elles sont équivalentes aux projections affines (Basri : « Para-perspective $\equiv$ Affine » [10]). Par inspection des paramètres intrinsèques de la caméra affine, nous pouvons remarquer qu'une projection para-perspective générale peut être modélisée par une caméra affine avec des pixels non rectangulaires ! Dans [164], Quan montre la relation entre les paramètres intrinsèques de la caméra affine avec une paramétrisation courante de la caméra para-perspective.

2.2.4 La caméra linéaire

Dans les paragraphes précédents, nous avons considéré des caméras effectuant des projections du monde 3D sur un *plan* image ou, pour des caméras physiques, sur une surface d'éléments photosensibles. Il existent d'autres types de caméras, dont celles avec une *droite* image, ou possédant un *vecteur* d'éléments photosensibles. Ces caméras sont, par exemple, utilisées dans l'imagerie satellitaire, où une image bidimensionnelle est obtenue en collant une série d'images linéaires (colonnes de pixels) ensemble, obtenues par la caméra tout en « balayant » la scène [84].

Un autre mode d'application condamne les caméras à une vie dans un plan : et leur centre de projection et la droite image bougent dans un plan et seuls les points du plan sont vus par la caméra. Nous désignons cette situation par la notion de **caméra linéaire**. Si besoin est de distinguer les caméras habituelles des caméras linéaires, nous employons les expressions **caméras planaires** ou **caméras 2D**. Une application de caméras linéaires est prévue par exemple dans le projet Praxitèle [145] : des voitures sont équipées avec des caméras linéaires qui leur permettront de se positionner par rapport à d'autres voitures afin de les suivre en conduite automatique. Le positionnement sera facilité à l'aide de lumières fixées à l'arrière de chaque voiture.

La caméra linéaire peut être modélisée par une projection sténopé du plan projectif $\mathcal{P}^2$ sur la droite projective $\mathcal{P}^1$. Les notations introduites pour la caméra planaire au 2.2.1 se traduisent sans changement au cas présent (par exemple, nous pourrions distinguer les caméras linéaires perspectives et affines). La position de la caméra linéaire est définie par la position du centre de projection dans le plan et l'orientation de la droite de projection, ce qui compte pour les **3 paramètres extrinsèques**. Il n'y a que **2 paramètres intrinsèques** : le point principal, représenté par une coordonnée u_0, et la distance focale f, mesurée relativement à k_u, l'inverse de la taille des pixels (dans la direction de la droite image).

En 6.3, nous décrivons une méthode d'auto-calibrage pour la caméra linéaire, pour laquelle nous exploitons son tenseur trifocal (voir 2.4.1). Le modèle de la caméra linéaire est utilisé par Quan pour la reconstruction affine de droites à partir de vues affines (cf. 2.4.1).

2.2.5 Modèles plus complexes

Pour des applications nécessitant une haute précision de mesure, le modèle sténopé est généralement insuffisant, l'approximation de la projection physique étant trop faible. Ce sont surtout les photogrammètres qui ont étudié les déviations par rapport à la projection perspective que l'on peut modéliser et qu'il faudrait prendre en compte pour une précision désirée [178]. Un modèle très répandu contient, outre les paramètres intrinsèques du modèle sténopé, des coefficients de distorsion radiale et de décentrage [28, 29]. Une alternative à la modélisation et au calibrage explicites des distorsions consiste en la correction des distorsions par des transformations locales des images, qui ont été déterminées à partir de l'image d'une mire spéciale [23]. Il existe une multitude d'autres phénomènes, comme le « line-jitter » [17] ou le vignetage [8], qui invalident le modèle sténopé (voir [178] pour une collection). Nous n'allons pas détailler cette liste, puisque nous adoptons le modèle de caméra sténopé dans cette thèse qui est suffisant pour la plupart des applications en vision par ordinateur.

2.3 La géométrie épipolaire

La géométrie épipolaire est surtout considérée en vue de deux objectifs : l'appariement de primitives image et l'orientation relative de deux vues. En ce qui concerne l'appariement, la contrainte épipolaire bien connue se dérive facilement à l'aide d'un dessin comme montré en figure 2.4. Un point 3D $\mathbf{Q}$ définit, avec les deux centres de projection, un plan que l'on appellera son **plan épipolaire**. Ce plan coupe les plans image en deux droites, appelées **droites épipolaires**. Au faisceau des plans épipolaires, dont la base est la droite reliant les centres optiques, correspond un faisceau de droites épipolaires dans chaque image. Le point de base d'un faisceau de droites épipolaires est appelé **épipôle**. L'épipôle dans une vue est l'image du centre de projection de l'autre vue. Nous désignons la projection du centre optique de la j^{e} caméra dans la i^{e} vue par $\mathbf{e}_{ij}$. Dans le cas de seulement deux vues, nous utilisons les notations simplifiées $\mathbf{e}_1 = \mathbf{e}_{12}$ et $\mathbf{e}_2 = \mathbf{e}_{21}$.

La correspondance bijective entre les plans épipolaires et les droites épipolaires dans chaque image induit une correspondance entre les deux faisceaux épipolaires : l'**homographie épipolaire** projette une droite épipolaire sur la droite épipolaire dans l'autre image qui se trouve dans le même plan épipolaire.

Considérons les projections $\mathbf{q}_1$ et $\mathbf{q}_2$ du point 3D $\mathbf{Q}$. Les deux points image se trouvent sur les deux droites épipolaires qui sont associées au plan épipolaire de $\mathbf{Q}$. On dira de ces deux droites qu'elles sont les droites épipolaires correspondantes, ou associées, à $\mathbf{q}_1$ et $\mathbf{q}_2$. Si $\mathbf{Q}$ se déplace dans son plan épipolaire, les positions de $\mathbf{q}_1$ et $\mathbf{q}_2$ dans les images peuvent changer, mais ils restent sur les mêmes droites épipolaires. Il est donc facile à voir que toute paire de points $\mathbf{q}_1$ et $\mathbf{q}_2$ qui se trouvent sur deux droites épipolaires correspondantes, sont les projections possibles d'un point 3D. Dans le cas contraire, c'est-à-dire si les droites épipolaires de points image $\mathbf{q}_1$ et $\mathbf{q}_2$ ne se correspondent pas, ces deux points ne peuvent pas être les projections communes d'un point 3D.

Nous venons de décrire la contrainte épipolaire pour l'appariement de points dans deux images. Elle est en général utilisée pour restreindre la recherche de points correspondants le long de paires de droites épipolaires correspondantes.

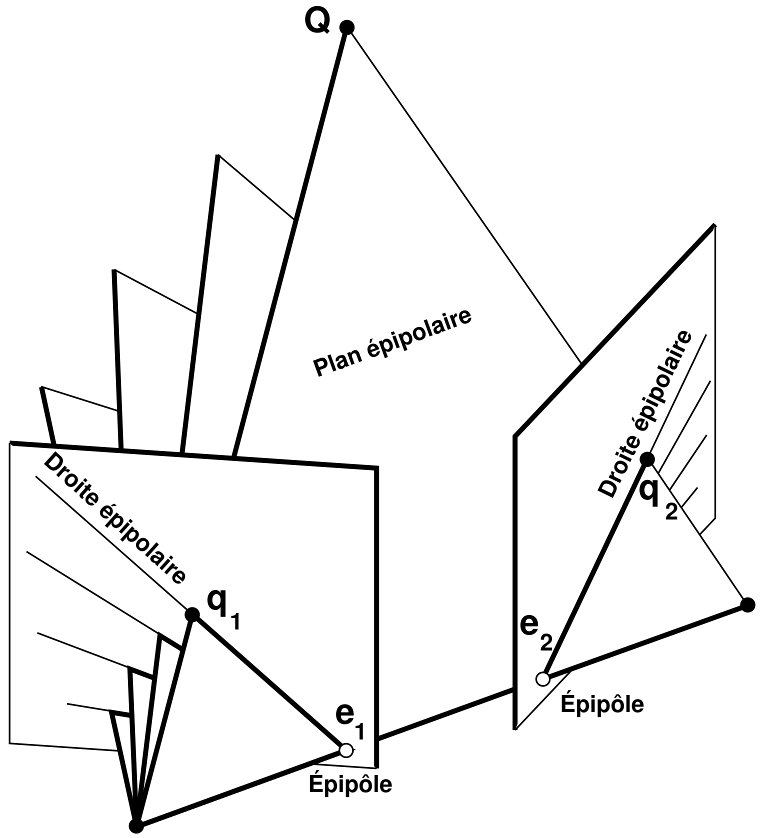


FIG. 2.4: *La géométrie épipolaire.*

Il existe une contrainte similaire pour les images de quadriques : les droites épipolaires qui sont tangentes aux coniques image sont des droites épipolaires correspondantes (à une permutation près – il y a deux paires de telles tangentes ; voir la figure 2.5). La même contrainte est valable pour les images de coniques 3D. On retrouve cette contrainte bifocale d'appariement de quadriques/coniques dans les équations de Kruppa pour l'auto-calibrage [108, 123] (voir 6.2.1.5).

2.3.1 Les matrices fondamentale et essentielle

Pour ce qui est de l'aspect orientation relative de deux vues, on distingue les géométries épipolaires « calibrées » et « non calibrées », selon qu'on considère des caméras calibrées ou non. À ce point, il est aisé d'introduire les notions de matrice essentielle et fondamentale. La **matrice fondamentale** F est une représentation algébrique de la géométrie épipolaire non calibrée [63, 72]. Elle est la matrice d'une corrélation qui transforme des points dans une image sur des droites dans l'autre image, le tout exprimé dans les repères pixels des images. Plus précisément, il s'agit de la corrélation transformant un point image sur la droite épipolaire correspondante dans l'autre image. Puisqu'il n'y a qu'un faisceau de droites épipolaires, la corrélation n'est pas de plein rang ce qui veut dire que le déterminant de la matrice fondamentale est nul. Le noyau de F est précisément l'épipôle.

Soit F_{12} la matrice fondamentale passant de l'image 1 à l'image 2. La matrice fondamentale F_{21} pour le sens inverse est juste sa transposée : $\mathsf{F}_{21} = \mathsf{F}_{12}^{\mathsf{T}}$. La contrainte épipolaire pour deux points

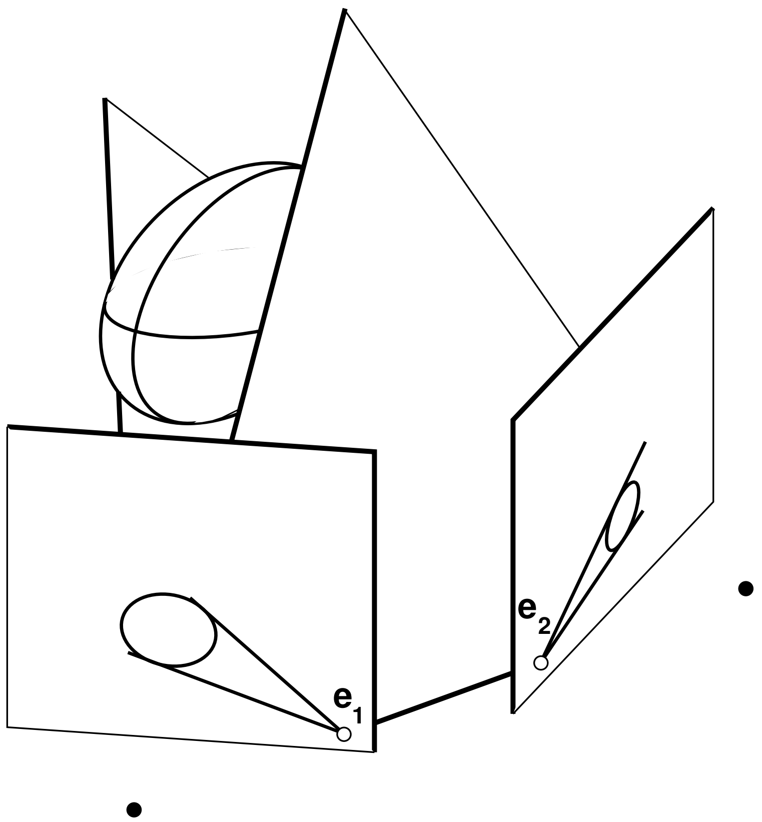


FIG. 2.5: *La contrainte épipolaire pour les quadriques : il y a deux plans épipolaires tangents à la quadrique et les droites épipolaires en induites sont tangentes aux coniques image. Donc, les droites épipolaires tangentes aux coniques image se correspondent.*

image $\mathbf{q}_1$ et $\mathbf{q}_2$ peut s'écrire comme :

$$\mathbf{q}_2^\mathsf{T} \mathsf{F}_{12} \mathbf{q}_1 = 0 \ .$$

Cette équation exprime que $\mathbf{q}_2$ se trouve sur la droite épipolaire correspondante à $\mathbf{q}_1$ et qui est donnée par $\mathbf{l}_2 = \mathsf{F}_{12}\mathbf{q}_1$. La transposition de l'équation donne le résultat analogue dans l'autre sens : $\mathbf{q}_1^\mathsf{T} \mathsf{F}_{12}^\mathsf{T} \mathbf{q}_2 = \mathbf{q}_1^\mathsf{T} \mathsf{F}_{21} \mathbf{q}_2 = 0$, c'est-à-dire que $\mathbf{q}_1$ se trouve sur la droite épipolaire correspondante à $\mathbf{q}_2$, $\mathbf{l}_1 = \mathsf{F}_{21}\mathbf{q}_2$.

Jusqu'alors nous avons travaillé dans les repères pixels. Si l'information du calibrage des deux vues est disponible, on peut remonter aux repères image. Soient K_1 et K_2 les deux matrices des paramètres intrinsèques. Transformer les points image via K_1^{-1} et K_2^{-1} implique la transformation suivante de la matrice fondamentale :

$$\mathsf{E} = \mathsf{K}_2^\mathsf{T} \ \mathsf{F} \ \mathsf{K}_1 \ .$$

La matrice E est appelée **matrice essentielle** ; elle représente la géométrie épipolaire calibrée. Elle contient toute l'information sur le mouvement des caméras, qui peut être extraite des seules images. Soient R_1, R_2, $\mathbf{t}_1$ et $\mathbf{t}_2$ les orientations et positions des deux caméras. La matrice essentielle est donnée par :

$$\mathsf{E} = [\mathsf{R}_2(\mathbf{t}_1 - \mathbf{t}_2)]_\wedge \mathsf{R}_2 \mathsf{R}_1^\mathsf{T} \ .$$

Le fait que E est le produit d'une matrice anti-symétrique et d'une autre orthogonale, impose, outre la nullité de son déterminant, la contrainte que ses deux valeurs singulières non nulles soient égales [101]. Il est possible d'extraire, de E, l'**orientation relative** entre les deux caméras, c'est-à-dire le déplacement rigide entre elles, mais, comme il est connu, seulement à un facteur d'échelle global près : on peut juste déterminer la *direction* de la translation entre les caméras, non son amplitude. Afin d'éviter des confusions concernant la notion classique de l'orientation relative, nous introduisons ici la **rotation relative**, qui décrit uniquement la partie rotationnelle du déplacement entre les deux caméras. La rotation relative $\mathsf{R} = \mathsf{R}_2\mathsf{R}_1^{\mathsf{T}}$ peut être déterminée à partir de la matrice essentielle. Deux solutions sont obtenues (étant donné une solution, la deuxième est obtenue en rajoutant une rotation de 180° autour de la droite de translation). En pratique on peut généralement rejeter une de ces 2 solutions.

La matrice essentielle a été introduite par Longuet-Higgins [112] et utilisée par beaucoup de chercheurs pour l'estimation du mouvement entre deux caméras calibrées [103, 112, 147, 212, 223, 240]. Une fois le mouvement déterminé, il est aisé d'établir deux matrices de projection afin de reconstruire la scène par triangulation. La matrice essentielle contient donc toute l'information nécessaire pour obtenir une reconstruction euclidienne de la scène. Son analogue non calibré, la matrice fondamentale, permet de trouver la structure projective de la scène [63, 72], ainsi que deux informations sur le calibrage des deux caméras [69, 77] (voir aussi 6.2.1.6).

2.3.2 Estimation de la matrice fondamentale

Il y a une multitude de méthodes d'estimation de la matrice fondamentale, presque toutes utilisant uniquement des correspondances de points image comme données [21, 46, 71, 110, 114, 116, 156, 204, 236, 237]. Ces méthodes se distinguent surtout dans la paramétrisation de F et le critère d'optimisation. Nous n'allons pas décrire ces méthodes en détail, mais renvoyons aux nombreuses références citées ci-dessus. Ici, nous allons juste étudier deux critères d'optimisation, afin de préparer la discussion sur l'optimalité d'une méthode de triangulation que nous proposons en chapitre 4.

Il est généralement admis que le mieux qu'on puisse faire pour l'estimation de la matrice fondamentale est d'utiliser une méthode itérative, en partant d'une bonne initialisation. Le critère de minimisation le plus souvent utilisé pour les méthodes itératives est la somme des carrés des distances des points image à leurs droites épipolaires [114] :

$$\widehat{\mathsf{F}} = \operatorname{argmin} \sum_{p=1}^{n} \left(d(\mathbf{q}_{1p}, \mathsf{F}^{\mathsf{T}}\mathbf{q}_{2p})^2 + d(\mathbf{q}_{2p}, \mathsf{F}\mathbf{q}_{1p})^2 \right) \quad , \tag{2.2}$$

éventuellement incluant une pondération des termes basée sur l'incertitude des coordonnées des points image.

Hartley a remarqué [71] que, sous l'hypothèse d'un bruit gaussien dans les coordonnées des points image, ce critère ne donne pas la matrice fondamentale statistiquement optimale (au sens du maximum de vraisemblance). Il faudrait plutôt déterminer et la matrice $\widehat{\mathsf{F}}$ et des points image corrigés $\widehat{\mathbf{q}}'_{ip}$ tels que ces points satisfassent exactement la contrainte épipolaire donnée par $\widehat{\mathsf{F}}$ et que la somme des carrés des distances des points corrigés à ceux mesurés soit minimale :

$$\begin{gathered} \{\widehat{\mathsf{F}}, \widehat{\mathbf{q}}'_{11}, \ldots, \widehat{\mathbf{q}}'_{1n}, \widehat{\mathbf{q}}'_{21}, \ldots, \widehat{\mathbf{q}}'_{2n}\} = \operatorname{argmin} \textstyle\sum_{p=1}^{n} \left(d(\mathbf{q}_{1p}, \mathbf{q}'_{1p})^2 + d(\mathbf{q}_{2p}, \mathbf{q}'_{2p})^2 \right) \\ \text{sous la contrainte} \quad \mathbf{q}'^{\mathsf{T}}_2 \mathsf{F} \mathbf{q}'_1 = 0 \quad . \end{gathered} \tag{2.3}$$

Les inconvénients du critère (2.3) sont le nombre élevé des inconnues et le besoin d'une méthode d'estimation itérative. En effet, minimiser (2.3) revient à effectuer, implicitement, une reconstruction projective optimale, c'est-à-dire qui minimise les erreurs de reprojection (voir 3.7).

Trivedi a déjà appliqué le critère (2.3) pour l'estimation de la matrice essentielle [211]. Shapiro *et al.* montrent dans [172] que, pour le cas de la géométrie épipolaire de caméras affines, la minimisation de (2.3) ne nécessite pas la prise en compte d'inconnues pour les points image corrigés. Nous redérivons ce résultat en 4.6.1, où nous considérons la triangulation de points 3D avec des caméras affines. Notamment, nous montrons une expression simple qui donne les points corrigés optimaux $\widehat{\mathbf{q}}'_{ip}$ en fonction des coefficients de la « matrice fondamentale affine ».

2.3.3 Parties symétrique et anti-symétrique de la matrice fondamentale

Les parties symétrique et anti-symétrique de la matrice fondamentale, $\mathsf{F}_s = \mathsf{F}+\mathsf{F}^\mathsf{T}$ et $\mathsf{F}_a = \mathsf{F}-\mathsf{F}^\mathsf{T}$, ont une interprétation géométrique intéressante, que nous décrivons dans la suite.

2.3.3.1 La partie symétrique

Considérons un point 3D $\mathbf{Q}$, dont les projections dans les deux images ont les mêmes coordonnées homogènes, $\mathbf{q}_1 \sim \mathbf{q}_2$. Nous pouvons alors écrire la contrainte épipolaire pour ces deux points comme $\mathbf{q}_1^\mathsf{T}\mathsf{F}\mathbf{q}_1 = 0$ ou $\mathbf{q}_1^\mathsf{T}\mathsf{F}^\mathsf{T}\mathbf{q}_1 = 0$. L'addition de ces équations donne :

$$\mathbf{q}_1^\mathsf{T}\mathsf{F}_s\mathbf{q}_1 = 0 \ .$$

Il est tentant d'interpréter la matrice symétrique F_s comme une conique. En effet, si l'on représente les deux épipôles et les faisceaux épipolaires dans le même repère, F_s est la matrice de la conique de Steiner (voir 2.7.4) construite par les faisceaux épipolaires (voir la figure 2.6). Donc, F_s est le lieu de tous les points image $\mathbf{q}$ pour lesquels il existe un point 3D $\mathbf{Q}$ tel que $\mathbf{q} \sim \mathsf{P}_1\mathbf{Q} \sim \mathsf{P}_2\mathbf{Q}$.

2.3.3.2 La partie anti-symétrique

Soit $\mathbf{q}_a$ le noyau de la matrice anti-symétrique F_a, c'est-à-dire que $(\mathsf{F}-\mathsf{F}^\mathsf{T})\mathbf{q}_a = \mathbf{0}$ ou, autrement dit, $\mathsf{F}_a = [\mathbf{q}_a]_\wedge$. Il en découle que $\mathsf{F}\mathbf{q}_a = \mathsf{F}^\mathsf{T}\mathbf{q}_a$. Nous pouvons interpréter $\mathbf{l}_a = \mathsf{F}\mathbf{q}_a$ comme une droite épipolaire dans la deuxième image, qui passe donc par le deuxième épipôle. De même, $\mathbf{l}_a = \mathsf{F}^\mathsf{T}\mathbf{q}_a$ représente une droite épipolaire dans la première vue, passant par le premier épipôle. $\mathbf{l}_a$ est donc la droite reliant les épipôles (algébriquement, et non pas au sens géométrique 3D).

Il est facile de montrer que le point $\mathbf{q}_a$ n'est rien d'autre que le pôle de $\mathbf{l}_a$ par rapport à la conique F_s (cf. la figure 2.6) :

$$\mathsf{F}_s\mathbf{q}_a = (\mathsf{F}+\mathsf{F}^\mathsf{T})\mathbf{q}_a = \mathsf{F}\mathbf{q}_a + \mathsf{F}^\mathsf{T}\mathbf{q}_a = 2\mathsf{F}\mathbf{q}_a \sim \mathbf{l}_a \ .$$

F_s et F_a sont une représentation presque unique de la géométrie épipolaire : les épipôles sont les deux intersections de $\mathbf{l}_a$ avec la conique F_s. Pour un point image donné, on détermine sa droite épipolaire dans l'autre image comme suit : on relie le point avec l'épipôle de son image et détermine le point d'intersection de cette droite avec F_s. La droite reliant le point d'intersection avec l'autre épipôle est la droite épipolaire cherchée. La seule indétermination est l'attribution des deux épipôles aux deux images.

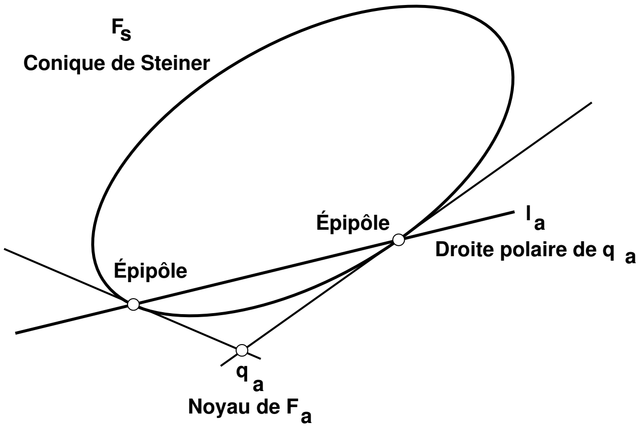


FIG. 2.6: *Interprétation géométrique des parties symétrique et anti-symétrique de la matrice fondamentale* F. *La partie symétrique* F_s *représente la conique de Steiner des faisceaux épipolaires et celle anti-symétrique,* F_a*, le pôle de la droite reliant les épipôles, par rapport à cette conique.*

2.3.4 La matrice fondamentale pour des mouvements particuliers

Dans ce paragraphe, nous considérons le cas du mouvement d'une caméra, incluant éventuellement un changement de focale ou une mise au point. Pour des mouvements particuliers, la matrice fondamentale peut prendre des formes particulières dépendant de moins de paramètres que la forme générale. Ces formes particulières de F ont déjà été étudiées par Torr *et al.* [204] et Viéville et Lingrand [215]. Torr propose une méthode robuste de choix de modèle pour la matrice fondamentale, afin de trouver la forme minimale de la matrice fondamentale, qui correspond bien à la nature du mouvement sous-jacent. Viéville et Lingrand étudient comment des formes particulières de F permettent de simplifier les équations de Kruppa [108, 123] et ainsi de simplifier et stabiliser l'auto-calibrage basé sur ces équations (voir 6.2.3.3).

Nous décrivons, dans la suite, des formes particulières de F pour certains types de mouvements. Dans la section 4.6, ces formes particulières sont exploitées afin de simplifier la triangulation de points 3D à partir de deux vues. Le tableau 2.1 sur la page 25 résume les différentes formes (d'autres formes particulières peuvent être trouvées dans [215]).

2.3.4.1 Translation pure

Pour une translation pure entre deux vues (les paramètres intrinsèques restent constants), F est anti-symétrique :

$$\mathsf{F} = [\mathbf{e}]_\wedge = \begin{pmatrix} 0 & -e_3 & e_2 \\ e_3 & 0 & -e_1 \\ -e_2 & e_1 & 0 \end{pmatrix} ,$$

où $\mathbf{e} = (e_1, e_2, e_3)^\mathsf{T}$ est l'épipôle, qui est commun aux deux images. La construction de F implique qu'elle est de rang 2.

2.3.4.2 Caméras affines

Pour une caméra affine [134], le plan à l'infini est projeté sur la droite à l'infini du plan image. Considérons une paire de caméras affines. Leurs centres optiques se trouvent à l'infini, donc les épipôles sont des points à l'infini dans les images. En outre, les droites à l'infini des plans image sont des droites épipolaires correspondantes. La **matrice fondamentale affine** a donc la forme particulière suivante [172] :

$$\mathsf{F} = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{pmatrix} \quad . \tag{2.4}$$

La construction de F implique qu'elle est de rang au plus 2.

2.3.4.3 Zoom et mise au point

Nous considérons le cas d'un changement de focale et éventuellement d'une mise au point d'une caméra qui ne se déplace pas. Les systèmes mécaniques et optiques sont complexes et relativement difficiles à modéliser, donc nous adoptons les hypothèses faites par Viéville et Lingrand [215] : la mise au point et le changement de focale peuvent être modélisés par un changement de quelques paramètres intrinsèques (point principal et distance focale) et d'une translation du centre optique. La taille des pixels, l'angle entre les axes des pixels et l'orientation de la caméra sont supposés rester constant.

Soient :

$$\mathsf{K}_1 = \begin{pmatrix} \tau\alpha_1 & s & u_1 \\ 0 & \alpha_1 & v_1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathsf{K}_2 = \begin{pmatrix} \tau\alpha_2 & s & u_2 \\ 0 & \alpha_2 & v_2 \\ 0 & 0 & 1 \end{pmatrix}$$

les matrices des paramètres intrinsèques avant et après le mouvement. Soit $\mathbf{t}$ le vecteur de translation du déplacement du centre optique. La matrice fondamentale se calcule d'après (voir la section 2.6) :

$$\mathsf{F} = \mathsf{K}_2^{-\mathsf{T}} \, [\mathbf{t}]_\wedge \, \mathsf{K}_1^{-1} \quad .$$

Cette forme de la matrice fondamentale est relativement compliquée, mais des contraintes entre ses coefficients permettent de dériver une forme générique qui ne dépend que de 5 paramètres :

$$\mathsf{F} \sim \begin{pmatrix} 0 & 1 & a \\ -1 & b & c \\ d & e & ae - cd + abd \end{pmatrix} \quad .$$

Nous considérons des combinaisons des cas particuliers suivants :

- ▷ les pixels sont rectangulaires ($\Theta = 90°$ ou $s = 0$) ;
- ▷ la translation du centre optique s'effectue le long de l'axe optique ($\mathbf{t} = (0, 0, t_3)^\mathsf{T}$) ;
- ▷ la translation du centre optique s'effectue dans le plan focal ($\mathbf{t} = (t_1, t_2, 0)^\mathsf{T}$) ;

▹ le point principal reste fixe.

Les différentes formes de F sont regroupées dans le tableau 2.1 sur la page 25. Notons que l'hypothèse de pixels *carrés* ($\tau = 1$) n'apporte pas de simplifications supplémentaires par rapport au cas de pixels *rectangulaires*.

2.3.4.4 Rotation autour d'un axe fixe

Nous considérons le cas d'une caméra avec des paramètres intrinsèques fixes, qui effectue une rotation pure, autour d'un axe qui ne passe *pas* par le centre de projection. C'est le mouvement de base pour le cas du mouvement planaire, c'est-à-dire où la caméra se déplace dans un plan et effectue des rotations autour d'axes perpendiculaires au plan de mouvement (voir 6.2.3.2 pour l'auto-calibrage basé sur de tels mouvements).

Les points 3D dont les projections dans les deux images sont identiques, sont montrés dans la figure 2.7 (a) : il s'agit des points sur l'axe de rotation et sur le cercle dans le plan de mouvement qui passe par les deux centres de projection et l'axe de rotation. La conique de Steiner représentée par la partie symétrique de F, dégénère en une paire de droites (voir la figure 2.7 (b)). Il en découle une contrainte supplémentaire sur la matrice fondamentale :

$$\det \mathsf{F}_s = \det(\mathsf{F} + \mathsf{F}^\mathsf{T}) = 0 \ . \tag{2.5}$$

Notons $\mathbf{l}_r$ l'image de l'axe de rotation, et $\mathbf{l}_h$ la « ligne d'horizon », l'image du plan de mouvement Π. Nous dérivons maintenant une paramétrisation pour la matrice fondamentale qui engendre la contrainte (2.5). Pour ce faire, nous suivons la construction géométrique de la corrélation représentée par F (cf. figure 2.7 (c)) : soit $\mathbf{q}_1$ un point dans la première image ; sa droite épipolaire dans la deuxième image est obtenue par :

$$\mathbf{l}_2 \sim \mathsf{F}\mathbf{q}_1 \sim \mathbf{e}_2 \wedge (\mathbf{l}_r \wedge (\mathbf{e}_1 \wedge \mathbf{q}_1)) \ .$$

En représentation matricielle ceci s'écrit comme :

$$\mathsf{F}\mathbf{q}_1 \sim [\mathbf{e}_2]_\wedge \, [\mathbf{l}_r]_\wedge \, [\mathbf{e}_1]_\wedge \, \mathbf{q}_1 \ .$$

Puisque ceci est valable pour tous les points $\mathbf{q}_1$, nous pouvons identifier la matrice fondamentale comme :

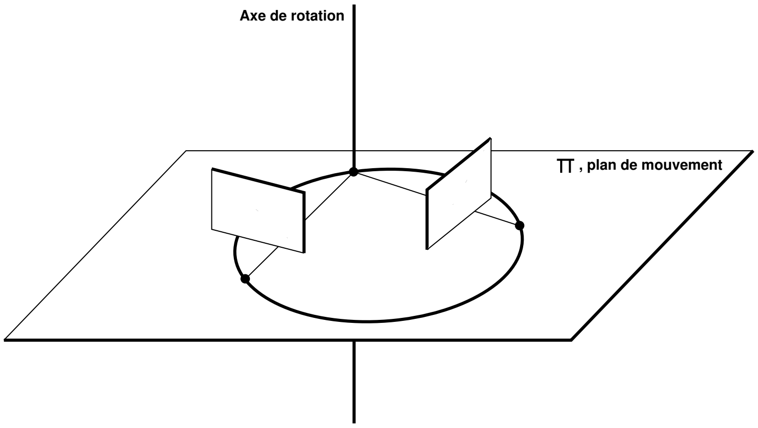
$$\mathsf{F} \sim [\mathbf{e}_2]_\wedge \, [\mathbf{l}_r]_\wedge \, [\mathbf{e}_1]_\wedge \ . \tag{2.6}$$

F peut donc être paramétrée par les deux épipôles et l'image de l'axe de rotation (il y a encore d'autres possibilités). Chacune de ces trois entités n'est définie qu'à un scalaire près, ce qui fait que F est définie par 6 paramètres. Ceci est en accord avec le fait que la contrainte (2.5) diminue de 1 le nombre de paramètres de F.

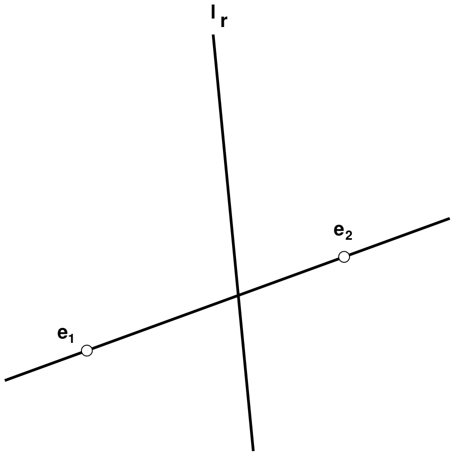
Plusieurs cas particuliers existent, dont le plus intéressant sera développé dans la suite. Considérons une caméra avec des pixels rectangulaires et supposons que l'axede rotation est parallèle au plan image et, de plus, à un des axes des pixels (voir la figure 2.8). Il s'agit d'une configuration stéréo typique que l'on retrouve fréquemment pour les « têtes stéréo ». Nous considérons le cas où l'axe de rotation est parallèle à l'axe des v.

La matrice fondamentale prend la forme suivante :

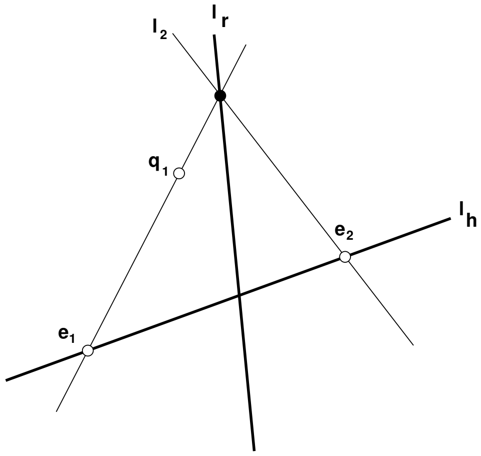
$$\mathsf{F} = \begin{pmatrix} 0 & -\cos\rho_2 & v_0 \cos\rho_2 \\ \cos\rho_1 & 0 & \alpha_u \sin\rho_1 - u_0 \cos\rho_1 \\ -v_0\cos\rho_1 & u_0\cos\rho_2 - \alpha_u \sin\rho_2 & v_0(\alpha_u(\sin\rho_2 - \sin\rho_1) + u_0(\cos\rho_1 - \cos\rho_2)) \end{pmatrix} \ .$$



(a) Les points 3D avec des projections fixes sont ceux sur l'axe de rotation et sur le cercle dans le plan du mouvement qui passe par les deux centres de projection et l'axe de rotation.



(b) La conique de Steiner qui contient les projections des points 3D de la figure (a) est une conique dégénérée, consistant de deux droites : $\mathbf{l}_r$ est l'image de l'axe de rotation et $\mathbf{l}_h$ la « ligne d'horizon », c'est-à-dire la projection du plan de mouvement. Les deux épipôles se trouvent évidemment sur $\mathbf{l}_h$.



(c) La construction de la droite épipolaire $\mathbf{l}_2$ correspondante au point $\mathbf{q}_1$. Relier $\mathbf{e}_2$ avec le point d'intersection de $\mathbf{l}_r$ et la droite $\langle \mathbf{e}_1 \mathbf{q}_1 \rangle$ donne $\mathbf{l}_2$.

FIG. 2.7: *Géométrie épipolaire pour une rotation autour d'un axe fixe.*

où u_0, v_0, α_u sont les paramètres intrinsèques de la caméra et ρ_1 et ρ_2 les angles de vergence. Nous

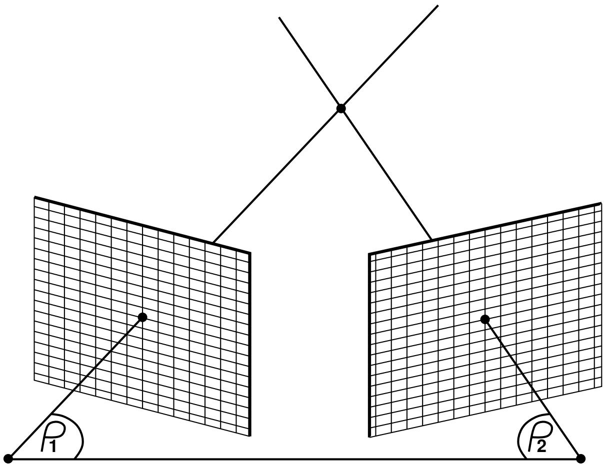


FIG. 2.8: *Configuration géométrique d'une tête stéréo. Les grilles de pixels sont esquissées pour illustrer qu'il n'y a pas de rotation autour des axes optiques.*

pouvons identifier les paramètres suivants :

$$\begin{aligned} a &= -\cos\rho_2 \\ b &= v_0 \\ c &= -\cos\rho_1 \\ d &= \alpha_u \sin\rho_1 - u_0\cos\rho_1 \\ e &= u_0\cos\rho_2 - \alpha_u\sin\rho_2 \ , \end{aligned}$$

avec lesquels la matrice fondamentale s'écrit comme :

$$\mathsf{F} = \begin{pmatrix} 0 & a & -ab \\ -c & 0 & d \\ bc & e & -b(d+e) \end{pmatrix} \ .$$

Dans le cas d'une « tête stéréo symétrique », c'est-à-dire où les angles de vergence sont opposés ($\rho_2 = -\rho_1$), la matrice fondamentale se simplifie davantage :

$$\mathsf{F} = \begin{pmatrix} 0 & \cos\rho & -v_0\cos\rho \\ \cos\rho & 0 & \alpha_u\sin\rho - u_0\cos\rho \\ -v_0\cos\rho & -\alpha_u\sin\rho - u_0\cos\rho & 2u_0v_0\cos\rho \end{pmatrix} \ .$$

Avec les paramètres :

$$\begin{aligned} a &= \cos\rho \\ b &= v_0 \\ c &= \alpha_u\sin\rho \\ d &= u_0\cos\rho \ , \end{aligned}$$

la matrice fondamentale s'écrit comme :

$$\mathsf{F} = \begin{pmatrix} 0 & a & -ab \\ a & 0 & c-d \\ -ab & -c-d & 2bd \end{pmatrix} \ .$$

Une forme similaire peut être dérivée pour un axe de rotation parallèle à l'axe des u. Li *et al.* proposent des méthodes pour estimer la matrice fondamentale pour ces configurations de type tête stéréo, prenant en compte sa forme particulière [111].

2.4 Tenseurs d'appariement

Dans la section précédente, nous avons décrit la contrainte épipolaire pour des points image correspondants, contrainte d'appariement qui est effective pour deux images. Cette contrainte peut être codée par la matrice fondamentale (ou essentielle) et se manifeste par une relation *bilinéaire* entre les coordonnées de deux points correspondants :

$$\mathbf{q}_2^\mathsf{T} \mathsf{F} \mathbf{q}_1 = 0 \ . \tag{2.7}$$

Durant les quelques années passées, des relations algébriques pour 3 vues ont été découvertes et puis étudiées systématiquement. Shashua a trouvé qu'il existe des relations *trilinéaires* entre les coordonnées de points correspondants dans trois images [175], résultat anticipé par Spetsakis et Aloimonos [181]. Shashua a également introduit le tenseur trifocal, l'analogue de la matrice fondamentale, pour trois vues. Hartley a dérivé le tenseur d'appariement de droites dans trois images et révélé les liens qui existent avec le tenseur des points [70]. Après ces découvertes, les relations algébriques des correspondances de points (et de droites) pour un nombre arbitraire d'images ont été étudiées plus systématiquement par Triggs [207, 208] et Faugeras et Mourrain [62]. Le principe de leurs développements peut se résumer comme suit (pour plus de détails et de résultats, nous renvoyons aux références citées). Soit $\mathbf{Q}$ un point 3D qui se projette dans m vues d'après :

$$\begin{array}{rcl}
\mathsf{P}_1 \mathbf{Q} & = & \lambda_1 \mathbf{q}_1 \\
\mathsf{P}_2 \mathbf{Q} & = & \lambda_2 \mathbf{q}_2 \\
\cdots & \cdots & \cdots \\
\mathsf{P}_m \mathbf{Q} & = & \lambda_m \mathbf{q}_m \ .
\end{array}$$

Nous pouvons regrouper ces équations en une équation matricielle :

$$\begin{pmatrix} \mathsf{P}_1 \\ \mathsf{P}_2 \\ \cdots \\ \mathsf{P}_m \end{pmatrix} \mathbf{Q} = \begin{pmatrix} \lambda_1 \mathbf{q}_1 \\ \lambda_2 \mathbf{q}_2 \\ \cdots \\ \lambda_m \mathbf{q}_m \end{pmatrix} \ .$$

Il en découle que :

$$\begin{pmatrix} \mathsf{P}_1 \\ \mathsf{P}_2 \\ \cdots \\ \mathsf{P}_m \end{pmatrix} \begin{pmatrix} \mathtt{I}_4 & | \mathbf{Q} \end{pmatrix} = \underbrace{\left(\begin{array}{c|c} \mathsf{P}_1 & \lambda_1 \mathbf{q}_1 \\ \mathsf{P}_2 & \lambda_2 \mathbf{q}_2 \\ \cdots & \cdots \\ \mathsf{P}_m & \lambda_m \mathbf{q}_m \end{array} \right)}_{\mathsf{M}} \ .$$

La matrice M est de dimension $3m \times 5$, mais son rang est au plus 4, puisque M est le produit d'une matrice de 4 colonnes et d'une de 4 lignes. Ceci signifie que tous les mineurs 5×5 de M s'annulent. En développant les mineurs par les coefficients de la dernière colonne, nous obtenons des équations homogènes et linéaires en les composantes des coordonnées image $\lambda_i \mathbf{q}_i$. Les coefficients de ces équations sont des déterminants 4×4 d'ensembles de lignes des matrices de projection P_i. Pour obtenir les

Type de mouvement	Forme générique de F	# Par.
Translation pure	$\begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}$	2
Caméras affines	$\begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{pmatrix}$	4
Zoom + mise au point	Forme générale $\begin{pmatrix} 0 & a^3 & a^2 b \\ -a^3 & a^2 c & a^2 d \\ a^2 e & a^2 f & abf - ade + bce \end{pmatrix}$	5
	Pixels rectangulaires $\begin{pmatrix} 0 & a^2 & ab \\ -a^2 & 0 & ac \\ ad & ae & be - cd \end{pmatrix}$	4
	Translation du centre optique dans le plan focal $\begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{pmatrix}$	4
	Translation dans le plan focal, pixels rectangulaires $\begin{pmatrix} 0 & 0 & ab \\ 0 & 0 & -bd \\ -ac & cd & e \end{pmatrix}$	3
	Translation le long de l'axe optique et point principal fixe $\begin{pmatrix} 0 & a & -ab \\ -a & c & ad - bc \\ ab & -ad - bc & b^2 c \end{pmatrix}$	3
	Translation le long de l'axe optique, point principal fixe, pixels rectangulaires $\begin{pmatrix} 0 & 1 & a \\ -1 & 0 & b \\ -a & -b & 0 \end{pmatrix}$	2
Rotation autour d'un axe fixe	Forme générale $[\mathbf{e}']_\wedge [\mathbf{l}_r]_\wedge [\mathbf{e}]_\wedge$	6
	Tête stéréo $\begin{pmatrix} 0 & a & -ab \\ -c & 0 & d \\ bc & e & -b(d+e) \end{pmatrix}$	4
	Tête stéréo symétrique $\begin{pmatrix} 0 & a & -ab \\ a & 0 & c - d \\ -ab & -c - d & 2bd \end{pmatrix}$	3

TAB. 2.1: *Des formes particulières de la matrice fondamentale pour quelques types de mouvements particuliers.*

contraintes d'appariement, il faut éliminer les facteurs d'échelle λ_i. On obtient des équations respectivement bilinéaires, trilinéaires ou quadrilinéaires si les déterminants sont composés de deux, trois ou quatre matrices de projection. Un ordre supérieur à 4 n'est évidemment pas possible. Par conséquent, les relations algébriques entre m vues sont divisées en des relations entre deux, trois ou quatre vues uniquement.

Les équations multi-linéaires pour un ensemble de deux, trois ou quatre vues peuvent être regroupées en des équations tensorielles entre des points image correspondants. Les coefficients des **tenseurs d'appariement** F, G (trifocal), H (quadrifocal) et aussi des épipôles sont des déterminants 4×4 déjà mentionnés, de 4 lignes prises dans les matrices de projection. Il existe des relations algébriques entre les coefficients des tenseurs, l'exemple le plus simple en est la nullité du déterminant de la matrice fondamentale.

Dans cette thèse, nous n'utilisons pas les tenseurs d'appariement pour plus de deux vues, à l'exception du cas de la caméra linéaire qui est décrit dans le paragraphe suivant. Nous ne détaillons donc pas plus les considérations des autres tenseurs et renvoyons le lecteur intéressé à [62, 207, 208].

2.4.1 Tenseur trifocal de la caméra linéaire

En 6.3, nous décrivons une méthode d'auto-calibrage pour la caméra linéaire qui est basée sur son tenseur trifocal, que nous considérons brièvement dans la suite. Notons tout d'abord qu'il n'existe que des relations trilinéaires entre des vues de caméras linéaires. Il n'y a pas de contrainte d'appariement bifocale : chaque point image définit un rayon de projection dans le plan ; donc pour n'importe quelle paire de points image, il existe un point 2D qui se projette sur eux – le point d'intersection des rayons de projection.

Le seul tenseur d'appariement est le tenseur trifocal G, de dimension $2 \times 2 \times 2$. G est défini à un facteur multiplicatif près, mais en dehors de ceci, il n'existe pas de contrainte supplémentaire sur ses coefficients, contrairement à l'existence de relations quadratiques pour les tenseurs de vues planaires.

Le tenseur trifocal de caméras linéaires a été utilisé par Quan pour la reconstruction affine de droites [165, 166] : la détermination des directions des droites y est formulée par la restriction des projections affines aux points du plan à l'infini. Ceci peut être exprimé en termes de caméras linéaires, leur monde 2D étant le plan à l'infini.

2.5 Homographies associées aux projections d'un plan

Considérons un plan Π et une projection perspective ou affine avec un centre de projection hors de Π. La restriction de la projection aux points de Π est une transformation bijective H_1. Une deuxième projection induit de manière analogue une bijection H_2.

Nous pouvons maintenant établir une correspondance entre les deux projections des points de Π. Soit $\mathbf{q}_1$ la première projection d'un point $\mathbf{Q} \in \Pi$. $\mathbf{Q}$ est donc donné par $\mathbf{Q} \sim \mathsf{H}_1^{-1}\mathbf{q}_1$. La reprojection de $\mathbf{Q}$ dans la deuxième image par H_2 donne le point $\mathbf{q}_2$ correspondant à $\mathbf{q}_1$. Il existe donc une correspondance bijective entre les projections des points de Π, qui peut s'écrire à l'aide de la transformation $\mathsf{H}_\pi = \mathsf{H}_2\mathsf{H}_1^{-1}$:

$$\mathbf{q}_2 \sim \mathsf{H}_2\mathsf{H}_1^{-1}\mathbf{q}_1 = \mathsf{H}_\pi\mathbf{q}_1 \quad .$$

H_π est une homographie et nous l'appelons l'**homographie associée aux projections du plan** Π, ou plus simplement l'**homographie de** Π.

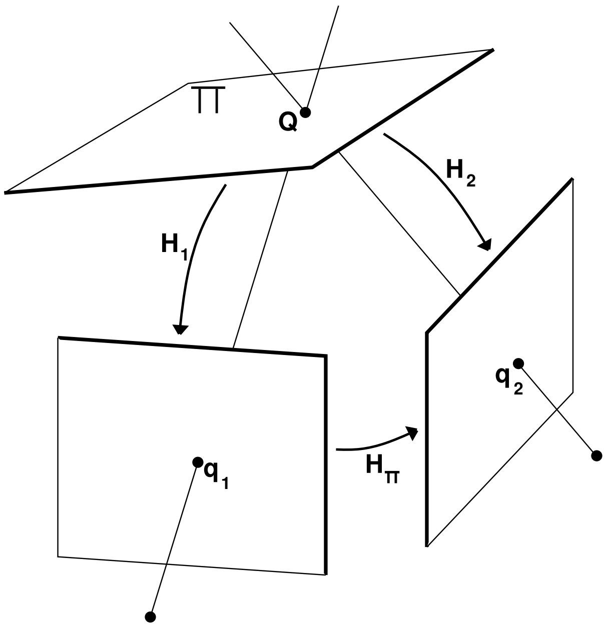


FIG. 2.9: *L'homographie associée aux projections d'un plan. L'homographie* H_π *relie les projetés de points du plan* Π *dans les deux images.*

L'homographie du plan à l'infini est appelée **homographie infinie** et désignée par H_∞. Au chapitre 3, nous décrivons le rôle que jouent le plan à l'infini et son homographie pour la reconstruction et le calibrage affines.

2.6 Quelques relations entre homographies, matrices fondamentales et matrices de projection

Nous résumons brièvement, dans cette section, quelques formules portant sur le calcul de la matrice fondamentale et les homographies de plans. Nous considérons deux caméras, avec des matrices des paramètres intrinsèques K_1 et K_2 et des matrices d'orientation et des vecteurs de position $\mathsf{R}_1, \mathsf{R}_2, \mathbf{t}_1$ et $\mathbf{t}_2$. Leurs matrices de projection sont donc $\mathsf{P}_i = \mathsf{K}_i\mathsf{R}_i(\ \mathtt{I}\ | - \mathbf{t}_i)$. Les centres de projection sont donnés par $\mathbf{C}_1^\mathsf{T} = (\mathbf{t}_1^\mathsf{T}, 1)$ et $\mathbf{C}_2^\mathsf{T} = (\mathbf{t}_2^\mathsf{T}, 1)$.

Les épipôles. L'épipôle dans une image est le projeté du centre de projection de l'autre vue. Les deux épipôles sont donc donnés par :

$$\mathbf{e}_1 \sim \mathsf{K}_1\mathsf{R}_1(\mathbf{t}_2 - \mathbf{t}_1) \quad (2.8)$$

$$\mathbf{e}_2 \sim \mathsf{K}_2\mathsf{R}_2(\mathbf{t}_1 - \mathbf{t}_2)\ . \quad (2.9)$$

L'homographie infinie. Les projections d'un point à l'infini $(\mathbf{Q}^\mathsf{T}, 0)^\mathsf{T}$ sont les points $\mathbf{q}_i \sim \mathsf{K}_i\mathsf{R}_i\mathbf{Q}$. Il en découle que l'homographie infinie entre la vue 1 et la vue 2 se calcule par :

$$\mathsf{H}_\infty \sim \mathsf{K}_2\mathsf{R}_2\mathsf{R}_1^\mathsf{T}\mathsf{K}_1^{-1}\ . \quad (2.10)$$

Notons que dans le cas d'une caméra effectuant une translation pure, en gardant ses paramètres intrinsèques constant, l'homographie infinie est alors l'identité.

Les homographies générales. L'homographie d'un plan quelconque $\Pi^\mathsf{T} = (\mathbf{n}^\mathsf{T}, -d)$ est donnée par :

$$\mathsf{H}_\pi \sim \mathsf{H}_\infty + \mathbf{e}_2 \frac{\mathbf{n}^\mathsf{T}}{\Pi^\mathsf{T}\mathbf{C}_1} \mathsf{R}_1^\mathsf{T}\mathsf{K}_1^{-1} \quad .$$

Au cas où le repère monde coïnciderait avec le repère de la première caméra, l'homographie deviendrait :

$$\mathsf{H}_\pi \sim \mathsf{H}_\infty + \mathbf{e}_2 \frac{\mathbf{n}^\mathsf{T}}{d} \mathsf{K}_1^{-1} \quad .$$

Il est facile de prouver que les deux épipôles se correspondent par toute homographie d'un plan, c'est-à-dire que $\mathbf{e}_2 \sim \mathsf{H}_\pi \mathbf{e}_1$ pour tout plan Π.

Les homographies pour des rotations pures. Considérons le cas particulier d'une caméra effectuant une rotation pure autour de son centre de projection. Dans ce cas, tous les plans ont la même homographie (sauf les plans contenant le centre optique, pour lesquels l'homographie n'est pas définie), et par conséquent toutes les homographies de plans sont égales à l'homographie infinie (voir l'équation (2.10)).

La matrice fondamentale et les homographies. La matrice fondamentale peut être déterminée par un épipôle et une homographie de plan quelconque par :

$$\mathsf{F} \sim [\mathbf{e}_2]_\wedge \mathsf{H}_\pi \quad . \tag{2.11}$$

L'interprétation géométrique de cette formule est simple : l'homographie H_π projette un point $\mathbf{q}_1$ sur le point $\mathbf{q}_2$ dans la deuxième image qui est la projection du même point 3D de Π. Puisque $\mathbf{q}_1$ et $\mathbf{q}_2$ sont des points image correspondants, $\mathbf{q}_2$ doit se trouver sur la droite épipolaire de $\mathbf{q}_1$ dans la deuxième image. Par conséquent, la droite $\langle \mathbf{e}_2\mathbf{q}_2 \rangle \sim [\mathbf{e}_2]_\wedge \mathsf{H}_\pi \mathbf{q}_1$ est cette droite épipolaire.

L'équation suivante établit une relation directe entre les homographies et la matrice fondamentale, qui peut servir à vérifier la cohérence d'une homographie estimée avec la géométrie épipolaire :

$$\mathsf{F}^\mathsf{T}\mathsf{H}_\pi + \mathsf{H}_\pi^\mathsf{T}\mathsf{F} = 0 \quad . \tag{2.12}$$

La matrice fondamentale et les matrices de projection. Nous donnons maintenant une formule pour calculer la matrice fondamentale à partir des matrices de projection des deux vues, données par $\mathsf{P}_1 = (\ \overline{\mathsf{P}}_1\ \ |\overline{\mathbf{p}}_1)$ et $\mathsf{P}_2 = (\ \overline{\mathsf{P}}_2\ \ |\overline{\mathbf{p}}_2)$. Le premier centre de projection est le point :

$$\mathbf{C}_1 = \begin{pmatrix} -\overline{\mathsf{P}}_1^{-1}\overline{\mathbf{p}}_1 \\ 1 \end{pmatrix} \quad .$$

Sa projection dans la deuxième image est le deuxième épipôle :

$$\mathbf{e}_2 \sim \overline{\mathsf{P}}_2\overline{\mathsf{P}}_1^{-1}\overline{\mathbf{p}}_1 - \overline{\mathbf{p}}_2 \quad .$$

L'homographie infinie est donnée par $\mathsf{H}_\infty \sim \overline{\mathsf{P}}_2\overline{\mathsf{P}}_1^{-1}$. Donc, d'après (2.11), nous obtenons pour la matrice fondamentale (cette formule n'est valable que si le centre de projection de P_1 n'est pas à l'infini) :

$$\mathsf{F} \sim [\overline{\mathsf{P}}_2\overline{\mathsf{P}}_1^{-1}\overline{\mathbf{p}}_1 - \overline{\mathbf{p}}_2]_\wedge \overline{\mathsf{P}}_2\overline{\mathsf{P}}_1^{-1} \quad . \tag{2.13}$$

Le chemin inverse, c'est-à-dire remonter de la matrice fondamentale aux matrices de projection, n'est évidemment pas entièrement défini. On peut pourtant donner l'ensemble des solutions. Les matrices de projection compatibles avec une matrice fondamentale F sont :

$$\begin{aligned} \mathsf{P}_1 &= (\ \mathtt{I} \ |\mathbf{0}) \\ \mathsf{P}_2 &= (\ \mathsf{H}_\pi \ |\mathbf{e}_2) \end{aligned}$$

et les matrices $\mathsf{P}_1\mathsf{T}, \mathsf{P}_2\mathsf{T}$, pour toute transformation $\mathsf{T}_{4\times4}$ non singulière. H_π est une homographie d'un plan quelconque.

2.7 Coniques, quadriques, cônes, ...

2.7.1 Coniques et quadriques

Une **quadrique** dans $\mathcal{P}^n$ est un ensemble de points satisfaisant une équation quadratique en leurs coordonnées homogènes. Dans cette thèse, nous ne considérons que des quadriques avec des coefficients réels. Toute quadrique peut être représentée par une matrice $(n+1) \times (n+1)$ symétrique.

Une **quadrique imaginaire** est une quadrique qui ne contient aucun point réel et une **quadrique propre** est une quadrique dont la matrice est non singulière. Les **coniques** sont les quadriques du plan projectif $\mathcal{P}^2$; dans la suite nous n'allons pas distinguer la conique de sa matrice. Une conique dans l'espace 3D – une **conique 3D** – est définie par son **plan support** et l'équation de la conique dans ce plan.

Toutes les coniques propres et imaginaires sont des coniques à centre [20] (le centre d'une conique est le pôle de la droite à l'infini). Les caractéristiques suivantes se réfèrent aux coniques imaginaires et propres. Leur forme normale euclidienne est la matrice diagonale des valeurs propres, définie à une permutation des valeurs propres près. Si les trois valeurs propres sont distinctes, la conique est une **ellipse imaginaire** et possède deux axes de symétrie. Les axes sont orthogonaux et passent par le centre de la conique. Si les valeurs propres ne sont pas distinctes, la conique est un **cercle imaginaire** et toutes les droites passant par le centre sont des axes de symétrie.

2.7.2 Cônes

Un cône (propre) est une quadrique de $\mathcal{P}^n$ de rang n. Nous nous restreignons, dans la suite, au cas de $\mathcal{P}^3$. Sauf mention contraire, nous comprenons par **cône** une quadrique de rang 3 dans $\mathcal{P}^3$, dont le **vertex** (le point singulier) ne se trouve *pas* sur le plan à l'infini. Un cône est défini de manière unique par son vertex et une section plane quelconque ne contenant pas le vertex, les sections planes d'un cône étant des coniques. Dans le chapitre 7, nous utilisons les cônes via la notion de **cône de projection** d'une conique 3D, c'est-à-dire le cône tracé par les rayons de projection des points sur la conique (voir la figure 2.10).

La forme normale euclidienne d'un cône est une matrice diagonale de ses valeurs propres, $\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, 0)$, avec des λ_i non nuls. Si tous les λ_i sont distincts, le cône est **elliptique** et possède trois **axes** mutuellement orthogonaux. Des rotations de $180°$ autour des axes laissent le cône globalement invariant. Si exactement deux des λ_i sont égaux, le cône est **circulaire** et invariant à des rotations de degré arbitraire autour de son **axe principal**. Un cône avec des valeurs propres non nulles identiques est un **cône isotropique**. Des rotations arbitraires autour du vertex laissent un cône isotropique globalement invariant. L'intersection d'un cône isotropique avec le plan à l'infini est la conique absolue (voir plus bas).

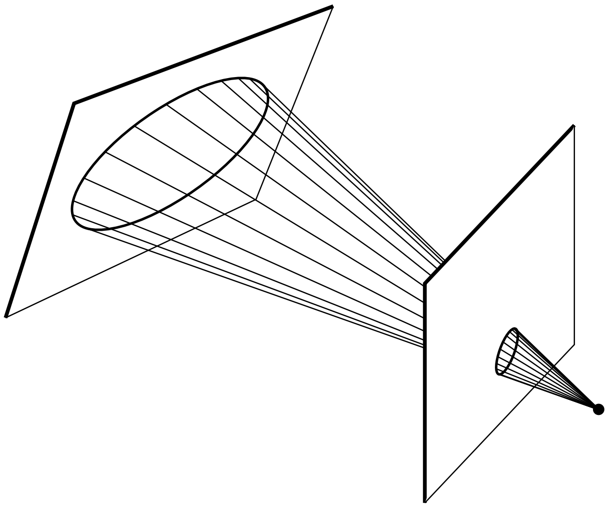


FIG. 2.10: *Cône de projection d'une conique 3D.*

2.7.3 Quadrique et conique absolues

La **quadrique absolue** de $\mathcal{P}^n$ est définie par les équations :

$$\sum_{i=1}^{n} X_i^2 = X_{n+1} = 0 \ .$$

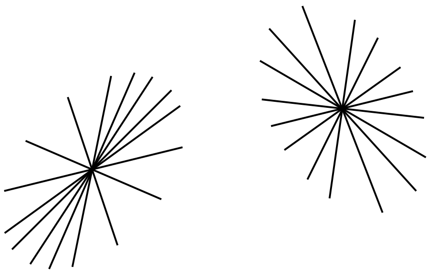
La **conique absolue** Ω_∞ est la quadrique absolue de $\mathcal{P}^3$ et celle du plan projectif $\mathcal{P}^2$ consiste en deux points imaginaires conjugués, connus sous le nom de **points cycliques**.

La quadrique absolue de $\mathcal{P}^n$ est une quadrique imaginaire sur l'hyperplan à l'infini et sa position définit la structure euclidienne de l'espace considéré ; par exemple, connaître la conique absolue nous donne la structure euclidienne de l'espace tridimensionnel.

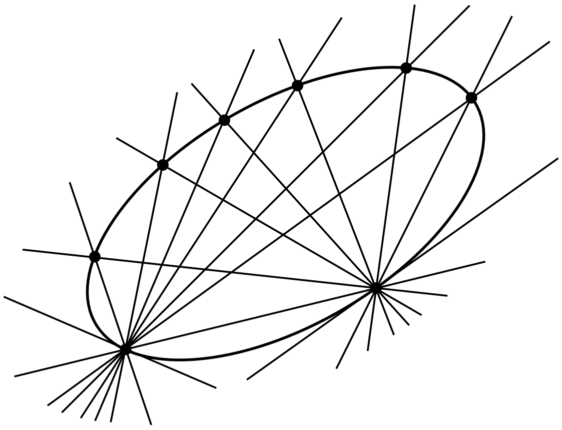
Le calibrage intrinsèque d'une caméra est équivalent à la détermination de l'image ω_∞ de Ω_∞ (ou de son dual ω^*_∞) [59, 123]. La relation $\omega^*_\infty \sim \mathsf{K}\mathsf{K}^\mathsf{T}$ permet de déterminer la matrice des paramètres intrinsèques K de manière unique par une décomposition de Cholesky.

2.7.4 La conique de Steiner

La notion de **conique de Steiner** se réfère à un principe de construction d'une conique : considérons deux faisceaux de droites, chacun paramétré projectivement. Les intersections des droites avec les mêmes coordonnées projectives dans leur faisceau forment une conique (voir la figure 2.11). Cette conique contient les points de base des deux faisceaux.



(a) Deux faisceaux de droites.



(b) Les points d'intersection des droites correspondantes forment une conique.

FIG. 2.11: *La conique de Steiner.*

3.

Stratification du calibrage et de la reconstruction

Dans ce chapitre, nous esquissons un scénario général de calibrage et de reconstruction, celui de l'ajustement de faisceaux. Le but est d'orienter les caméras (de les calibrer) et de déterminer des points 3D tels que les rayons de projection des points image correspondants se coupent simultanément aux points 3D. Si les seules données sont des points image, les résultats de l'ajustement de faisceaux sont un calibrage et une reconstruction projectifs. Afin d'obtenir des résultats plus riches en information, il faut introduire des connaissances sur la scène à reconstruire ou sur le calibrage des caméras. Nous décrivons plusieurs scénarios, selon le type des connaissances disponibles. En particulier, nous décrivons les cas de reconstruction et de calibrage projectifs, affines et euclidiens.

Le but de ce chapitre est de placer les chapitres suivants, qui traitent des problèmes de reconstruction projective et d'auto-calibrage, dans un cadre plus général. La distinction des niveaux projectif, affine et euclidien pour la reconstruction et le calibrage suit les idées de stratification présentées par Faugeras [59], Luong et Viéville [120] et Koenderink et Van Doorn [105].

3.1 Introduction

Considérons les caméras comme des instruments de mesure. Notre but est de les utiliser pour la perception tridimensionnelle et en particulier pour la détermination de la *structure* tridimensionnelle d'un objet ou d'une scène, ce que nous appelons aussi *reconstruction (tridimensionnelle)* de cet objet ou de cette scène. Les applications en sont nombreuses ; nous renvoyons au chapitre d'introduction pour des exemples. Outre la structure d'un objet on peut aussi s'intéresser au *mouvement* de celui-ci ou d'une caméra. Cette tâche est en effet complémentaire à la détermination de la structure de l'objet et quand nous parlons de reconstruction, nous sous-entendons généralement qu'elle consiste en la détermination de la structure *et* du mouvement.

Le scénario pour la reconstruction est le suivant : une ou plusieurs images d'une scène sont prises par autant de caméras. L'ensemble des caméras est appelé **système de caméras**. Les caméras ne sont pas nécessairement toutes différentes, i.e. une caméra peut prendre plusieurs images, tout en se déplaçant entre les prises d'image, mais cela ne fait pas de différence pour la suite.

Comme tout instrument de mesure, les caméras peuvent être calibrées. La notion classique de calibrage se réfère à la détermination des propriétés physiques qui sont intrinsèques aux caméras. Au sens de notre scénario de reconstruction, nous parlons de calibrage avec l'idée de calibrer le *système complet* de caméras, que nous considérons comme un seul instrument de mesure. Le calibrage du système de caméras consiste à établir à la fois les propriétés intrinsèques des caméras ainsi que les relations géométriques qui décrivent les positions de prise de vue. Nous allons voir que plusieurs niveaux de calibrage existent, qui donnent lieu à des reconstructions de différents niveaux d'information. Il est souhaitable d'obtenir une reconstruction euclidienne, c'est-à-dire qui contient des informations métriques sur la scène : des distances entre des points, des angles entre des segments de droite ou entre des plans... Nous appelons alors le niveau de calibrage nécessaire un calibrage euclidien ou métrique.

Si seulement une partie des paramètres du calibrage peut être déterminée, la structure 3D reconstruite sera éventuellement moins riche que métrique. Parmi les calibrages partiels, nous distinguons les calibrages affine et projectif, qui permettent d'obtenir respectivement une reconstruction affine ou projective de la scène. Ces reconstructions contiennent des informations moins riches que les reconstructions métriques, mais peuvent néanmoins s'avérer utiles. Une reconstruction affine préserve le parallélisme et permet de mesurer les rapports de distances entre des points sur des segments de droite parallèles. Une reconstruction projective préserve la collinéarité de points et on peut y mesurer les birapports de points collinéaires.

3.2 Reconstruction

Comment faire pour déterminer la structure 3D d'un objet ? Considérons d'abord le cas où nous disposons d'une seule image de l'objet. Il est connu qu'en général une seule image ne révèle pas la profondeur de la scène. Pourtant, des connaissances a priori sur l'objet permettent de combler ce manque. Par exemple, une personne peut estimer la distance entre elle et une porte, même avec un seul œil ouvert, tout en s'appuyant sur la connaissance qu'une porte mesure en général 2 mètres de haut. L'utilisation de telles connaissances nécessite que l'objet portant cette information puisse être identifié au préalable. Nous touchons ici à un autre vaste domaine de recherche en vision par ordinateur, celui de la *reconnaissance d'objets*. Il est hors de la portée de cette thèse d'approfondir ce sujet. Sauf mention contraire, nous supposerons dans la suite de ne disposer d'aucune connaissance a priori sur l'objet à reconstruire. La seule hypothèse que nous ferons sera sur la nature rigide des objets considérés.

Une seule image n'est donc pas suffisante pour reconstruire la structure 3D de la scène. Si nous rajoutons une deuxième image, prise d'un point de vue différent, la situation change. La reconstruction se fait souvent individuellement pour des *primitives* de la scène qui peuvent être extraites dans les images. Considérons par exemple un point d'intérêt (un coin par exemple) d'un objet 3D et ses deux projetés. Chacun des deux points image permet de limiter les positions possibles du point 3D à reconstruire comme étant les points sur le rayon de projection associé. La position du point 3D peut donc en général être déterminée comme étant l'intersection des deux rayons de projection (voir la figure 3.1).

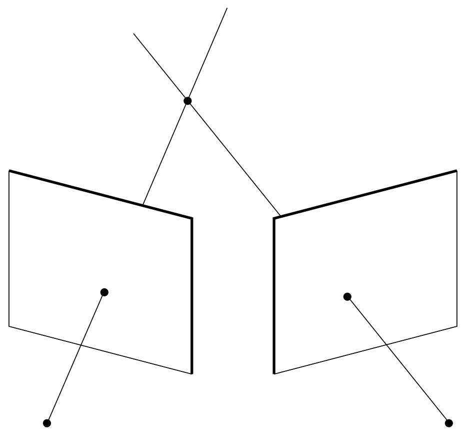


FIG. 3.1: *Reconstruction d'un point par intersection de rayons de projection.*

Ce procédé requiert la détermination préalable des primitives dans les images qui sont les projetés de la *même* primitive de l'espace 3D. Ceci est le problème de la *mise en correspondance* ou de l'*appariement* de primitives image. Nous supposons dans la suite que la mise en correspondance a été établie et nous nous concentrons sur la reconstruction proprement dite.

3.3 Calibrage

Une autre condition pour pouvoir établir la reconstruction par intersection de rayons de projection est qu'il faut connaître la géométrie du système de caméras. D'une part cela veut dire qu'il faut être capable de représenter les rayons de projection dans un repère 3D commun. Ceci requiert la connaissance de l'*orientation relative* des caméras entre elles. D'autre part il faut connaître, pour chacune des caméras, la relation géométrique entre la position des primitives dans le plan image et celle du centre de projection. Cette information seulement permet de calculer les rayons de projection des primitives. Cette relation géométrique est intrinsèque aux caméras en question et on l'appellera *orientation intrinsèque*. Elle peut être représentée à l'aide des *paramètres intrinsèques* des caméras (voir 2.2.2). L'orientation relative des caméras est, de manière analogue, appelée *orientation extrinsèque*. Nous utilisons souvent l'expression *calibrage intrinsèque* à la place d'orientation intrinsèque ou même tout court *calibrage*, s'il est clair qu'il s'agit du calibrage des paramètres intrinsèques. Au cas où nous considérons une seule caméra en mouvement, on parlera aussi de *mouvement* au lieu d'orientation extrinsèque.

Le calibrage de tout un système de caméras consiste en la détermination des orientations intrinsèques *et* extrinsèques de toutes les caméras. Une représentation convenable du calibrage d'un système de caméras est l'ensemble des matrices de projection, données par rapport à un repère du monde qui reste à définir. La matrice de projection d'une caméra contient ses orientations intrinsèque et extrinsèque (par rapport au repère du monde). S'il est nécessaire, on peut extraire ces informations de la matrice de projection (voir 2.2.2), mais sinon nous préférons d'utiliser directement la matrice elle-même pour représenter le calibrage.

3.4 Reconstruction et calibrage par ajustement de faisceaux

Dans les deux paragraphes précédents, nous avons décrit les problèmes de la reconstruction et du calibrage séparément. Dans la suite, nous allons les réunir dans une formulation plus générale. Rappelons que les données premières sont des primitives dans les images et que l'on dispose de l'appariement de ces primitives dans ces images. Les points permettant des expressions analytiques convenables, nous nous restreindrons dans la suite à ceux-ci comme primitives, bien que toutes les définitions soient valables pour n'importe quel type de primitive. Supposons que nous disposions des projetés de n points dans m images, $\mathbf{q}_{ip}$ étant la i^{e} image du p^{e} point [1].

Un problème combiné de reconstruction et de calibrage consiste alors à déterminer n points 3D $\mathbf{Q}_p$ et m matrices de projection P_i qui collent bien aux données, c'est-à-dire les points image :

$$\forall i = 1, ..., m \; \forall p = 1, ..., n : \mathsf{P}_i \mathbf{Q}_p \sim \mathbf{q}_{ip} \quad . \tag{3.1}$$

L'idée géométrique liée à cette équation est qu'on tente simultanément d'orienter les caméras et de déterminer des points 3D tels que les rayons de projection des points image correspondants se coupent aux points 3D correspondants. On parle donc aussi d'**ajustement de faisceaux** (de rayons de projection) ou **bundle adjustment** en anglais.

Nous avons jusqu'ici considéré les points image comme seules données. Comme nous allons le voir plus bas, cette information est suffisante pour obtenir le calibrage et la structure 3D, mais seulement à un niveau projectif. Pour obtenir une structure 3D et un calibrage plus riches en informations (par exemple au niveau affine ou euclidien), nous devons introduire des données supplémentaires. Celles-ci peuvent être de plusieurs types :

- ▷ connaissances sur la structure 3D de la scène ;
- ▷ connaissances sur l'orientation intrinsèque des caméras ;
- ▷ connaissances sur l'orientation extrinsèque des caméras.

Par exemple, si la structure *euclidienne* de la scène est connue et les orientations intrinsèques et extrinsèques sont à déterminer, nous retrouvons le problème classique du calibrage (euclidien) de caméras.

Lors d'un ajustement de faisceaux, les informations sur la structure 3D de la scène et l'orientation des caméras sont introduites sous la forme de contraintes que la solution doit respecter pour être valable. Nous étendons alors la définition de l'ajustement de faisceaux faite ci-dessus : l'ajustement de faisceaux consiste à déterminer des points 3D $\mathbf{Q}_p$ et des matrices de projection P_i tels que l'équation (3.1) et toutes les contraintes sur la structure 3D de la scène et l'orientation des caméras soient satisfaites [2].

Une contrainte que nous imposons partout, même si ce n'est pas dit explicitement, est que la scène est supposée être rigide durant la prise des images. La présence ou absence d'autres contraintes de

1. Nous supposons ici que tous les points sont visibles dans toutes les vues. Cette contrainte n'est pas nécessaire, mais elle permet une formulation plus simple.

2. Nous supposons que les contraintes ne sont pas contradictoires.

types énumérés plus haut donne lieu à de différents scénarios de reconstruction et de calibrage. Nous allons décrire quelques-uns de ces scénarios en section 3.6.

3.5 Stratification du calibrage et de la reconstruction

Dans l'introduction de ce chapitre nous parlons de plusieurs niveaux de calibrage, qui donnent lieu à de différents niveaux de reconstruction. Une façon de voir les choses est de considérer le repère par rapport auquel les matrices de projection sont données. S'il s'agit d'un repère euclidien, la reconstruction sera euclidienne. Souvent la position absolue du repère n'est pas importante, pourvu qu'il reste un repère euclidien – la structure euclidienne d'un objet est indépendante du repère dans laquelle elle est représentée. Si le repère d'une reconstruction peut être choisi arbitrairement, nous disons que la reconstruction est à une transformation euclidienne (inconnue) près.

Si le repère est affine (respectivement projectif), la reconstruction sera du niveau analogue, c'est-à-dire affine (respectivement projective). Ceci veut dire qu'il y a une transformation affine (respectivement projective) entre la vraie structure et la reconstruction. Bien que cette transformation ne préserve en général pas les propriétés métriques de la scène, les propriétés affines (respectivement projectives) sont préservées et peuvent donc être extraites de la reconstruction. Comme dans le cas euclidien, le repère affine (respectivement projectif) peut normalement être choisi de manière arbitraire, et la reconstruction sera définie à une transformation affine (respectivement projective) près.

Si les matrices de projection sont données par rapport à un repère euclidien (respectivement affine ou projectif), nous disons que nous disposons d'un calibrage euclidien (respectivement affine ou projectif) du système de caméras. Maintenant se pose une question très importante : quelles sont les informations nécessaires pour obtenir un calibrage d'un type donné ?

D'abord, nous nous intéressons au cas où *aucune* information supplémentaire n'est disponible.

3.5.1 Calibrage et reconstruction projectifs

Faugeras et indépendamment Hartley *et al.* ont montré que même sans aucune connaissance sur la structure de la scène ou sur l'orientation des caméras (extrinsèque et intrinsèque), le calibrage projectif[3] d'un système de caméras est possible [57, 72] ! Or, la structure 3D de la scène peut être reconstruite à une transformation projective inconnue près. La reconstruction projective révèle des caractéristiques de la scène, qui sont invariantes aux transformations projectives, comme la collinéarité et les birapports.

Il est donc possible de trouver des matrices de projection P_i et des points 3D $\mathbf{Q}_p$ tels que :

$$\mathbf{q}_{ip} \sim \mathsf{P}_i \mathbf{Q}_p$$

pour tout i et p. Si l'on ne dispose pas de connaissances supplémentaires sur les P_i et les $\mathbf{Q}_p$, on ne peut pas obtenir de reconstruction et de calibrage plus riches que projectifs : pour toute transformation projective non singulière $\mathsf{T}_{4\times 4}$, les matrices de projection $(\mathsf{P}_i\mathsf{T}^{-1})$ et les points 3D $(\mathsf{T}\mathbf{Q}_p)$ sont également des solutions possibles pour le calibrage et la reconstruction, puisque :

$$(\mathsf{P}_i\mathsf{T}^{-1})(\mathsf{T}\mathbf{Q}_p) = \mathsf{P}_i\mathbf{Q}_p \sim \mathbf{q}_{ip} \ .$$

Considérons maintenant le cas de deux images. Une représentation minimale du calibrage projectif d'un système de deux caméras est donnée par sa géométrie épipolaire : sa connaissance permet de

3. Souvent aussi appelé *calibrage faible*.

calculer toutes les matrices de projection représentant un calibrage projectif du système. Une possibilité pour ce faire est la suivante. Soit la géométrie épipolaire représentée par la matrice fondamentale, F. Nous définissons des matrices de projection par :

$$\begin{aligned} \mathsf{P}_1 &= \left(\begin{array}{c|c} \mathtt{I} & \mathbf{0} \end{array} \right) \\ \mathsf{P}_2 &= \left(\begin{array}{c|c} \mathsf{H}_\pi & \mathbf{e}_2 \end{array} \right) , \end{aligned}$$

où H_π est une homographie quelconque qui doit être cohérente avec F (cf. l'équation (2.12) dans la section 2.6). Les matrices de projection de tous les calibrages projectifs possibles peuvent être obtenues en multipliant P_1 et P_2 par des transformations projectives non singulières.

3.5.2 Calibrage et reconstruction affines

Nous supposons qu'un calibrage projectif est donné, c'est-à-dire que nous avons des matrices de projection dans un repère projectif du monde. Pour obtenir un calibrage affine, il faut effectuer un changement de repère approprié, qui transforme le repère projectif en repère affine. La difficulté est de trouver un repère qui soit affine. Pour ce faire, il faut identifier le plan à l'infini qui, dans le repère projectif, n'est pas en position canonique (qui est représentée par les coordonnées de plan $(0,0,0,1)^{\mathsf{T}}$). Appliquer une transformation projective quelconque qui projette le plan à l'infini sur cette position canonique revient ensuite à effectuer le changement de repère désiré.

La connaissance du plan à l'infini permet de déterminer les caractéristiques affines de la scène et réciproquement, la connaissance de suffisamment de caractéristiques affines permet de déterminer le plan à l'infini. Parmi les caractéristiques affines se trouvent les rapports de distances de points sur des segments de droite parallèles et le parallélisme de droites et de plans. Considérons par exemple trois paires de droites parallèles. Chaque paire se coupe en un point à l'infini. Les trois points à l'infini, s'ils ne sont pas collinéaires, définissent le plan à l'infini [167]. D'autres types de connaissances qui permettent de trouver le plan à l'infini seront décrits en section 3.6.

Comme pour le cas projectif, nous discutons plus en détail le cas de deux vues. Étant donné la géométrie épipolaire, une représentation convenable du calibrage affine d'un système de deux caméras est donnée par l'homographie infinie H_∞ : sa connaissance permet de calculer toutes les matrices de projection représentant un calibrage affine du système. Définissons des matrices de projection par :

$$\begin{aligned} \mathsf{P}_1 &= \left(\begin{array}{c|c} \mathtt{I} & \mathbf{0} \end{array} \right) \\ \mathsf{P}_2 &= \left(\begin{array}{c|c} \mathsf{H}_\infty & \mathbf{e}_2 \end{array} \right) . \end{aligned}$$

Les matrices de projection de tous les calibrages affines possibles peuvent être obtenues en multipliant P_1 et P_2 par des transformations affines non singulières.

3.5.3 Calibrage et reconstruction euclidiens

Nous supposons qu'un calibrage affine est donné, c'est-à-dire que nous avons des matrices de projection dans un repère affine du monde. Pour obtenir un calibrage euclidien, il faut effectuer un changement de repère approprié, qui transforme le repère affine en repère euclidien. Pour trouver un repère euclidien, il faut identifier la conique absolue qui, dans le repère affine, n'est pas en position canonique (cf. la section 2.7.3). Appliquer une transformation affine quelconque qui projette la conique absolue sur sa position canonique revient ensuite à effectuer le changement de repère désiré.

Il y a une petite imprécision dans les propos précédents : la conique absolue ne spécifie pas complètement la géométrie euclidienne, mais seulement à un changement d'échelle uniforme près. Plus précisément, la conique absolue permet de mesurer les angles, mais les distances ne peuvent être calculées qu'à un facteur d'échelle près. Si besoin est de calculer des distances absolues, il suffit de fixer cette échelle, ce qui est possible si une seule distance absolue est connue (par exemple, la distance entre deux points reconstruits ou entre deux caméras). Dans la suite, quand nous parlons de reconstructions euclidiennes, nous sous-entendons que celles-ci sont en fait définies à une transformation de similitude près.

Comme caractéristiques euclidiennes, nous avons déjà mentionné les angles et les distances. D'autres types de connaissances qui permettent de trouver la conique absolue sont décrits en section 3.6.

Comme pour les cas projectif et affine, nous discutons le cas de deux vues. Étant donné la géométrie épipolaire et l'homographie H_∞, des représentations convenables du calibrage euclidien sont données par les deux images $\omega_{\infty 1}$ et $\omega_{\infty 2}$ de la conique absolue Ω_∞ ou par les deux matrices des paramètres intrinsèques K_1 et K_2. Ces représentations sont équivalentes ; elles sont reliées entre elles par les relations $\omega^*_{\infty 1} \sim \mathsf{K}_1\mathsf{K}_1^\mathsf{T}$ et $\omega^*_{\infty 2} \sim \mathsf{K}_2\mathsf{K}_2^\mathsf{T}$ (cf. 2.7.3). Les images de la conique absolue peuvent donc être calculées à partir des matrices des paramètres intrinsèques et réciproquement (par décomposition de Cholesky des matrices définies positives $\omega_{\infty i}$). Nous décrivons maintenant comment calculer toutes les matrices de projection représentant un calibrage euclidien du système. Définissons des matrices de projection par :

$$\begin{aligned} \mathsf{P}_1 &= \left(\begin{array}{c|c} \mathsf{K}_1 & \mathbf{0} \end{array} \right) \\ \mathsf{P}_2 &= \left(\begin{array}{c|c} \mathsf{H}_\infty \mathsf{K}_1 & \mathbf{e}_2 \end{array} \right) \ . \end{aligned}$$

Les matrices de projection de tous les calibrages euclidiens possibles peuvent être obtenues en multipliant P_1 et P_2 par des transformations euclidiennes.

Pour la construction de la matrice P_2 nous avons profité du fait que l'homographie infinie est donnée par $\mathsf{H}_\infty \sim \mathsf{K}_2\mathsf{R}\mathsf{K}_1^{-1}$ où R est la rotation relative entre les deux caméras. Ainsi, la sous-matrice $\mathsf{H}_\infty\mathsf{K}_1$ de P_2 est égale à $\mathsf{K}_2\mathsf{R}$, comme on s'y attend.

3.6 Scénarios de reconstruction et de calibrage

Nous avons vu que sans aucune connaissance sur la scène à reconstruire et sur le calibrage des caméras, le calibrage projectif des caméras et la reconstruction projective de la scène sont possibles. Pour obtenir une reconstruction et un calibrage plus riches en information, il faut introduire des connaissances (cf. la discussion dans les sections précédentes). Les types et le nombre de ces connaissances déterminent le niveau de reconstruction et de calibrage qu'il est possible d'atteindre. Dans la suite, nous décrirons plusieurs scénarios typiques en vision par ordinateur tridimensionnelle ou stéréoscopique. Nous les distinguons selon le niveau géométrique des résultats attendus (projectif, affine ou euclidien) et selon les connaissances disponibles a priori. Pour les scénarios considérés plus en détail dans cette thèse nous indiquons les liens vers les paragraphes correspondants.

3.6.1 Reconstruction et calibrage projectifs

C'est le scénario où aucune connaissance sur la scène à reconstruire et sur le calibrage des caméras n'est disponible ou utilisée. Deux chapitres de cette thèse sont principalement dédiés à ce domaine. Dans le chapitre 4, nous considérons le problème de la reconstruction à partir de deux images, étant donné le calibrage projectif des caméras. Nous proposons des méthodes dont l'utilisation n'est pas bornée à la reconstruction projective mais qui sont particulièrement bien adaptées à ce cas de reconstruction. Dans le chapitre 5, nous considérons le problème de la reconstruction et du calibrage projectifs à partir de plusieurs vues.

3.6.2 Reconstruction et calibrage affines

Nous décrivons plusieurs types de connaissances qui permettent de s'élever au niveau affine.

3.6.2.1 Utilisation de connaissances sur le calibrage

Caméras affines. Le seul fait que les images aient été obtenues par des projections affines suffit à déterminer le plan à l'infini (à partir de deux images). Ce résultat est facile à démontrer : nous supposons qu'une reconstruction projective et un calibrage projectif ont déjà été obtenus. Pour identifier le plan à l'infini dans le repère projectif, il suffit de trouver trois points à l'infini. Les centres de projection de caméras affines étant des points à l'infini, nous disposons donc déjà de deux points. Un ou plusieurs autres points à l'infini peuvent être construits de la manière suivante : nous savons que les droites à l'infini des plans image sont des droites épipolaires correspondantes et que leur plan épipolaire est le plan à l'infini. Toute paire de points à l'infini dans les deux plans image est la projection d'un point 3D du plan à l'infini. Il suffit alors de prendre une telle paire et de reconstruire le point 3D pour obtenir le troisième point à l'infini qui permet d'identifier le plan à l'infini (cf. la figure 3.2).

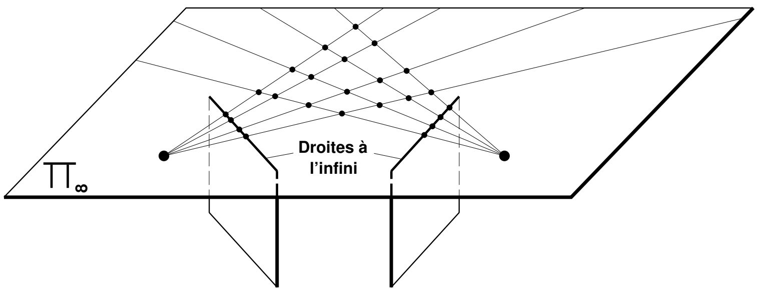


FIG. 3.2: *Reconstruction affine à partir de vues affines. Une reconstruction projective fournit deux points à l'infini via les deux centres de projection. Il suffit alors de prendre une quelconque paire de points à l'infini dans les deux images et de reconstruire le point 3D pour obtenir le troisième point qui permet de déterminer le plan à l'infini.*

Translations pures. Deux images prises par une caméra effectuant une translation pure sans changer ses paramètres intrinsèques permettent de trouver la structure affine de la scène et le calibrage affine [131]. Une façon de voir ceci est de constater que l'homographie infinie pour ce type de mouvement est l'identité (cf. l'équation (2.10)). Or, après la détermination de la géométrie épipolaire (l'épipôle suffirait), nous pouvons construire des matrices de projection représentant le calibrage affine (voir 3.5.2). La reconstruction avec ce calibrage sera donc affine.

Paramètres intrinsèques constants. Dans ce cas, les homographies infinies entre des paires de vues ont une forme particulière – elles sont conjuguées à des matrices orthogonales (cf. l'équation (2.10)). Ceci implique que les valeurs propres de chaque homographie infinie ont le même module. Cette contrainte est appelée la *contrainte du module* [120, 150, 153] et peut être utilisée pour identifier le plan à l'infini à partir de 3 vues. Nous considérons cette contrainte plus en détail en 6.2.1.4 et 7.10.3.

Système stéréo. Plusieurs travaux considèrent des systèmes de deux caméras qui se déplacent ensemble et sans changer leurs paramètres intrinsèques. Quan a été le premier à parler de calibrage affine [161] : puisque l'orientation relative entre les caméras ne change pas lors de déplacements du système, il suffit de calculer une fois pour toutes l'homographie infinie des deux caméras. Ensuite, même après un déplacement arbitraire du système, on pourra établir une reconstruction affine de la scène. Dans [190, 191], nous décrivons des méthodes de calibrage affine qui sont basées sur un déplacement translationnel du système des caméras. Beardsley et Zisserman, dans [15], se basent sur un mouvement planaire des caméras (translation dans un plan et rotation autour d'axes perpendiculaires à ce plan). Zisserman *et al.* ont finalement montré que le calibrage affine est possible même avec un mouvement général du système stéréo [242]. Cette dernière approche est décrite un peu plus en détail en section 6.2.9.

3.6.2.2 Utilisation de connaissances sur la scène

La connaissance de suffisamment de propriétés affines de la scène permet de déterminer le plan à l'infini. Au 3.5.2, nous avons déjà évoqué le cas de trois paires de droites parallèles dans la scène. D'autres possibilités sont l'utilisation de rapports de distance entre des points collinéaires et la connaissance du parallélisme de cercles [191].

3.6.3 Reconstruction et calibrage euclidiens

3.6.3.1 Utilisation de connaissances sur le calibrage

Estimation du mouvement basée sur le calibrage intrinsèque. C'est le scénario classique d'estimation du mouvement. Au 2.3.1, nous décrivons le cas de deux images.

Calibrage intrinsèque basé sur la connaissance du mouvement. La configuration typique est une caméra montée sur un bras de robot, dont on peut mesurer le mouvement ou, si besoin est, effectuer des mouvements contrôlés [95, 97]. Le calibrage de la caméra se base sur l'hypothèse que sa position relative au bras de robot reste fixe. En section 6.2.5, nous décrivons quelques méthodes de calibrage intrinsèque d'une caméra dont le mouvement est, du moins partiellement, connu.

Calibrage basé sur des mouvements particuliers. Il existe des méthodes de calibrage qui ne nécessitent pas la connaissance complète du mouvement (angles de rotation ou distances de translation), mais seulement la connaissance du *type* de mouvement. Par exemple, le calibrage intrinsèque d'une caméra est possible si elle effectue des rotations pures autour du centre de projection [82]. Au chapitre 6, nous décrivons ce scénario et quelques autres plus en détail.

Auto-calibrage. Le scénario classique d'auto-calibrage concerne une caméra en mouvement, mais sans changement des paramètres intrinsèques. L'invariance des paramètres intrinsèques (et ainsi de l'image de la conique absolue) et la rigidité de la scène sont suffisantes pour déterminer le calibrage (intrinsèque et extrinsèque) et la structure euclidiens : la conique absolue peut être déterminée comme étant la seule conique dans la scène dont les projetées sont identiques dans toutes les images. Différentes méthodes d'auto-calibrage qui existent dans la littérature sont décrites au chapitre 6. Ce chapitre contient aussi une description de méthodes d'auto-calibrage pour des systèmes stéréo.

Au lieu d'auto-calibrage, nous parlons aussi de *reconstruction euclidienne non calibrée*, puisque souvent la reconstruction est le but et le calibrage joue juste un rôle intermédiaire.

3.6.3.2 Utilisation de connaissances sur la scène

Nous trouvons ici le scénario classique de calibrage de caméras qui est basé sur des images de « mires de calibrage ». Il s'agit typiquement d'objets avec des « cibles » facilement détectables dans les images. Les coordonnées tridimensionnelles de ces cibles sont connues (par rapport à un repère attaché à la mire) avec une haute précision. Des méthodes de calibrage qui utilisent des mires sont décrites par exemple dans [65, 222].

En photogrammétrie et surtout pour la reconstruction à partir d'images aériennes, l'utilisation de points de contrôle est courante [178] : les coordonnées de quelques points de la scène sont mesurés, par rapport à un repère donné (manuellement ou à l'aide d'autres capteurs que des caméras). Cette information sert à initialiser le calibrage des caméras (souvent le calibrage intrinsèque est déjà donné). Ensuite, les autres points de la scène peuvent être reconstruits.

D'autres connaissances euclidiennes sur la scène permettent de transformer une reconstruction projective en une euclidienne ou de calibrer des paramètres intrinsèques. Boufama *et al.* [22] prennent en compte des distances entre des points 3D et des angles droits pour transformer une reconstruction projective en une euclidienne. Des images de sphères peuvent être utilisées pour déterminer le rapport d'échelle τ [146]. Une méthode appliquée surtout par les photogrammètres permet de calculer des distorsions optiques à partir d'images de segments de droite (« plumb lines ») [29]. Dans [191], nous décrivons comment calibrer une caméra à partir de deux images de trois paires de cercles parallèles.

3.7 Prise en compte du bruit

Lors de la définition du problème de la reconstruction et du calibrage par ajustement de faisceaux en section 3.4, nous avons négligé les effets du bruit sur les positions des primitives image. En outre, les connaissances qui sont disponibles sur la structure de la scène ou sur le calibrage sont sujet à une certaine incertitude (par exemple, les coordonnées des points caractéristiques d'une mire de calibrage sont seulement connues avec une précision finie). La présence de bruit et de connaissances incertaines fait qu'il y a en général pas de solution exacte pour l'ajustement de faisceaux : et l'équation (3.1) et les contraintes ne pourront être satisfaites exactement ; il restera toujours une erreur de résidu. On tente alors de minimiser cette erreur afin d'obtenir une reconstruction et un calibrage optimaux.

La question d'optimalité dépend du choix d'un critère d'optimisation. Nous décrivons dans la suite le critère qui est généralement adopté. Il s'agit du **critère des erreurs de reprojection** dont le but est de minimiser la somme des carrés des distances entre les points image mesurés (les données) et des points obtenus par reprojection de la reconstruction :

$$\{\widehat{\mathsf{P}}_1, ..., \widehat{\mathsf{P}}_m, \widehat{\mathbf{Q}}_1, ..., \widehat{\mathbf{Q}}_n\} = \text{argmin} \sum_{i,p} d(\mathbf{q}_{ip}, \mathsf{P}_i \mathbf{Q}_p)^2 \quad , \tag{3.2}$$

où $d(*, *)$ désigne la distance euclidienne.

Ce critère n'est pas choisi au hasard : si les coordonnées des points image sont affectées par un bruit gaussien (voir 2.1), ce critère donne la solution au maximum de vraisemblance [178].

3.8 Résumé et liens vers les chapitres suivants

Nous avons brièvement décrit les principes de la stratification de la reconstruction et du calibrage dans un formalisme d'ajustement de faisceaux. Les niveaux projectif, affine et euclidien ont été passés en revue et nous avons décrit pour chaque niveau quel type d'information est nécessaire pour l'atteindre. L'appariement de primitives image suffit pour la reconstruction et pour le calibrage projectifs. Les deux chapitres suivants sont principalement dédiés au développement de méthodes numériques pour cette tâche.

Ensuite, nous nous intéressons aux problèmes de la reconstruction euclidienne non calibrée et de l'auto-calibrage. Nous considérons surtout le cas où aucune information sur la structure de la scène ou sur le mouvement des caméras n'est disponible. Les seules connaissances sont des contraintes sur les paramètres intrinsèques telles que leur invariance pour une caméra en mouvement ou le fait que les pixels sont rectangulaires. Ces cas et quelques autres sont décrits dans le chapitre 6.

Finalement, au chapitre 7, nous décrivons des mouvements de caméra qui causent des dégénérescences pour la reconstruction euclidienne non calibrée et l'auto-calibrage.

4.

Triangulation – reconstruction à partir de 2 vues

Dans ce chapitre, nous considérons le problème de la reconstruction de points à partir de deux vues. D'abord, nous passons en revue des méthodes classiques de reconstruction à partir de vues calibrées, donc des méthodes de reconstruction euclidienne. Ensuite, nous nous concentrons sur le cas de la reconstruction projective, à partir de vues non calibrées et de la reconstruction affine, étant donné un calibrage affine. Nous montrons que les méthodes « calibrées » sont inappropriées dans ces cas. La partie principale du chapitre est la description d'une méthode de reconstruction, applicable aux cas projectif, affine et euclidien, et qui fournit le résultat optimal, d'après le critère de minimisation des erreurs de reprojection.

Pour cette méthode, nous décrivons des simplifications qui sont possibles pour certains types de mouvement, impliquant des formes particulières de la matrice fondamentale. Nous présentons également d'autres méthodes et discutons leur applicabilité aux différents cas de reconstruction. Les résultats d'expériences évaluant toutes ces méthodes sont donnés en fin de chapitre.

Les résultats de ce chapitre ont été obtenus en grande partie en collaboration avec Richard Hartley. Une partie des résultats a été publiée dans [73, 74, 75].

4.1 Le problème de la triangulation

Nous supposons qu'un point $\mathbf{Q}$ de $\mathcal{R}^3$ est visible dans deux vues, dont les matrices de projection P_1 et P_2 sont connues. Soient $\mathbf{q}_1$ et $\mathbf{q}_2$ les projections de $\mathbf{Q}$ dans les deux vues. Il est facile de déterminer les rayons de projection correspondant aux deux points image. Le **problème de la triangulation** consiste en la détermination de l'intersection des deux rayons de projection. À première vue, ce problème semble être trivial puisque l'intersection de deux droites ne représente aucune difficulté. Pourtant, en présence de bruit dans les coordonnées des points image, les rayons de projection ne se coupent généralement pas, et il devient alors nécessaire de trouver une meilleure solution, sous l'hypothèse d'un certain modèle de bruit. La méthode classique consiste à choisir le point minimisant la somme des distances aux rayons de projection non incidents – il s'agit du point milieu de la perpendiculaire commune aux rayons. Cette solution s'appuie sur des notions euclidienne (perpendicularité) et affine (point de milieu) et n'a donc pas de sens si nous considérons la reconstruction projective ou même affine (pour la perpendicularité) où le monde 3D n'est connu qu'à une transformation projective ou affine près. Nous proposons dans la suite l'utilisation d'un autre critère de minimisation, qui est basé sur des distances *dans les images*. Ceci a l'avantage qu'on combat le bruit là où il se manifeste – dans la position des points image – et que le critère est valide pour tous les niveaux de reconstruction – euclidien, affine et projectif.

Organisation de ce chapitre. Nous introduisons d'abord la notion d'invariance de méthodes de triangulation par rapport à des transformations de l'espace 3D d'un groupe donné. Ensuite, dans la section 4.3 nous rappelons brièvement les méthodes classiques de reconstruction à partir de deux vues. Dans la section 4.4, nous proposons un critère de minimisation dans les images. Une méthode originale de triangulation qui résout ce critère de manière optimale est développée en 4.5. Ensuite, nous étudions des simplifications de la complexité algébrique de notre méthode dues à des mouvements particuliers entre les deux vues. En 4.7, nous décrivons d'autres méthodes, quelques-unes originales, qui ont été comparées avec la méthode optimale et nous présentons des résultats d'expériences en 4.8. Le chapitre est clos par un ensemble de conclusions.

4.2 Invariance des méthodes de triangulation

Les années passées, un intérêt considérable a été porté au sujet des reconstructions projective ou affine [57, 72, 104, 130, 157, 172, 173], où la scène est reconstruite à une transformation inconnue du groupe donné près. Généralement, dans une telle situation, au lieu de connaître les « vraies » matrices de projection P_1 et P_2, on dispose de matrices $\mathsf{P}_1\mathsf{T}^{-1}$ et $\mathsf{P}_2\mathsf{T}^{-1}$ où T est une transformation inconnue du groupe de transformations considéré.

Une caractéristique désirable d'une méthode de reconstruction est son invariance aux transformations du groupe considéré. Soit τ une méthode de triangulation, qui reconstruit un point 3D $\mathbf{Q}$ à partir de ses projections $\mathbf{q}_1$ et $\mathbf{q}_2$ et des matrices de projection correspondantes, P_1 et P_2. Nous écrivons ceci comme :

$$\mathbf{Q} = \tau(\mathbf{q}_1, \mathbf{q}_2, \mathsf{P}_1, \mathsf{P}_2) \ .$$

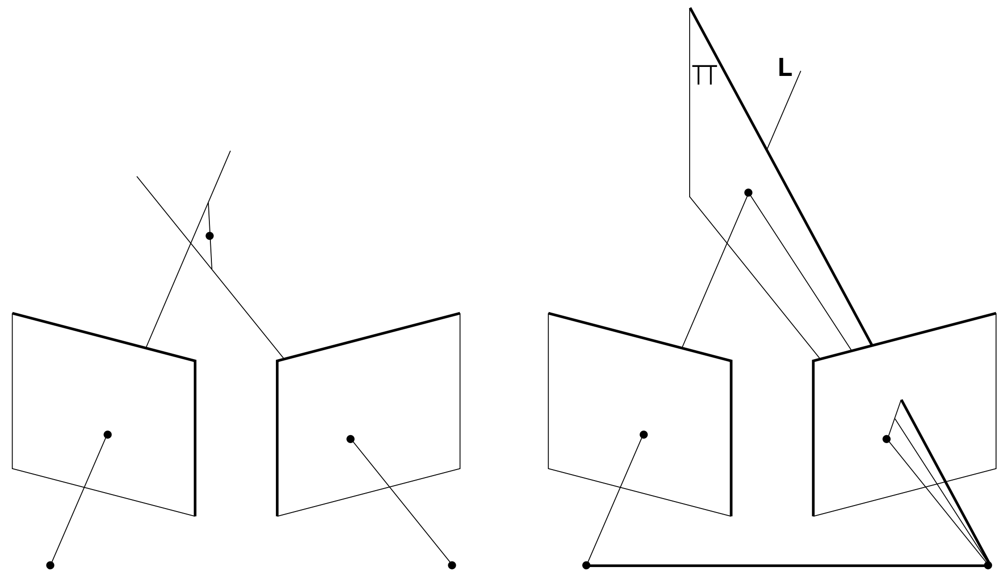
La méthode de triangulation est dite **invariante à une transformation** T si :

$$\tau(\mathbf{q}_1, \mathbf{q}_2, \mathsf{P}_1, \mathsf{P}_2) = \mathsf{T}^{-1}\tau(\mathbf{q}_1, \mathbf{q}_2, \mathsf{P}_1\mathsf{T}^{-1}, \mathsf{P}_2\mathsf{T}^{-1}) \ .$$

Ceci veut dire qu'un changement de repère (transformation des matrices de projection) résulte en le point 3D, transformé de manière analogue. Si nous nous trouvons dans le contexte de reconstruction projective ou affine, c'est-à-dire si les matrices de projection sont connues à une transformation projective ou affine près, il est donc clairement recommandé d'utiliser une méthode de triangulation qui est invariante aux transformations projectives respectivement affines.

4.3 La méthode du point milieu et d'autres approches classiques

Les travaux classiques de reconstruction à partir de 2 vues d'une scène s'inscrivent dans un contexte de stéréo-vision où les caméras sont supposées être calibrées. Le problème principal considéré dans ces travaux est généralement la détermination du mouvement entre les deux caméras (souvent sous la forme d'une *ego-motion estimation* ou *orientation relative*) [1, 3, 25, 30, 55, 64, 103, 112, 147, 158, 182, 194, 195, 212, 221, 223, 224, 230, 240]. La partie de reconstruction proprement dite, la triangulation, ne reçoit généralement que peu d'attention : la méthode du POINT MILIEU est couramment utilisée, c'est-à-dire qu'on choisit le point milieu de la perpendiculaire commune aux deux rayons de projection [13, 35, 205, 220, 223, 229].



(a) Le point milieu de la perpendiculaire commune aux rayons de projection.

(b) Une autre possibilité proposée par Toscani et Faugeras (voir le texte).

FIG. 4.1: *Deux méthodes classiques pour la triangulation.*

Toscani et Faugeras proposent deux solutions alternatives [206]. La première correspond en fait à notre formulation de la triangulation (voir 4.4), c'est-à-dire le point reconstruit est celui qui minimise l'erreur de reprojection. Ceci a été appliqué aussi par nombre d'autres chercheurs, pas dans le sens de la triangulation des points 3D, mais plutôt dans une formulation du style « ajustement de faisceaux » (bundle adjustment), avec le but d'estimer et les paramètres du mouvement et les points 3D. Cette estimation est effectuée par des méthodes itératives, tandis que nous donnons dans la suite une solution directe pour le sous-problème de la triangulation. La deuxième solution alternative proposée dans [206] a l'air moins intuitive que le choix du point milieu, mais des résultats plus stables ont été rapportés. Le point reconstruit est choisi comme étant l'intersection du rayon de projection $\mathbf{L}$ de l'un des points image avec le plan Π contenant l'autre rayon de projection et qui est perpendiculaire au plan épipolaire de $\mathbf{L}$ (voir la figure 4.1 (b)).

Dans leurs travaux, Weng *et al.* adoptent en général le point milieu comme point reconstruit [220, 223]. Dans [221], page 873, une petite discussion est donnée sur la solution directe du critère de reprojection, ce qui est dans la même lignée que notre approche. Pourtant, il est supposé que le meilleur point se trouve sur la perpendiculaire commune des rayons de projection des points image mesurés. Ceci n'est en général pas correct. Sous cette hypothèse, le point minimisant l'erreur de reprojection peut être déterminé en résolvant un polynôme quartique.

Dans le cas d'une reconstruction projective ou affine, les matrices de projection sont données par rapport à un repère projectif ou affine. Ici, les concepts euclidiens, comme les angles et donc la perpendicularité, n'ont pas de sens. Quant à la reconstruction projective, même les notions affines comme le rapport de distances entre points collinéaires et donc le concept de milieu, perdent leur valeur. Or, toutes les méthodes citées ci-dessus sont basées sur les notions de perpendicularité, ou, pour la méthode du POINT MILIEU, de milieu. Ces méthodes ne sont donc ni projectivement ni affinement invariantes et, comme nos expériences le montrent, sont en effet un mauvais choix pour effectuer des reconstructions projectives et même affines.

Des méthodes de reconstruction projective à partir de deux vues non calibrées ont été proposées par exemple par Faugeras [57], Hartley *et al.* [72], Shashua [173, 174] et Ponce *et al.* [157]. Ces travaux concernent principalement la détermination des matrices de projection dans un repère projectif du monde. La triangulation est partout effectuée en résolvant des systèmes d'équations linéaires, donc en minimisant une erreur algébrique, ce qui est certainement sous-optimal. Aucune de ces méthodes n'est projectivement invariante.

Dans [13, 14], Beardsley *et al.* considérons plusieurs critères pour la triangulation. Ils proposent d'utiliser un calibrage approximatif des caméras pour rendre la situation « quasi-euclidienne ». Ainsi, l'application de la méthode du POINT MILIEU est possible, mais elle est tout de même sous-optimale. Beardsley *et al.* argumentent, comme nous, qu'il faudrait minimiser l'erreur de reprojection. Ils proposent donc une solution approximative : un point image est laissé fixe, tandis que son correspondant est corrigé en le projetant perpendiculairement sur la droite épipolaire du premier point. Ainsi, on obtient deux points qui se correspondent exactement et qui permettent d'intersecter les rayons de projection sans erreur. Cette approche est évidemment sous-optimale, mais elle va dans la bonne direction.

4.4 Un meilleur critère de minimisation

Le bruit intervient dans les mesures, c'est-à-dire les positions des primitives dans les images, dû aux erreurs de numérisation et d'extraction de primitives dans les images. Or, pour le combattre, il faudrait s'appuyer sur un critère d'optimisation qui est défini dans les images. Dans cette section, nous définissons un tel critère.

Il est courant de supposer un bruit gaussien pour l'éloignement des primitives de leur position exacte. Dans la suite, nous adoptons ce modèle de bruit. La meilleure reconstruction au sens du maximum de vraisemblance est alors celle qui minimise l'erreur de reprojection – la somme des carrés des distances entre les points reprojetés et les points image mesurés [178] (cf. la section 3.7) :

$$\widehat{\mathbf{Q}} = \text{argmin}\ \left(d(\mathbf{q}_1, \mathsf{P}_1\mathbf{Q})^2 + d(\mathbf{q}_2, \mathsf{P}_2\mathbf{Q})^2\right)\ , \tag{4.1}$$

où $\mathbf{Q}$ est le point reconstruit et $d(*,*)$ la distance euclidienne.

Soit F la matrice fondamentale associée aux matrices de projection P_1 et P_2 (voir 2.6 pour un moyen de calculer F). Les points $\mathbf{q}_1$ et $\mathbf{q}_2$ peuvent être les projections d'un point 3D uniquement s'ils satisfont la contrainte épipolaire, c'est-à-dire si $\mathbf{q}_2^\mathsf{T} \mathsf{F} \mathbf{q}_1 = 0$. Nous pouvons maintenant transformer le problème de l'estimation du point 3D $\widehat{\mathbf{Q}}$ en celui de l'estimation de ses projections : nous cherchons les points $\widehat{\mathbf{q}}'_1$ et $\widehat{\mathbf{q}}'_2$ tels que :

$$
\begin{aligned}
\{\widehat{\mathbf{q}}'_1, \widehat{\mathbf{q}}'_2\} &= \text{argmin} \ \ (d(\mathbf{q}_1, \mathbf{q}'_1)^2 + d(\mathbf{q}_2, \mathbf{q}'_2)^2) \\
\text{sous la contrainte} &\quad \mathbf{q'}_2^\mathsf{T} \mathsf{F} \mathbf{q}'_1 = 0 \ .
\end{aligned}
\tag{4.2}
$$

On peut voir la détermination de $\widehat{\mathbf{q}}'_1$ et $\widehat{\mathbf{q}}'_2$ comme une correction des points image, avec le but de les transformer en des projections possibles d'un point 3D. Sous l'hypothèse d'un bruit gaussien, le critère (4.2) donne les points corrigés qui sont le plus vraisemblablement les « vraies » correspondances des points image. En effet, l'optimisation de ce critère inclut une reconstruction 3D implicite de $\mathbf{Q}$, optimale au sens du critère des erreurs de reprojection (voir 3.7). Mais, au lieu de paramétrer la reconstruction par les 3 coordonnées du point 3D, nous le faisons via une seule variable (voir la suite), qui définit ses projections.

Après avoir déterminé $\widehat{\mathbf{q}}'_1$ et $\widehat{\mathbf{q}}'_2$, le point 3D $\widehat{\mathbf{Q}}$ peut être calculé par n'importe quelle méthode de triangulation puisque les rayons de projection de ces deux points se coupent *précisément*. Numériquement, bien sûr, ceci n'est pas valide, en raison des erreurs d'arrondi. Mais, en pratique, nous n'avons pas pu constater de défaillances causées par ceci.

4.5 Une méthode optimale pour la triangulation

Dans cette section, nous décrivons une méthode de triangulation qui trouve le minimum global de la fonction de coût (4.2) : la solution requiert la détermination des racines d'un polynôme de degré 6. C'est pourquoi nous appelons cette méthode la méthode POLY. Si le modèle de bruit gaussien est correct, cette méthode est alors optimale.

4.5.1 Reformulation du problème

Pour des points image mesurés $\mathbf{q}_1$ et $\mathbf{q}_2$, nous cherchons une paire de points corrigés $\widehat{\mathbf{q}}'_1$ et $\widehat{\mathbf{q}}'_2$ qui minimise la somme des carrés des distances et respecte la contrainte épipolaire (critère (4.2)). Deux points image satisfaisant la contrainte épipolaire doivent se trouver sur des droites épipolaires correspondantes. En particulier, le point optimal $\widehat{\mathbf{q}}'_1$ se trouve sur une droite épipolaire $\widehat{\mathbf{l}}_1$ et $\widehat{\mathbf{q}}'_2$ est sur la droite épipolaire correspondante $\widehat{\mathbf{l}}_2$.

Considérons maintenant une paire de droites épipolaires correspondantes $\mathbf{l}_1$ et $\mathbf{l}_2$. De tous les points sur $\mathbf{l}_1$ et $\mathbf{l}_2$, c'est naturellement la paire des projections orthogonales de $\mathbf{q}_1$ et $\mathbf{q}_2$ sur respectivement $\mathbf{l}_1$ et $\mathbf{l}_2$, qui minimise la somme des carrés des distances. Soit $(\overline{\mathbf{q}}_1, \overline{\mathbf{q}}_2)$ la paire de ces projections orthogonales. Écrivons $d(\mathbf{q}_1, \overline{\mathbf{q}}_1) = d(\mathbf{q}_1, \mathbf{l}_1)$, où $d(\mathbf{q}_1, \mathbf{l}_1)$ représente la distance perpendiculaire du point $\mathbf{q}_1$ de la droite $\mathbf{l}_1$. Une expression similaire existe pour $d(\mathbf{q}_2, \overline{\mathbf{q}}_2)$.

Au vu de la section précédente, nous pouvons reformuler le problème de minimisation comme suit. Nous cherchons la paire de droites épipolaires correspondantes selon :

$$
\{\widehat{\mathbf{l}}_1, \widehat{\mathbf{l}}_2\} = \text{argmin} \ \ (d(\mathbf{q}_1, \mathbf{l}_1)^2 + d(\mathbf{q}_2, \mathbf{l}_2)^2) \ . \tag{4.3}
$$

Supposons qu'on ait déterminé cette paire de droites épipolaires. Alors, les points recherchés sont simplement les projections orthogonales de $\mathbf{q}_1$ et $\mathbf{q}_2$ sur $\widehat{\mathbf{l}}_1$ respectivement $\widehat{\mathbf{l}}_2$.

Notre stratégie pour minimiser (4.3) est la suivante.

1. Paramétrer le faisceau des droites épipolaires dans la première image en fonction d'un seul paramètre t. Une droite épipolaire dans la première image peut être écrite comme $\mathbf{l}_1(t)$.
2. Paramétrer le faisceau des droites épipolaires dans la deuxième image en fonction du même paramètre t (en utilisant la matrice fondamentale F), ce qu'on peut donc écrire comme $\mathbf{l}_2(t)$.
3. Exprimer la fonction de coût $d(\mathbf{q}_1, \mathbf{l}_1(t))^2 + d(\mathbf{q}_2, \mathbf{l}_2(t))^2$ explicitement en t.
4. Trouver la valeur de t qui minimise cette fonction globalement.

Ainsi, le problème est réduit à la détermination du minimum d'une fonction en une seule variable t. Dans la section suivante, nous montrons que, avec une paramétrisation appropriée du faisceau épipolaire, la fonction de coût est une fonction rationnelle en t. En utilisant des techniques de calcul élémentaires, le problème de minimisation peut être ramené à la détermination des racines réelles d'un polynôme de degré 6.

4.5.2 Algorithme

4.5.2.1 Développements

Si les deux points image mesurés coïncident avec les épipôles, le point 3D se trouve sur la droite reliant les deux centres de projection. Dans ce cas, il est impossible de déterminer sa position seulement à partir des deux vues disponibles. Le cas où un seul des points image coïncide avec un épipôle, n'est pas réalisable. Par conséquent, nous supposons pour la suite qu'aucun des deux points image ne coïncide avec un épipôle. Les développements suivants sont illustrés par la figure 4.2.

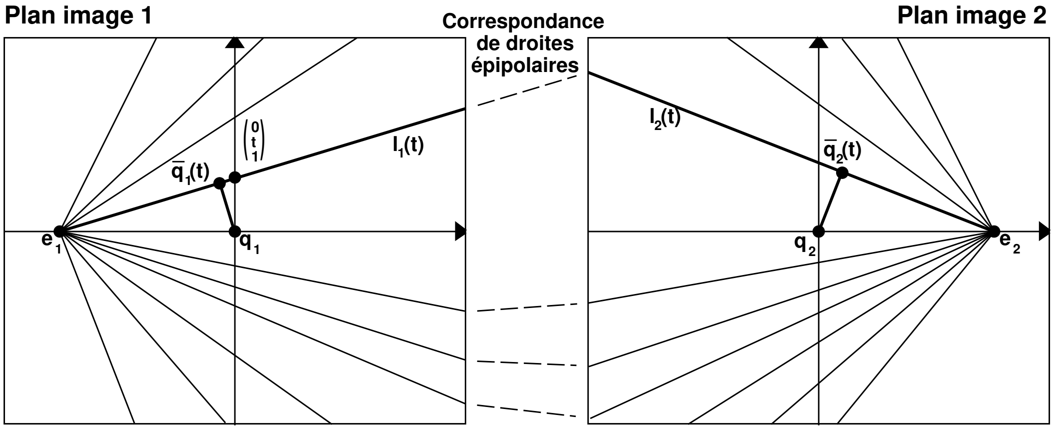


FIG. 4.2: *Principe de notre méthode de triangulation.*

Nous pouvons simplifier le problème en appliquant des transformations rigides aux images qui envoient les points image $\mathbf{q}_1$ (respectivement $\mathbf{q}_2$) sur l'origine, $(0,0,1)^\mathsf{T}$ en coordonnées homogènes. En outre, les épipôles peuvent être envoyés sur les axes des u de chaque image, sur des points $(1,0,f_1)^\mathsf{T}$ (respectivement $(1,0,f_2)^\mathsf{T}$). Une valeur $f_i = 0$ signifie que l'épipôle se trouve à l'infini. Les détails de la détermination de ces transformations rigides sont donnés plus bas dans la section 4.5.2.2. Dans la suite, nous supposons que, en coordonnées homogènes :

$$\mathbf{q}_1 \sim \mathbf{q}_2 \sim (0,0,1)^\mathsf{T}$$

et que les épipôles se trouvent sur les axes des u :

$$\mathbf{e}_1 \sim (1, 0, f_1)^\mathsf{T} \qquad \mathbf{e}_2 \sim (1, 0, f_2)^\mathsf{T} \ .$$

L'application de ces transformations rigides n'a pas d'effet sur la valeur de la fonction de coût (4.2), puisque celle-ci est basée sur des distances, qui sont conservées par les transformations rigides. Pourtant, la matrice fondamentale doit être adaptée pour tenir compte des transformations dans les images. La nouvelle matrice fondamentale F doit satisfaire les équations :

$$\mathsf{F} \begin{pmatrix} 1 \\ 0 \\ f_1 \end{pmatrix} = \mathsf{F}^\mathsf{T} \begin{pmatrix} 1 \\ 0 \\ f_2 \end{pmatrix} = \mathbf{0} \ .$$

F a donc une forme particulière (un moyen de calculer F à partir de la matrice fondamentale originale est décrit dans la section 4.5.2.2) :

$$\mathsf{F} = \begin{pmatrix} f_1 f_2 d & -f_2 c & -f_2 d \\ -f_1 b & a & b \\ -f_1 d & c & d \end{pmatrix} \ . \tag{4.4}$$

Considérons une droite épipolaire dans la première image qui coupe l'axe des v au point $(0, t, 1)^\mathsf{T}$. Notons cette droite $\mathbf{l}_1(t)$. Le vecteur représentant $\mathbf{l}_1(t)$ est donné par le produit vectoriel suivant :

$$\mathbf{l}_1(t) \sim \begin{pmatrix} 0 \\ t \\ 1 \end{pmatrix} \wedge \begin{pmatrix} 1 \\ 0 \\ f_1 \end{pmatrix} = \begin{pmatrix} t f_1 \\ 1 \\ -t \end{pmatrix} \ .$$

Le carré de la distance de $\mathbf{l}_1(t)$ au point $\mathbf{q}_1$ (l'origine) s'évalue donc comme :

$$d(\mathbf{q}_1, \mathbf{l}_1(t))^2 = \frac{t^2}{1 + (t f_1)^2} \ .$$

La droite épipolaire correspondante dans la deuxième image est donnée par :

$$\mathbf{l}_2(t) \sim \mathsf{F} \begin{pmatrix} 0 \\ t \\ 1 \end{pmatrix} = \begin{pmatrix} -f_2(ct + d) \\ at + b \\ ct + d \end{pmatrix} \ .$$

Le carré de la distance carrée de $\mathbf{l}_2(t)$ au point $\mathbf{q}_2$ est égale à :

$$d(\mathbf{q}_2, \mathbf{l}_2(t))^2 = \frac{(ct + d)^2}{(at + b)^2 + f_2^2(ct + d)^2} \ .$$

La somme totale des carrés des distances (c-à-d la fonction de coût (4.3)) est donc donnée par :

$$s_1(t) = \frac{t^2}{1 + (t f_1)^2} + \frac{(ct + d)^2}{(at + b)^2 + f_2^2(ct + d)^2} \ . \tag{4.5}$$

Le but est de trouver la valeur de t qui minimise cette fonction. Nous trouvons le minimum en appliquant des techniques de calcul élémentaires. D'abord, calculons la dérivée première de la fonction $s_1(t)$:

$$s_1'(t) = \frac{2t}{(1+(tf_1)^2)^2} - \frac{2(ad-bc)(at+b)(ct+d)}{((at+b)^2+f_2^2(ct+d)^2)^2} \quad . \tag{4.6}$$

Pour déterminer les minima de $s_1(t)$, nous nous intéressons aux racines de $s_1'(t)$. Collecter les deux termes de $s_1'(t)$ sur leur dénominateur commun et annuler le numérateur mène à la condition :

$$\begin{aligned} r_1(t) &= t((at+b)^2+f_2^2(ct+d)^2)^2 - (ad-bc)(1+(tf_1)^2)^2(at+b)(ct+d) \\ &= t^6(acf_1^4(bc-ad)) + \\ &\quad t^5(f_1^4(b^2c^2-a^2d^2)+c^2f_2^4(c^2+2a^2)+a^4) + \\ &\quad t^4(bdf_1^4(bc-ad)+2acf_1^2(bc-ad)+4c^3df_2^4+4acf_2^2(ad+bc)+4a^3b) + \\ &\quad t^3(2f_1^2(b^2c^2-a^2d^2)+6c^2d^2f_2^4+2f_2^2(a^2d^2+b^2c^2+4abcd)+6a^2b^2) + \\ &\quad t^2(2bdf_1^2(bc-ad)+4cd^3f_2^4+4bdf_2^2(ad+bc)+4ab^3+ac(bc-ad)) + \\ &\quad t^1(d^4f_2^4+2b^2d^2f_2^2+b^4+b^2c^2-a^2d^2) + \\ &\quad t^0(bd(bc-ad)) \\ &= 0 \quad . \end{aligned} \tag{4.7}$$

Les minima et maxima de $s_1(t)$ se trouvent aux racines de ce polynôme. Le polynôme est de degré 6 et peut donc avoir jusqu'à 6 racines réelles, qui correspondent à 3 minima et 3 maxima de $s_1(t)$. Le minimum global peut être trouvé tout simplement en évaluant $s_1(t)$ aux racines réelles de $r_1(t)$. On devrait aussi tester la valeur asymptotique de $s_1(t)$ pour $t \to \infty$, pour voir si la distance minimale apparaît quand $t = \infty$. Ceci correspond à une droite épipolaire verticale dans la première image. Normalement, cette situation n'a pas d'intérêt : le point corrigé $\widehat{\mathbf{q}}_1'$ serait la projection orthogonale de l'origine sur la droite verticale passant par $\mathbf{e}_1$, qui n'est rien d'autre que l'épipôle $\mathbf{e}_1$ même (voir la figure 4.2). Pourtant, il pourrait être possible que la valeur $s_1(\infty)$ soit plus petite que celles aux racines de r_1, puisque celles-ci sont déterminées par une méthode numérique, qui donnera des racines peu précises à proximité de l'infini. Ce cas indiquerait alors une certaine instabilité de la reconstruction du point considéré.

La valeur asymptotique de s_1 est donnée par :

$$s_1(\infty) = \frac{1}{f_1^2} + \frac{c^2}{a^2+f_2^2c^2} \quad .$$

4.5.2.2 Détails

Détermination des transformations rigides des images. Nous calculons la transformation rigide pour une seule image. Celle de l'autre image est déterminée de manière analogue.

Nous effectuons d'abord une translation T qui ramène le point $\mathbf{q}$ à l'origine. Soit $\mathbf{q} = (u,v,w)^\mathsf{T}$. Alors :

$$\mathsf{T} = \begin{pmatrix} 1 & 0 & -\frac{u}{w} \\ 0 & 1 & -\frac{v}{w} \\ 0 & 0 & 1 \end{pmatrix} \quad .$$

Nous plaçons l'épipôle (après l'avoir déplacé par T) sur l'axe des u par une rotation R autour de l'origine. Soit $\mathbf{e} = (u_e, v_e, w_e)^\mathsf{T}$ l'épipôle. Alors, il faut que :

$$\mathsf{RT}\mathbf{e} = \begin{pmatrix} \cos\Theta & -\sin\Theta & 0 \\ \sin\Theta & \cos\Theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\frac{u}{w} \\ 0 & 1 & -\frac{v}{w} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_e \\ v_e \\ w_e \end{pmatrix} \sim \begin{pmatrix} 1 \\ 0 \\ f \end{pmatrix} \quad ,$$

pour $f \in \mathcal{R}$. Le fait que la deuxième coordonnée doive s'annuler nous donne une équation qui permet de déterminer l'angle de rotation Θ :

$$\sin\Theta(u_e w - w_e u) + \cos\Theta(v_e w - w_e v) = 0 \quad .$$

La transformation rigide complète est alors donnée par $\mathsf{X} = \mathsf{RT}$.

Transformation de la matrice fondamentale. La matrice fondamentale doit être adaptée pour être valide pour les images transformées. Soient $\mathsf{X}_1 = \mathsf{R}_1\mathsf{T}_1$ et $\mathsf{X}_2 = \mathsf{R}_2\mathsf{T}_2$ les transformations rigides des deux images et F_0 la matrice fondamentale avant l'application des transformations. La matrice fondamentale F pour les images transformées est calculée ainsi :

$$\begin{aligned} \mathsf{F} &= \mathsf{X}_2{}^{-\mathsf{T}}\mathsf{F}_0\mathsf{X}_1{}^{-1} \\ &= \mathsf{R}_2\mathsf{T}_2{}^{-\mathsf{T}}\mathsf{F}_0\mathsf{T}_1{}^{-1}\mathsf{R}_1^\mathsf{T} \quad . \end{aligned}$$

La matrice F aura la forme particulière donnée dans l'équation (4.4).

Détermination des points $\widehat{\mathbf{q}}'_1$ et $\widehat{\mathbf{q}}'_2$. Après avoir trouvé le minimum global de $s_1(t)$, nous pouvons aisément calculer les points $\widehat{\mathbf{q}}'_1$ et $\widehat{\mathbf{q}}'_2$ par :

$$\begin{aligned} \widehat{\mathbf{q}}'_1 &= \begin{pmatrix} t^2 f_1 \\ t \\ 1 + t^2 f_1^2 \end{pmatrix} \\ \widehat{\mathbf{q}}'_2 &= \begin{pmatrix} f_2(ct+d)^2 \\ -(at+b)(ct+d) \\ f_2^2(ct+d)^2 + (at+b)^2 \end{pmatrix} \quad . \end{aligned}$$

4.5.2.3 Algorithme complet

Préalablement à l'application de l'algorithme de triangulation proprement dit, nous devons calculer la matrice fondamentale F_0. Dans la formulation du problème de triangulation, nous avons supposé que les matrices de projection des deux vues sont données en entrée. De ces matrices, on peut calculer la matrice fondamentale. En pratique, le procédé est souvent inversé : la matrice fondamentale est calculée à partir de correspondances de points image et des matrices de projection sont ensuite dérivées de celle-ci (voir [121] pour des formules).

Pour chaque correspondance de points image nous effectuons les opérations suivantes.

Algorithme : Triangulation

1. Déterminer les transformations rigides X_1 et X_2, corriger la matrice fondamentale et l'exprimer sous la forme de l'équation (4.4).
2. Construire les polynômes $s_1(t)$ et $r_1(t)$.
3. Déterminer les racines de $r_1(t)$ par une méthode numérique quelconque (voir par exemple [159]).
4. Évaluer $s_1(t)$ aux racines réelles et à l'abscisse $t = \infty$ et choisir le t qui donne la valeur minimale.
5. Déterminer les points $\widehat{\mathbf{q}}_1'$ et $\widehat{\mathbf{q}}_2'$.
6. Défaire les transformations rigides X_1 et X_2 en appliquant leurs inverses à $\widehat{\mathbf{q}}_1'$ et $\widehat{\mathbf{q}}_2'$.
7. Trianguler les points obtenus par une méthode quelconque (des exemples sont donnés dans la section 4.7).

4.5.3 Sur les minima locaux

Le fait que $r_1(t)$ dans l'équation (4.7) est un polynôme de degré 6 a pour conséquence que $s_1(t)$ peut avoir jusqu'à 3 minima. Ceci est en effet possible comme nous le montrons à l'aide d'un exemple. Soient $f_1 = f_2 = 1$ et $a = 2, b = 3, c = 3, d = 4$. La matrice fondamentale est donc donnée par :

$$\mathsf{F} = \begin{pmatrix} 4 & -3 & -4 \\ -3 & 2 & 3 \\ -4 & 3 & 4 \end{pmatrix} \quad ,$$

et pour la fonction de coût $s_1(t)$ on obtient :

$$s_1(t) = \frac{t^2}{1+t^2} + \frac{(3t+4)^2}{(2t+3)^2 + (3t+4)^2} \quad .$$

Le graphe de $s_1(t)$ est montré dans la figure 4.3 (a). Les 3 minima sont clairement visibles.

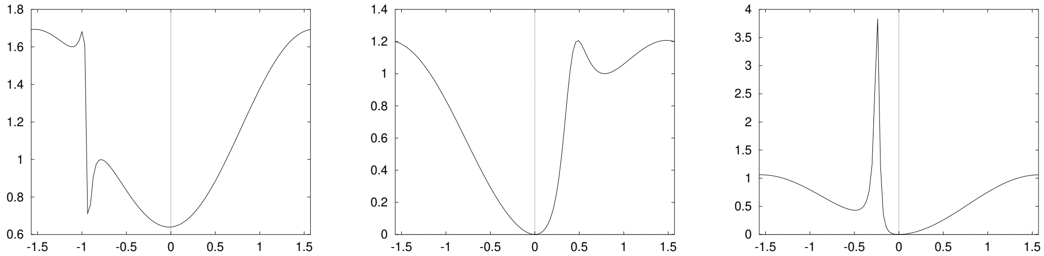
Comme deuxième exemple, considérons les valeurs $f_1 = f_2 = 1$ et $a = 2, b = -1, c = 1, d = 0$ ce qui nous donne :

$$\mathsf{F} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 2 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad ,$$

et :

$$s_1(t) = \frac{t^2}{t^2+1} + \frac{t^2}{t^2 + (2t-1)^2} \quad .$$

Les deux termes de la fonction de coût s'annulent à $t = 0$ ce qui veut dire que les points correspondants $\mathbf{q}_1$ et $\mathbf{q}_2$ satisfont exactement la contrainte épipolaire. Ceci peut être vérifié en observant



(a) Exemple d'une fonction de coût avec 3 minima.

(b) Une fonction de coût pour une correspondance de points parfaite (le minimum global est à l'abscisse 0), mais qui néanmoins possède un autre minimum local.

(c) Un autre exemple pour une correspondance parfaite, avec un minimum local séparé du minimum global par un pic très prononcé (obtenu avec les valeurs $f_1 = 1, f_2 = 0.5, a = 4, b = c = 1, d = 0$).

FIG. 4.3: *Exemples de fonctions de coût avec des minima locaux. Pour montrer la totalité des fonctions de coût, nous substituons t par* $\tan(\alpha)$ *et affichons les graphes en fonction de* α *dans l'intervalle* $-\pi/2 \leq \alpha \leq \pi/2$.

que $\mathbf{q}_2^\top \mathsf{F} \mathbf{q}_1 = 0$. Le graphe de $s_1(t)$ est illustré dans la figure 4.3 (b). On peut constater que, à côté du minimum global à $t = 0$, il y a aussi un minimum local à $t = 1$ ($\alpha = \pi/4$). Donc, même dans le cas de correspondances parfaites, des minima locaux peuvent exister. Ces exemples démontrent qu'un algorithme qui tente à minimiser les fonctions de coût (4.2) ou (4.3) par une recherche itérative, est susceptible de trouver un minimum local, même dans le cas d'une correspondance parfaite !

Rothwell *et al.* ont « réimplémenté » notre méthode pour effectuer des expériences de reconstruction projective [169]. Leur implémentation a deux défauts : une recherche itérative est utilisée pour trouver le minimum global et cette recherche est initialisée avec le résultat de la méthode du POINT MILIEU qui est l'une des pires solutions pour la reconstruction projective. Ceci est une explication possible du fait que les résultats qui sont présentés dans [169] semblent être plus mauvais que les nôtres, quant à la comparaison avec d'autres méthodes.

4.5.4 Minimisation de la somme des distances absolues

Au lieu de minimiser la somme des carrés des distances entre les points mesurés et les points reprojetés, il est possible d'adapter notre méthode de manière à minimiser la somme des distances absolues. Nous appelons ceci la méthode POLY-ABS.

La quantité à minimiser est :

$$d(\mathbf{q}_1, \mathbf{l}_1(t)) + d(\mathbf{q}_2, \mathbf{l}_2(t)) \tag{4.8}$$

ou, explicitement en fonction de t :

$$s_2(t) = \frac{|t|}{\sqrt{1 + (tf_1)^2}} + \frac{|ct + d|}{\sqrt{(at + b)^2 + f_2^2(ct + d)^2}} \quad .$$

Sa dérivée première est de la forme :

$$s_2'(t) = \omega_1 \frac{1}{(1+(tf_1)^2)^{3/2}} - \omega_2 \frac{(ad-bc)(at+b)}{((at+b)^2 + f_2^2(ct+d)^2)^{3/2}} \quad ,$$

où ω_1 et ω_2 sont égaux à -1 ou 1, selon les signes de t et $(ct+d)$. Le minimum global de $s_2(t)$ se trouve à une racine de $s_2'(t)$. Nous essayons donc de déterminer les racines de $s_2'(t)$ et déterminons le minimum global en évaluant s_2 à toutes les racines et en retenant celle qui donne la valeur minimale. La condition $s_2'(t) = 0$ peut s'écrire comme :

$$\omega_1 \frac{1}{(1+(tf_1)^2)^{3/2}} = \omega_2 \frac{(ad-bc)(at+b)}{((at+b)^2 + f_2^2(ct+d)^2)^{3/2}} \quad .$$

Nous élevons les deux côtés de cette équations au carré, afin d'éliminer les ω_1, ω_2 et les racines carrées. La soustraction des termes obtenus donne :

$$r_2(t) = \frac{1}{(1+(tf_1)^2)^3} - \frac{(ad-bc)^2(at+b)^2}{((at+b)^2 + f_2^2(ct+d)^2)^3} \quad . \tag{4.9}$$

Le numérateur commun de r_2 est un polynôme de degré 8 en t. Comme pour la méthode POLY, nous évaluons $s_2(t)$ aux racines de ce polynôme afin de trouver son minimum global.

4.5.5 Et pour 3 vues ?

L'extension de la méthode pour 3 vues ou plus serait désirable et nous avons exploré cette voie, d'abord pour le cas général et ensuite pour le cas de trois vues alignées.

Cas général. La minimisation directe du critère de reprojection ne semble être praticable qu'avec deux vues. Pour 3 vues, le problème ne peut plus être paramétré par une seule variable, mais il en faut trois. Une paramétrisation analogue à celle choisie pour le cas de deux vues serait de répartir les paramètres de la façon suivante. Deux paramètres définissent la position du point $\widehat{\mathbf{q}}_1'$ dans la première image. Le deuxième point doit être sur la droite épipolaire du premier et peut donc être positionné par un paramètre. La position du point dans la troisième image est entièrement déterminée par les relations trifocales. Au lieu de paramétrer les points image, nous pourrions donc directement et simplement utiliser les coordonnées 3D du point reconstruit. Les trois dérivées partielles de la fonction de coût, qui est la somme des résidus des reprojections, sont des polynômes de degré 8. Même s'il était algébriquement possible d'engendrer un polynôme en une seule variable, celui-ci serait certainement de degré beaucoup trop élevé pour être résolu avec une précision raisonnable.

Nous avons tenté de réduire la complexité des formules en appliquant des transformations rigides appropriées aux images. Par exemple, deux éléments d'une matrice de projection peuvent facilement être annulés en appliquant une translation appropriée dans l'image : soit P une matrice de projection et $\overline{\mathsf{P}}$ sa première sous-matrice 3×3. L'application d'une translation de vecteur $\mathbf{t}$ dans l'image donne lieu à la matrice de projection transformée P' avec :

$$\overline{\mathsf{P}}' = \begin{pmatrix} 1 & 0 & t_1 \\ 0 & 1 & t_2 \\ 0 & 0 & 1 \end{pmatrix} \overline{\mathsf{P}} = \begin{pmatrix} P_{11} + t_1 P_{31} & P_{12} + t_1 P_{32} & P_{13} + t_1 P_{33} \\ P_{21} + t_2 P_{31} & P_{22} + t_2 P_{32} & P_{23} + t_2 P_{33} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \quad .$$

Il est facile de voir comment choisir t_1 et t_2 pour annuler deux éléments de $\overline{\mathsf{P}}'$. Ce procédé permet de simplifier la forme de la fonction à minimiser, mais pas assez pour pouvoir résoudre le problème comme au cas de deux vues.

Trois vues alignées. Pour des vues alignées (centres de projections alignés), les plans épipolaires des paires de vues coïncident. Ceci rend possible d'exprimer le problème par un seul paramètre, comme dans le cas de deux vues. On peut appliquer le même schéma que dans le paragraphe 4.5.2.1. Par rapport à la fonction de coût (4.5), un troisième terme s'ajoute, pour la distance du point image à la droite épipolaire dans la troisième image. On obtient finalement un polynôme de degré 10, dont il faut déterminer les racines.

Il serait facile d'obtenir des résultats pour des mouvements particuliers entre les vues, comme nous le faisons pour le cas de deux vues dans le paragraphe suivant, mais nous nous arrêtons ici.

Caméra linéaire. Nous avons également étudié le problème de triangulation pour trois vues de caméras linéaires, qui se présente mieux, voir l'annexe A.2.

4.6 Triangulation pour des mouvements particuliers

Récemment, l'étude de mouvements particuliers a suscité quelque intérêt dans le domaine de la vision non calibrée. Pour citer quelques exemples, des méthodes de calibrage pour des rotations pures [82] et des mouvements planaires [6, 226] ou de reconstruction affine à partir de translations pures [131] ont été proposées (voir 6.2.3 pour une discussion approfondie). Dans le contexte de la triangulation, nous nous intéressons à des mouvements particuliers entre deux vues d'une même caméra. Torr *et al.* [204] et Viéville & Lingrand [215] ont étudié des formes particulières que prend la matrice fondamentale lors de mouvements particuliers (voir la section 2.3.4).

Nous considérons ces formes particulières de F en vue de nos méthodes de triangulation POLY et POLY-ABS. Pour quelques mouvements particuliers, nous montrons comment la recherche du minimum global de la fonction de coût (4.2) (respectivement (4.8)) se simplifie, par une réduction du degré du polynôme $r_1(t)$ (respectivement $r_2(t)$) à résoudre (cf. les équations (4.7) et (4.9)).

Dans les deux paragraphes suivants, nous développons le cas des caméras affines et celui d'une caméra (perspective) effectuant une rotation autour d'un axe fixe. Deux autres cas, translation pure et changement de focale ou de mise au point, sont traités dans l'annexe A. Dans cette annexe, nous considérons également la triangulation à partir de trois vues d'une caméra linéaire. Si à la place de la matrice fondamentale nous considérons une homographie associée aux projections d'un plan dans les deux vues, nous obtenons un problème similaire à la triangulation. Dans l'annexe B, nous proposons un schéma de correction de points image de manière à ce qu'ils se correspondent exactement par l'homographie donnée.

Le tableau 4.1 résume tous ces cas.

4.6.1 Caméras affines

Après l'application des mêmes transformations rigides que dans la section 4.5.2, les épipôles sont les points à l'infini des axes des u et la matrice fondamentale se simplifie encore davantage :

$$\mathsf{F} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & b' \\ 0 & d' & e' \end{pmatrix} \quad .$$

Poly. La fonction de coût est :

$$s_1(t) = t^2 + \frac{(d't+e')^2}{b'^2} = (1 + \frac{d'^2}{b'^2})t^2 + 2\frac{d'e'}{b'^2}t + \frac{e'^2}{b'^2} \quad ,$$

Type de mouvement/caméra	Poly	Poly-Abs
Général	6	8
Translation pure		
– épipôles finis	2	2
– épipôles infinis	1	1
Caméras affines	1	1
Zoom + mise au point		
– forme générale	6	8
– pixels rectangulaires	2	2
– translation dans plan focal et pixels rectangulaires	1	1
Rotation autour d'un axe fixe		
– forme générale	6	8
– tête stéréo	6	8
– tête stéréo symétrique	2	2
3 vues de caméras linéaires (voir section A.2)	15	6
3 vues de caméras habituelles		
– forme générale	grand	grand
– vues alignées	10	grand
Homographie d'un plan (voir annexe B)	8	grand
Homographie affine d'un plan	1	2

TAB. 4.1: *La complexité des méthodes* POLY *et* POLY-ABS *pour différents scénarios. Les chiffres correspondent aux degrés maximaux des polynômes à résoudre.*

et pour $r_1(t)$ nous obtenons :

$$r_1(t) = (b'^2 + d'^2)t + d'e' \quad .$$

Le minimum global de $s_1(t)$ peut donc être calculé directement. Il est atteint en :

$$t = -\frac{d'e'}{b'^2 + d'^2} \quad .$$

La dérivée seconde est positive, donc il s'agit vraiment d'un minimum.

Poly-Abs. La fonction de coût est :

$$s_2(t) = |t| + \frac{|d't + e'|}{|b'|} \quad .$$

Il s'agit de la somme des valeurs absolues de deux fonctions linéaires. Le minimum global se trouve alors à une des deux racines, 0 ou $-\frac{e'}{d'}$, ou il occupe tout l'intervalle entre ces deux racines.

Solution directe. Au lieu d'appliquer des transformations rigides préalablement à la formulation de la fonction de coût, nous pouvons omettre cette étape ici et résoudre le problème pour la matrice fondamentale originale, donnée par :

$$\mathsf{F} = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{pmatrix} \quad .$$

Les épipôles sont les points à l'infini $\mathbf{e}_1 = (d, -c, 0)^\mathsf{T}$ et $\mathbf{e}_2 = (b, -a, 0)^\mathsf{T}$. Les droites épipolaires dans chaque image sont parallèles, avec les directions normales $\mathbf{n}_1 = (c, d, 0)^\mathsf{T}$ et $\mathbf{n}_2 = (a, b, 0)^\mathsf{T}$. Les points corrigés $\widehat{\mathbf{q}}'_1$ et $\widehat{\mathbf{q}}'_2$ étant les projections orthogonales des points image mesurés sur des droites épipolaires correspondantes, nous pouvons formuler le problème en utilisant un paramètre t :

$$\mathbf{q}'_1 = \mathbf{q}_1 + t\mathbf{n}_1 = \mathbf{q}_1 + t \begin{pmatrix} c \\ d \\ 0 \end{pmatrix} \quad .$$

Le point corrigé dans la deuxième image est la projection orthogonale de $\mathbf{q}_2$ sur la droite épipolaire de $\mathbf{q}'_1$:

$$\mathbf{q}'_2 = (\mathsf{F}\mathbf{q}'_1) \wedge (\mathbf{q}_2 \wedge \mathbf{n}_2) \quad .$$

Nous omettons les détails et donnons directement la solution unique de t qui minimise la somme des carrés des distances entre points corrigés et points mesurés :

$$t = -\frac{aq_{2_1} + bq_{2_2} + cq_{1_1} + dq_{1_2} + e}{a^2 + b^2 + c^2 + d^2} \quad .$$

Le minimum global de la somme des carrés des distances s'évalue d'après :

$$\widehat{s}_1 = \frac{(aq_{2_1} + bq_{2_2} + cq_{1_1} + dq_{1_2} + e)^2}{a^2 + b^2 + c^2 + d^2} \quad .$$

Les points corrigés optimaux sont :

$$\begin{aligned}
\widehat{\mathbf{q}}'_1 &\sim \begin{pmatrix} q_{1_1}(a^2 + b^2 + d^2) - c(q_{1_2}d + q_{2_1}a + q_{2_2}b + e) \\ q_{1_2}(a^2 + b^2 + c^2) - d(q_{1_1}c + q_{2_1}a + q_{2_2}b + e) \\ a^2 + b^2 + c^2 + d^2 \end{pmatrix} \quad , \\
\widehat{\mathbf{q}}'_2 &\sim \begin{pmatrix} q_{2_1}(b^2 + c^2 + d^2) - a(q_{1_1}c + q_{1_2}d + q_{2_2}b + e) \\ q_{2_2}(a^2 + c^2 + d^2) - b(q_{1_1}c + q_{1_2}d + q_{2_1}a + e) \\ a^2 + b^2 + c^2 + d^2 \end{pmatrix} \quad .
\end{aligned}$$

On constate que les points corrigés sont directement donnés en fonction des coefficients de la matrice fondamentale et des points image mesurés. Nous pouvons en tirer profit pour l'estimation optimale de la matrice fondamentale d'après le critère (2.3) : la minimisation ne nécessite pas des inconnues supplémentaires pour les points image corrigés, mais peut se faire directement en fonction des coefficients de F. Le critère de minimisation devient donc :

$$\widehat{\mathsf{F}}(a, b, c, d, e) = \text{argmin} \; \frac{1}{a^2 + b^2 + c^2 + d^2} \sum_{p=1}^{n} (aq_{2_1} + bq_{2_2} + cq_{1_1} + dq_{1_2} + e)^2 \quad . \tag{4.10}$$

L'optimalité de ce critère a déjà été démontré par Shapiro *et al.* [172], mais par d'autres moyens. Il existe un lien intéressant entre l'estimation optimale de la matrice fondamentale affine et la méthode de reconstruction affine par factorisation de Tomasi et Kanade [202]. Les deux méthodes fournissent la reconstruction affine optimale, c'est-à-dire qui minimise l'erreur de reprojection (voir 3.7), mais par des moyens très différents : la première effectue une reconstruction optimale implicite via l'estimation itérative de la matrice fondamentale affine, tandis que la deuxième méthode est basée sur une décomposition en valeurs singulières d'une grande matrice contenant toutes les coordonnées image. La factorisation offre tout de même l'avantage d'être applicable pour plusieurs vues simultanément.

4.6.2 Rotation autour d'un axe fixe

Le seul cas parmi ceux décrits au 2.3.4.4 pour lequel nous avons réussi à baisser le degré des polynômes à résoudre, est celui de la « tête stéréo symétrique ». La matrice fondamentale a la forme :

$$\mathsf{F} = \begin{pmatrix} 0 & a & -ab \\ a & 0 & c-d \\ -ab & -c-d & 2bd \end{pmatrix} .$$

Les épipôles sont $\mathbf{e}_1 = (\frac{d-c}{a}, b, 1)^\mathsf{T}$ et $\mathbf{e}_2 = (\frac{d+c}{a}, b, 1)^\mathsf{T}$ et leur point milieu est $\mathbf{q} = (\frac{d}{a}, b, 1)^\mathsf{T}$. Appliquer la même translation dans les deux images, qui ramène $\mathbf{q}$ à l'origine, déplace les épipôles sur les axes des u respectifs, sur des points opposés par rapport à l'origine. La matrice fondamentale transformée est de la forme :

$$\mathsf{F} \sim \begin{pmatrix} 0 & a & 0 \\ a & 0 & c \\ 0 & -c & 0 \end{pmatrix} ,$$

et les épipôles sont les points $(-\frac{c}{a}, 0, 1)^\mathsf{T}$ et $(\frac{c}{a}, 0, 1)^\mathsf{T}$. Contrairement à la méthode générale, nous ne plaçons pas les points image à l'origine, ce qui doit être pris en compte lors de la construction de la fonction de coût. Soient $\mathbf{q}'_1$ et $\mathbf{q}'_2$ les points image transformés par les translations décrites.

Poly. La fonction de coût est :

$$s_1(t) = \frac{(cq'_{12} - aq'_{11}t - ct)^2 + (cq'_{22} + aq'_{21}t - ct)^2}{a^2t^2 + c^2}$$

et le polynôme $r_1(t)$ est une fonction quadratique en t :

$$\begin{aligned} r_1(t) &= t^2a^2c(a(q'_{11}q'_{12} - q'_{21}q'_{22}) + c(q'_{12} + q'_{22})) + \\ &\quad tc^2(2c^2 + a^2({q'_{11}}^2 + {q'_{21}}^2 - {q'_{12}}^2 - {q'_{22}}^2) + 2ac(q'_{11} - q'_{21})) - \\ &\quad c^3(a(q'_{11}q'_{12} - q'_{21}q'_{22}) + c(q'_{12} + q'_{22})) \ . \end{aligned}$$

Nous obtenons donc directement les deux solutions candidates possibles. Il est tout de même nécessaire d'évaluer la fonction de coût pour $t = \infty$, pour traiter la situation où $\widehat{\mathbf{q}}'_1$ se trouverait sur la même droite verticale que $\mathbf{e}_1$. Nous obtenons la valeur $s_1(\infty) = \frac{2c^2+a^2({q'_{11}}^2+{q'_{21}}^2)+2ac(q'_{11}-q'_{21})}{a^2}$.

Poly-Abs. La fonction de coût est :

$$s_2(t) = \frac{|cq'_{12} - aq'_{11}t - ct| + |cq'_{22} + aq'_{21}t - ct|}{\sqrt{a^2t^2 + c^2}} \ .$$

Le polynôme r_2 est quadratique, tout comme r_1 :

$$r_2(t) = ac^2(c(q'_{11} + q'_{21}) + a(q'_{12} - q'_{22})t)(2c^2 + a^2(q'_{12} + q'_{22})t + ac(q'_{11} - q'_{21})) \ ,$$

et nous pouvons tirer les mêmes conclusions pour ce cas.

4.7 Autres méthodes de triangulation

Dans cette section, nous présentons quelques autres méthodes de triangulation qui ont été comparées avec les méthodes POLY et POLY-ABS (voir les résultats expérimentaux dans la section 4.8).

4.7.1 Méthodes linéaires

En dehors de la méthode du POINT MILIEU, les méthodes linéaires présentées dans la suite sont les plus populaires. Considérons la projection d'un point 3D $\mathbf{Q}$ sur un point $\mathbf{q}$ dans l'image :

$$\mathbf{q} \sim \mathsf{P}\mathbf{Q} \ .$$

Nous écrivons en coordonnées homogènes $\mathbf{q} = w(u, v, 1)^\mathsf{T}$, où (u, v) sont les coordonnées observées et w est un facteur d'échelle inconnu. En notant $\mathbf{p}_i^\mathsf{T}$ la i^e ligne de la matrice P, cette équation peut s'écrire comme :

$$wu = \mathbf{p}_1^\mathsf{T}\mathbf{Q} \quad , \quad wv = \mathbf{p}_2^\mathsf{T}\mathbf{Q} \quad , \quad w = \mathbf{p}_3^\mathsf{T}\mathbf{Q} \ .$$

En éliminant w nous obtenons :

$$\begin{aligned} u\mathbf{p}_3^\mathsf{T}\mathbf{Q} &= \mathbf{p}_1^\mathsf{T}\mathbf{Q} \\ v\mathbf{p}_3^\mathsf{T}\mathbf{Q} &= \mathbf{p}_2^\mathsf{T}\mathbf{Q} \ . \end{aligned} \tag{4.11}$$

Avec deux vues, nous obtenons au total 4 équations linéaires par rapport aux coordonnées de $\mathbf{Q}$, qui peuvent être mises ensemble en un système d'équations de la forme $\mathsf{A}\mathbf{Q} = \mathbf{0}$, où A est une matrice 4×4. Ces équations définissent $\mathbf{Q}$ seulement à un facteur d'échelle près et nous cherchons une solution non-nulle pour $\mathbf{Q}$. Bien entendu, en présence de bruit dans les données, les équations ne seront pas satisfaites exactement, et nous cherchons une « meilleure » solution. Nous décrivons maintenant deux méthodes pour ce faire.

La première consiste à trouver $\mathbf{Q}$ qui minimise $||\mathsf{A}\mathbf{Q}||$, sous la contrainte $||\mathbf{Q}|| = 1$. La solution est le vecteur propre unitaire qui correspond à la plus petite valeur propre de la matrice $\mathsf{A}^\mathsf{T}\mathsf{A}$. Il peut être déterminé en utilisant une décomposition en valeurs singulières (SVD) ou la méthode de Jacobi pour trouver les valeurs propres de matrices symétriques [159]. Nous appelons ce procédé la méthode LIN-PROPRE.

Une deuxième approche consiste à fixer une coordonnée de $\mathbf{Q}$, par exemple en écrivant :

$$\mathbf{Q} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \ .$$

Ainsi, nous réduisons l'ensemble des équations homogènes $\mathsf{A}\mathbf{Q} = \mathbf{0}$ à un système de 4 équations non-homogènes en 3 inconnues. La solution aux moindres carrés peut être trouvée par la méthode des pseudo-inverses ou par une décomposition en valeurs singulières. Nous appelons cette approche la méthode LIN-MC.

Les deux méthodes sont similaires, mais elles ont par contre des propriétés assez distinctes en présence de bruit dans les données. La méthode LIN-MC suppose que le point $\mathbf{Q}$ n'est pas à l'infini puisque sa quatrième coordonnée est fixée à 1. Ceci peut être désavantageux dans le cas d'une reconstruction projective où des points reconstruits peuvent se trouver sur le plan à l'infini.

Outre ceci, aucune des deux méthodes n'est appropriée à la reconstruction projective puisqu'elles ne sont pas projectivement invariantes. Afin de se rendre compte de ceci, supposons que les matrices de projection P_1 et P_2 sont remplacées par $\mathsf{P}_1\mathsf{T}^{-1}$ et $\mathsf{P}_2\mathsf{T}^{-1}$. La matrice du système d'équations A devient alors $\mathsf{A}\mathsf{T}^{-1}$. Un point $\mathbf{Q}$ tel que $\mathsf{A}\mathbf{Q} = \epsilon$ pour le problème original correspond à un point $\mathsf{T}\mathbf{Q}$ satisfaisant $(\mathsf{A}\mathsf{T}^{-1})(\mathsf{T}\mathbf{Q}) = \epsilon$ pour le problème transformé. Donc, il y a une correspondance bijective entre les points $\mathbf{Q}$ et les points $\mathsf{T}\mathbf{Q}$ qui mène à la même erreur. Pourtant, ni la condition $||\mathbf{Q}|| = 1$ ni

$Q_4 = 1$ ne sont invariantes par application d'une transformation projective T. Par conséquent, si un point $\mathbf{Q}$ résout le problème original, le point transformé $\mathsf{T}\mathbf{Q}$ ne sera généralement pas la solution du problème transformé.

La situation est différente pour les transformations affines. Bien que la condition $||\mathbf{Q}|| = 1$ ne soit pas préservée par les transformations affines, la deuxième paramétrisation, $\mathbf{Q} = (X, Y, Z, 1)^{\mathsf{T}}$ l'est, puisque pour une transformation affine T : $\mathsf{T}(X, Y, Z, 1)^{\mathsf{T}} = (X', Y', Z', 1)^{\mathsf{T}}$. Ceci veut dire qu'il y a une correspondance bijective entre les vecteurs $\mathbf{Q} = (X, Y, Z, 1)^{\mathsf{T}}$ (avec $\mathsf{A}(X, Y, Z, 1)^{\mathsf{T}} = \epsilon$) et les vecteurs $\mathsf{T}\mathbf{Q} = (X', Y', Z', 1)^{\mathsf{T}}$ (avec $(\mathsf{A}\mathsf{T}^{-1})(X', Y', Z', 1)^{\mathsf{T}} = \epsilon$). L'erreur est la même pour des points correspondants. Donc, les points qui minimisent l'erreur $||\epsilon||$ se correspondent, ce qui signifie que la méthode LIN-MC est invariante aux transformations affines tandis que LIN-PROPRE ne l'est pas. Ces conclusions sont confirmées par les résultats des expériences.

4.7.2 Méthodes linéaires itératives

Un inconvénient des méthodes LIN-MC et LIN-PROPRE est que la valeur minimisée $||\mathsf{A}\mathbf{Q}||$ n'a pas de signification géométrique. Surtout, elle ne correspond pas au critère des erreurs de reprojection (4.1). En outre, une pondération individuelle des équations homogènes (lignes de A) changera la solution. L'idée de base pour les méthodes présentées ci-dessous est de changer les poids des équations de manière itérative afin de faire correspondre la valeur minimisée à l'erreur de reprojection.

Considérons la première des équations (4.11). En général, le point $\mathbf{Q}$ trouvé ne satisfera pas cette équation – mais engendre une erreur $\epsilon = u\mathbf{p}_3^{\mathsf{T}}\mathbf{Q} - \mathbf{p}_1^{\mathsf{T}}\mathbf{Q}$. La quantité à minimiser est la différence entre les coordonnées u du point mesuré et de celui reprojeté ($\mathbf{p}_1^{\mathsf{T}}\mathbf{Q}/\mathbf{p}_3^{\mathsf{T}}\mathbf{Q}$). Concrètement, nous voulons minimiser :

$$\epsilon' = \epsilon/\mathbf{p}_3^{\mathsf{T}}\mathbf{Q} = u - \mathbf{p}_1^{\mathsf{T}}\mathbf{Q}/\mathbf{p}_3^{\mathsf{T}}\mathbf{Q} \quad .$$

Ceci signifie que si l'équation avait été pondérée par le facteur $1/w$, où :

$$w = \mathbf{p}_3^{\mathsf{T}}\mathbf{Q} \quad ,$$

l'erreur résultante aurait été précisément ce que nous voulons minimiser. Le même poids $1/w$ est à attribuer à la deuxième équation de (4.11). Le poids correct pour les équations associées à la deuxième image, $1/w'$, est déterminé de manière analogue.

Bien entendu, nous ne pouvons pas pondérer les équations au préalable parce que les poids dépendent des coordonnées du point à reconstruire $\mathbf{Q}$ que nous ne connaissons qu'après avoir résolu les équations ! C'est pourquoi nous procédons de manière itérative pour adapter les poids. Nous commençons en mettant $w_0 = w'_0 = 1$, et utilisons une des méthodes LIN-PROPRE ou LIN-MC pour trouver une solution initiale $\mathbf{Q}_0$. Ensuite, nous pouvons mettre à jour les poids.

Ce processus est répété plusieurs fois, multipliant à la i^{e} itération les équations (4.11) par $1/w_i$ (respectivement $1/w'_i$) où w_i (respectivement w'_i) est déterminé à l'aide de la solution $\mathbf{Q}_{i-1}$ de l'itération précédente. Après quelques itérations, ce processus va converger (on l'espère), auquel cas nous aurons $\mathbf{Q}_i = \mathbf{Q}_{i-1}$ et alors $w_i = \mathbf{p}_3^{\mathsf{T}}\mathbf{Q}_i$. L'erreur (pour la première équation de (4.11) par exemple) approchera l'erreur de reprojection : $\epsilon_i \approx u - \mathbf{p}_1^{\mathsf{T}}\mathbf{Q}_i/\mathbf{p}_3^{\mathsf{T}}\mathbf{Q}_i$.

Cette approche peut se baser et sur LIN-PROPRE et sur LIN-MC. Nous appelons les méthodes résultantes IT-PROPRE et IT-MC. L'avantage de ces méthodes par rapport à d'autres méthodes de moindres carrés non linéaires comme Levenberg-Marquardt (LM) [159] est qu'elles sont très faciles à implémenter. En effet, leur implémentation ne requiert qu'une adaptation triviale des méthodes linéaires. Il n'y a pas besoin d'une méthode d'initialisation séparée comme c'est souvent nécessaire pour LM. En outre, la décision du moment auquel on peut arrêter les itérations (convergence) est simple : on termine le processus quand le changement des poids devient petit. Le moment d'arrêt n'est pas crucial puisque le point reconstruit $\mathbf{Q}$ n'est pas très sensible à des petits changements des poids.

L'inconvénient de ces méthodes est une divergence occasionnelle. Dans des situations instables, surtout quand les points sont proches des épipôles, ce phénomène se manifeste suffisamment souvent pour être un problème (dans environ 5 % des cas lors de nos simulations). Si ces méthodes sont utilisées dans de telles circonstances, une méthode de secours est indispensable. Pour nos expériences, nous avons utilisé la méthode optimale, POLY, lorsqu'il n'y avait pas convergence après 10 itérations. Ainsi, l'évaluation statistique des résultats n'est pas biaisée par quelques très mauvais résultats dus à une divergence occasionnelle.

Les méthodes itératives héritent des méthodes linéaires correspondantes leurs propriétés d'invariance ou de non-invariance à des transformations. Notamment, IT-MC et IT-PROPRE ne sont pas projectivement invariantes. Pourtant, les expériences ci-dessous montrent que ces méthodes sont relativement insensibles aux transformations projectives. La méthode IT-MC est, tout comme LIN-MC, invariante aux transformations affines. Les expériences montrent que les méthodes itératives sont considérablement meilleures que les méthodes linéaires non-itératives sous-jacentes.

4.8 Résultats expérimentaux

4.8.1 Expériences sur des données synthétiques

Un grand nombre d'expériences ont été menées pour évaluer les différentes méthodes de triangulation décrites ci-dessus. Nous nous sommes concentrés sur deux configurations.

Configuration 1. La première configuration simule une situation qui peut apparaître par exemple, quand un robot longe un couloir en regardant droit devant. Ceci est esquissé dans la figure 4.4 (a). Dans ce cas, les deux épipôles sont très proches des centres des images. Cette situation est connue pour être intrinsèquement instable pour la reconstruction : la profondeur de points qui se trouvent *exactement* sur la droite reliant les centres optiques ne peut être déterminée, et la reconstruction des points *proches* de cette droite devient très sensible même à un petit bruit dans les coordonnées image. Des expériences ont été effectuées simulant des objets à des distances variées des caméras.

Les grandeurs suivantes sont utilisées :

- ▷ la distance entre les deux caméras est de 1 unité ;
- ▷ les points 3D sont placés aléatoirement (distribution uniforme) dans une sphère d'un rayon de 0,05 unités ;
- ▷ la distance entre le centre de la sphère des points et le centre optique de la caméra frontale est de 0,15 ou 0,55 unités. Le centre de la sphère est sur la droite des centres optiques des caméras ;
- ▷ les deux caméras ont les mêmes paramètres intrinsèques :

$$\mathsf{K} = \begin{pmatrix} 700 & 0 & 0 \\ 0 & 700 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

Configuration 2. La deuxième configuration correspond à peu près à une tête stéréo, observant un objet équidistant des deux caméras (voir la figure 4.4 (b)). La plupart des méthodes donnent de relativement bons résultats pour cette configuration.

Les grandeurs suivantes sont utilisées :

- ▷ la distance entre les deux caméras est de 1 unité ;
- ▷ le rayon de la sphère des points 3D est de 1 unité ;
- ▷ la distance entre le centre de la sphère et la droite reliant les centres optiques varie entre 1 et 9 unités ;
- ▷ les caméras ont les mêmes paramètres intrinsèques comme dans la configuration 1 ;
- ▷ les axes optiques des caméras forment un angle de 80° avec la droite reliant les centres optiques.

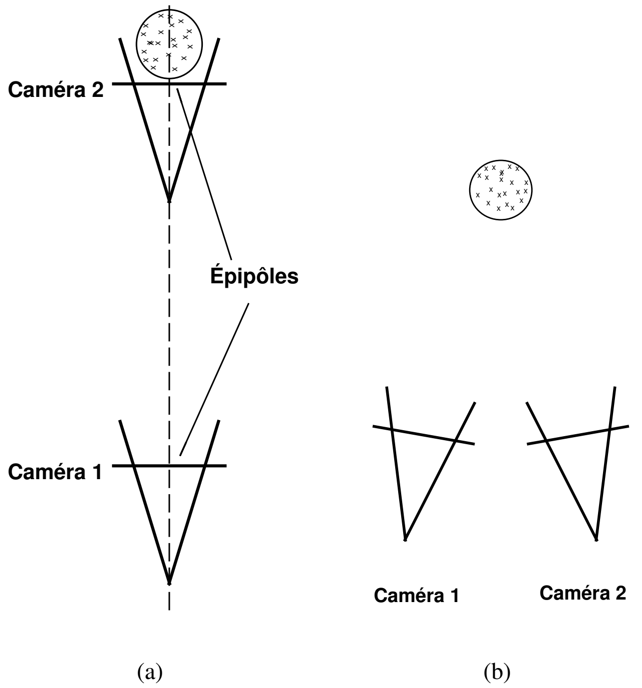


FIG. 4.4: *Les configurations pour les simulations.*

Pour chaque ensemble d'expériences, 50 points sont choisis aléatoirement dans le champ de vue commun. Les projections des points dans les images sont perturbées par un bruit gaussien dont l'écart-type varie entre 1 et 10 pixels (la résolution de l'image est de 700×700). Pour chaque niveau de bruit, les points sont perturbés et ensuite reconstruits à 100 reprises. Pour chaque point reconstruit nous mesurons l'erreur 3D et le résidu de la reprojection. Dans les graphes des pages suivantes sont représentées les valeurs médianes de ces erreurs. Les moyennes sont aussi calculées, mais les graphes ne sont pas représentés – ils ont la même forme et donnent lieu aux mêmes conclusions, bien qu'ils soient moins lisses que les graphes montrés, dû à d'occasionnelles erreurs grossières.

Pour évaluer l'invariance à des transformations, une transformation affine ou projective a été appliquée aux matrices de projection. Ces transformations ont été choisies pour que l'une des matrices de projection ait la forme $(\ \mathtt{I} \ |\mathbf{0})$. C'est la forme normalisée couramment utilisée pour la reconstruction projective. Elle représente une distorsion significative puisque la vraie matrice de projection est $(\ \mathsf{M} \ |\mathbf{0})$, où M est la matrice diagonale $\mathrm{diag}(700, 700, 1)$.

La situation la plus instable est la configuration 1, où les épipôles sont au centre des images et les points image proches des épipôles. Nous présentons des résultats pour cette situation puisqu'elle constitue un défi sérieux pour les algorithmes. Nous donnons des résultats pour les deux cas où les points 3D se trouvent à une distance de 0,15 unités (*points proches*) ou de 0,55 unités (*points éloignés*) de la caméra frontale. Les résultats sont présentés en forme de graphes commentés. Les erreurs mesurées sont étiquetées comme *erreurs 2D* (résidu de la reprojection) ou *erreurs 3D*. Chaque point dans un graphe représente la moyenne de 100 erreurs médianes pour 50 points. Les axes horizontaux des graphes montrent le niveau de bruit et les axes verticaux représentent l'erreur mesurée, en pixels pour les erreurs 2D ou en unités 3D pour les erreurs 3D.

Dans la suite, nous présentons des résultats pour des reconstructions euclidiennes, projectives et affines, en forme de graphes commentés. Pour les reconstructions affines et projectives, les matrices de projection ont été transformées comme décrit ci-dessus. Ensuite, la triangulation a été effectuée et finalement les points reconstruits ont été retransformés dans le repère original pour les comparer aux points originaux. Pour les reconstructions euclidiennes, aucune transformation n'a été appliquée.

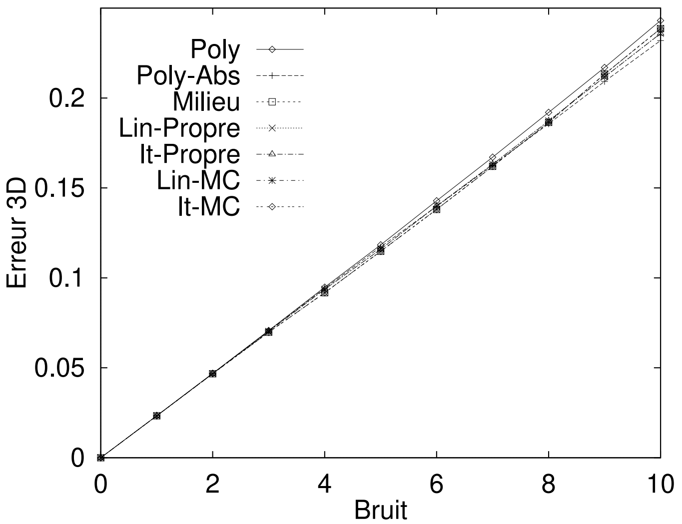


FIG. 4.5: *Erreur 3D pour la reconstruction euclidienne (points proches). La méthode* POLY *est légèrement plus mauvaise que les autres. Elle est conçue pour minimiser l'erreur 2D, donc elle n'optimise pas obligatoirement l'erreur 3D. La reconstruction euclidienne est le seul cas où la méthode du* POINT MILIEU *a donné des résultats meilleurs que* POLY *et* POLY-ABS.

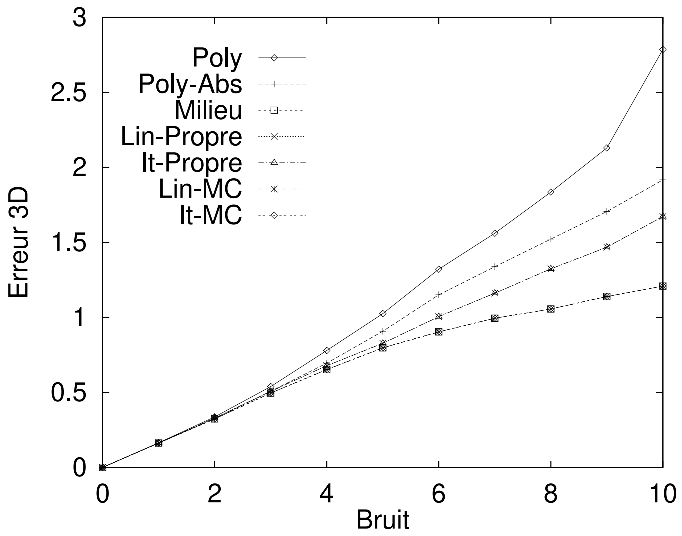


FIG. 4.6: *Erreur 3D pour la reconstruction euclidienne (points éloignés). La configuration est la même que pour le graphe 4.5, si ce n'est que les points sont plus éloignés des caméras. Sont visibles du bas vers le haut :* LIN-MC, IT-MC *et* POINT MILIEU *qui sont presque indistincts ;* LIN-PROPRE *et* IT-PROPRE *sont aussi identiques ; puis* POLY-ABS *et* POLY.

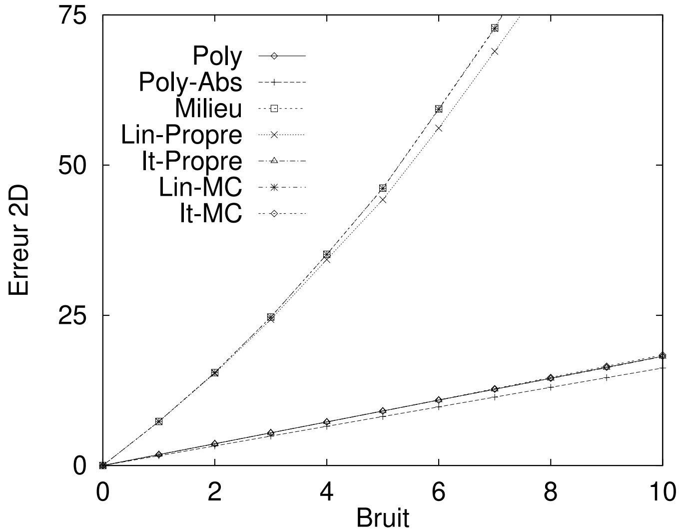


FIG. 4.7: *Erreur médiane 2D pour la reconstruction euclidienne (points proches). La configuration est la même que pour le graphe 4.5. Bien entendu,* POLY-ABS *est la meilleure méthode puisqu'elle est spécialement conçue pour minimiser l'erreur 2D.* POLY, IT-MC *et* IT-PROPRE *sont légèrement plus mauvaises (les graphes se superposent). Les méthodes* LIN-PROPRE, LIN-MC *et* POINT MILIEU *donnent de très mauvais résultats (notons que l'axe vertical compte 75 pixels d'erreur). Ce graphe, en comparaison avec le graphe 4.5, montre que les erreurs 2D et 3D ne sont pas directement liées : malgré de grandes erreurs 2D pour quelques méthodes, toutes les méthodes ont une bonne performance dans le 3D.*

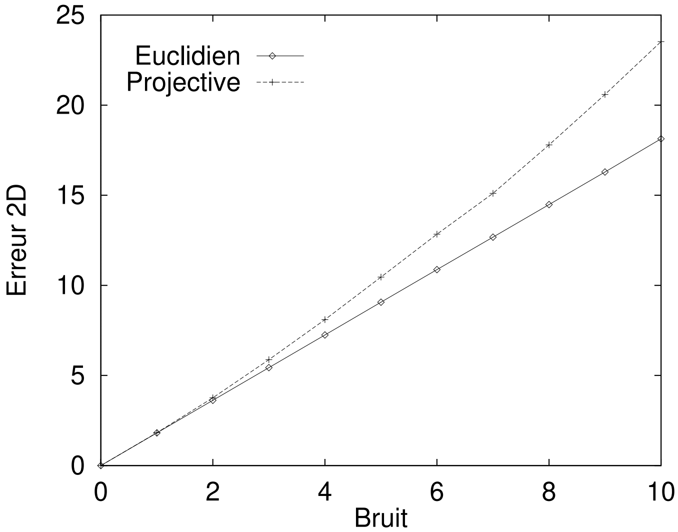


FIG. 4.8: *Comparaison de l'erreur 2D pour les reconstructions euclidiennes et projectives. La méthode utilisée est* IT-PROPRE. *Le graphe montre que cette méthode est presque projectivement invariante (c'est-à-dire que les deux courbes sont très proches l'une de l'autre).* IT-PROPRE *est une excellente méthode, à l'exception de problèmes de convergence dans des situations très instables (dans environ 1 % des essais pour un niveau de bruit de 2 pixels). Les essais où il n'y avait pas convergence ne sont pas inclus dans le graphe. Si les points ne sont pas proches des épipôles, il n'y a plus de problème de divergence. La méthode* IT-MC *(non montrée) a un comportement légèrement plus mauvais, mais tout de même similaire tandis que* POLY *est exactement projectivement invariante (les deux courbes se superposent).*

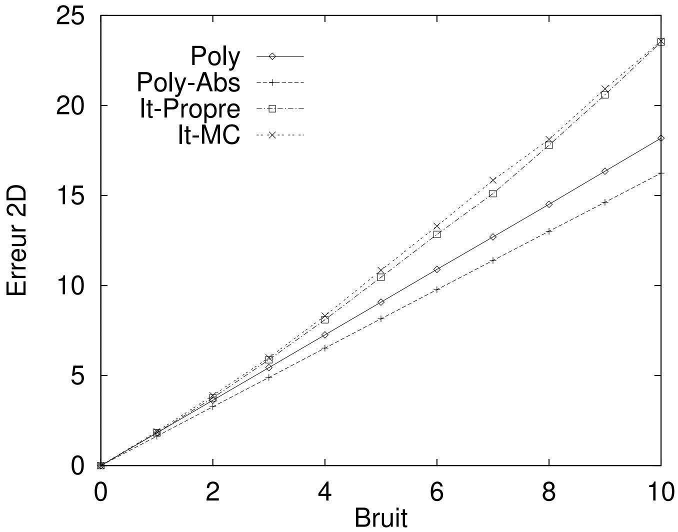


FIG. 4.9: *Erreur 2D pour la reconstruction projective (points proches). C'est la même configuration où toutes les méthodes ont donné de bons résultats pour la reconstruction euclidienne. Le graphe montre les résultats des méthodes (du bas vers le haut)* POLY-ABS, POLY, IT-PROPRE *et* IT-MC.

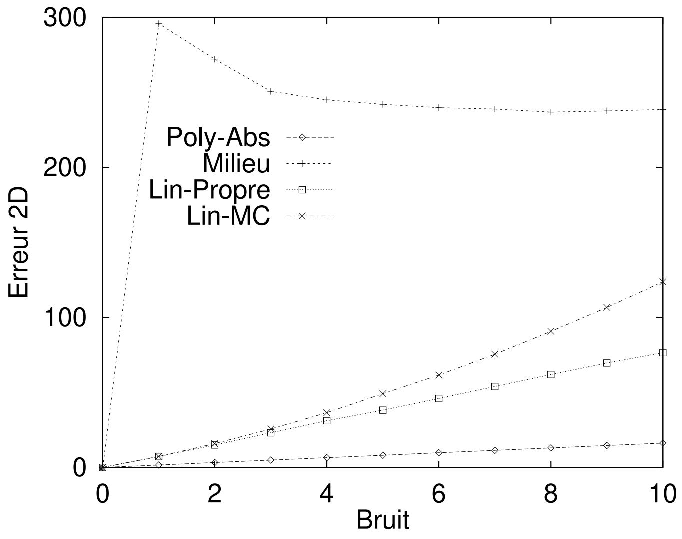


FIG. 4.10: *Erreur 2D pour la reconstruction projective (points proches), suite. Ici les mauvaises méthodes pour la même configuration comme pour le graphe 4.9 sont montrées. Les courbes représentent (du bas vers le haut)* POLY-ABS *(comme référence),* LIN-PROPRE, LIN-MC *et* POINT MILIEU. *Le graphe démontre la sensibilité des méthodes au problème de non-invariance à des transformations.*

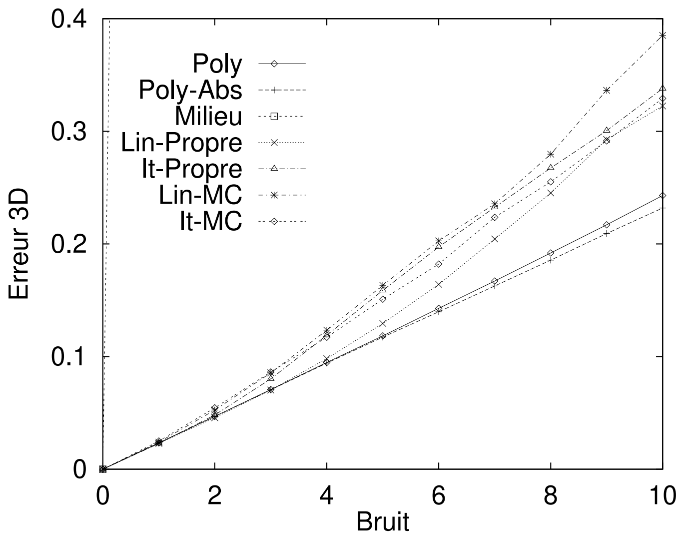


FIG. 4.11: *Erreur 3D pour la reconstruction projective (points proches). Graphe pour la même configuration que pour les graphes 4.9 et 4.10.* POLY-ABS *est un peu meilleure que* POLY. *Suivent les méthodes* LIN-PROPRE, IT-MC, IT-PROPRE *et* LIN-MC. *La courbe pour* POINT MILIEU *sort du graphe déjà pour un niveau de bruit de 1 pixel.*

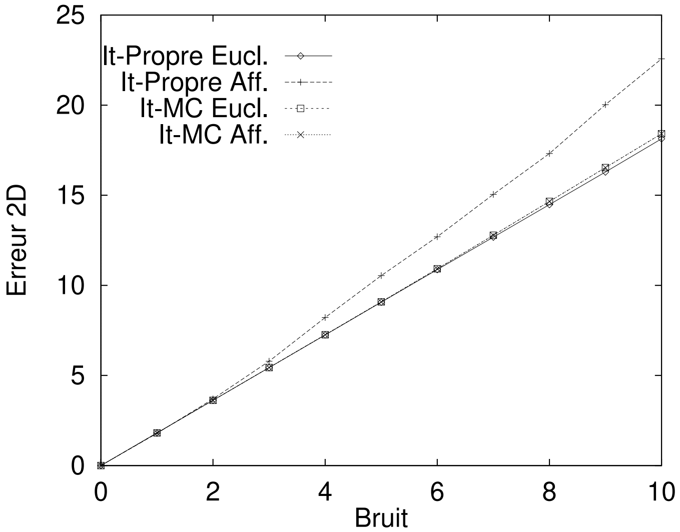


FIG. 4.12: *Invariance à des transformations affines. Les courbes montrées représentent* IT-PROPRE *(reconstruction euclidienne),* IT-MC *(les courbes pour les reconstructions euclidienne et affine se superposent) et* IT-PROPRE *(reconstruction affine). Nous pouvons constater que, comme il a été prédit théoriquement, la méthode* IT-MC *est invariante aux transformations affines, contrairement à* IT-PROPRE*, qui l'est presque.*

4.8.2 Expériences sur des données réelles

Nous avons testé les méthodes avec des images de la séquence du château [1] (voir la figure 4.13 pour une image de la séquence). Elle est constituée de 11 images et de 28 correspondances de points image. Les images sont prises des points de vue affichés dans la figure 4.14, avec des axes optiques parallèles. La vérité de terrain pour les coordonnées tridimensionnelles des points est disponible ce qui nous permet de quantifier la qualité de la reconstruction.

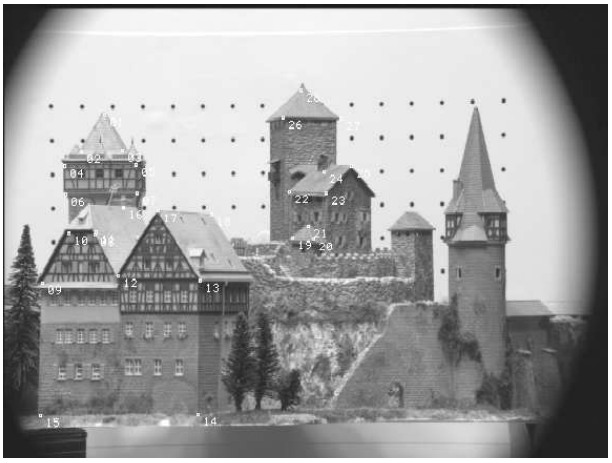


FIG. 4.13: *Une image de la séquence du château.*

1. Nous remercions le « Calibrated Imaging Laboratory » de l'Université Carnegie Mellon qui a fourni cette séquence avec le soutien de ARPA, NSF et NASA. La séquence est disponible à http://www.cs.cmu.edu/ cil/cil-ster.html.

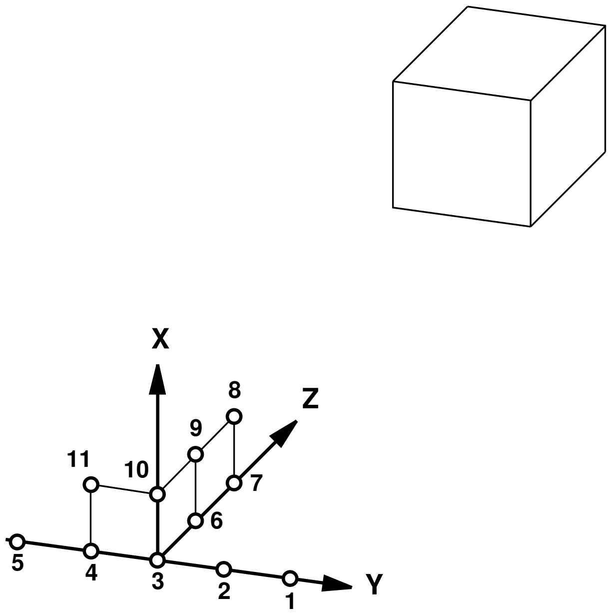


FIG. 4.14: *Les points de vue de la séquence du château (la figure ne montre pas les distances corrects entre les points de vue et l'objet).*

Nous avons effectué l'expérience suivante pour chaque paire de vues de la séquence.

1. Calcul de la matrice fondamentale (méthode des 8 points [71]).
2. Calcul de deux matrices de projection à partir de la matrice fondamentale.
3. Reconstruction projective des 28 points avec différentes méthodes de triangulation.
4. Alignement de chaque reconstruction projective avec le modèle euclidien.
5. Calcul de l'erreur des reconstructions projectives comme étant la moyenne des distances absolues entre points alignés et points du modèle.

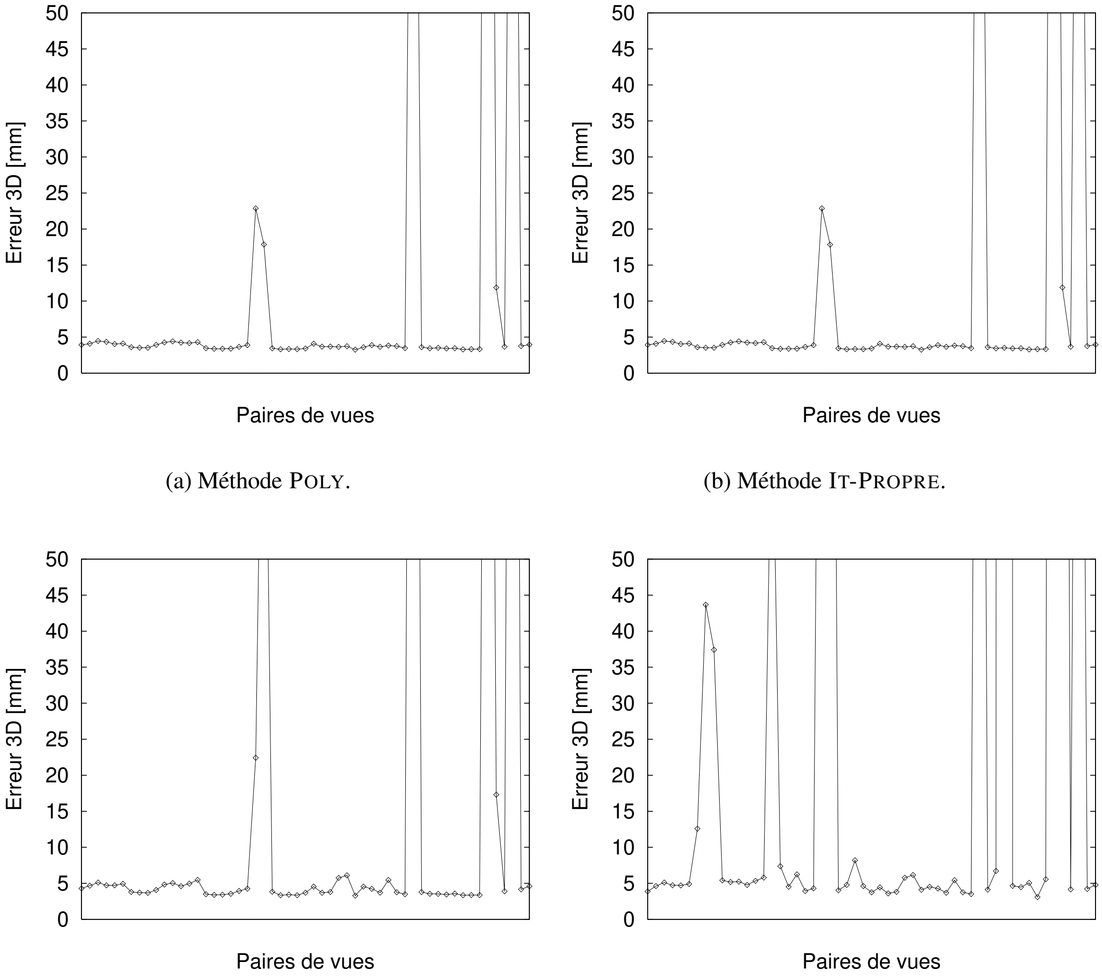
Les résultats des méthodes suivantes sont affichés dans la figure 4.15 : POLY, IT-PROPRE, IT-MC et POINT MILIEU.

Les méthodes POLY et IT-PROPRE ont un comportement quasiment identique. Pour quelques paires de vues, les deux méthodes échouent. De gauche à droite (par rapport aux graphes dans la figure 4.15), il s'agit des paires :

$$(3,6) \quad (3,7) \quad (6,7) \quad (8,9) \quad (8,10) \quad (9,10) \ .$$

Si l'on compare avec la figure 4.14, on constate qu'il s'agit exactement des paires de vues qui constituent une translation pure de la caméra. Nous avons déjà mentionné que ce cas est toujours difficile pour la reconstruction. Pour les autres mouvements, les résultats sont comparables avec ceux d'une reconstruction projective en tenant compte de *toutes* les 11 images (cf. les résultats de notre méthode de factorisation en section 5.8.2.2).

La méthode IT-MC donne des résultats légèrement moins bons que POLY et IT-PROPRE. Pour ce qui est de la méthode POINT MILIEU, la reconstruction pour la plupart des paires de vues est bonne, mais le taux d'échec est clairement supérieur à celui des autres méthodes présentées.



(a) Méthode POLY. (b) Méthode IT-PROPRE.

(c) Méthode IT-MC. (d) Méthode POINT MILIEU.

FIG. 4.15: *Résultats de reconstructions projectives avec la séquence du château.*

4.8.3 Discussion des résultats

Avant d'analyser les résultats précédents, nous rappelons, dans le tableau 4.2, les abréviations des méthodes utilisées ainsi que très brièvement leur procédé.

Toutes les méthodes se comportent bien pour la reconstruction euclidienne, en termes d'erreurs 3D. Quant aux erreurs 2D, seules les méthodes POLY, POLY-ABS, IT-MC et IT-PROPRE donnent des résultats acceptables. Les deux dernières ont l'inconvénient de diverger occasionnellement. La méthode POLY-ABS semble être légèrement meilleure que POLY, au sens des erreurs 3D, mais toutes les deux sont des méthodes excellentes. Leur inconvénient majeur est qu'elles ne sont pas directement généralisables à plus de 2 vues (voir la discussion dans la section 4.5.5), contrairement aux autres approches. Elles sont également plus lentes, mais toujours assez rapides pour pouvoir reconstruire des milliers de points par seconde ce qui devrait être suffisant pour la plupart des applications envisageables.

POLY	Minimise la somme des carrés des distances entre points image mesurés et points reprojetés
POLY-ABS	Minimise la somme des distances absolues entre points image mesurés et points reprojetés
POINT MILIEU	Choisit le point milieu de la perpendiculaire commune des rayons de projection associés aux points image
LIN-MC	Détermine le point 3D $\mathbf{Q}$ en résolvant un système de 4 équations linéaires et homogènes en 4 inconnues, utilisant la contrainte $\|\|\mathbf{Q}\|\| = 1$
IT-MC	Variante itérative de la méthode précédente ; adapte les poids des équations linéaires
LIN-PROPRE	Détermine le point 3D $\mathbf{Q}$ en résolvant un système de 4 équations linéaires en 3 inconnues, ayant fixé une coordonnée de $\mathbf{Q}$
IT-PROPRE	Variante itérative de la méthode précédente ; adapte les poids des équations linéaires

TAB. 4.2: *Récapitulatif des méthodes de triangulation utilisées.*

En résumé :

POLY C'est la méthode recommandée si seulement 2 images sont traitées et si le temps de calcul n'est pas critique. Ses performances sont supérieures à celles des autres méthodes, exceptée POLY-ABS. En fait, elle est optimale sous l'hypothèse d'un bruit gaussien. La méthode est projectivement invariante.

POLY-ABS Cette méthode garantit de trouver le minimum global de la somme des erreurs absolues de reprojection. Ceci est peut-être un meilleur modèle pour le bruit dans les images ; elle est moins sensible à de larges erreurs. Elle donne des reconstructions légèrement meilleures que POLY, mais sinon les deux méthodes se comportent de manière très similaire. POLY-ABS est également projectivement invariante.

POINT MILIEU C'est la méthode classique pour la reconstruction euclidienne. Malgré une bonne performance dans ce cas, elle n'est pas meilleure que les autres méthodes linéaires, et pour la reconstruction projective, les résultats sont inutilisables. La méthode n'est invariante ni aux transformations projectives ni aux transformations affines.

LIN-PROPRE Son avantage majeur est la vitesse et la simplicité. Elle n'est invariante ni aux transformations projectives ni aux transformations affines.

LIN-MC Cette méthode est invariante aux transformations affine, mais ne devrait pas être utilisée pour une reconstruction projective.

IT-PROPRE Cette méthode donne de très bons résultats, considérablement meilleurs que la méthode non-itérative correspondante, LIN-PROPRE. Elle n'est pas aussi bonne que POLY ou POLY-ABS, mais son grand avantage est qu'elle est facilement généralisable au cas de plusieurs images. L'inconvénient de la méthode est une divergence occasionnelle ce qui a pour conséquence le besoin d'une méthode de secours.

IT-MC Elle est similaire à LIN-PROPRE du point de vue performance et propriétés. Elle est invariante aux transformations affines, mais ne devrait pas être appliquée pour la reconstruction projective puisque les points à l'infini ne peuvent pas être traités.

4.9 Conclusion

Dans ce chapitre, nous avons examiné le problème de la reconstruction à partir de deux vues, étant données les matrices de projection. Nous avons montré que les méthodes classiques sont inadaptées au cas de la reconstruction projective ou affine, dû à l'utilisation de principes qui ne sont valables que dans l'espace euclidien. Un critère de minimisation pour la reconstruction a été proposé qui combat le bruit là où il se manifeste – dans la position des primitives dans les images. Nous avons développé une méthode (POLY) qui minimise ce critère globalement, ainsi donnant une reconstruction optimale.

Pour ce qui est de l'optimalité de notre méthode, il faut préciser qu'elle donne le résultat qui est optimal pour les matrices de projection données. Une optimisation non linéaire ultérieure, portant sur les matrices de projection et les points 3D simultanément, va en général encore améliorer la reconstruction (cf. aussi le paragraphe sur l'estimation de la matrice fondamentale en 2.3.2).

La nouvelle méthode a été testée en concurrence avec d'autres méthodes. Elle donne les résultats les plus stables, mais les méthodes linéaires itératives sont également très performantes. L'atout de POLY est son optimalité statistique. Nous avons aussi décrit des simplifications de notre méthode pour des mouvements particuliers entre les deux vues. Contrairement aux méthodes linéaires, elle n'est pas généralisable à plus de deux images.

Des opérations similaires à la correction des points image dans notre méthode sont éventuellement possibles pour d'autres primitives. Dans [39], Cross et Zisserman effectuent une correction de la position de coniques dans les images comme étape préalable à la reconstruction de quadriques. Leur méthode n'est pas optimale mais les coniques sont données avec une haute précision ce qui fait qu'une correction sous-optimale peut être suffisante. Une étude plus profonde de ce cas serait intéressante à mener.

5.

Reconstruction projective multi-image

Dans ce chapitre, nous présentons une méthode de reconstruction (et calibrage) projective à partir de plusieurs images. La méthode est basée sur la factorisation d'une matrice contenant les coordonnées de tous les points image vus dans toutes les images. Notre méthode est une extension des méthodes de Tomasi, Poelman et Kanade [202, 149] des projections affines aux projections perspectives. Contrairement au cas affine, la factorisation n'est possible que si les coordonnées homogènes des points image sont correctement échelonnées. L'une des contributions de notre travail est une méthode pour la détermination des facteurs d'échelle nécessaires, à partir des seules matrices fondamentales. Nous décrivons également des détails numériques qui contribuent de manière significative à la bonne performance de notre méthode de reconstruction. Les résultats d'expériences évaluant la méthode sont donnés en fin de chapitre.

Le travail décrit dans ce chapitre a été effectué en collaboration avec Bill Triggs et a été publié dans [192, 193].

5.1 Introduction

Le problème de la reconstruction projective a principalement été étudié pour le cas de deux vues (voir chapitre 4 pour une revue de quelques méthodes) ou de trois vues [81, 162, 175]. Il existe quelques méthodes pour des séquences d'images que nous passons en revue dans la suite. Grossièrement, on peut distinguer quatre types de méthodes. La plus simple, proposée par Hartley [79], consiste à reconstruire la structure projective à partir de deux vues et de déterminer les matrices de projection des autres vues en utilisant les points reconstruits qui sont visibles dans ces vues. Ensuite, les points non encore reconstruits peuvent être triangulés.

Mohr *et al.* et Szeliski & Kang décrivent des méthodes d'ajustement itératif non linéaire (ajustement de faisceaux) qui prennent en compte toutes les observations [129, 130, 196, 198]. Ce type de méthode nécessite généralement une bonne initialisation des paramètres, ici à la fois de la structure et du mouvement projectifs. Il semble que pour de courtes séquences d'images, une initialisation ad hoc peut être suffisante pour faire converger l'estimation itérative. Le comportement de ce procédé pour des séquences plus longues n'est pas clair. Un ajustement itératif peut toujours être appliqué comme étape finale de tout algorithme de reconstruction afin d'optimiser le résultat.

Laveau propose de fusionner des paires de reconstructions obtenues à partir de triplets de vues, en estimant l'homographie entre les nuages de points [109] (voir la section 5.7).

Finalement, il existe des approches récursives de filtrage qui sont initialisées par une reconstruction à partir de deux ou trois vues [14, 126, 217, 218].

Plus récemment, Heyden a proposé un formalisme qui permet de reconstruire simultanément plusieurs points à partir de plusieurs vues [85, 87]. La clé de ce formalisme est l'affectation de coordonnées canoniques à trois points, dans l'espace (reconstruction) et dans toutes les images. Les matrices fondamentales et les tenseurs trifocaux prennent alors une forme réduite (« reduced fundamental matrix/tensor »). Les matrices de projection correspondantes dépendent de seulement 5 paramètres et il est possible de les estimer simultanément avec la structure 3D par des méthodes linéaires. Un autre avantage de cette méthode est qu'elle ne nécessite pas que tous les points soient visibles dans toutes les vues. Par contre, le résultat est fortement biaisé par le choix des points de référence. En outre, l'affectation de coordonnées canoniques à ces points se fait par des transformations affines ce qui peut rendre la méthode de résolution mal conditionnée.

En résumé, plusieurs des approches mentionnées ont l'inconvénient de fixer le repère de reconstruction en privilégiant quelques points ou des vues, auxquels sont associées des coordonnées canoniques. La qualité de la reconstruction dépend de la qualité du repère choisi, par exemple de la précision de la localisation des projections de 5 points de référence nécessaires pour fixer le repère projectif 3D. Généralement, on peut dire qu'il n'existe pas de méthode qui prenne en compte toutes les données (les coordonnées image) simultanément et de manière unifiée.

De telles méthodes existent pour la reconstruction multi-image supposant un modèle de projection affine ; il s'agit là des méthodes de factorisation bien connues de Tomasi, Poelman et Kanade [202, 149] [1]. Le schéma est simple : les coordonnées de tous les points image sont stockées dans une seule matrice qui, ce qui est la clé de la méthode, est idéalement de rang 3. Cette contrainte permet de factoriser cette matrice en deux matrices, l'une contenant les coefficients des matrices de projection, l'autre contenant les coordonnées des points 3D reconstruits de manière affine. Ce procédé est décrit plus en détail en section 5.3.

Dans la suite du chapitre, nous proposons une méthode de reconstruction projective qui applique

1. Une méthode similaire pour la reconstruction affine de droites a été développée par Quan et Kanade [165, 166] ; voir aussi les méthodes de factorisation incrémentielle de Morita et Kanade [132, 133] et la méthode de Costeira et Kanade [38] pour des objets avec des mouvements différents.

le schéma de factorisation aux projections perspectives. La méthode est plus compliquée que dans le cas affine, parce que des facteurs d'échelle pour les coordonnées image doivent être déterminés avant la factorisation. Nous décrivons maintenant brièvement le formalisme qui permettra de résoudre ce problème.

Dans le cadre d'une étude systématique des relations algébriques entre plusieurs vues, Triggs a introduit un nouveau formalisme de reconstruction projective [207, 208]. Les coordonnées des projections d'un point 3D sont regroupées en un seul vecteur de longueur $3m$ (dans le cas de m projections perspectives sur des plans), appelé **vecteur image composite** (« joint image vector ») et appartenant à l'**espace image composite** (« joint image space »). Pour un ensemble de vues données, les vecteurs image composite formés des projections de tous les points 3D forment un sous-espace de l'espace image composite de dimension 4, appelé l'**image composite** (« joint image »). Cet espace est caractérisé par les contraintes d'appariement multi-linéaires entre les vues.

Malheureusement, les facteurs d'échelle des coordonnées homogènes des points image sont perdus lors de la projection (les points d'intérêt dans l'image sont donnés en coordonnées non homogènes). Un choix arbitraire pour ces facteurs a pour conséquence que les vecteurs image composites ne se trouvent plus dans l'image composite. Dans ce contexte, la reconstruction projective consiste à trouver les bonnes échelles pour les coordonnées homogènes des points image afin de placer les vecteurs image composite dans l'image composite. Ceci constitue une reconstruction implicite, mais il est possible d'expliciter la structure en forme de coordonnées tridimensionnelles habituelles.

La méthode de reconstruction projective proposée dans ce chapitre est basée sur ce formalisme. Il ne sera pas nécessaire de le maîtriser pour bien comprendre le fonctionnement de la méthode. L'étape essentielle de la méthode est la détermination de facteurs d'échelle pour les coordonnées image, appelés *profondeurs projectives*, à partir de matrices fondamentales. Nous montrons que, avec des profondeurs projectives correctes, la matrice contenant toutes les coordonnées image est de rang 4. La reconstruction explicite est effectuée par une factorisation de cette matrice en structure et mouvement projectifs. Donc, notre méthode peut être considérée comme l'extension des méthodes de Tomasi, Poelman et Kanade [202, 149] des projections affines aux projections perspectives.

Organisation de ce chapitre. Nous définissons d'abord le problème de la reconstruction projective. En 5.3, nous décrivons la méthode de factorisation pour la reconstruction affine avec des caméras affines, et nous montrons comment l'étendre au cas de vues perspectives. En 5.4, nous expliquons comment déterminer les profondeurs projectives – les facteurs d'échelle qui sont nécessaires pour pouvoir faire la factorisation. La section 5.5 décrit la méthode complète de reconstruction projective. Ensuite, nous discutons de différentes méthodes pour traiter des données incomplètes, c'est-à-dire des points qui ne sont pas vus ou appariés dans toutes les images d'une séquence. Le chapitre se termine avec des résultats d'expériences utilisant notre méthode de reconstruction, suivis de conclusions.

5.2 Le problème de la reconstruction projective

La reconstruction projective consiste, comme son nom l'indique, en la détermination de la structure projective de la scène observée, c'est-à-dire de la structure de la scène à une transformation projective (inconnue) près. Elle est possible avec des caméras absolument pas calibrées et sans aucune information a priori sur la scène ou le mouvement des caméras [57, 72]. Généralement, quand on parle de reconstruction projective, on suppose implicitement qu'aucune de ces informations n'est disponible. Les seules données sont des correspondances de primitives image. Comme la plupart des approches existantes, la méthode que nous proposons est conçue pour utiliser des points.

Les seules données pour la reconstruction projective sont donc des points image appariés $\mathbf{q}_{ip}$. Le **problème de la reconstruction projective** est la détermination de matrices de projection $(\mathsf{P}_i)_{i=1}^m$ et de points 3D $(\mathbf{Q}_p)_{p=1}^n$ qui satisfont l'équation :

$$\forall i,p \quad : \quad \mathsf{P}_i\mathbf{Q}_p \sim \mathbf{q}_{ip} \ . \tag{5.1}$$

À l'exception de situations dégénérées (surfaces critiques, plusieurs solutions pour des données minimales, caméra stationnaire), il existe une seule famille de solutions satisfaisant l'équation (5.1). Elles sont reliées entre elles par des transformations projectives : si $\{(\mathsf{P}_i)_{i=1}^m, (\mathbf{Q}_p)_{p=1}^n\}$ est une reconstruction projective, alors, pour toute transformation projective non singulière T, l'ensemble $\{(\mathsf{P}_i\mathsf{T}^{-1})_{i=1}^m, (\mathsf{T}\mathbf{Q}_p)_{p=1}^n\}$ en est aussi une.

En présence de bruit affectant la position des points image, une solution exacte n'est normalement pas possible et il faut définir un critère selon lequel on peut déterminer une reconstruction optimale. Le critère généralement admis est la minimisation de la somme des carrés des distances entre les points reprojetés et ceux mesurés. Il s'agit du critère de maximum de vraisemblance sous hypothèse d'un bruit gaussien (voir 3.7). L'utilisation de ce critère engendre une formulation non linéaire de la fonction de coût à minimiser et l'estimation se fait avec des algorithmes numériques adéquates. Il s'agit d'une méthode standard, introduite par Mohr *et al.* [129, 130], qui est la spécialisation de l'ajustement de faisceaux au cas non calibré. Le problème habituel des méthodes d'optimisation non linéaire est le besoin d'une bonne initialisation des paramètres. La recherche se concentre donc sur le développement de méthodes ne nécessitant pas une telle initialisation, qui sont sous-optimales mais qui fournissent un bon point de départ pour une éventuelle phase d'optimisation.

Nous proposons dans la suite une méthode de reconstruction projective qui prend en compte simultanément toutes les données et qui ne s'appuie pas sur une préférence de quelques vues ou points pour fixer le repère projectif de reconstruction.

5.3 Reconstruction par factorisation

5.3.1 Caméras affines

La reconstruction 3D par factorisation a été introduite par Tomasi et Kanade, dans le cas d'une séquence d'images orthographiques [202] (voir aussi l'approche contemporaine de Debrunner et Ahuja [43]). Le principe de leur méthode est très simple ; considérons l'équation de projection réduite d'une caméra affine (cf. 2.2.3.1) :

$$\begin{pmatrix} \mathbf{q}_{ip} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathsf{P}_i & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{Q}_p \\ 1 \end{pmatrix} \ ,$$

où $\mathbf{q}_{ip}$ et $\mathbf{Q}_p$ sont les coordonnées non homogènes des points image et 3D respectivement[2], et P_i est une matrice de projection de dimension 2×3. On peut regrouper les projections de n points dans m vues dans une seule grande équation matricielle :

$$\mathsf{W} \equiv \begin{pmatrix} \mathbf{q}_{11} & \mathbf{q}_{12} & \cdots & \mathbf{q}_{1n} \\ \mathbf{q}_{21} & \mathbf{q}_{22} & \cdots & \mathbf{q}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_{m1} & \mathbf{q}_{m2} & \cdots & \mathbf{q}_{mn} \end{pmatrix}_{2m\times n} = \begin{pmatrix} \mathsf{P}_1 \\ \mathsf{P}_2 \\ \vdots \\ \mathsf{P}_m \end{pmatrix}_{2m\times 3} \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 & \cdots & \mathbf{Q}_n \end{pmatrix}_{3\times n} \ .$$

2. Afin d'obtenir l'équation de projection réduite, les points dans chaque image et les points 3D doivent être transformés tels que leurs centres de gravité se trouvent à l'origine des repères correspondants.

La matrice W est appelée la **matrice des mesures** parce qu'elle contient toutes les données – les coordonnées des points image. W est le produit de la matrice de projection composite et de la matrice de structure composite, toutes deux de rang au plus 3 (ayant seulement 3 colonnes ou 3 lignes). Par conséquent, W est aussi de rang au plus 3. C'est l'observation clé pour la méthode de reconstruction proposée. Considérons une décomposition en valeurs singulières de W :

$$\mathsf{W} = \mathsf{U}_{2m\times n}\Sigma_{n\times n}\mathsf{V}^{\mathsf{T}}_{n\times n} \quad ,$$

où Σ est la matrice diagonale des valeurs singulières dont au plus 3 sont non nulles. Si l'on recompose W par multiplication des matrices du côté droit de l'équation précédente, il y a donc seulement 3 colonnes de U et 3 lignes de V^{T} qui interviennent dans les calculs (voir aussi la figure 5.1). Soient U' et V'^{T} les matrices composées de ces colonnes ou lignes et Σ' la matrice diagonale des 3 valeurs singulières non nulles. On observe alors l'égalité suivante :

$$\mathsf{W} = \mathsf{U}'_{2m\times 3}\Sigma'_{3\times 3}\mathsf{V}'^{\mathsf{T}}_{3\times n} \quad .$$

Soit $\Sigma' = \Sigma''\Sigma'''$ une décomposition quelconque de Σ'. Alors, avec les matrices $\mathsf{U}'' = \mathsf{U}'\Sigma''$ et $\mathsf{V}''^{\mathsf{T}} = \Sigma'''\mathsf{V}'^{\mathsf{T}}$, nous avons :

$$\mathsf{W} = \mathsf{U}''_{2m\times 3}\mathsf{V}''^{\mathsf{T}}_{3\times n} \quad .$$

Cette équation signifie que U'' et $\mathsf{V}''^{\mathsf{T}}$ constituent une reconstruction admissible de la scène : U'' joue le rôle de matrice de projection composite et $\mathsf{V}''^{\mathsf{T}}$ celui de matrice de structure composite. Cette interprétation est admissible puisque la reprojection des points 3D reproduit les points image (qui sont contenus dans W). La reconstruction n'est qu'affine puisque, pour toute transformation $\mathsf{T}_{3\times 3}$ non singulière, $\mathsf{U}''\mathsf{T}^{-1}$ et $\mathsf{T}\mathsf{V}''^{\mathsf{T}}$ interprètent aussi bien les données.

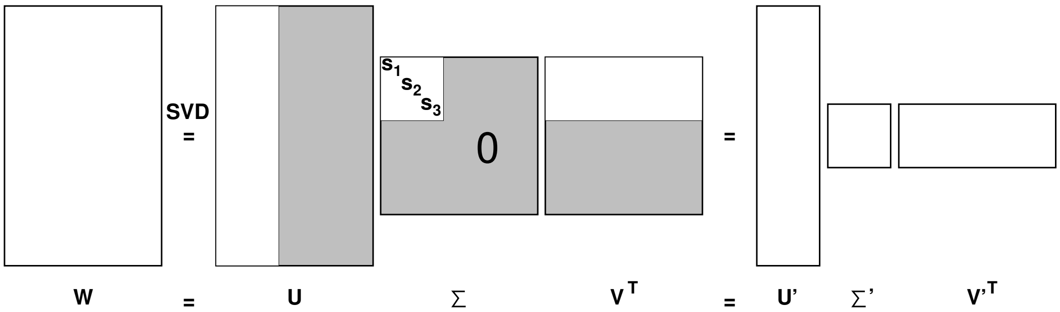


FIG. 5.1: *Principe de la factorisation. La matrice* W *est de rang 3 ce qui se manifeste par l'existence de seulement trois valeurs singulières non nulles* (s_1, s_2 *et* s_3)*. Les parties ombrées des matrices* U, Σ *et* V^{T} *sont donc « inutiles ».*

Jusqu'ici nous n'avons pas parlé de bruit dans les coordonnées image. En présence de bruit, la matrice des mesures ne sera plus de rang 3, mais généralement de rang plein. Pour combattre le bruit, on détermine alors une matrice $\widehat{\mathsf{W}}$ de rang 3 qui approche bien la matrice des mesures W. Ensuite on peut factoriser $\widehat{\mathsf{W}}$ comme décrit ci-dessus.

En pratique, la détermination de $\widehat{\mathsf{W}}$ et sa factorisation se font en une seule étape : on effectue une décomposition en valeurs singulières de W et on annule toutes les valeurs singulières sauf les trois plus grandes. On retient seulement les colonnes et lignes de U et V^{T} associées à ces trois valeurs singulières. Si l'on multiplie les matrices U'' et $\mathsf{V}''^{\mathsf{T}}$ obtenues, on obtient la matrice $\widehat{\mathsf{W}}$ de rang 3, qui approche au mieux W, au sens de la norme spectrale [68] et même de la norme de Frobenius [168] :

$$\widehat{\mathsf{W}} = \underset{\mathsf{W}',\text{rang}(\mathsf{W}')=3}{\text{argmin}} ||\mathsf{W} - \mathsf{W}'||_F \ .$$

Ce constat signifie que la reconstruction $\{(\mathsf{P}_i), (\mathbf{Q}_p)\}$ obtenue en factorisant W, est *optimale*, au sens du critère des erreurs de reprojection (voir la section 3.7), puisque $||\mathsf{W} - \widehat{\mathsf{W}}||_F = \sum_{i,p} d(\mathbf{q}_{ip}, \mathsf{P}_i\mathbf{Q}_p)^2$.

5.3.2 Caméras perspectives

Nous considérons maintenant le cas des caméras perspectives, où l'équation de projection prend la forme :

$$\lambda_{ip}\mathbf{q}_{ip} = \mathsf{P}_i\mathbf{Q}_p \ .$$

Contrairement au cas affine, nous ne pouvons pas éliminer les facteurs d'échelle λ_{ip}. En pratique, les points extraits d'une image sont donnés en coordonnées non homogènes, c'est-à-dire implicitement multipliées avec un scalaire tel que la troisième coordonnée est égale à 1 ; donc, le « vrai » facteur d'échelle n'est pas mesurable. Nous allons appeler les facteurs d'échelle inconnus λ_{ip} les **profondeurs projectives**[3]. Ce nom a été choisi parce que ces facteurs contiennent toute l'information nécessaire pour la reconstruction projective. Leur détermination est en effet équivalente à une reconstruction projective dans le formalisme de l'image composite [207, 208].

Les profondeurs projectives sont liées aux « vraies » profondeurs comme suit. Si $\mathbf{Q}_p$ et $\mathbf{q}_{ip}$ sont représentés en coordonnées étendues (1 comme dernière coordonnée) et si P_i est normalisée tel que le sous-vecteur $(\mathsf{P}_{i_{31}}, \mathsf{P}_{i_{32}}, \mathsf{P}_{i_{33}})$ est unitaire, alors la profondeur projective λ_{ip} est la vraie profondeur de $\mathbf{Q}_p$, c'est-à-dire sa distance du plan focal de la caméra.

L'ensemble des équations de projection peuvent être regroupées en une seule équation matricielle de dimension $3m \times n$:

$$\mathsf{W} \equiv \begin{pmatrix} \lambda_{11}\mathbf{q}_{11} & \lambda_{12}\mathbf{q}_{12} & \cdots & \lambda_{1n}\mathbf{q}_{1n} \\ \lambda_{21}\mathbf{q}_{21} & \lambda_{22}\mathbf{q}_{22} & \cdots & \lambda_{2n}\mathbf{q}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1}\mathbf{q}_{m1} & \lambda_{m2}\mathbf{q}_{m2} & \cdots & \lambda_{mn}\mathbf{q}_{mn} \end{pmatrix}_{3m\times n} = \begin{pmatrix} \mathsf{P}_1 \\ \mathsf{P}_2 \\ \vdots \\ \mathsf{P}_m \end{pmatrix}_{3m\times 4} \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 & \cdots & \mathbf{Q}_n \end{pmatrix}_{4\times n} \ . \tag{5.2}$$

En analogie au cas affine, nous appelons W la **matrice des mesures**, bien qu'elle contienne les facteurs d'échelle λ_{ip} qui ne sont pas mesurés. Il est clair que W est de rang au plus 4. Ceci n'est pourtant vrai que si les profondeurs projectives correctes sont utilisées. Malheureusement, les λ_{ip} sont perdues lors de la projection et regrouper les coordonnées image étendues (1 comme 3$^{\text{e}}$ coordonnée) résulte en une matrice W qui est généralement de rang plein et ne permet pas sa factorisation en mouvement et structure.

Si nous étions capables de déterminer les profondeurs projectives correctes, nous pourrions construire une matrice des mesures avec des coordonnées image correctement échelonnées, qui sera de rang 4. Ensuite, la structure 3D de la scène et le calibrage projectif des caméras peuvent être déterminés par le même schéma de factorisation que pour la caméra affine. Dans le cas présent, on obtient

3. Cette désignation n'a pas le même sens que les « profondeurs projectives » de Shashua, qui sont des birapports le long de rayons de projection [174].

une factorisation $\widehat{\mathsf{W}} = \mathsf{U}''_{3m\times 4}\mathsf{V}''^{\mathsf{T}}_{4\times n}$ qui est définie à une transformation projective 4×4 près : pour toute transformation T non singulière, la factorisation $\widehat{\mathsf{W}} = (\mathsf{U}''\mathsf{T}^{-1})(\mathsf{T}\mathsf{V}''^{\mathsf{T}})$ est aussi valable. Donc, la reconstruction obtenue par factorisation est une reconstruction projective.

Contrairement au cas de la factorisation affine, la reconstruction projective ne minimise pas les erreurs de reprojection et n'est donc pas optimale au sens de ce critère. La raison en est que la matrice W contient les coordonnées *homogènes* des points image, donc ce qui est minimisé est la somme des différences au carré des coordonnées homogènes, entre points mesurés et reprojetés. Ceci ne correspond pas à la distance géométrique. Les expériences montrent quand même un comportement très stable et des résultats précis, si des coordonnées normalisées sont utilisées (voir 5.5.1).

Dans le paragraphe suivant, nous décrivons une méthode d'estimation des profondeurs projectives à partir de matrices fondamentales et d'épipôles, qui, eux, sont estimés à partir des données image.

5.4 Détermination des profondeurs projectives

Nous décrivons comment les profondeurs projectives introduites dans le paragraphe précédent peuvent être déterminées à partir des tenseurs d'appariement, spécialement des matrices fondamentales et des épipôles. Notons d'abord qu'il n'y a pas un choix unique pour les mn profondeurs projectives : individuellement, les profondeurs projectives sont arbitraires, puisqu'elles dépendent des facteurs d'échelle des P_i et $\mathbf{Q}_p$, qui eux sont arbitraires puisqu'il s'agit d'entités homogènes. Pourtant, une fois ces $m + n$ facteurs d'échelle choisis, les mn profondeurs projectives sont fixées. Donc, prises dans leur ensemble, les profondeurs projectives ne sont pas arbitraires, mais doivent être cohérentes et contiennent donc réellement de l'information.

Réciproquement, il y a $m + n - 1$ degrés de liberté dans la détermination des profondeurs projectives. Considérons l'équation (5.2) : multiplier toutes les profondeurs associées aux images d'un point $\mathbf{Q}_p$ par le même scalaire a pour effet de multiplier les coordonnées de $\mathbf{Q}_p$ par ce scalaire. De manière analogue, multiplier le triplet de lignes de W associé à une vue i par un scalaire revient à multiplier la matrice de projection P_i par le même scalaire. Donc, les profondeurs projectives peuvent être déterminées modulo des rééchelonnements vue-par-vue (triplets de lignes de W) et point-par-point (colonnes de W). Cette liberté est utilisée pour améliorer la stabilité numérique de la factorisation (voir 5.5.2).

Nous décrivons maintenant des relations entre des matrices fondamentales, des épipôles, des points image et des profondeurs projectives, qui permettent de calculer ces dernières. L'équation de projection $\lambda_{ip}\mathbf{q}_{ip} = \mathsf{P}_i\mathbf{Q}_p$ implique que la matrice 6×5 :

$$\left(\begin{array}{c|c} \mathsf{P}_i & \lambda_{ip}\mathbf{q}_{ip} \\ \mathsf{P}_j & \lambda_{jp}\mathbf{q}_{jp} \end{array}\right)$$

est de rang au plus 4 (voir aussi la discussion dans 2.4) :

$$\left(\begin{array}{c|c} \mathsf{P}_i & \lambda_{ip}\mathbf{q}_{ip} \\ \mathsf{P}_j & \lambda_{jp}\mathbf{q}_{jp} \end{array}\right) = \left(\begin{array}{c|c} \mathsf{P}_i & \mathsf{P}_i\mathbf{Q}_p \\ \mathsf{P}_j & \mathsf{P}_j\mathbf{Q}_p \end{array}\right) = \left(\begin{array}{c} \mathsf{P}_i \\ \mathsf{P}_j \end{array}\right)\left(\begin{array}{c|c} \mathtt{I}_4 & \mathbf{Q}_p \end{array}\right) .$$

Donc, tous ses mineurs 5×5 s'annulent. En développant les mineurs par leurs éléments de dernière colonne, nous obtenons des équations homogènes et linéaires par rapport aux composantes de $\lambda_{ip}\mathbf{q}_{ip}$ et $\lambda_{jp}\mathbf{q}_{jp}$. Les coefficients de ces équations sont des déterminants 4×4 d'ensembles de lignes des matrices de projection. Ces déterminants, avec un certain choix de facteurs d'échelle relatifs, se révèlent n'être rien d'autre que les coefficients de la matrice fondamentale et des épipôles des deux vues considérées [62, 208]. Concrètement, si (abc) et $(a'b'c')$ sont des permutations paires de (123)

et si $\mathbf{p}_i^a$ désigne la a^{e} ligne de P_i, nous avons :

$$[\mathsf{F}_{ij}]_{aa'} = \begin{vmatrix} \mathbf{p}_i^b \\ \mathbf{p}_i^c \\ \mathbf{p}_j^{b'} \\ \mathbf{p}_j^{c'} \end{vmatrix} \qquad [\mathbf{e}_{ij}]^a = \begin{vmatrix} \mathbf{p}_i^a \\ \mathbf{p}_j^1 \\ \mathbf{p}_j^2 \\ \mathbf{p}_j^3 \end{vmatrix} \quad . \tag{5.3}$$

L'application de ces relations aux 3 mineurs de 2 lignes de l'image i et 3 lignes de l'image j résulte en la relation :

$$(\mathsf{F}_{ij}\mathbf{q}_{jp})\,\lambda_{jp} = (\mathbf{e}_{ij} \wedge \mathbf{q}_{ip})\,\lambda_{ip} \quad . \tag{5.4}$$

Cette relation n'exprime, tout d'abord, rien d'autre que la contrainte épipolaire. Pourtant, si les profondeurs projectives correctes sont utilisées, l'équation (5.4) est non seulement valable à un facteur multiplicatif près, mais elle devient aussi une égalité *exacte*. C'est ce nouveau résultat qui permet de déterminer les profondeurs projectives à partir des matrices fondamentales et des épipôles. Des résultats analogues pour des tenseurs d'appariement d'ordre supérieur sont donnés dans [207].

Notre stratégie pour la détermination des profondeurs projectives est relativement simple. L'équation (5.4) relie les profondeurs projectives des projetés d'un point 3D vus dans deux images. Ayant estimé un nombre suffisant de matrices fondamentales et d'épipôles, nous pouvons établir un système d'équations linéaires homogènes qui permettent de trouver l'ensemble complet des profondeurs projectives d'un point 3D, à un facteur global près. Il faut un minimum de $m-1$ matrices fondamentales qui relient les m images en un graphe connecté (voir des exemples dans la figure 5.2). Si d'autres matrices fondamentales sont disponibles, les équations deviennent redondantes ce qui, a priori, apporte une plus grande robustesse. À la limite, toutes les $m(m-1)/2$ matrices fondamentales et toutes les $m(m-1)$ équations pourraient être utilisées pour trouver les m profondeurs inconnues, mais ceci serait coûteux en temps de calcul et de toute façon non réalisable avec des séquences d'images réalistes.

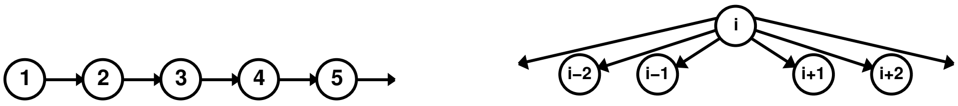
Actuellement, nous considérons deux façons d'utiliser un nombre minimal d'équations (5.4). Les images sont traitées de manière séquentielle ou parallèle, c'est-à-dire que nous utilisons les paires d'images $(1,2),(2,3),\ldots,(m-1,m)$, ou bien $(i,1),(i,2),\ldots,(i,i-1),(i,i+1),\ldots,(i,m)$, où l'image i peut par exemple être choisie comme l'image au milieu de la séquence (voir la figure 5.2). Ces approches ne sont probablement pas les plus robustes possibles, mais elles permettent de déterminer les profondeurs projectives de manière triviale, en résolvant $m-1$ équations de manière récursive.

On commence par l'affectation d'une valeur arbitraire à une des profondeurs projectives, par exemple $\lambda_{1p} = 1$ (rappelons que les profondeurs projectives associées aux projetés d'un point 3D sont définies à un facteur global près). Ensuite, les autres profondeurs projectives associées au point $\mathbf{Q}_p$ peuvent être déterminées une-par-une en résolvant l'équation (5.4) aux moindres carrés :

$$\lambda_{ip} = \frac{(\mathbf{e}_{ij} \wedge \mathbf{q}_{ip})^{\mathsf{T}}\,(\mathsf{F}_{ij}\mathbf{q}_{jp})}{\|\mathbf{e}_{ij} \wedge \mathbf{q}_{ip}\|^2}\,\lambda_{jp} \quad . \tag{5.5}$$

Si l'on veut utiliser plus que le nombre minimal d'équations (5.4) nécessaires pour déterminer les profondeurs projectives, il faut prendre en compte les relations quadratiques qui existent entre les matrices fondamentales « redondantes » [207]. Considérons un petit exemple où les trois équations associées à trois images sont utilisées (nous omettons l'indice p) :

$$\begin{aligned}
(\mathsf{F}_{12}\mathbf{q}_2)\,\lambda_2 &= (\mathbf{e}_{12} \wedge \mathbf{q}_1)\,\lambda_1 \\
(\mathsf{F}_{23}\mathbf{q}_3)\,\lambda_3 &= (\mathbf{e}_{23} \wedge \mathbf{q}_2)\,\lambda_2 \\
(\mathsf{F}_{31}\mathbf{q}_1)\,\lambda_1 &= (\mathbf{e}_{31} \wedge \mathbf{q}_3)\,\lambda_3 \quad .
\end{aligned}$$



(a) Schéma séquentiel. (b) Schéma parallèle.

FIG. 5.2: *Deux possibilités pour l'enchaînement des équations de détermination des profondeurs projectives.*

A priori, les matrices fondamentales et les épipôles sont définis seulement à des facteurs d'échelle individuels près. Pourtant, des facteurs d'échelle arbitraires prohibent en général une solution cohérente pour les λ_i. Avant d'utiliser des équations redondantes, il est donc indispensable de choisir des facteurs d'échelle cohérents pour les matrices fondamentales et les épipôles. Leur détermination peut être basée sur les relations quadratiques entre les tenseurs d'appariement (pour plus de détails voir [207]).

Même si seules des équations non redondantes sont utilisées, les facteurs d'échelle des F_{ij} et $\mathbf{e}_{ij}$ ont un impact sur l'ordre de grandeur des profondeurs projectives, ce qui devient évident en regardant l'équation (5.5). Avec des échelles mal équilibrées pour les matrices fondamentales et les épipôles, les profondeurs projectives peuvent prendre des valeurs extrêmement grandes ou petites lors de l'enchaînement récursif des équations. Théoriquement, ceci ne constitue pas un problème mais la phase de factorisation numérique pour la reconstruction explicite peut devenir mal conditionnée. Pour éviter cette situation, nous équilibrons les profondeurs projectives avant la factorisation, par des multiplications scalaires vue-par-vue respectivement point-par-point, ainsi profitant des degrés de liberté dans la détermination des profondeurs projectives, décrits au début de cette section. Ce processus d'équilibrage est décrit en 5.5.2.

5.5 Une méthode de reconstruction projective multi-image basée sur la factorisation

Nous avons développé un algorithme pratique pour la reconstruction projective multi-image, basé sur les observations faites dans les sections précédentes. À côté des deux étapes principales, la détermination des profondeurs projectives et la factorisation de la matrice des mesures, des procédures numériques sont incorporées à l'algorithme, qui augmentent considérablement la stabilité et la précision des résultats. Nous décrivons d'abord ces dernières, avant de donner le schéma complet de l'algorithme.

5.5.1 Normalisation des coordonnées image

Afin d'obtenir un bon conditionnement numérique, nous travaillons avec des coordonnées normalisées, comme Hartley l'a proposé pour l'estimation de la matrice fondamentale [71]. Cette normalisation consiste à appliquer une similitude T_i à toute image i telle que les points image transformés soient centrés autour de l'origine et que leur distance moyenne à l'origine soit égale à $\sqrt{2}$. Concrètement, les transformations sont réalisées par la composition d'une translation et d'un changement d'échelle. Cette normalisation a, comme effet, un équilibrage des ordres de grandeur des coordonnées image – au lieu d'avoir des coordonnées homogènes de l'ordre $(\mathrm{O}(100), \mathrm{O}(100), 1)^\mathsf{T}$ (la notation O indique un ordre de grandeur), un point typique normalisé aura des coordonnées $(\mathrm{O}(1), \mathrm{O}(1), 1)^{\mathsf{T}}$. Ainsi, les éléments de la matrice des mesures seront, en moyenne, d'ordres de grandeur similaires, ce qui améliore le conditionnement numérique de cette dernière et stabilise sa factorisation.

Toutes les étapes de l'algorithme sont effectuées en coordonnées normalisées. Comme nous calculons les mouvement et structure projectifs correspondants aux points image transformés, $\mathsf{T}_i \mathbf{q}_{ip}$, nous devons « dé-normaliser » les matrices de projection $\widehat{\mathsf{P}}_i$ estimées : $\widehat{\mathsf{P}}'_i = \mathsf{T}_i^{-1} \widehat{\mathsf{P}}_i$. Les $\widehat{\mathsf{P}}'_i$ et $\widehat{\mathbf{Q}}_p$ représentent alors les mouvement et structure projectifs correspondants aux points image originaux $\mathbf{q}_{ip}$.

Nos expériences montrent que cette simple normalisation améliore de manière tellement drastique les résultats de la reconstruction projective, qu'elle devient une partie indispensable de l'algorithme.

5.5.2 Équilibrage de la matrice des mesures

Considérons la factorisation de la matrice des mesures rééchelonnées W :

$$\mathsf{W} = \begin{pmatrix} \lambda_{11}\mathbf{q}_{11} & \lambda_{12}\mathbf{q}_{12} & \cdots & \lambda_{1n}\mathbf{q}_{1n} \\ \lambda_{21}\mathbf{q}_{21} & \lambda_{22}\mathbf{q}_{22} & \cdots & \lambda_{2n}\mathbf{q}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1}\mathbf{q}_{m1} & \lambda_{m2}\mathbf{q}_{m2} & \cdots & \lambda_{mn}\mathbf{q}_{mn} \end{pmatrix} = \begin{pmatrix} \widehat{\mathsf{P}}_1 \\ \widehat{\mathsf{P}}_2 \\ \vdots \\ \widehat{\mathsf{P}}_m \end{pmatrix} \begin{pmatrix} \widehat{\mathbf{Q}}_1 & \widehat{\mathbf{Q}}_2 & \cdots & \widehat{\mathbf{Q}}_n \end{pmatrix} .$$

La multiplication de la colonne p de W avec un scalaire non nul ν_p correspond à (ou implique) la multiplication de $\widehat{\mathbf{Q}}_p$ avec le même facteur. De manière analogue, multiplier les lignes associées à la vue i (les lignes $(3i-2, 3i-1, 3i)$) avec un scalaire non nul μ_i correspond à multiplier la matrice de projection $\widehat{\mathsf{P}}_i$ avec μ_i. Donc, des opérations d'équilibrage point-par-point et vue-par-vue appliquées à W, n'affectent pas la reconstruction projective obtenue.

Cette considération n'est valable qu'en absence de bruit. En présence de bruit dans les coordonnées image, W ne sera qu'approximativement[4] de rang 4. Les opérations décrites ci-dessus, *vont effectivement affecter* les résultats de la factorisation et nous pouvons donc essayer de profiter des degrés de liberté qui nous sont donnés via les facteurs d'échelle arbitraires des $\widehat{\mathsf{P}}_i$ et $\widehat{\mathbf{Q}}_p$, afin d'améliorer le conditionnement numérique de la factorisation de W ou de répartir les effets du bruit. En appliquant le schéma itératif suivant, que nous appelons *équilibrage*, nous essayons d'amener les éléments de W à un ordre de grandeur commun.

Algorithme : Équilibrage de la matrice des mesures

1. Multiplier chaque colonne p de W par un ν_p tel que $\sum_{i=1}^{3m} (\nu_p \mathsf{W}_{ip})^2 = 1$.
2. Multiplier chaque triplet de lignes $(3i-2, 3i-1, 3i)$ par un μ_i tel que $\sum_{p=1}^{n} \sum_{l=3i-2}^{3i} (\mu_i \mathsf{W}_{lp})^2 = 1$.
3. Si le changement des éléments de W est important, répéter les étapes 1 et 2.

En pratique, une ou deux itérations sont suffisantes.

Nous avons décrit l'équilibrage de la matrice W entière. En effet, puisque les coordonnées image $\mathbf{q}_{ip}$ sont déjà normalisées, il serait suffisant d'équilibrer seulement la matrice $m \times n$ des profondeurs projectives, (λ_{ip}).

4. Nous parlons du rang numérique ici.

5.5.3 Algorithme complet

Algorithme : Reconstruction projective par factorisation

1. **Normalisation** des coordonnées des points image, par application de similitudes T_i.
2. **Estimation des matrices fondamentales** et des épipôles (nous utilisons l'algorithme de 8 points [71]).
3. **Détermination des profondeurs projectives** λ_{ip} à l'aide de l'équation (5.5).
4. **Construction de la matrice des mesures** W contenant les coordonnées image rééchelonnées.
5. **Équilibrage** de W par des multiplications scalaires colonne-par-colonne et par triplets de lignes.
6. **Factorisation** :
   (a) Décomposition en valeurs singulières de W : $\mathsf{W} = \mathsf{U}\Sigma\mathsf{V}^\mathsf{T}$.
   (b) Annulation des valeurs singulières à l'exception des 4 plus grandes. Soient Σ' la matrice diagonale de ces 4 valeurs, U' et ${\mathsf{V}'}^\mathsf{T}$ les sous-matrices des 4 premières colonnes de U et V^T.
   (c) Extraction des matrices composites de projection et de structure :
   $\mathsf{P} = \mathsf{U}'\Sigma''$ et $\mathsf{Q} = \Sigma'''{\mathsf{V}'}^\mathsf{T}$, où $\Sigma''\Sigma'''$ est une décomposition quelconque de Σ'.
7. **Dé-normalisation** : adaptation des matrices de projection pour défaire la normalisation effectuée en l'étape 1 (multiplication de gauche par T_i^{-1}).

5.6 Traitement de données incomplètes

L'un des problèmes majeurs des méthodes de factorisation apparaît quand les données sont incomplètes, dû, dans notre cas, à des points non visibles ou non détectables dans quelques vues ou dû à l'échec de l'appariement. Dans ce cas, les entrées concernées de la matrice des mesures ne sont pas définies et la décomposition en valeurs singulières ne peut pas être appliquée directement. Ce problème ne semble pas avoir de solution simple et les méthodes proposées dans la littérature ont toutes des inconvénients prononcés. Nous décrivons, dans la suite, trois méthodes différentes apportant des solutions à ce problème.

Dans la section suivante, nous proposons une approche de reconstruction à partir de longues séquences d'images qui peut traiter des données incomplètes. Nous appliquons une autre approche que celles présentées dans la suite.

Hallucination. Tomasi et Kanade, dans [202], proposent une méthode qu'on pourrait appeler « hallucination ». Le principe est simple : la factorisation est initialement effectuée pour une sous-matrice sans données manquantes. Ceci résulte en la reconstruction de quelques points et la détermination des matrices de projection de quelques vues. L'information récupérée est ensuite propagée vue-par-vue respectivement point-par-point, tout en remplissant les « trous » de la matrice des mesures, par reprojection virtuelle de points reconstruits dans les vues où il n'y a pas d'observation de ces points.

Le choix de la sous-matrice factorisée initialement peut être crucial pour le succès de la méthode. Idéalement, la sous-matrice sans données manquante la plus grande (par rapport au nombre de vues *et* au nombre de points) devrait être choisie, mais ceci représente un problème NP-complet [102]. La méthode propage potentiellement les erreurs de la factorisation initiale et il n'est pas clair qu'un ajustement non linéaire final puisse converger vers une solution globalement optimale.

Analyse en composantes principales de données incomplètes. Cette méthode, initialement proposée indépendamment d'applications en vision par Wiberg [225], a été appliquée par Shum *et al.* pour la modélisation d'objets composés de facettes planes à partir d'images de profondeur (range images) [176, 177]. La formulation originale du problème de factorisation est bilinéaire (voir l'équation (5.2)), donc linéaire par rapport à chacun des deux ensembles d'inconnues, mouvement et structure. Étant donné le mouvement ou la structure, l'ensemble complémentaire d'inconnues peut être déterminé par une estimation aux moindres carrés linéaire. L'absence de données ne pose évidemment pas de problème ici. La méthode complète consiste à estimer la structure puis le mouvement alternativement, jusqu'à convergence (plusieurs critères de convergence sont imaginables). Shum *et al.* montrent dans [176, 177] comment inclure des incertitudes sur les mesures.

Cette méthode nécessite l'initialisation de l'un des ensembles d'inconnues, celui du mouvement ou celui de la structure. Ce problème n'est pas bien discuté dans [176, 177] et il n'est pas facile de savoir comment les auteurs ont procédé lors de leurs expériences. Si l'on suppose que l'on dispose d'une initialisation d'un ensemble d'inconnues, pourquoi alors ne pas simplement utiliser, après avoir estimé linéairement l'autre partie des inconnues, une méthode non linéaire qui estimera simultanément les inconnues, au lieu de procéder par des itérations linéaires ? Shum *et al.* donnent comme argument contre une méthode non linéaire le nombre élevé d'inconnues. Cet argument n'est pas toujours justifié, puisque souvent le nombre d'inconnues est dominé par un des deux ensembles et donc la somme des inconnues n'est pas beaucoup plus grande que le nombre des inconnues de l'ensemble dominant. Un autre argument est la structure moins complexe de la formulation linéaire par rapport à un algorithme non linéaire, ce qui facilite son implémentation.

Des approches similaires sont les méthodes itératives pour l'estimation de la position ou la reconstruction euclidienne calibrée proposées par Dementhon et Davis [45] et étendues par Horaud, Christy *et al.* [36, 94].

L'approche de Jacobs. Récemment, Jacobs a présenté une méthode d'estimation d'un modèle linéaire avec des données incomplètes [102]. Nous décrivons brièvement le principe de cette approche. Le problème est de déterminer une matrice $\widehat{\mathsf{W}}$ de rang r qui approche au mieux une matrice W donnée, dont des éléments peuvent être bruités ou même manquants. Les colonnes de $\widehat{\mathsf{W}}$ peuvent être considérées comme les coordonnées de points dans un espace de dimension m. Le problème posé peut donc être vu comme la recherche d'un sous-espace $\mathcal{L}$ de dimension r de cet espace de colonnes, qui approche au mieux les colonnes données de la matrice W. En l'absence de bruit, tout r-uplet de colonnes de W sans élément manquant constitue généralement une base de $\mathcal{L}$. Si parmi les r colonnes il y en a avec des éléments manquants, ce n'est plus le cas. Remplir les éléments manquants avec toutes les valeurs possibles résulte en un espace engendré des colonnes complétées, de dimension supérieure à r, mais contenant $\mathcal{L}$. L'intersection d'un nombre suffisant de tels espaces engendrés par des r-uplets de colonnes de W est alors le sous-espace $\mathcal{L}$ cherché de dimension r. Pour combattre l'effet du bruit dans les données, on doit prendre en compte plus que le nombre minimal de r-uplets pour déterminer $\mathcal{L}$. Pour déterminer l'intersection d'espaces décrite ci-dessus, Jacobs propose de calculer le complément de l'espace engendré par les compléments des espaces de départ. Le complément d'un espace qui est représenté par des vecteurs de base peut être calculé par une décomposition en valeurs singulières de la matrice contenant ces vecteurs.

Cette méthode est assez élégante et la seule parmi les trois décrites ayant simultanément deux caractéristiques désirées : prise en compte simultanée de toutes ou au moins de beaucoup de données et aucun besoin d'initialisation. Pourtant, le grand inconvénient est sa complexité – la détermination du complément décrite ci-dessus se fait par une décomposition en valeurs singulières d'une matrice qui, en général, est beaucoup plus grande que la matrice de départ W. Soient $m \times nr$ les dimensions de W et supposons que $p\%$ de ses éléments sont non occupés et que ces éléments sont dispersés sur la matrice de manière uniforme. Pour chaque r-uplet de colonnes, on va rajouter en moyenne $m - r - m(1 - (1 - p)^r) = m(1 - p)^r - r$ colonnes à la matrice à décomposer. La prise en compte de chaque colonne dans exactement un r-uplet signifie qu'il faut décomposer une matrice de $mn(1 - p)^r - rn$ colonnes ! Par exemple, considérons une matrice de mesures pour 100 vues et 500 points, dont 30 % des éléments sont non occupés. Au lieu de factoriser une matrice des dimensions 300×500, on a à faire avec une des dimensions $300\times \mathrm{O}(8500)$. Seulement si le taux de non occupation dépasse les 60 %, la matrice sera plus petite que l'originale. Ce compte n'est certainement pas tout à fait réaliste, mais il reflète l'ordre de grandeur du problème. De toute façon, en pratique, les « trous » dans la matrice W seront assez regroupés (des points non visibles dans des vues voisines). Dans ce cas, notre calcul n'est qu'une sous-estimation de la taille du problème. Il faut donc clairement avoir de bonnes heuristiques pour le choix des colonnes.

Une autre complication, en rapport avec notre méthode de reconstruction, réside dans la détermination des profondeurs projectives. L'absence de points image dans une ou quelques vues nécessite de pouvoir « sauter » par dessus ces vues et pour ce faire, de calculer des matrices fondamentales entre des vues éventuellement éloignées. Ce problème ne se pose pas avec la factorisation affine.

5.7 Reconstruction à partir de longues séquences d'images

Nous décrivons brièvement une implémentation récente d'une méthode de reconstruction projective à partir d'une longue séquence d'images. Elle est essentiellement basée sur notre méthode de factorisation. La méthode est conçue pour faire face à un problème majeur avec de longues séquences d'images, c'est-à-dire le problème de données incomplètes, où les points ne sont pas visibles dans toutes les vues. Il s'agit de combiner des méthodes existantes de manière appropriée.

Bien souvent, les images sont prises lors d'un mouvement continu d'une caméra ce qui veut dire que des images consécutives dans la séquence sont prises de points de vue proches. Nous pouvons alors supposer que les images de la séquence sont ordonnées. Ainsi, il existe des sous-séquences suffisamment courtes pour pouvoir appliquer la factorisation, qui contiennent un ensemble de points communs. Ceci nous permet d'appliquer avec succès notre méthode de factorisation pour des sous-séquences d'images, en n'utilisant que les points communs aux images traitées.

Le schéma global de la reconstruction de la séquence d'images entière est le suivant. La reconstruction est d'abord effectuée pour des sous-séquences en utilisant la méthode de factorisation. Ceci donne de bons résultats localement. Ensuite, nous fusionnons des reconstructions projectives deux-par-deux, de manière hiérarchique. Des étapes d'optimisation non linéaire intermédiaires permettent de stabiliser la fusion des reconstructions. En cours de la fusion, nous intégrons les points qui n'ont pas été reconstruits par factorisation puisqu'ils n'ont pas été visibles dans toutes les vues d'une sous-séquence. Le résultat final sera une reconstruction projective de toute la séquence.

Une approche similaire a déjà été décrite par Laveau [109]. Il l'homographie entre deux reconstructions ayant des points en commun, afin de les fusionner, et des procédures d'optimisation pour augmenter la précision de la reconstruction globale. La différence principale est que nous initialisons la reconstruction à partir de plusieurs images tandis que Laveau en utilise trois à la fois. Aussi, nous appliquons la fusion de manière récursive à travers plusieurs niveaux ce qui nous permet de traiter des séquences d'images très longues (cf. les expériences préliminaires rapportées dans le paragraphe 5.8.2).

Notre approche combine les méthodes suivantes.

▷ Reconstruction projective par factorisation. Ceci sert à initialiser la reconstruction pour des sous-séquences.

▷ Calcul de l'homographie 4×4 entre deux reconstructions projectives. Cette homographie permet de fusionner les deux reconstructions. Ensuite, nous transformons les points de la deuxième reconstructions dans le repère de la première. Les points qui sont contenus dans les deux reconstructions sont laissés à la position donnée par la première reconstruction.

  Nous avons implémenté trois méthodes pour le calcul de l'homographie, dont deux sont linéaires. La première méthode linéaire minimise la différence des coordonnées des points 3D correspondants. Ceci n'est pas une distance géométrique et les résultats avec cette méthode sont assez mauvais.

  La deuxième méthode linéaire minimise l'erreur de reprojection après transformation du deuxième des points 3D dans le repère de la première reconstruction. La reprojection est effectuée seulement en utilisant les matrices de projection associées à la première reconstruction. Ainsi, la méthode n'est pas « symétrique ». Nous effectuons la méthode dans les deux directions et retenons l'homographie qui donne l'erreur de reprojection globale minimale.

  La troisième méthode minimise l'erreur de reprojection globale, où l'homographie est appliquée dans les deux sens. Ceci requiert la paramétrisation de l'inverse de l'homographie. Nous avons implémenté une approximation au premier ordre de l'inverse. Cette méthode effectue une minimisation non linéaire, que nous initialisons avec le résultat de la deuxième méthode. Nous avons constaté en pratique que le résultat de la deuxième méthode constitue un minimum local pour la troisième méthode et le résultat n'est pas amélioré par la minimisation globale.

▷ Optimisation non linéaire sur les matrices de projection et les points 3D. Nous avons implémenté un algorithme de type ajustement de faisceaux, basé sur une méthode numérique similaire à Levenberg-Marquardt [160]. Incorporé dans l'algorithme est un mécanisme de rejet d'erreurs grossières. Si la fonction de coût (erreur de reprojection) n'atteint pas un certain seuil après plusieurs itérations, nous calculons le résidu pour chaque point image et enlevons tous les points image dont le résidu est grand. Le choix du seuil dépend de la précision de l'extraction des points dans les images, qui est en général connue approximativement.

▷ Triangulation de points. Ceci sert à reconstruire les points qui n'ont pas été reconstruits par factorisation ou qui ont été éliminés de l'optimisation non linéaire puisque les résidus de tous les points 2D correspondants ont été trop grands.

Nous combinons ces outils de base pour former deux « méta-outils » :

Initialisation Nous effectuons les étapes suivantes.

1. Factorisation.
2. Optimisation non linéaire.

3. Triangulation.
4. Deuxième optimisation non linéaire.

Fusion Nous effectuons les étapes suivantes.

1. Calcul de l'homographie et transformation des reconstructions dans un repère commun.
2. Optimisation non linéaire.
3. Triangulation.
4. Deuxième optimisation non linéaire.

Ces deux procédures peuvent être combinés de beaucoup de façons différentes. Un mécanisme de décision automatique prendra en compte plusieurs facteurs : le nombre de points communs entre des vues consécutives, les nombres minimal et maximal de vues qui donnent de bons résultats pour la factorisation, le nombre minimal nécessaire pour un calcul stable de l'homographie...

Nous n'avons pas encore mis au point un tel schéma automatique, mais le schéma ad hoc que nous avons appliqué pour nos expériences, n'a posé aucun problème. Nous avons initialisé la structure pour des blocs de 20 images. Souvent, il est conseillé de choisir des blocs contenant des images communes. Pourtant, dans les séquences que nous avons traitées, les points sont visibles dans toutes les vues. Ainsi, nous prenons des blocs sans superposition (les blocs des images 1 – 20, 21 – 40, 41 – 60, etc.). Les reconstructions sont ensuite fusionnées selon un schéma d'arbre binaire, jusqu'à ce qu'on arrive à la reconstruction qui prend en compte toute la séquence d'images.

Ce schéma pour la reconstruction projective à partir de longues séquences d'images est provisoire et certainement pas optimal, mais il constitue une première étape de développement de méthodes plus sophistiquées. Des résultats d'expériences préliminaires sont rapportés dans la section suivante. Dans les séquences traitées tous les points sont visibles dans toutes les vues. Ainsi, ces séquences ne constituent pas un test très sérieux de la méthode. Pourtant, avec 181 images, l'une des séquences est suffisamment longue pour que la reconstruction ne soit pas une affaire triviale.

5.8 Expériences

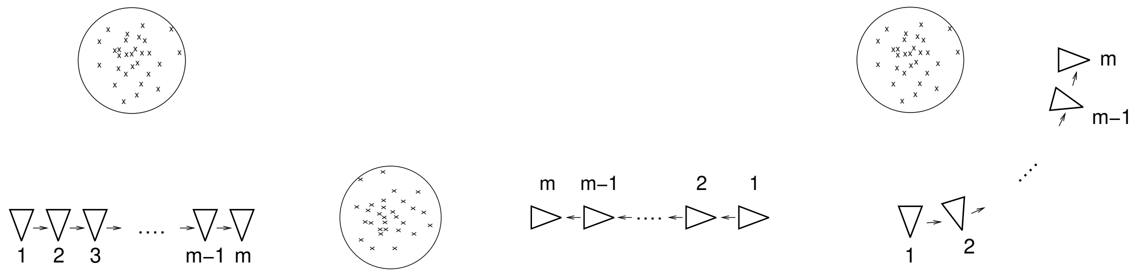
5.8.1 Expériences sur des données synthétiques

Nous avons conduit un certain nombre d'expériences sur des données synthétiques afin de quantifier la performance de l'algorithme. Trois configurations ont été testées (cf. la figure 5.3) :

- ▹ une caméra effectuant des mouvements latéraux par rapport à l'objet à reconstruire ;
- ▹ une caméra se dirigeant droit vers l'objet ;
- ▹ une caméra tournant autour de l'objet tout en le fixant.

La configuration 2 est connue pour être mal conditionnée pour la reconstruction : la détermination de la profondeur des points proches de la droite de mouvement est très sensible au bruit dans la position des points image. Dans ce cas, notre méthode fournit aussi de mauvais résultats.

Pour la configuration 3, le mouvement de la caméra achève un quart de cercle autour de l'objet. Pour chaque expérience, la longueur des trajectoires pour les deux autres configurations est toujours égale à la longueur de l'arc de mouvement de la configuration 3. Les m points de vue pour chaque séquence sont répartis de manière uniforme le long de la trajectoire.



(a) Configuration 1 : mouvement latéral

(b) Configuration 2 : mouvement vers l'objet

(c) Configuration 3 : mouvement orbital

FIG. 5.3: *Les configurations pour les simulations.*

La matrice des paramètres intrinsèques pour toutes les caméras est donnée par $\mathrm{diag}(1000, 1000, 1)$. L'objet est constitué de 50 points, distribués de manière uniforme dans une sphère de rayon 100 unités. Afin de créer des effets perspectifs importants, la distance entre la caméra et le centre de la sphère objet est égale à 200 unités (pour la configuration 2, c'est la distance de la caméra la plus proche à l'objet – la vue m dans la figure 5.3 (b)).

Pour chaque configuration, l'expérience suivante est effectuée 50 fois.

1. Création aléatoire de 50 points dans la sphère objet.

2. Projection des points objet dans les m images.

3. Perturbation des coordonnées image par un bruit gaussien de variance $0.0, 0.5, 1.0, 1.5$ ou 2.0.

4. Reconstruction projective avec notre méthode.

5. Calcul de l'erreur de reprojection (erreur 2D) : $\frac{1}{mn}\sum_{i=1}^{m}\sum_{p=1}^{n} d(\widehat{\mathsf{P}}_i\widehat{\mathbf{Q}}_p, \mathbf{q}_{ip})^2$.

6. Calcul de l'erreur tridimensionnelle de la reconstruction :

   ▷ alignement de la reconstruction projective avec le modèle euclidien de l'objet, en déterminant la transformation projective qui transforme la reconstruction tel qu'elle approche au mieux le modèle euclidien ;

   ▷ calcul de l'erreur 3D comme la moyenne de la somme des distances absolues des points reconstruits alignés aux points du modèle.

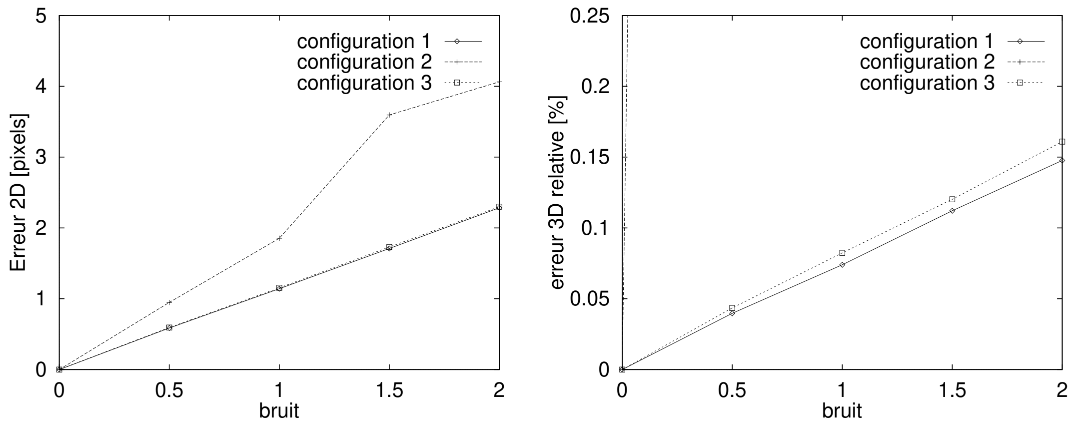
Dans les paragraphes suivants, les résultats de ces expériences sont analysés par rapport à plusieurs variables : sensibilité au bruit ; impact du nombre de vues ; influence de la normalisation des coordonnées image et de l'équilibrage de la matrice des mesures ; stabilité de la factorisation. Chaque point dans les graphes montrés représente la valeur moyenne des résultats de 50 expériences. Nous avons aussi calculé les valeurs médianes des résultats. Les graphes de celles-ci sont similaires aux graphes des valeurs moyennes et ne sont donc pas montrés.

Les erreurs 2D sont données en pixels. Les erreurs 3D sont des erreurs relatives à la taille de l'objet (le diamètre de la sphère objet) ; elles sont affichées en pourcents.

5.8.1.1 Sensibilité au bruit

La figure 5.4 montre le comportement de l'algorithme par rapport au niveau de bruit dans les coordonnées des points image. Pour cette expérience, la reconstruction est effectuée à partir de 10 vues.

L'erreur 3D pour la configuration 2 dépasse les 6 % pour un bruit de 2 pixels, malgré une erreur 2D relativement petite. Ceci montre que les erreurs 2D et 3D ne sont pas toujours bien corrélées. Les résultats pour les configurations 1 et 3 sont presque identiques. L'algorithme se comporte très bien pour ces configurations : les graphes sont quasi-linéaires et les erreurs de reprojection sont seulement légèrement supérieures au bruit.



(a) Erreur 2D (b) Erreur 3D

FIG. 5.4: *Sensibilité au bruit. Les courbes de l'erreur 2D pour les configurations 1 et 3 sont superposées. L'erreur 3D pour la configuration 2 dépasse l'échelle du graphe déjà pour un bruit de 0.5 pixels.*

5.8.1.2 Impact du nombre de vues

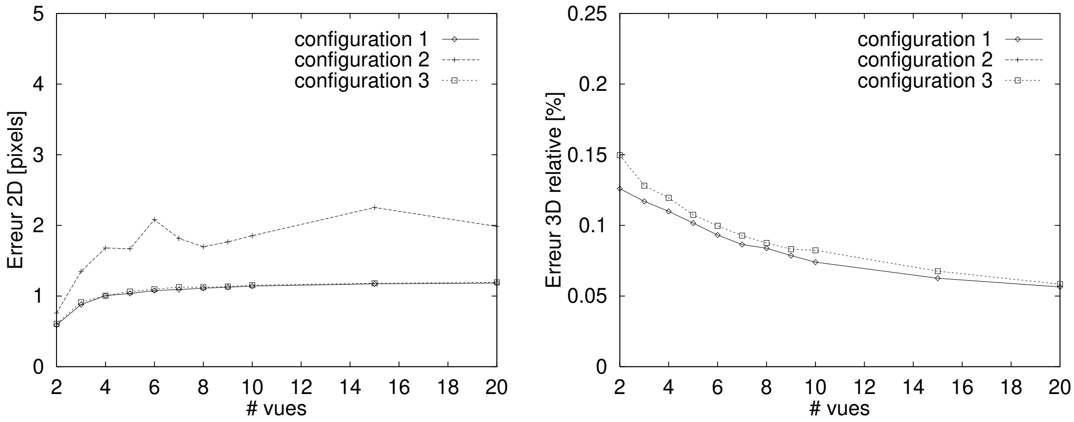
Pour cette expérience, le bruit dans les coordonnées des points image est de 1 pixel.

Les graphes montrent le comportement expecté : le plus de vues sont utilisée pour la reconstruction, le plus précisément la structure est reconstruite. L'erreur 2D augmente, ce qui est également attendu, mais elle a un comportement clairement asymptotique. Il est important de noter que l'utilisation de 20 vues réduit l'erreur 3D de plus de la moitié par rapport au cas de 2 vues.

5.8.1.3 Influence de la normalisation et de l'équilibrage

Les graphes précédents montrent les résultats de la méthode telle qu'elle est décrite en 5.5.3. Afin d'évaluer l'importance de l'utilisation de coordonnées image normalisées nous effectuons l'algorithme aussi avec des coordonnées non normalisées. De même, nous examinons l'apport de l'équilibrage de la matrice des mesures à la précision des résultats.

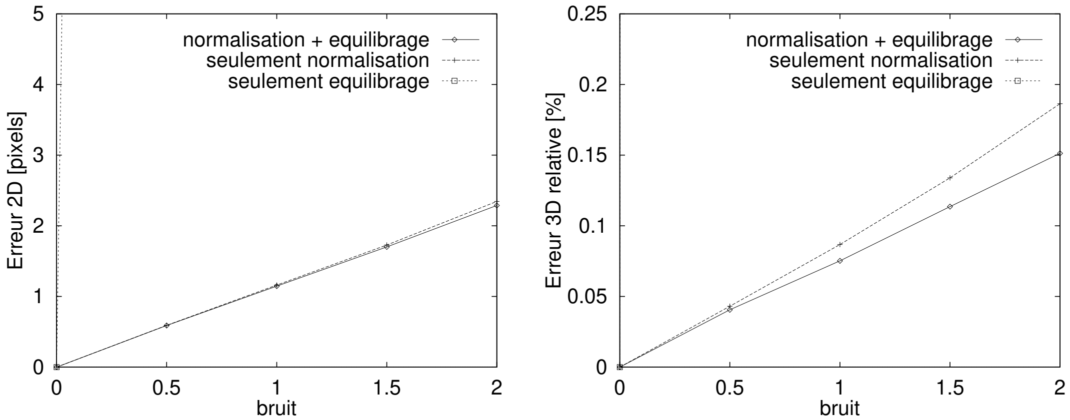
Si des coordonnées non normalisées sont utilisées, l'erreur est très grande même avec un bruit de 0.5 pixels. Pour ce qui est de l'équilibrage, on peut conclure qu'il améliore l'erreur 3D d'environ 20



(a) Erreur 2D (b) Erreur 3D

FIG. 5.5: *Impact du nombre de vues. Les courbes de l'erreur 2D pour les configurations 1 et 3 sont quasiment superposées. L'erreur 3D pour la configuration 2 est toujours supérieure à 5 % ; la courbe d'erreur n'est donc pas visible dans le graphe.*

%. Puisque l'équilibrage n'est pas très coûteux en temps de calcul, il devrait toujours faire partie de l'algorithme.



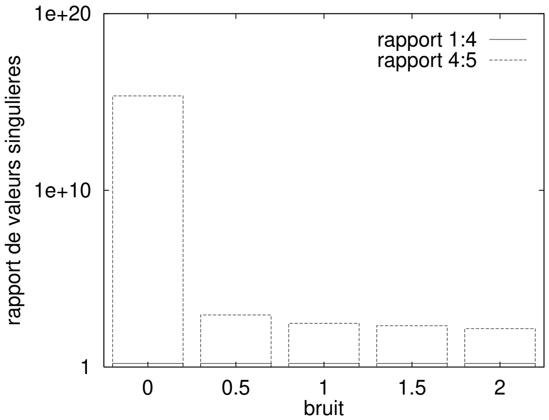
(a) Erreur 2D (b) Erreur 3D

FIG. 5.6: *Influence de la normalisation et de l'équilibrage. Les résultats présentés sont ceux pour la configuration 1. La courbe de l'erreur 2D sans normalisation (« seulement équilibrage ») dépasse l'échelle du graphe même pour un bruit de 0.5 pixels et l'erreur 3D est tellement grande que la courbe n'est pas visible !*

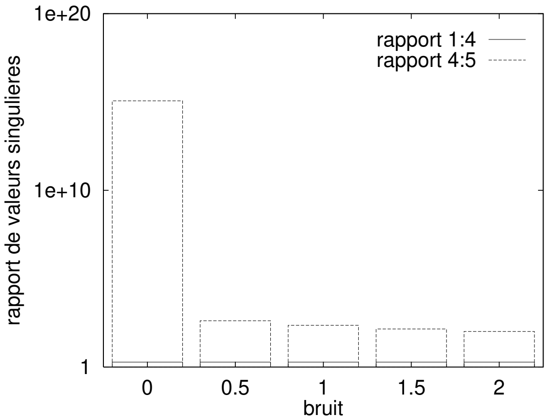
5.8.1.4 Stabilité de la factorisation

L'applicabilité de notre méthode de factorisation s'appuie sur le fait que la matrice des mesures W est de rang 4, si les profondeurs projectives correctes sont utilisées. Pour tester la robustesse de cette propriété en présence de bruit, nous évaluons la proximité de W au rang 4. Une matrice est approximativement de rang 4 si le rapport σ_4/σ_5 des 4^{e} et 5^{e} valeurs singulières et grand par rapport au rapport σ_1/σ_4 des 1^{re} et 4^{e} valeurs singulières. Dans les figures 5.7 et 5.8, ces rapports sont affichés, pour les configurations 1 et 2 et pour les cas de 2 et 20 vues. Notons que les axes verticaux ont une échelle logarithmique.

Nous pouvons constater que pour la configuration 1, la matrice est toujours très proche du rang 4 : le rapport σ_1/σ_4 est inférieur à 2 tandis que le rapport σ_4/σ_5 est supérieur à 100. Pour ce qui est de la configuration 2, les graphes reflètent la mauvaise performance pour la reconstruction 3D : en présence de bruit, le rapport σ_1/σ_4 est jusqu'à 10 fois supérieur à σ_4/σ_5.



(a) 2 vues



(b) 20 vues

FIG. 5.7: *Rapports de valeurs singulières pour la configuration 1.*

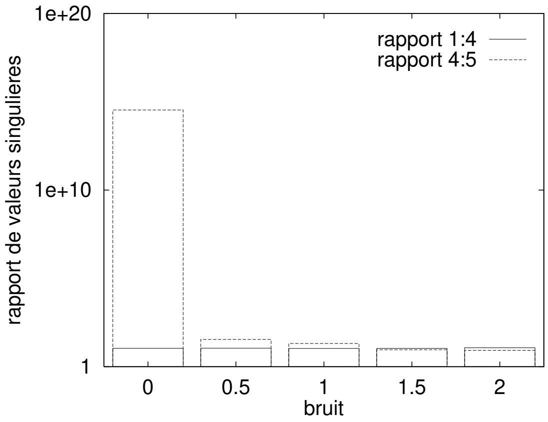
5.8.2 Expériences sur des données réelles

Nous avons testé notre méthode avec des séquences d'images réelles. Nous montrons les résultats pour quatre séquences.

5.8.2.1 La maison MOVI

La figure 5.9 montre la première et la dernière image d'une séquence de 6 images de la maison MOVI. Nous disposons de l'appariement de 38 points sur toute la séquence ; les points image sont extraits avec une précision de l'ordre du pixel.

Afin d'évaluer la qualité de la reconstruction projective, nous l'avons alignée avec un modèle euclidien qui a été obtenu par reconstruction avec des vues calibrées. La reconstruction alignée est visualisée dans la figure 5.10. Quelques segments de droite sont affichés pour faciliter la visualisation. À l'œil, les angles droits sont bien conservés et les fenêtres sont coplanaires. L'artefact sur le côté gauche du toit est dû au fait que le toit surplombe le mur ce qui fait que le coin entre le mur et le toit est occulté dans la dernière image.



(a) 2 vues

(b) 20 vues

FIG. 5.8: *Rapports de valeurs singulières pour la configuration 2.*

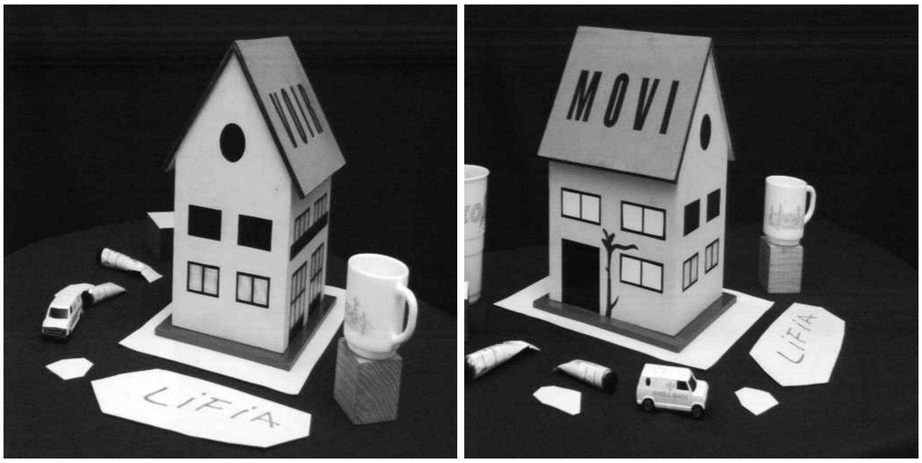


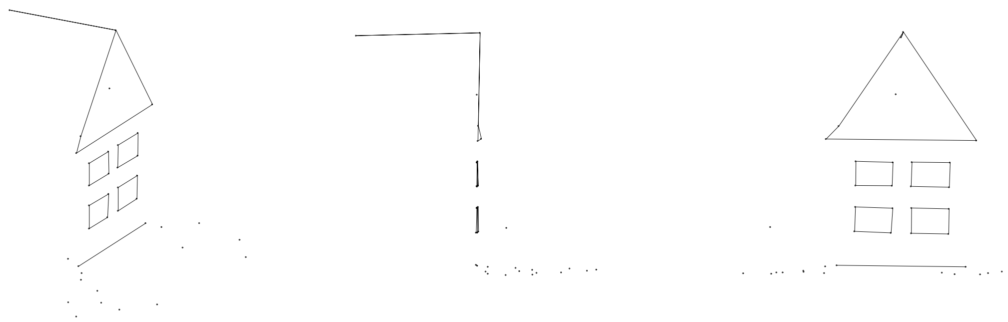
FIG. 5.9: *Deux images de la séquence avec la maison MOVI.*

5.8.2.2 La séquence du château

La figure 5.11 montre une image de la séquence du château[5]. La séquence est constituée de 11 images et de 28 correspondances de points image. La vérité de terrain pour les coordonnées tridimensionnelles des points est disponible ce qui nous permet de quantifier la qualité de la reconstruction projective. Après l'alignement de la reconstruction projective avec le modèle euclidien, nous avons obtenu une erreur de $3,7$ mm (moyenne de la somme des distances absolues). L'objet remplit un cube de dimension environ 220 mm $\times$ 210 mm $\times$ 280 mm. Si l'on compare l'erreur avec le diamètre maximal de l'objet (413 mm), nous obtenons une erreur relative de moins de 1 %. L'erreur de reprojection est d'environ $0,02$ pixels.

Nous avons effectué une minimisation non linéaire du style ajustement de faisceaux, avec les résultats de notre méthode comme initialisation. Les erreurs de reprojection sont légèrement améliorées

5. Nous remercions le « Calibrated Imaging Laboratory » de l'Université Carnegie Mellon qui a fourni cette séquence avec le soutien de ARPA, NSF et NASA. La séquence est disponible à http://www.cs.cmu.edu/ cil/cil-ster.html.



(a) Vue oblique (b) Vue de côté (c) Vue de face

FIG. 5.10: *Trois vues de la reconstruction.*

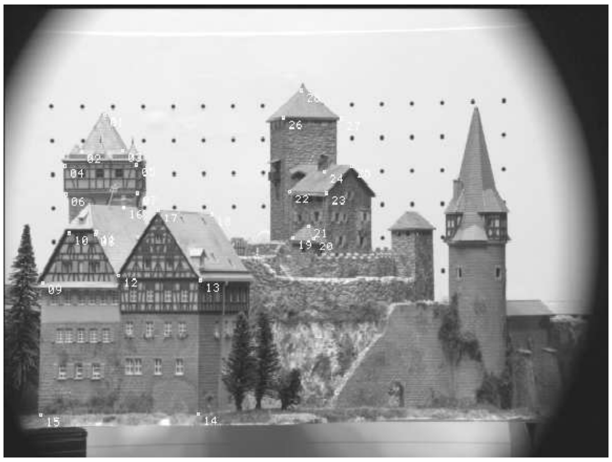


FIG. 5.11: *Une image de la séquence du château.*

tandis que la qualité de la reconstruction ne change pas de manière significative.

5.8.2.3 La séquence de l'hôtel

La séquence est constituée de 181 images et de 197 correspondances de points image. Nous avons appliqué la méthode décrite dans la section 5.7. Nous ne disposons pas de vérité de terrain pour cette séquence ; pour donner une idée de la qualité de la reconstruction, nous présentons quelques vues d'un modèle 3D texturé crées par reprojection à partir de nouveaux points de vue. La figure 5.12 montre une image originale de la séquence, sur laquelle sont superposés les points image utilisés pour la reconstruction ainsi que les triangles qui ont servi à la création du modèle texturé.

Quelques vues du modèle reconstruit sont montrées dans les figures 5.13, 5.14 et 5.15. Pour avoir des vues à peu près réalistes, nous avons transformé la reconstruction projective de manière appropriée.

FIG. 5.12: *Une image de la séquence de l'hôtel.*

5.8.2.4 La séquence de la cuisine

La séquence est constituée de 40 images et de 102 correspondances de points image. Nous avons appliqué la méthode décrite dans la section 5.7. Une image originale de la séquence est montrée dans la figure 5.16 et quelques vues du modèle reconstruit sont affichées dans les figures 5.17 à 5.20.

5.9 Conclusion et perspectives

Nous avons développé une méthode de factorisation pour la reconstruction projective à partir de plusieurs images non calibrées. La méthode est très élégante – elle détermine la structure de la scène et le calibrage projectif des caméras par factorisation d'une matrice qui contient tous les points de toutes les images[6].

La factorisation n'est possible que si des facteurs d'échelle corrects (profondeurs projectives) pour les coordonnées homogènes des points image sont utilisés. Nous avons proposé un moyen de déterminer ces profondeurs projectives, qui n'utilise que des matrices fondamentales et des épipôles, estimés à partir des données dans les images.

La méthode donne de bons résultats avec des images réelles. L'évaluation quantitative par simulation numérique confirme la robustesse de la factorisation et la bonne performance par rapport au bruit. Les résultats montrent qu'il est essentiel de travailler avec des coordonnées normalisées.

L'inconvénient majeur de la méthode est qu'elle ne peut pas traiter le cas de points non visibles dans quelques images. Nous avons présenté une approche encore préliminaire de reconstruction pro-

6. Des méthodes similaires ont été développées récemment au sein du groupe de Lund [16, 86, 180].

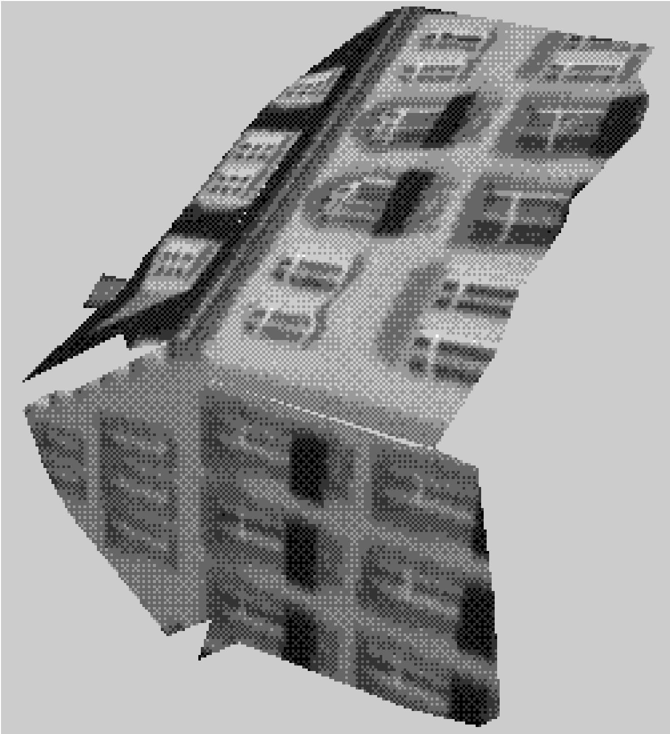


FIG. 5.13: *Première vue du modèle de l'hôtel reconstruit.*

jective pour de longues séquences et qui peut prendre en compte des données incomplètes. Cette approche utilise la factorisation pour initialiser la reconstruction projective pour des blocs de la séquence d'images et fusionne ensuite les résultats, tout en raffinant la reconstruction par minimisation non linéaire.

Un autre inconvénient est le temps de calcul pour la décomposition en valeurs singulières, qui peut être assez élevé si beaucoup d'images et beaucoup de points doivent être traités simultanément. Des méthodes de factorisation approximative, qui ne déterminent que les valeurs et vecteurs singuliers dominants, peuvent y remédier [209].

Quelques autres aspects méritent d'être examinés plus profondément. Dans l'implémentation actuelle, les profondeurs projectives sont déterminées en utilisant le nombre minimal d'équations (5.4). Il serait intéressant d'étudier dans quelle mesure l'utilisation d'équations redondantes peut améliorer les résultats. La prise en compte d'incertitudes sur les matrices fondamentales pour le choix des équations est aussi envisageable.

L'une des choses les plus importantes à faire est la conception et le développement de méthodes de reconstruction pour des séquences d'images très longues. De telles méthodes permettront la modélisation d'objets à grande échelle. En outre, elles constitueront un bon point de départ pour la détermination de la structure euclidienne de la scène.

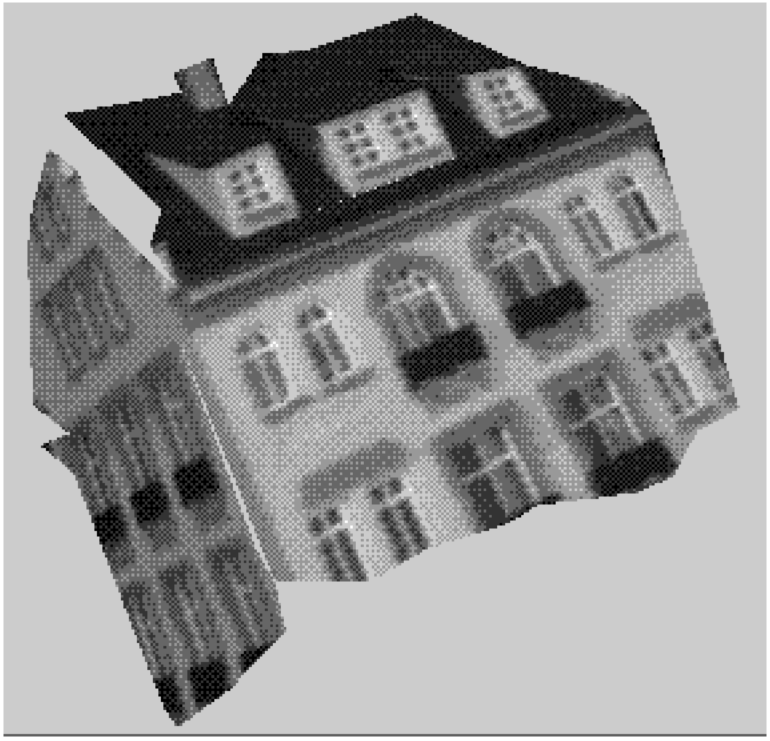


FIG. 5.14: *Deuxième vue du modèle de l'hôtel reconstruit.*

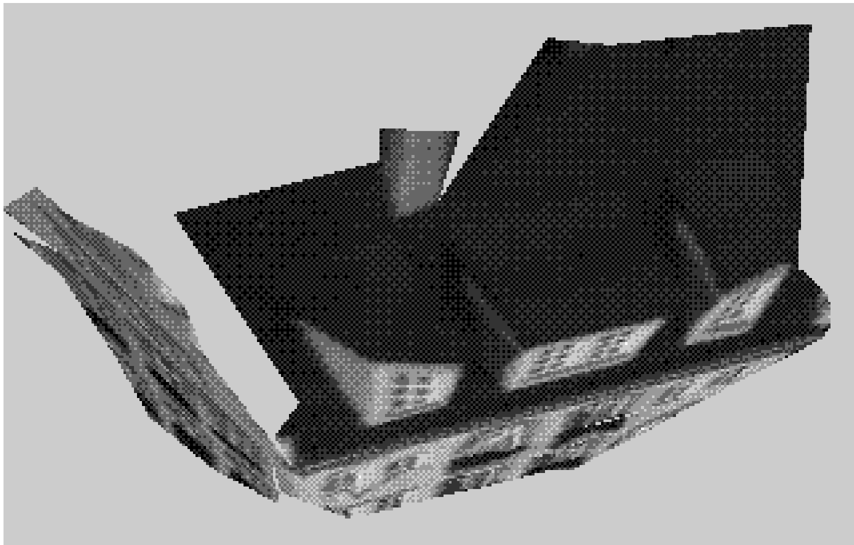


FIG. 5.15: *Troisième vue du modèle de l'hôtel reconstruit.*

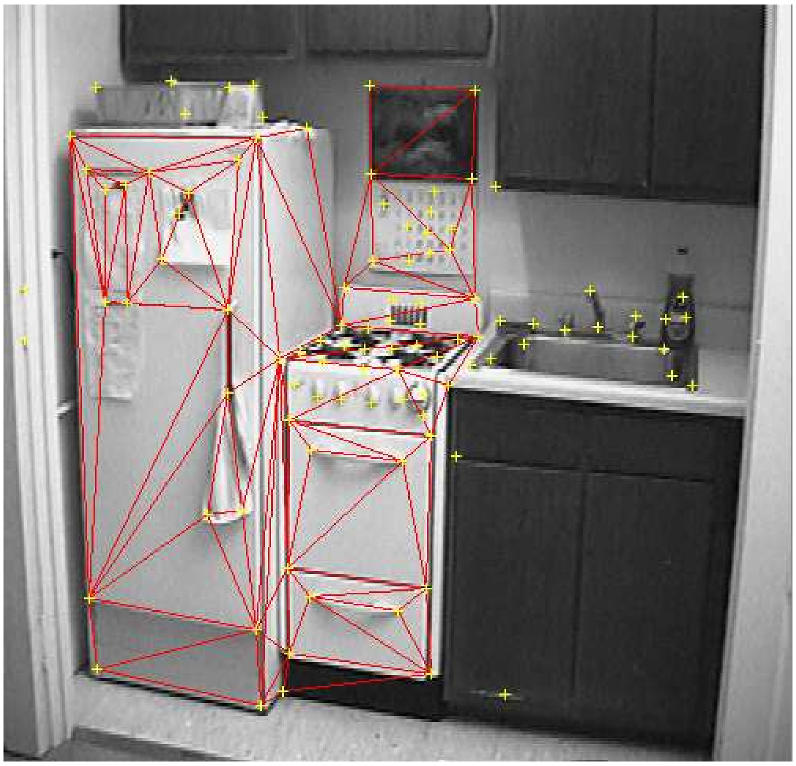


FIG. 5.16: *Une image de la séquence de la cuisine.*

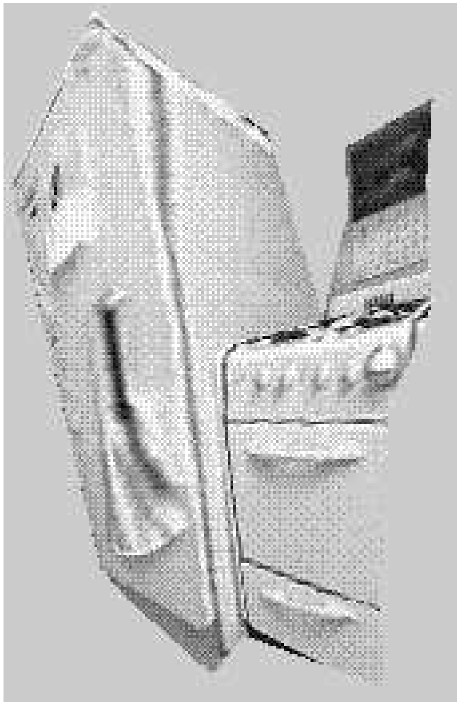


FIG. 5.17: *Première vue du modèle de la cuisine reconstruite.*

FIG. 5.18: *Deuxième vue du modèle de la cuisine reconstruite.*

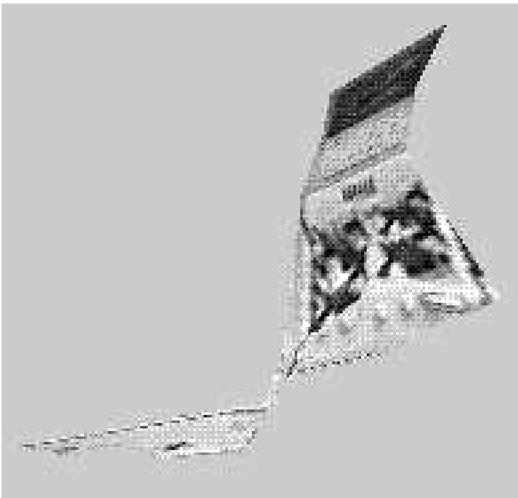


FIG. 5.19: *Troisième vue du modèle de la cuisine reconstruite.*

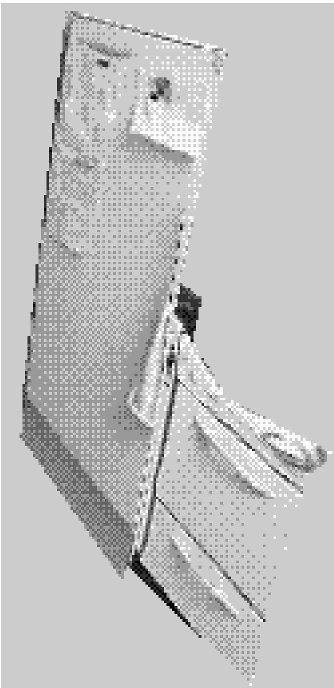


FIG. 5.20: *Quatrième vue du modèle de la cuisine reconstruite.*

6. Auto-calibrage et reconstruction euclidienne non calibrée

Une grande partie de ce chapitre contient un état de l'art des méthodes d'auto-calibrage. Nous décrivons principalement des méthodes pour le scénario classique d'une caméra en mouvement et avec des paramètres intrinsèques constants. D'autres approches, basées sur des mouvements particuliers ou sur un modèle de projection affine, sont également présentées.

Nos contributions sont les suivantes. Nous proposons une méthode d'auto-calibrage pour les caméras linéaires, qui est basée sur le tenseur trifocal. Ensuite, nous examinons les interdépendances qui existent entre les paramètres intrinsèques pour un objectif à focale variable. Nous proposons une méthode d'auto-calibrage qui est basée sur le pré-calibrage de ces interdépendances. Ce travail a été publié dans [185, 186, 189].

Une autre contribution au sujet de l'auto-calibrage est une étude théorique des séquences de mouvements critiques. Celle-ci est présentée dans le chapitre 7.

6.1 Introduction

Aux chapitres précédents, nous avons vu que sans aucune information en dehors de correspondances de primitives image une reconstruction projective est possible. Pour certaines applications, l'information projective peut être suffisante. Néanmoins, souvent il sera préférable d'obtenir la « vraie » structure de la scène ou le « vrai » mouvement des caméras, c'est-à-dire l'information métrique.

Cette information peut être obtenue dès lors qu'on dispose des paramètres intrinsèques des caméras. L'approche classique consiste à effectuer un calibrage hors ligne. Ceci est non seulement fastidieux mais aussi peu réaliste pour des applications où les changements de paramètres de la caméra sont fréquents ou où l'équipement de calibrage n'est pas disponible. La nouvelle tendance consiste à calibrer les caméras en ligne – nous parlons alors d'auto-calibrage – ou d'établir le calibrage par d'autres moyens, par exemple en effectuant des mouvements appropriés avec la caméra.

Dans ce chapitre, nous passons en revue des méthodes d'auto-calibrage. Le cas le plus étudié est le calibrage euclidien d'une caméra en mouvement, dont on sait que ses paramètres intrinsèques sont invariants. De telles méthodes et d'autres, s'appuyant sur des informations sur le mouvement de la caméra, sont décrites dans l'état de l'art de la section suivante. Un résumé de toutes les méthodes décrites est donné dans le tableau 6.1.

En section 6.3, nous proposons une méthode pour l'auto-calibrage d'une caméra linéaire, qui est basée sur le tenseur trifocal.

Ensuite, en 6.4, nous considérons l'auto-calibrage pour une caméra avec un objectif à focale variable. On peut observer que des changements de focale font varier les paramètres intrinsèques, mais pas de manière indépendante. Il est possible de pré-calibrer ces interdépendances des paramètres intrinsèques, qui ensuite permettent de simplifier la formulation du problème de l'auto-calibrage. Nous proposons une méthode d'auto-calibrage qui s'appuie sur le pré-calibrage de la caméra.

Le chapitre est clos par des conclusions.

6.2 État de l'art des méthodes d'auto-calibrage

6.2.1 Mouvements généraux et paramètres intrinsèques constants

C'est le scénario classique pour l'auto-calibrage et depuis le travail initial de Maybank et Faugeras [123], plusieurs approches ont été proposées. Le cas de l'auto-calibrage d'une caméra avec des paramètres intrinsèques constants nous intéresse particulièrement – au chapitre 7 nous étudions les séquences de mouvements de caméra qui sont critiques pour cette tâche. Dans la suite, nous passons en revue les différentes approches et évoquons les liens qui existent entre elles.

La notion d'auto-calibrage se réfère d'abord uniquement à la détermination du calibrage intrinsèque de la caméra, mais nous y comprenons plutôt la détermination de la structure euclidienne de la scène (et du mouvement « euclidien » de la caméra) ce qui inclut l'auto-calibrage. Afin de pouvoir décrire les différentes approches dans un même formalisme, nous supposons donc que l'auto-calibrage a pour but de rectifier la structure projective de manière à retrouver la structure euclidienne. Une reconstruction projective est donnée par des matrices de projection P_i^P et des points 3D $\mathbf{Q}_p^P$ qui collent bien aux données, c'est-à-dire les points image $\mathbf{q}_{ip}$:

$$\mathsf{P}_i^P \mathbf{Q}_p^P \sim \mathbf{q}_{ip} \quad .$$

La transition vers une reconstruction euclidienne se fait par une transformation projective T_P^E à déterminer : $\mathsf{P}_i^E \sim \mathsf{P}_i^P \mathsf{T}_P^E$ et $\mathbf{Q}_p^E \sim (\mathsf{T}_P^E)^{-1} \mathbf{Q}_p^P$. Le critère pour trouver la bonne transformation (à

Type de mouvement et informations requises	Approches
Mouvements généraux, paramètres intrinsèques variables	Heyden et Åström [89] Pollefeys et Van Gool [151]
Mouvements généraux, paramètres intrinsèques constants	Maybank, Faugeras, Luong [63, 114, 117, 118, 123] Zeller et Faugeras [233, 234] Heyden et Åström [88] Triggs [210] Pollefeys et Van Gool [150, 152, 153] Hartley [79, 80]
Modèles de calibrage simplifiés	Hartley [77] Brooks *et al.* [26, 27, 139, 143, 144] Ponce *et al.* [155] Hippisley-Cox et Porrill [91] Azarbayejani et Pentland [9]
Mouvements mesurés – Connaissance des angles de rotation – Connaissance de la translation	 Viéville [214] Stein [183] McLauchlan et Murray [127] Dron [48]
Mouvements particuliers – Rotations pures – Mouvements planaires – Mouvements particuliers entre vues consécutives	 Hartley [82] Basu [11, 12] Du et Brady [49] Daniilidis et Ernst [41] Armstrong *et al.* [6] Wiles et Brady [226] Viéville et Lingrand [216]
Mouvements initiaux particuliers – Une translation – Une translation et un zoom	 Armstrong *et al.* [5] Pollefeys *et al.* [154]
Pré-calibrage	Sturm [186, 189]
Mise à jour d'un calibrage initial	Crowley *et al.* [40] Enciso *et al.* [51, 52, 53]
Système stéréo	Brooks *et al.* [26, 27, 139, 143, 144] Zhang, Luong, Faugeras [115, 238, 239] Zisserman *et al.* [242] Devernay et Faugeras [47] Horaud et Csurka [96]
Caméra affine	Quan [164]
Caméra linéaire	voir section 6.3

TAB. 6.1: *Des méthodes d'auto-calibrage et de calibrage à partir de mouvements particuliers*

une transformation euclidienne près) est que les matrices de projection transformées doivent avoir les mêmes matrices des paramètres intrinsèques, K. C'est-à-dire qu'elles doivent pouvoir s'écrire de la manière suivante :

$$\mathsf{P}_i^E \sim \mathsf{K}\mathsf{R}_i(\quad \mathtt{I} \quad | - \mathbf{t}_i) \ , \tag{6.1}$$

pour des matrices de rotation R_i, des vecteurs $\mathbf{t}_i$ et une matrice des paramètres intrinsèques K. Toutes les méthodes rencontrées sont basées implicitement sur ce critère.

Nous simplifions la formulation en suivant Hartley [79, 80] et Heyden et Åström [88] en contraignant les repères des reconstructions projective et euclidienne qui a priori sont arbitraires : pour la reconstruction projective, le repère peut toujours être choisie de manière à ramener l'une des matrices de projection (nous en choisissant la première) à la forme suivante :

$$\mathsf{P}_1^P = (\quad \mathtt{I} \quad | \mathbf{0}) \ .$$

Nous attachons le repère de la reconstruction euclidienne également à la première caméra, en supposant que $\mathsf{R}_1 = \mathtt{I}$ et $\mathbf{t}_i = \mathbf{0}$. Donc :

$$\mathsf{P}_1^E = (\quad \mathsf{K} \quad | \mathbf{0}) \ .$$

Ainsi, la transformation T_P^E est contrainte d'avoir la forme :

$$\mathsf{T}_P^E \sim \begin{pmatrix} \mathsf{K} & \mathbf{0} \\ (a,b,c) & d \end{pmatrix} \ . \tag{6.2}$$

Puisque T_P^E et les unités 3D sont définies à un facteur près et puisque d doit être non nul pour que la transformation soit non singulière, nous pouvons imposer $d = 1$. Notons $\mathbf{a} = (a,b,c)^\mathsf{T}$. La matrice K est la matrice des paramètres intrinsèques à trouver et le vecteur $\mathbf{a}$ représente le plan à l'infini. Concrètement, $(\mathbf{a}^\mathsf{T}, 1)^\mathsf{T}$ sont les coordonnées du plan dans le monde euclidien qui, dans la reconstruction projective, a les coordonnées $(0,0,0,1)^\mathsf{T}$. Le vecteur $(\mathbf{v}^\mathsf{T}, 1)^\mathsf{T} = (-\mathbf{a}^\mathsf{T}\mathsf{K}^{-1}, 1)^\mathsf{T}$ contient les coordonnées du *vrai* plan à l'infini, dans la reconstruction projective. Une approche stratifiée, explicitant une structure affine intermédiaire, tâchera donc d'estimer $\mathbf{v}$ plutôt que $\mathbf{a}$.

Les matrices de projection P_i^P pour $i > 1$ sont transformées d'après :

$$\mathsf{P}_i^P \mathsf{T}_P^E \sim \left(\quad \overline{\mathsf{P}}_i^P \mathsf{K} + \mathbf{p}_i^P \mathbf{a}^\mathsf{T} \quad \middle| \, \mathbf{p}_i^P \right) \ ,$$

où $\mathsf{P}_i^P = (\quad \overline{\mathsf{P}}_i^P \quad | \mathbf{p}_i^P)$. La contrainte (6.1) est satisfaite s'il existe des matrices de rotation R_i, des vecteurs $\mathbf{t}_i$ et des scalaires non nuls λ_i tels que :

$$\overline{\mathsf{P}}_i^P \mathsf{K} + \mathbf{p}_i^P \mathbf{a}^\mathsf{T} = \lambda_i \mathsf{K}\mathsf{R}_i \qquad \text{et} \qquad \mathbf{p}_i^P = -\lambda_i \mathsf{K}\mathsf{R}_i \mathbf{t}_i$$

(remarquons que, étant donné le choix du repère projectif, cette contrainte est directement satisfaite pour la première vue). S'il existe de telles R_i et K, il en est de même pour les $\mathbf{t}_i$: $\mathbf{t}_i = -\frac{1}{\lambda_i}\mathsf{R}_i^\mathsf{T}\mathsf{K}^{-1}\mathbf{p}_i^P$. Donc, la contrainte pour trouver la structure euclidienne est l'existence de R_i et K avec :

$$\overline{\mathsf{P}}_i^P \mathsf{K} + \mathbf{p}_i^P \mathbf{a}^\mathsf{T} \sim \mathsf{K}\mathsf{R}_i \ . \tag{6.3}$$

Les différentes méthodes d'auto-calibrage ne font, en effet, rien d'autre que d'exploiter cette contrainte d'une façon ou d'une autre. Toutes les méthodes éliminent la matrice de rotation R_i de l'équation (6.3). Nous explicitons deux procédés d'élimination. Le premier permet à Pollefeys et Van Gool de formuler la « contrainte du module » (voir plus bas). Le deuxième procédé élimine R_i en multipliant les deux côtés de l'équation par leurs transposées respectives ce qui donne :

$$\mathsf{P}_i^P \begin{pmatrix} \mathsf{K}\mathsf{K}^\mathsf{T} & \mathsf{K}\mathbf{a} \\ \mathbf{a}^\mathsf{T}\mathsf{K}^\mathsf{T} & \mathbf{a}^\mathsf{T}\mathbf{a} \end{pmatrix} (\mathsf{P}_i^P)^\mathsf{T} \sim \mathsf{K}\mathsf{K}^\mathsf{T} \quad . \tag{6.4}$$

Dans les paragraphes suivants, nous décrivons plusieurs méthodes d'auto-calibrage. Nous montrons les liens qui existent entre les méthodes, via l'équation (6.4).

6.2.1.1 L'approche de Heyden et Åström

Heyden et Åström [88] estiment K et $\mathbf{a}$ en minimisant le critère :

$$\sum_{i=2}^{m} ||\lambda_i \mathsf{P}_i^P \begin{pmatrix} \mathsf{K}\mathsf{K}^\mathsf{T} & \mathsf{K}\mathbf{a} \\ \mathbf{a}^\mathsf{T}\mathsf{K}^\mathsf{T} & \mathbf{a}^\mathsf{T}\mathbf{a} \end{pmatrix} (\mathsf{P}_i^P)^\mathsf{T} - \mathsf{K}\mathsf{K}^\mathsf{T}||_F \quad ,$$

où $|| \cdot ||_F$ désigne la norme de Frobenius. Pour tenir compte du fait que l'équation (6.4) n'est valable qu'à un facteur près, des facteurs d'échelle λ_i sont introduits pour chaque vue comme inconnues supplémentaires. L'initialisation du processus d'optimisation est basée sur un calibrage approximatif K_0 à partir duquel des valeurs initiales pour $\mathbf{a}$ et les λ_i sont calculées linéairement.

Pollefeys et Van Gool suivent l'approche de Heyden et Åström, en évitant les inconnues λ_i supplémentaires, au prix d'équations de degré supérieur [152]. Comme paramètres du plan à l'infini, le vecteur $\mathbf{v} = \mathsf{K}^{-\mathsf{T}}\mathbf{a}$ est utilisé (notation différente de [152]) et il est montré comment exprimer les λ_i en fonction des éléments de $\mathbf{v}$ et des matrices de projection P_i^P.

6.2.1.2 L'approche de Hartley

Hartley a été le premier à proposer une approche d'auto-calibrage « stratifiée », passant d'une reconstruction projective à une euclidienne, via une reconstruction affine [79, 80]. Comme Pollefeys et Van Gool, il estime le plan à l'infini via $\mathbf{v} = \mathsf{K}^{-\mathsf{T}}\mathbf{a}$. Sa méthode est un peu moins directe que celle de Heyden et Åström : à chaque étape du processus d'estimation, une décomposition QR de la matrice gauche de l'équation (6.3) est effectuée (analytiquement) : $(\bar{\mathsf{P}}_i^P + \mathbf{p}_i^P \mathbf{v}^\mathsf{T})\mathsf{K} = \mathsf{K}'_i \mathsf{R}'_i$, où K et $\mathbf{v}$ sont les estimations courantes. Avec les bonnes valeurs K et $\mathbf{v}$, les K'_i devraient être égales à K. Donc, le critère de minimisation est basé sur la différence entre ces deux matrices. Plus exactement, les matrices $\mathsf{X}_i = \mathsf{K}^{-1}\mathsf{K}'_i$ sont calculées et normalisées par des facteurs α_i pour que la somme des carrés des éléments diagonaux de X_i soit égale à 3. Le critère de minimisation est finalement le suivant : $\sum_{i=2}^m ||\alpha_i \mathsf{X}_i - \mathtt{I}||^2$.

Hartley a aussi décrit comment, étant donné (une estimation du) le plan à l'infini, on peut retrouver le calibrage en résolvant des équations linéaires, suivi par la décomposition de Cholesky d'une matrice définie positive (nous décrivons cette méthode dans l'annexe C). Ceci est utilisé dans sa méthode d'auto-calibrage pour l'initialisation des inconnues K et $\mathbf{v}$, évitant ainsi le besoin d'un calibrage approximatif : l'initialisation est basée sur la recherche de vecteurs $\mathbf{v}$ qui représentent des plans à l'infini potentiels. L'espace de recherche est borné par les « contraintes de cheiralité » [78] qui imposent que le plan à l'infini ne coupe pas la scène observée. Seuls des $\mathbf{v}$ sont admis qui permettent de déterminer une matrice des paramètres intrinsèques (c-à-d qui donnent lieu à une matrice X définie positive ; cf. l'annexe C).

En pratique, Hartley utilise plusieurs $\mathbf{v}$ choisis arbitrairement pour augmenter les chances de convergence vers la bonne solution. La détermination d'un tel vecteur $\mathbf{v}$ est désignée par le terme de reconstruction « quasi-affine » ce qui reflète que $\mathbf{v}$ représente un plan à l'infini *potentiel*. À partir d'une reconstruction quasi-affine, la reconstruction affine n'est pas obtenue de manière isolée, mais simultanément avec la détermination des paramètres intrinsèques et donc de la reconstruction euclidienne.

6.2.1.3 L'approche de Triggs

Triggs propose une approche similaire à celle de Heyden et Åström [210]. L'équation de base utilisée correspond à (6.4). La différence essentielle réside dans la paramétrisation différente de la matrice symétrique 4×4 :

$$\Lambda = \begin{pmatrix} \mathsf{K}\mathsf{K}^\mathsf{T} & \mathsf{K}\mathbf{a} \\ \mathbf{a}^\mathsf{T}\mathsf{K}^\mathsf{T} & \mathbf{a}^\mathsf{T}\mathbf{a} \end{pmatrix} ,$$

qui apparaît dans l'équation (6.4). Triggs a montré que cette matrice représente le cône dual de la conique absolue, cette dernière vue comme conique dans l'espace 3D (cf. la figure 6.1).

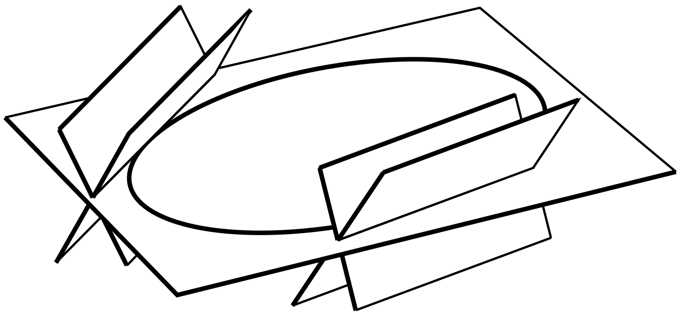


FIG. 6.1: *Le cône dual à une conique 3D est constitué de l'ensemble des plans tangents à la conique. Le plan singulier du cône est le plan support de la conique.*

Dans [210], ce cône dual est appelé « quadrique absolue » ce qui entre en conflit avec la définition donnée dans la section 2.7.3. Nous préférerions de le nommer « cône dual absolu ». Triggs paramètre Λ non pas par K et $\mathbf{a}$ mais le représente par les 10 éléments d'une matrice symétrique générique. Ainsi, le degré de la formulation en les inconnues est baissé. La paramétrisation en K et $\mathbf{a}$ implique que Λ est de rang 3, tandis que Triggs impose cette condition par une estimation sous contrainte ($\det \Lambda = 0$).

La deuxième matrice inconnue dans l'équation (6.4), $\mathsf{K}\mathsf{K}^\mathsf{T}$, est également représentée par une matrice symétrique générique, de dimension 3×3, ce qui amène le nombre d'inconnues à 16 (10+6). Les facteurs d'échelle λ_i sont éliminés par multiplication croisée. Une autre différence par rapport aux approches précédentes est que les repères de reconstruction ne sont pas fixés. Ainsi, la matrice T_P^E n'aura pas la forme (6.2). La contrainte que la première sous-matrice 3×3 de $\mathsf{P}_i^P \mathsf{T}_P^E$ doit être égale à $\mathsf{K}\mathsf{R}_i$ reste tout de même valable, ce qui peut s'écrire comme :

$$\mathsf{P}_i^P \mathsf{T}_P^E \begin{pmatrix} \mathtt{I}_3 \\ \mathbf{0}^\mathsf{T} \end{pmatrix} \sim \mathsf{K}\mathsf{R}_i \ .$$

L'élimination de R_i par multiplication des deux côtés de l'équation avec leurs transposées respectives mène à :

$$\mathsf{P}_i^P\ \mathsf{T}_P^E\ \begin{pmatrix} \mathtt{I} \\ \mathbf{0}^\mathsf{T} \end{pmatrix} \left(\ \mathtt{I}\ |\mathbf{0}\right) (\mathsf{T}_P^E)^\mathsf{T}\ (\mathsf{P}_i^P)^\mathsf{T}\ \sim\ \mathsf{K}\mathsf{K}^\mathsf{T}$$

ou :

$$\mathsf{P}_i^P \, \mathsf{T}_P^E \underbrace{\begin{pmatrix} \mathtt{I} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 0 \end{pmatrix}}_{\Lambda_\infty} (\mathsf{T}_P^E)^\mathsf{T} \, (\mathsf{P}_i^P)^\mathsf{T} \; \sim \; \mathsf{K}\mathsf{K}^\mathsf{T} \quad .$$

La matrice Λ_∞ représente le cône dual absolu dans le monde euclidien, et $\Lambda = \mathsf{T}_P^E \, \Lambda_\infty \, (\mathsf{T}_P^E)^\mathsf{T}$ est sa représentation dans le repère de la reconstruction projective.

Triggs a développé deux méthodes d'auto-calibrage, une quasi-linéaire et une non linéaire. En 7.10, nous décrivons une dégénérescence artificielle de la méthode linéaire qui est due à la négligence de la contrainte de rang 3 de Λ.

6.2.1.4 L'approche de Pollefeys et Van Gool : utilisation de la contrainte du module

Les approches précédentes estiment simultanément la position du plan à l'infini et le calibrage de la caméra, en passant directement d'une reconstruction projective à une euclidienne. Nous décrivons dans ce paragraphe et le suivant deux façons d'éliminer l'un des deux ensembles d'inconnues, le calibrage K ou bien le plan à l'infini $\mathbf{a}$.

Multiplier l'équation (6.3) de droite par K^{-1} donne :

$$\overline{\mathsf{P}}_i^P + \mathbf{p}_i^P \mathbf{a}^\mathsf{T} \mathsf{K}^{-1} \sim \mathsf{K}\mathsf{R}_i\mathsf{K}^{-1} \quad .$$

En introduisant $\mathbf{v} = \mathsf{K}^{-\mathsf{T}}\mathbf{a}$ et en remarquant que le côté droit est l'homographie infinie, nous avons :

$$\overline{\mathsf{P}}_i^P + \mathbf{p}_i^P \mathbf{v}^\mathsf{T} \sim \mathsf{H}_{\infty 1i} \quad .$$

Le but est l'estimation du plan à l'infini, représenté par $\mathbf{v}$. Une contrainte sur $\mathbf{v}$ est que la matrice $\overline{\mathsf{P}}_i^P + \mathbf{p}_i^P \mathbf{v}^\mathsf{T}$ doit être conjuguée à un multiple d'une matrice de rotation. Il en découle que ses valeurs propres doivent avoir le même module, ce qui constitue la *contrainte du module* [120]. Elle peut être exprimée par une équation quartique en les éléments de $\mathbf{v}$. Dans [153], $\mathbf{v}$ est estimé à partir de ces équations.

Nous avons jusque-là considéré uniquement les homographies infinies entre la première vue et les autres. La contrainte du module est bien sûr aussi valable entre deux vues $i, j \neq 1$. Dans [150], il est montré que ceci donne également des équations quartiques et que le plan à l'infini peut être estimé à partir d'un minimum de 3 vues. Une fois le plan à l'infini estimé, le calibrage peut être déterminé par la méthode de Hartley [79, 80] (voir l'annexe C). Cette approche de Pollefeys et Van Gool est la seule qui est complètement stratifiée, c'est-à-dire passant par une reconstruction intermédiaire affine (à l'exception de méthodes comme par exemple [60] qui s'appuie sur des informations affines extraites de scènes urbaines).

En 7.10 nous montrons que la contrainte du module n'est pas une caractérisation complète du plan à l'infini. Son utilisation entraîne, en effet, des ambiguïtés artificielles pour la reconstruction affine.

6.2.1.5 L'approche de Faugeras, Maybank, Luong et Zeller : les équations de Kruppa

Dans plusieurs articles [63, 114, 123, 232, 233, 234], ces auteurs ont d'abord montré que l'auto-calibrage est possible et ont ensuite proposé des méthodes pratiques. Leur formalisme repose sur les équations de Kruppa dont nous montrons dans la suite le lien avec l'équation (6.3). Une dérivation alternative des équations de Kruppa est décrite en 6.4.3.

Les équations de Kruppa. Puisque $\mathsf{P}_1^P = (\mathtt{I} \quad |\mathbf{0})$, la matrice $\overline{\mathsf{P}}_i^P$, pour $i > 1$, représente l'homographie d'un plan arbitraire, par rapport aux vues 1 et i, et $\mathbf{p}_i^P$ est l'épipôle $\mathbf{e}_{i1}$ [232]. Donc, nous avons la relation suivante avec la matrice fondamentale : $\mathsf{F}_{1i} \sim [\mathbf{p}_i^P]_\wedge \overline{\mathsf{P}}_i^P$ (voir 2.6). Multiplier l'équation (6.4) de gauche et de droite par $[\mathbf{p}_i^P]_\wedge$ mène à :

$$\begin{pmatrix} [\mathbf{p}_i^P]_\wedge \overline{\mathsf{P}}_i^P & [\mathbf{p}_i^P]_\wedge \mathbf{p}_i^P \end{pmatrix} \begin{pmatrix} \mathsf{K}\mathsf{K}^\mathsf{T} & \mathsf{K}\mathbf{a} \\ \mathbf{a}^\mathsf{T}\mathsf{K}^\mathsf{T} & \mathbf{a}^\mathsf{T}\mathbf{a} \end{pmatrix} \begin{pmatrix} \overline{\mathsf{P}}_i^{P\mathsf{T}} [\mathbf{p}_i^P]_\wedge \\ \mathbf{p}_i^{P\mathsf{T}} [\mathbf{p}_i^P]_\wedge \end{pmatrix} \sim [\mathbf{p}_i^P]_\wedge \mathsf{K}\mathsf{K}^\mathsf{T} [\mathbf{p}_i^P]_\wedge$$

$$\begin{pmatrix} \mathsf{F}_{1i} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathsf{K}\mathsf{K}^\mathsf{T} & \mathsf{K}\mathbf{a} \\ \mathbf{a}^\mathsf{T}\mathsf{K}^\mathsf{T} & \mathbf{a}^\mathsf{T}\mathbf{a} \end{pmatrix} \begin{pmatrix} \mathsf{F}_{1i}^\mathsf{T} \\ \mathbf{0}^\mathsf{T} \end{pmatrix} \sim [\mathbf{p}_i^P]_\wedge \mathsf{K}\mathsf{K}^\mathsf{T} [\mathbf{p}_i^P]_\wedge$$

$$\mathsf{F}_{1i}\mathsf{K}\mathsf{K}^\mathsf{T}\mathsf{F}_{1i}^\mathsf{T} \sim [\mathbf{p}_i^P]_\wedge \mathsf{K}\mathsf{K}^\mathsf{T} [\mathbf{p}_i^P]_\wedge \quad . \tag{6.5}$$

L'équation matricielle (6.5) est une forme des équations de Kruppa. Par rapport aux approches précédentes, non seulement le plan à l'infini a été éliminé, mais aussi la reconstruction projective explicite : les matrices de projection P_i^P ont disparues et seules les matrices fondamentales et les épipôles (les $\mathbf{p}_i^P$) sont présents ! D'un côté, ceci simplifie l'auto-calibrage parce qu'il n'y a pas besoin d'établir une reconstruction projective initiale et on peut estimer l'auto-calibrage plus directement. D'un autre côté, l'estimation des matrices fondamentales est un processus local entre deux vues, tandis que la reconstruction projective prendra en compte toutes les vues et représentera mieux la structure 3D et les caméras.

La méthode de Zeller et Faugeras. Plusieurs méthodes d'auto-calibrage basées sur les équations de Kruppa ont été proposées. La méthode la plus pratique semble être celle proposée par Zeller et Faugeras [232, 233, 234]. Les paramètres intrinsèques sont estimés en minimisant un critère non linéaire basé sur les équations de Kruppa qui sont pondérées par une estimation de l'incertitude des matrices fondamentales.

Trois méthodes d'initialisation sont considérées, dont la mieux adaptée au cas multi-image repose sur une estimation robuste de α_u et α_v : sous l'hypothèse de pixels rectangulaires et du point principal au centre de l'image, α_u et α_v peuvent être calculés à partir d'une seule matrice fondamentale. Ceci est fait pour toutes les matrices fondamentales disponibles et les valeurs obtenues sont ensuite évaluées par une statistique robuste. Pour ce faire, les paires des valeurs de (α_u, α_v) sont affichées comme points dans un graphe. A priori, il devrait y avoir un point d'accumulation prononcé parmi ces points, à proximité des vraies valeurs. Zeller a constaté lors de ses expériences, qu'au contraire, les points sont dispersés le long d'une droite. Il lui a donc été impossible d'obtenir des valeurs initiales fiables pour α_u et α_v.

Tout de même, la droite permet de déterminer le rapport d'échelle, qui est en effet égal à la pente. Donc, il est proposé d'estimer de manière robuste la droite passant par les points (α_u, α_v) et d'en extraire la pente pour initialiser le rapport d'échelle. Il n'est pas dit explicitement, mais nous le supposons, que pour l'initialisation des valeurs individuelles de α_u et α_v, à l'un des deux est attribuée une valeur standard et le deuxième calculé en utilisant le rapport d'échelle estimé.

Dans [232, 233, 234], aucune explication n'est donnée pour le phénomène que les solutions pour α_u et α_v sont distribuées le long d'une droite, avec le rapport d'échelle comme pente. En 7.10, nous montrons de manière géométrique qu'il s'agit d'une dégénérescence artificielle du problème de l'auto-calibrage. Celle-ci est introduite par les équations de Kruppa pour la configuration spéciale des caméras utilisée par Zeller, notamment des caméras distribuées sur une sphère et dirigées vers le centre de la sphère (l'axe optique passe par le centre de la sphère). Les équations de Kruppa, par paires de vues, sont très instables à proximité d'une telle configuration, ce qui se manifeste par la dispersion des (α_u, α_v).

Il semble pourtant, que l'ensemble des équations, lors de l'estimation globale non linéaire, soit plus robuste et permette de trouver des valeurs α_u et α_v relativement proches des vraies valeurs, même à proximité de la configuration dégénérée. Nous pouvons supposer que seul le fait que les caméras ne sont pas parfaitement convergentes, permet de réussir l'auto-calibrage avec les équations de Kruppa. La méthode d'initialisation de Zeller pourrait être adaptée afin de traiter des configurations de caméras non convergentes. À ce moment-là, il devient éventuellement possible d'extraire non seulement le rapport d'échelle, mais aussi des valeurs individuelles pour α_u et α_v.

Une méthode d'optimisation globale par Luong et Faugeras. Luong et Faugeras décrivent dans [118] une méthode d'optimisation globale pour l'auto-calibrage, du style ajustement de faisceaux avec auto-calibrage (voir 6.2.10), c'est-à-dire une estimation simultanée des paramètres intrinsèques et extrinsèques, mais sans impliquer les points 3D. Le critère de minimisation est la somme des carrés des distances des points image et des droites épipolaires correspondantes. Il s'agit du critère standard pour l'estimation de la matrice fondamentale (voir 2.3.2), mais avec la différence qu'ici, les matrices fondamentales sont paramétrées par les paramètres intrinsèques et extrinsèques, via la matrice essentielle :

$$\mathsf{F}_{ij} = \mathsf{K}^{-\mathsf{T}}[\mathbf{t}_{ij}]_\wedge \mathsf{R}_{ij}\mathsf{K}^{-1} \quad .$$

Le processus de minimisation est initialisé par le résultat de l'auto-calibrage basé sur les équations de Kruppa (pour K) et suivi de la décomposition de matrices essentielles (pour $\mathbf{t}_{ij}$ et R_{ij}).

D'autres approches basées sur les équations de Kruppa. En dehors du groupe de Faugeras, les équations de Kruppa ont été utilisées pour l'auto-calibrage par Niini [140, 141] et Soatto et Perona [179] qui les ont incorporées dans un schéma d'estimation récursive. Hartley montre dans [83] comment établir les coefficients des équations de Kruppa numériquement, par une décomposition en valeurs singulières de la matrice fondamentale.

6.2.1.6 Modèles de calibrage simplifiés

Le modèle sténopé général comprend cinq paramètres intrinsèques, dont quelques-uns peuvent en pratique souvent être supposés connus a priori. Typiquement, l'angle entre les axes des pixels est en général très proche de 90° ; la matrice des paramètres intrinsèques prend alors la forme :

$$\mathsf{K} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad .$$

En outre, le point principal se trouve normalement proche du centre de l'image, du moins pour des caméras avec un objectif fixe. Une troisième simplification est due à des pixels carrés.

Les modèles simplifiés sont d'une part intéressants parce que les équations d'auto-calibrage deviennent moins complexes, d'autre part parce qu'ils permettent souvent d'initialiser l'auto-calibrage suffisamment bien pour le faire converger vers la bonne solution.

Hartley montre dans [77] que, à partir de la matrice fondamentale, deux paramètres intrinsèques des deux vues peuvent être déterminés. Il suppose que les pixels sont rectangulaires, que le rapport d'échelle est connu et que les points principaux se trouvent aux centres des images. Ainsi, il ne reste plus qu'un seul paramètre intrinsèque pour chacune des deux caméras, la distance focale, mesurée en pixels. Hartley dérive une formule pour calculer les deux distances focales (voir [26, 27, 139, 143, 144] pour d'autres approches similaires, développées par Brooks *et al.*).

Une méthode similaire consiste à estimer la distance focale *et* le rapport d'échelle, cette fois-ci en supposant que les deux vues sont prises par la même caméra, sans changement de focale. Les deux paramètres à estimer sont donc deux mesures de la distance focale, par rapport aux dimensions horizontale et verticale des pixels. Ponce *et al.* proposent une solution pour ce problème [155], mais qui est plus compliquée que nécessaire. Hippisley-Cox et Porrill décrivent une solution sous la forme d'une équation quartique [91]. Cette solution est appliquée par Zeller pour initialiser l'auto-calibrage (voir plus haut et [234]).

Azarbayejani et Pentland ont développé un algorithme de reconstruction euclidienne récursif qui peut prendre en compte le fait que la distance focale change entre les vues [9]. Les auteurs rapportent pourtant des difficultés dans l'estimation de la focale et les expériences sont finalement conduites avec une caméra calibrée qui ne change pas de focale pendant le mouvement.

6.2.2 Mouvements généraux et paramètres intrinsèques variables

Récemment, Heyden et Åström ont considéré l'auto-calibrage d'une caméra avec des pixels rectangulaires dont le rapport d'échelle est connu [89]. Il reste trois paramètres intrinsèques, la distance focale et les coordonnées du point principal, qui peuvent varier entre chaque prise d'image. Les auteurs donnent une preuve que l'auto-calibrage dans ces conditions est possible : la structure euclidienne ainsi que les trois paramètres intrinsèques de chaque vue peuvent être déterminés, si le mouvement de la caméra est suffisamment général. Heyden et Åström écrivent les équations de l'ajustement de faisceaux avec auto-calibrage pour ce cas particulier, mais aucune méthode d'initialisation n'est proposée. La conclusion principale de [89] est donc que l'ajustement de faisceaux *peut* converger vers la bonne solution, s'il est suffisamment bien initialisé.

Le résultat de Heyden et Åström a été étendu par Pollefeys *et al.* [151] qui prouvent que la seule contrainte de pixels rectangulaires suffit pour faire de l'auto-calibrage (en théorie). Ceci veut dire qu'on peut auto-calibrer une caméra en mouvement qui peut changer arbitrairement son calibrage intrinsèque entre les vues ou, ce qui est encore plus intéressant, de calibrer un système de caméras où, à la limite, on ne dispose que d'une seule image par caméra ! Le minimum de vues nécessaire pour le faire est 8. Pollefeys *et al.* ne proposent pas d'algorithme pratique pour ce cas général, mais ils considèrent pour la pratique que le rapport d'échelle est connu et que le point principal est proche du centre de l'image.

6.2.3 Méthodes basées sur des mouvements particuliers

6.2.3.1 Rotations pures

Basu [11, 12] et Du et Brady [49] proposent des schémas de « calibrage actif » qui sont basés sur des *petites* rotations d'élévation ou de vergence de la caméra, c'est-à-dire des rotations haut-bas ou gauche-droite. Les paramètres intrinsèques calculés sont α_u et α_v ainsi que, quant à l'approche de Basu, le point principal. Du et Brady estiment également la distorsion radiale (un coefficient).

Daniilidis et Ernst [41] appliquent une approche similaire à celle de Basu et Du/Brady. Outre des rotations autour d'axes connus (tilt ou pan), un point de fuite, donné par un ensemble de droites parallèles, est utilisé pour trouver le calibrage intrinsèque. La différence par rapport à la méthode de Basu est que les calculs ne sont pas basés sur l'hypothèse de petites rotations.

Hartley a proposé le premier une méthode de calibrage basée sur des rotations pures et générales autour du centre de projection [82]. La clé de la méthode est le fait que les images sont reliées par des homographies H_{ij} et que ces homographies sont égales aux homographies infinies (cf. 2.6). Donc, pour trouver K, la méthode de l'annexe C peut être appliquée.

6.2.3.2 Mouvements planaires

Comme mouvements planaires nous considérons les translations dans un plan Π et les rotations autour d'axes perpendiculaires à Π. Armstrong *et al.* ont développé une méthode d'auto-calibrage pour ce type de mouvement [6]. Les mouvements planaires sont dégénérés pour l'auto-calibrage de tous les paramètres intrinsèques et la reconstruction euclidienne, mais la solution peut tout de même être déterminée à un paramètre près (cf. le chapitre 7). La structure affine est déterminée entièrement ainsi que la structure euclidienne dans le plan de mouvement Π. Nous décrivons brièvement le principe de la méthode de Armstrong *et al.* Les projections des entités géométriques suivantes restent fixes lors de mouvements planaires :

- ▹ les projections du point à l'infini commun aux axes de rotation ;
- ▹ les projections des points cycliques de Π.

Les projections de ces entités peuvent être déterminées à partir de matrices fondamentales et de tenseurs trifocaux. Le point à l'infini des axes de rotation et les deux points cycliques de Π engendrent le plan à l'infini ce qui donne la structure affine. Les points cycliques donnent évidemment la structure euclidienne dans Π (et dans les plans parallèles à Π). La détermination des points cycliques peut être formulée en termes d'une caméra linéaire (voir 6.3).

Wiles et Brady ont proposé plusieurs modèles de caméra pour le mouvement planaire (« ground plane motion camera models ») [226]. Ils proposent une méthode de reconstruction qui est basée sur ces modèles et qui mène bien sûr au même type de structure que la méthode de Armstrong.

6.2.3.3 Mouvements particuliers entre des vues consécutives

Viéville et Lingrand ont étudié l'utilisation de mouvements particuliers entre des vues consécutives pour l'auto-calibrage [216]. Concrètement, les formes particulières de la matrice fondamentale pour quelques mouvements (voir 2.3.4) permettent de simplifier les équations de Kruppa et ainsi l'auto-calibrage. Ces considérations sont spécialement utiles pour l'auto-calibrage de caméras montées sur un robot manipulateur, qui effectue souvent des mouvements « réguliers » tels que des rotations autour d'un axe fixe. Viéville et Lingrand proposent un schéma de décision automatique sur la forme de la matrice fondamentale pour des images consécutives. Il s'agit d'une méthode de sélection de modèle, qui est basée sur un compromis entre la taille du modèle (nombre de paramètres libres) et sa qualité (résidu) ; voir [204] pour une méthode similaire.

6.2.4 Méthodes basées sur des mouvements initiaux particuliers

L'approche de Armstrong *et al.* [5] est basée sur un mouvement initial purement translationnel, qui rend possible la reconstruction affine de la scène [131]. Une fois la structure affine acquise, la détermination du calibrage devient relativement simple [79] (voir aussi l'annexe C).

Pollefeys et Van Gool [154] étendent l'approche de Armstrong *et al.* pour pouvoir traiter des changements de focale. Pour ce faire, deux mouvements initiaux sont demandés : avec un changement de focale, le point principal est déterminé comme étant le point d'intersection commun des « trajectoires » des points image. Puis, comme Armstrong, une translation pure sert à récupérer la structure affine. Ensuite, des déplacements arbitraires, avec changement de focale éventuel, permettent de trouver la structure euclidienne et le calibrage des vues. Cette dernière étape peut être considérée comme l'extension de l'équation (C.1), prenant en considération des distances focales différentes avant et après le mouvement.

6.2.5 Méthodes basées sur des mesures des déplacements de la caméra

Dron considère le problème du calibrage d'une caméra montée sur un support dont la composante translationnelle du mouvement est connue très précisément [48]. Elle dérive des équations linéaires, en supposant que soit la rotation relative entre deux vues est petite, ou bien que le déplacement entre deux vues est beaucoup plus grand que la distance entre la caméra et son support. Dron n'estime en effet pas la matrice des paramètres intrinsèques de la caméra, mais le produit KR_0, où R_0 est la rotation relative entre la caméra et son support.

Viéville décrit comment on peut calibrer une caméra effectuant des rotations autour d'un axe fixe, si les angles de rotation sont connus [214]. Stein [183] et McLauchlan et Murray [127] ont développé des méthodes pour le même scénario. La méthode de Stein inclut l'estimation des distorsions radiales (deux coefficients et deux coordonnées pour le point de symétrie ou centre de distorsion).

6.2.6 Méthodes basées sur un pré-calibrage

Dans la section 6.4, nous décrivons comment l'auto-calibrage d'une caméra avec une focale variable peut être simplifié en exploitant les résultats d'un pré-calibrage de la caméra [186, 189].

6.2.7 Méthodes basées sur un calibrage initial

Crowley *et al.* proposent de mettre à jour la matrice de projection lors de mises au point, de changements de diaphragme (aperture) ou de rotations de la caméra [40]. Le calibrage n'est pas directement mis à jour puisque la correction porte sur la matrice de projection perspective entière, sans en extraire les paramètres intrinsèques et extrinsèques. La mise à jour se fait par des transformations affines ou projectives, calculées à partir d'un suivi de points d'intérêt.

Enciso *et al.* font une chose similaire, mais leur but est d'adapter directement les paramètres intrinsèques de la caméra, lors de mises au point ou de changements de focale [51, 52, 53]. Ils ont montré qu'une transformation affine à 3 paramètres est suffisante pour modéliser le changement du calibrage. Ce modèle n'est valable que si la caméra ne se déplace pas lors de la mise au point ou du changement de focale.

6.2.8 Auto-calibrage d'une caméra affine

Quan a proposé une méthode élégante pour l'auto-calibrage d'une caméra affine [164] qui repose sur la notion, introduite dans le même article, des paramètres intrinsèques affines. Il s'agit d'une méthode stratifiée dont la comparaison avec les méthodes ultérieures pour la caméra perspective révèle des fortes similitudes. Les approches de Heyden et Åström [88] et Triggs [210] suivent le même schéma que celle de Quan. Toutes ces approches partent d'une reconstruction non calibrée, obtenue en supposant des caméras arbitraires – pour la caméra affine, il s'agit d'une reconstruction affine et pour les perspectives, elle est projective. Ensuite, la reconstruction non calibrée est rectifiée en euclidienne en cherchant une transformation projective respectivement affine T qui transforme les matrices de projection en des matrices ayant la même matrice des paramètres intrinsèques (plus de détails sont donnés en 6.2.1). Ceci peut s'écrire comme :

$$\mathsf{P}_i \mathsf{T}\mathsf{T}^\mathsf{T} \mathsf{P}_i^\mathsf{T} \sim \mathsf{K}\mathsf{K}^\mathsf{T} \ .$$

Les matrices ont les dimensions suivantes pour le cas affine (projectif) : les matrices de projection P_i sont 2×3 (3×4), la transformation 3D T est 3×3 (4×4), et la matrice des paramètres intrinsèques K 2×2 (3×3). A priori, les mêmes méthodes peuvent être utilisées pour trouver T et K dans les deux cas. Pourtant, dans le cas affine, il existe une simplification. Rappelons l'interprétation géométrique de la matrice symétrique $\mathsf{T}\mathsf{T}^\mathsf{T}$ dans le cas perspectif (la matrice Λ dans le paragraphe 6.2.1.3) : elle représente le cône dual à la conique absolue, dans la reconstruction projective [210]. Un cône dual à une conique 3D est une quadrique de plans de rang 3. La contrainte du rang 3 sur $\mathsf{T}\mathsf{T}^\mathsf{T}$ doit être considérée dans l'estimation (Heyden garantit cette contrainte explicitement par la paramétrisation du problème, tandis que Triggs impose la contrainte numériquement).

Dans le cas affine, $\mathsf{T}\mathsf{T}^\mathsf{T}$ n'est rien d'autre que la conique duale à la conique absolue, dans le plan à l'infini ! C'est donc une conique de droites, mais *non* dégénérée, contrairement au cône dual dans le cas perspectif. Donc, il n'y a pas de contrainte supplémentaire sur T à respecter. La projection de la conique duale à la conique absolue est une conique dégénérée de deux points imaginaires conjugués qui se trouvent sur la droite à l'infini du plan image. Ces points sont les images des points cycliques du plan image (voir figure 6.2).

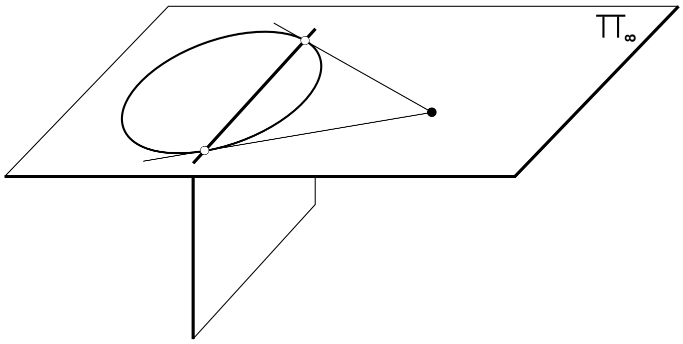


FIG. 6.2: *Projection de la conique absolue pour une caméra affine. Le centre de projection est le point directeur du plan image. Donc l'intersection du plan image avec le plan à l'infini est la droite polaire du centre de projection, par rapport à la conique absolue. La conique absolue duale se projette via les deux tangentes du centre de projection à la conique absolue. Ces tangentes touchent la conique absolue dans les points cycliques du plan image.*

6.2.9 Auto-calibrage d'un système stéréo

Brooks *et al.* ont développé des méthodes d'auto-calibrage pour des têtes stéréo, en adoptant un modèle de calibrage où seule la distance focale est inconnue (voir 6.2.1.6).

L'approche de Zhang, Luong et Faugeras [115, 238, 239] décompose l'auto-calibrage du système stéréo en deux étapes. D'abord, les deux caméras sont auto-calibrées séparément avec une approche basée sur les équations de Kruppa (voir 6.2.1.5). Ensuite, l'orientation relative des deux caméras est calculée, ce qui est similaire à la détermination du mouvement d'une caméra calibrée.

Zisserman *et al.* ont développé une méthode d'auto-calibrage qui exploite la rigidité du système stéréo [242]. C'est le premier travail qui considère véritablement un *auto-calibrage stéréo*. Des approches très similaires ont été proposées par Devernay et Faugeras [47] et Horaud et Csurka [96]. Ces méthodes sont basées sur des reconstructions projectives de la scène, obtenues des deux vues prises de chaque position du système. La clé est que les transformations entre deux reconstructions projectives de la même scène sont alors conjuguées au déplacement rigide du système stéréo. La forme particulière des transformations projectives permet de déterminer dans un premier temps le plan à l'infini dans les reconstructions projectives. Ayant ainsi trouvé la structure affine, les paramètres intrinsèques peuvent être déterminés comme décrit dans l'annexe C.

6.2.10 Ajustement de faisceaux avec auto-calibrage

Nous avons déjà décrit l'ajustement de faisceaux en 3.4. Une spécialisation de cette approche est l'*ajustement de faisceaux avec auto-calibrage* [18, 29, 37, 50, 54, 66, 67, 106, 125, 170, 178, 241]. Cette désignation met en évidence que l'ajustement de faisceaux se fait non seulement en estimant les orientations extrinsèques et les points 3D mais aussi les orientations intrinsèques des caméras. Un tel processus d'estimation globale nécessite généralement une bonne initialisation des paramètres à estimer. Un ajustement de faisceaux avec auto-calibrage peut être effectué pour optimiser les reconstruction et calibrage initiaux fournis par une méthode quelconque. En raison du besoin d'une bonne initialisation, nous ne considérons pas l'ajustement de faisceaux avec auto-calibrage comme une méthode autonome d'auto-calibrage.

6.3 Auto-calibrage d'une caméra linéaire

Ce problème a été évoqué par Olivier Faugeras et Long Quan et nous donnons ici une solution relativement simple.

Nous développons une équation qui peut être considérée comme l'analogue des équations de Kruppa de vues habituelles. Rappelons que le calibrage d'une caméra est acquis si l'image de la quadrique absolue de l'espace analogue est déterminée. Pour une caméra habituelle, il s'agit de la projection de la conique absolue, qui est aussi une conique dans l'image. Pour ce qui est des caméras linéaires, la quadrique absolue (du plan) est une conique de deux points, les points cycliques, tout comme sa projection. Les équations de Kruppa pour déterminer l'image de la conique absolue sont basées sur la géométrie épipolaire et ne sont en effet rien d'autre que des contraintes d'appariement bifocales de coniques, spécialisées aux cas de coniques identiques dans les vues (cf. 6.4.3). Quant aux caméras linéaires, il n'existe pas de contrainte d'appariement entre deux vues, mais un tenseur trifocal très simple (voir 2.4.1).

L'équivalent des équations de Kruppa serait une contrainte d'appariement de coniques (de deux points) entre trois vues, basée sur le tenseur trifocal. Le transfert de coniques de deux vues linéaires vers une troisième est relativement direct, mais il existe deux solutions : les coniques sont composées de deux points chacune ce qui donne lieu à deux possibilités pour l'appariement des points des coniques entre deux vues. À ce point, nous devons noter une propriété des points cycliques qui va simplifier le problème : non seulement l'image de la *conique* formée des deux points cycliques reste fixe pour une caméra en mouvement, mais aussi la projection des points cycliques individuels. Donc, il n'est pas nécessaire de considérer un transfert de coniques. Tout ce qu'il faut faire est de chercher les images des points cycliques parmi les points fixes dans les trois vues. Ceci se fait simplement en exprimant la relation trilinéaire pour trois points image avec les mêmes coordonnées et de résoudre cette équation pour ces coordonnées image :

$$\mathsf{G}_{2\times 2\times 2} \begin{pmatrix} u \\ 1 \end{pmatrix} \begin{pmatrix} u \\ 1 \end{pmatrix} \begin{pmatrix} u \\ 1 \end{pmatrix} = 0 \quad . \tag{6.6}$$

Nous avons normalisé la deuxième coordonnée à 1, ce qui n'engendre aucun problème puisque les images des points cycliques doivent être des points finis (il y a un seul point à l'infini sur la droite image, et celui-ci est un point réel). L'équation (6.6) est cubique en u :

$$G_{111}u^3 + (G_{112} + G_{121} + G_{211})u^2 + (G_{122} + G_{212} + G_{221})u + G_{222} = 0 \ .$$

Dans le cas général, cette équation a exactement trois racines. Parmi ces racines il y a les coordonnées des images des points cycliques, qui sont des points imaginaires conjugués. Les images des points cycliques peuvent donc être déterminées, puisque la troisième racine est réelle. Cette troisième racine correspond à un point réel dont l'image reste fixe dans les trois vues (voir la figure 6.3). Ce point n'a pas de signification pour le problème de l'auto-calibrage. Le même point est déjà apparu dans l'approche de Armstrong *et al.* pour l'auto-calibrage à partir de mouvements planaires [6] (cf. 6.2.3.2). La détermination de la structure euclidienne du plan de mouvement (et des plans parallèles) est une sous-étape de l'approche de Armstrong et peut être considérée comme l'auto-calibrage de la caméra linéaire obtenue par restriction des projections du monde 3D sur les points du plan de mouvement.

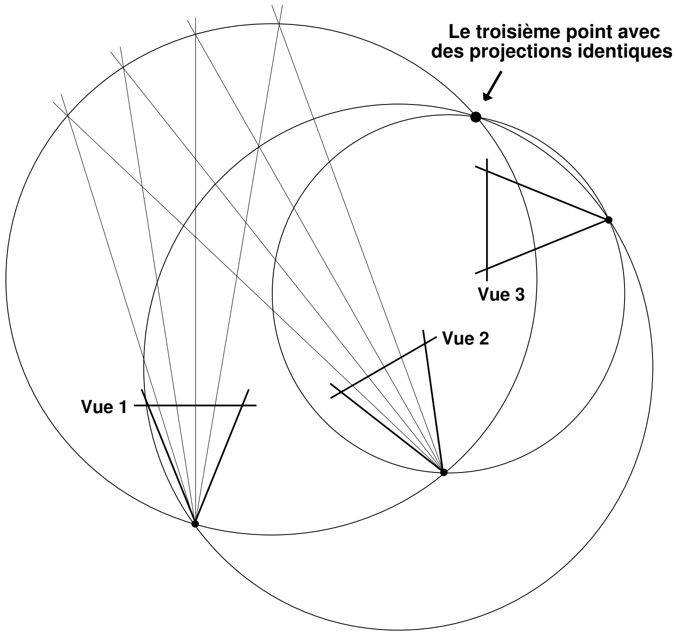


FIG. 6.3: *Le troisième point avec des projections identiques, en dehors des deux points cycliques. Pour toute paire de vues, les points 2D avec des projections identiques se trouvent sur une conique de Steiner. La construction de cette conique pour deux vues est esquissée par quelques points d'intersection de rayons de projection de points image avec les mêmes coordonnées. Les projections des points cycliques sont identiques dans toutes les vues. Les points cycliques sont alors sur toutes les coniques de Steiner considérées ; par conséquent, ces coniques sont des cercles. Les trois cercles, associés aux trois paires de vues, ont trois points en commun : les deux points cycliques et un troisième point qui est réel.*

6.3.1 Dégénérescence de l'équation d'auto-calibrage

Nous évoquons brièvement les situations où l'équation (6.6) dégénère. Puisqu'il y a toujours deux solutions complexes conjuguées (les images des points cycliques), il ne peut y avoir que deux possibilités pour la dégénérescence : le degré de l'équation se réduit à 2 – dans ce cas l'auto-calibrage est toujours possible – ou bien l'équation devient une tautologie. Dans le dernier cas, tous les couples de points imaginaires conjugués peuvent être confondues avec les images des points cycliques. Il y a donc au moins une famille de solutions pour l'auto-calibrage. L'équation devient une tautologie, si tous les termes disparaissent et c'est exactement le cas si les conditions suivantes sont satisfaites :

$$\begin{aligned} G_{111} &= 0 \\ G_{112} + G_{121} + G_{211} &= 0 \\ G_{122} + G_{212} + G_{221} &= 0 \\ G_{222} &= 0 \ . \end{aligned}$$

Il est certainement possible de déterminer les mouvements de caméra qui correspondent à ces équations. Dans la section 7.8, nous poursuivons une autre approche pour déterminer les *séquences de mouvements critiques*, pour un nombre *quelconque* de vues.

L'auto-calibrage d'une caméra linéaire avec plusieurs vues pourrait être effectué de manière analogue aux différentes approches pour les caméras habituelles, décrites au 6.2.1.

6.3.2 Calibrage pour une rotation pure

Comme les caméras habituelles, les caméras linéaires peuvent être calibrées à l'aide d'images prises en rotation pure (cf. 6.2.3.1). Tandis que le calibrage de caméras habituelles requiert au moins deux rotations (avec des axes différentes), la caméra linéaire peut être calibrée après une seule rotation (cf. la dérivation des séquences critiques en 7.8).

Les équations pour le calibrage sont analogues au cas de vues habituelles (voir l'annexe C). Nous ne répétons pas ces développements ici, mais donnons une solution explicite pour le calibrage. Soit $\mathsf{H}_{2\times 2}$ l'homographie entre les deux vues prises en rotation. Nous imposons que son déterminant soit égal à 1, ce qui fait que nous pouvons l'écrire sous la forme :

$$\mathsf{H} = \begin{pmatrix} a & b \\ c & \frac{1}{a}(1+bc) \end{pmatrix} \ .$$

L'équation de calibrage (cf. (C.1)) est :

$$\mathsf{HKK}^\mathsf{T}\mathsf{H}^\mathsf{T} = \mathsf{KK}^\mathsf{T} \ .$$

La solution de cette équation pour la matrice symétrique KK^T donne :

$$\mathsf{KK}^\mathsf{T} = \begin{pmatrix} -\frac{b}{c} & \frac{a^2-bc-1}{2ac} \\ \frac{a^2-bc-1}{2ac} & 1 \end{pmatrix} \ .$$

Nous pouvons en extraire directement la matrice des paramètres intrinsèques :

$$\mathsf{K} = \begin{pmatrix} \frac{\sqrt{-(a^2+2a+bc+1)(a^2-2a+bc+1)}}{2ac} & \frac{a^2-bc-1}{2ac} \\ 0 & 1 \end{pmatrix} \ .$$

6.4 Auto-calibrage pour un objectif à focale variable basé sur un pré-calibrage

6.4.1 Introduction

Dans cette section, nous considérons le problème de l'auto-calibrage d'une caméra en mouvement, équipée d'un objectif à focale variable (zoom). En pratique, les paramètres intrinsèques ne sont pas indépendants pour un tel système. Nous proposons d'effectuer une étape de *pré-calibrage* qui servira à modéliser l'interdépendance des paramètres intrinsèques. Nous montrons que l'auto-calibrage se réduit alors à l'estimation d'un seul paramètre. Nous développons une méthode qui exploite ceci et qui ne nécessite pas d'initialisation de la distance focale.

La méthode est basée sur les équations de Kruppa (voir 6.2.1.5). Les méthodes basées sur ces équations sont connues pour être assez sensibles au bruit dans les images [114]. Afin de modérer cet effet, plusieurs chercheurs proposent de fixer certains paramètres. Cela est justifié dans le cas du rapport d'échelle et de l'angle entre les axes des pixels, des entités qui sont très stables sur de longues périodes d'utilisation d'une caméra. Par contre, la position du point principal dépend de la position du zoom et de la mise au point de la caméra (voir la figure 6.4[1] [227]). Ce phénomène est dû à de mauvais alignements dans le système caméra – zoom et il peut être assez fort pour que les coordonnées du point principal changent jusqu'à 100 pixels lors de la variation de la distance focale (cf. la figure 6.8 (a)) !

D'une part, ceci démontre que le choix du point principal fixe peut perturber les résultats de l'auto-calibrage, d'autre part, la corrélation des paramètres intrinsèques semble être simple et stable. Or, si l'interdépendance entre les paramètres peut être modélisée de manière analytique, cette connaissance pourra être incluse dans le processus d'auto-calibrage afin de réduire le nombre d'inconnues à estimer et ainsi améliorer et robustifier les résultats.

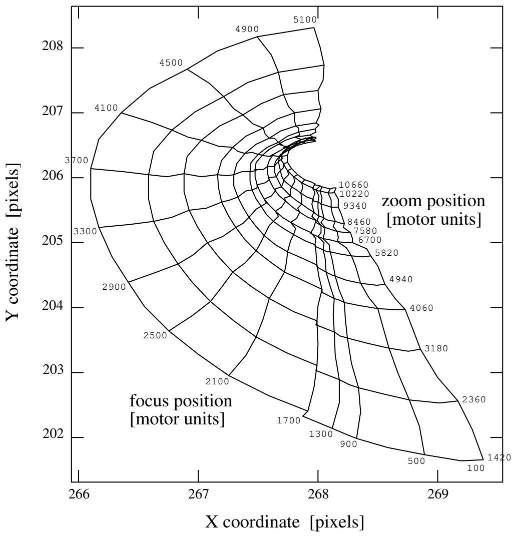


FIG. 6.4: *Position du point principal en fonction du zoom et de la mise au point [227].*

Dans cette section, nous examinons l'apport de l'interdépendance des paramètres au processus de l'auto-calibrage. Nous le faisons en étudiant la forme des équations de Kruppa par rapport à la

1. Nous remercions Reg Willson pour la permission de reproduire la figure 6.4.

modélisation de cette interdépendance. En 6.4.3 nous dérivons les équations de Kruppa. Ensuite, en 6.4.4, nous décrivons notre modélisation de l'interdépendance entre les paramètres intrinsèques et nous réécrivons les équations de Kruppa, en un seul paramètre. Une méthode pratique pour l'auto-calibrage basée sur cette modélisation est proposée en 6.4.5 et des résultats d'expériences avec cette méthode sont donnés en 6.4.6.

6.4.2 Stabilité et interdépendance des paramètres intrinsèques

Quelques-uns des paramètres intrinsèques sont très stables sur de longues périodes d'utilisation d'une caméra. Particulièrement, le rapport d'échelle $\tau = \frac{\alpha_u}{\alpha_v}$ et l'angle Θ entre les axes des pixels ne changent pas en général. En pratique, Θ est souvent très proche de $90°$. Ainsi, ce paramètre est généralement supposé connu et n'est pas estimé lors d'un calibrage.

Quant au point principal, il n'est souvent pas stable, plutôt sa position varie si la distance focale ou la mise au point de la caméra sont changées. Willson et Shafer ainsi que Burner ont observé que pour beaucoup d'objectifs à focale variable le point principal effectue un mouvement de translation lors d'un changement progressif de la focale [33, 227]. Nous avons retrouvé le même résultat lors de nos expériences (cf. la figure 6.8 (a)). Ce phénomène est dû à de mauvais alignements dans le système optique ou mécanique de la caméra et de l'objectif. Enciso *et al.* ont constaté un comportement différent [51, 52, 53]. La raison en est probablement une mécanique différente des objectifs : les changements de focale sont effectués par un vissage de lentilles tandis que notre objectif effectue un mouvement translationnel d'un bloc de lentilles.

Nous supposons dans la suite que pour des applications à courte distance, les changements de mise au point sont généralement de petite amplitude. Nous supposons alors que l'influence sur la position du point principal vient uniquement des changements de focale (en 6.4.7, nous proposons un moyen de prendre en compte aussi les effets de mise au point).

Malgré la révélation de l'influence du zoom sur les paramètres intrinsèques α_u, α_v, u_0 et v_0, l'interdépendance de ces paramètres n'a pas encore été utilisée dans le contexte de l'auto-calibrage. Afin de profiter de ces interdépendances nous tentons d'obtenir un modèle analytique qui exprime les paramètres intrinsèques en fonction de l'un parmi eux. Ici, nous choisissons α_v comme référence (Θ est considéré constant et connu) : les distances focales satisfont toujours la relation $\alpha_u(\alpha_v) = \tau \frac{\alpha_v}{\sin \Theta}$, tandis que $u_0(\alpha_v)$ et $v_0(\alpha_v)$ sont plus difficilement modélisables.

Nous appelons **pré-calibrage** la détermination hors ligne du rapport d'échelle τ, des relations fonctionnelles $u_0(\alpha_v)$ et $v_0(\alpha_v)$ ainsi que de l'angle Θ. Si une caméra est pré-calibrée, l'auto-calibrage revient alors à l'estimation de l'unique paramètre α_v. En 6.4.4, nous décrivons le processus de pré-calibrage et en 6.4.5 nous proposons une méthode d'auto-calibrage qui utilise un pré-calibrage.

6.4.3 Les équations de Kruppa

Pour effectuer l'auto-calibrage, nous utilisons les équations de Kruppa que nous avons déjà décrites en 6.2.1.5. Dans cette section, nous dérivons les équations de Kruppa de manière plus géométrique. Ces équations représentent en fait une contrainte d'appariement de coniques entre deux vues (cf. la section 2.3). La même contrainte est valable si les coniques dans les images sont les projections d'une conique 3D ou d'une quadrique.

Soient ω_1 et ω_2 les deux coniques image. Comme nous l'avons déjà décrit en 2.3, la géométrie épipolaire impose une contrainte d'appariement pour ces coniques : les droites épipolaires qui sont tangentes aux coniques sont des droites épipolaires correspondantes (à une permutation près – il y a deux paires de telles tangentes ; voir la figure 2.5). Ceci peut s'écrire comme :

$$(\mathbf{e}_1 \wedge \mathbf{q})^\mathsf{T} \omega_1^* (\mathbf{e}_1 \wedge \mathbf{q}) = 0 \quad \Leftrightarrow \quad (\mathsf{F}\mathbf{q})^\mathsf{T} \omega_2^* (\mathsf{F}\mathbf{q}) = 0 \ . \tag{6.7}$$

Ici, $\mathbf{e}_1 \wedge \mathbf{q}$ est une droite épipolaire dans la première image et $\mathsf{F}\mathbf{q}$ la droite épipolaire correspondante dans la deuxième image. Une droite $\mathbf{l}$ est tangente à une conique ϕ si elle est sur la conique duale ϕ^*, c'est-à-dire si $\mathbf{l}^\mathsf{T}(\phi^*)\mathbf{l} = 0$, d'où l'équation (6.7).

Les projections de la conique absolue sont identiques si les paramètres intrinsèques de la caméra ne changent pas entre les vues. Si ω_∞ désigne la projection de la conique absolue, nous avons donc :

$$(\mathbf{e}_1 \wedge \mathbf{q})^\mathsf{T} \omega_\infty^* (\mathbf{e}_1 \wedge \mathbf{q}) = 0 \quad \Leftrightarrow \quad (\mathsf{F}\mathbf{q})^\mathsf{T} \omega_\infty^* (\mathsf{F}\mathbf{q}) = 0 \ . \tag{6.8}$$

Si l'épipôle $\mathbf{e}_1$ n'est pas à l'infini, nous pouvons paramétrer le faisceau épipolaire par un paramètre t, via les points à l'infini dans l'image : $(1, t, 0)^\mathsf{T}$. L'équation (6.8) devient :

$$\left(\mathbf{e}_1 \wedge (1,t,0)^\mathsf{T}\right)^\mathsf{T} \omega_\infty^* (\mathbf{e}_1 \wedge (1,t,0)^\mathsf{T}) = 0 \quad \Leftrightarrow \quad \left(\mathsf{F}(1,t,0)^\mathsf{T}\right)^\mathsf{T} \omega_\infty^* (\mathsf{F}(1,t,0)^\mathsf{T}) = 0 \ .$$

Les deux termes de cette équivalence sont des polynômes quadratiques en t. Nous les écrivons comme :

$$c_2 t^2 + c_1 t + c_0 = 0 \quad \Leftrightarrow \quad c_2' t^2 + c_1' t + c_0' = 0 \ ,$$

où les coefficients c_i et c_i' dépendent de F, $\mathbf{e}_1$ et ω_∞^*. Cette équivalence signifie que les coefficients des deux polynômes sont égaux à un facteur multiplicatif près. Ceci s'exprime comme :

$$c_2 c_1' = c_1 c_2' \qquad c_2 c_0' = c_0 c_2' \qquad c_1 c_0' = c_0 c_1' \ . \tag{6.9}$$

Ces trois équations sont les **équations de Kruppa** (seules deux d'entre elles sont algébriquement indépendantes).

Si la géométrie épipolaire, i.e. F et $\mathbf{e}_1$, est connue, les équations (6.9) sont des contraintes sur ω_∞^*. Puisque ω_∞^* nous donne les paramètres intrinsèques de la caméra (cf. 2.7.3), les équations de Kruppa peuvent être utilisées pour l'auto-calibrage. L'approche habituelle consiste à minimiser un critère du style [114] :

$$\sum_{p=1}^{\text{\#paires de vues}} \left(\frac{c_{p2}}{c_{p2}'} - \frac{c_{p1}}{c_{p1}'}\right)^2 + \left(\frac{c_{p2}}{c_{p2}'} - \frac{c_{p0}}{c_{p0}'}\right)^2 + \left(\frac{c_{p1}}{c_{p1}'} - \frac{c_{p0}}{c_{p0}'}\right)^2 \ , \tag{6.10}$$

éventuellement incluant une pondération par l'incertitude sur les matrices fondamentales [233].

6.4.4 Pré-calibrage

Dans cette section, nous examinons comment les interdépendances des paramètres intrinsèques peuvent être modélisées et exprimons les équations de Kruppa prenant en compte cette modélisation.

6.4.4.1 Modélisation de l'interdépendance entre les paramètres intrinsèques

Nous supposons que le rapport des distances focales est donné par $\alpha_u = \tau \alpha_v$, où τ est connu et constant. En outre, l'angle Θ est supposé constant et égal à $90°$ et ne dépend donc pas d'autres paramètres. Quant aux coordonnées du point principal, nous essayons de les exprimer par rapport à α_v. Dans la suite, nous nous restreignons à des modèles polynômiaux :

$$\begin{aligned} u_0(\alpha_v) &= k_m\alpha_v^m + \ldots + k_1\alpha_v + k_0 \\ v_0(\alpha_v) &= l_n\alpha_v^n + \ldots + l_1\alpha_v + l_0 \ . \end{aligned} \tag{6.11}$$

Afin de déterminer les coefficients k_i et l_i de ces modèles, nous effectuons une série de calibrages avec une méthode classique utilisant une mire. Des calibrages doivent être effectués pour un jeu complet de positions du zoom. Ceci donne un tableau de valeurs des paramètres. Ensuite, on détermine les k_i et l_i de façon à approcher au mieux les valeurs obtenues lors de la série de calibrages. Pour construire un modèle suffisamment riche, mais tout de même de degré minimal m (resp. n, cf. les équations (6.11)), une méthode de régression robuste peut être appliquée [128].

6.4.4.2 Influence sur les équations de Kruppa

Les équations de Kruppa ont la forme suivante (α_u a été remplacé par $\tau\alpha_v$; les coefficients des termes ne sont pas indiqués dans cette équation) :

$$\begin{aligned} 0 \;=\; & \alpha_v^4 + \\ & \alpha_v^2(u_0^2 + v_0^2 + u_0v_0 + u_0 + v_0 + 1) + \\ & u_0^4 + u_0^3v_0 + u_0^2v_0^2 + u_0v_0^3 + v_0^4u_0^3 + u_0^2v_0 + u_0v_0^2 + v_0^3 + u_0^2 + u_0v_0 + v_0^2 + u_0 + v_0 + 1 \ . \end{aligned}$$

Si u_0 et v_0 sont modélisés comme dans les équations (6.11), alors les équations de Kruppa deviennent des polynômes en un seul paramètre α_v et de degré $\max\{4, 4{*}\max\{m,n\}\}$. Dans la suite, nous discutons les cas particuliers ($m = n = 0$) et ($m = n = 1$), puis le cas général.

$\mathbf{m = n = 0}$: point principal fixe. Ce cas a déjà été considéré par Hartley [77]. Il est intéressant de noter que les équations de Kruppa ont alors la forme $\alpha_v^4 + \alpha_v^2 + 1 = 0$. La détermination de α_v revient alors à trouver la solution de polynômes quadratiques, ce qui permet d'obtenir une solution analytique du problème.

$\mathbf{m, n \leq 1}$: point principal effectuant un mouvement de translation. Les équations de Kruppa sont des polynômes de degré 4 en α_v. Il existe alors aussi une solution analytique pour trouver des candidats pour α_v. Ce cas est souvent rencontré pour les objectifs à focale variable courants [227].

$\mathbf{m > 1}$ ou $\mathbf{n > 1}$. Les équations de Kruppa sont de degré supérieur à 5 et il n'y a pas de solution analytique. Cependant, puisqu'il s'agit de polynômes en une variable, la détermination des racines peut être faite avec des techniques numériques de manière stable.

6.4.5 Une méthode d'auto-calibrage basée sur un pré-calibrage

Nous montrons comment le processus de l'auto-calibrage peut être simplifié si les expressions analytiques de l'interdépendance des paramètres intrinsèques ont été déterminées par pré-calibrage. Trois équations de Kruppa (dont 2 algébriquement indépendantes) sont associées à chaque paire de vues. Soit k_{pi} la i^{e} équation de Kruppa de la p^{e} paire de vues. Les k_{pi} sont des polynômes en une seule variable α_v (voir 6.4.4.2). Le paramètre α_v recherché est donc la racine commune à toutes les équations k_{pi}.

Dans un premier temps, les racines des polynômes sont calculées séparément. Puisque α_v doit être réel et positif, seules les racines ayant ces propriétés sont retenues. Soit $S_{pi} = \{r_{pij} \mid j = 1, \ldots, n_{pi}\}$ l'ensemble des racines réelles et positives de l'équation k_{pi}. Nous recherchons la meilleure approximation de la racine qui est commune à toutes les équations (i.e. qui se trouve dans tous les ensembles S_{pi}). Dans ce contexte, nous définissons la « distance » d'un nombre r et d'un ensemble $S = \{r_1, \ldots, r_n\}$ comme :

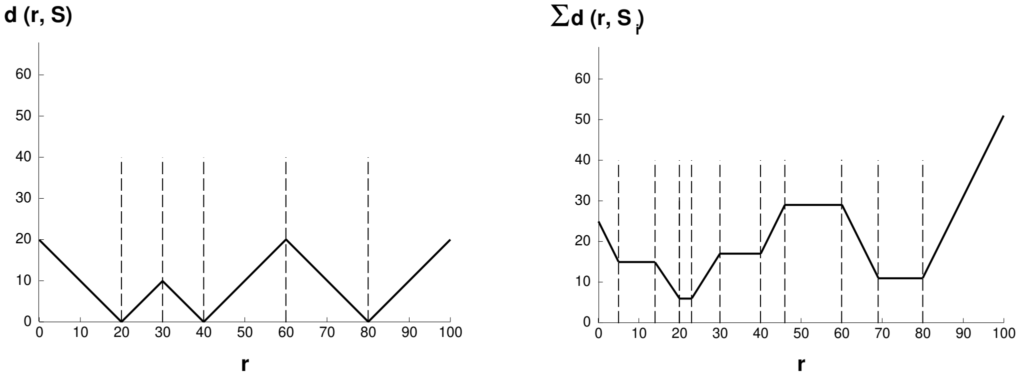
$$d(r, S) = \min_{j=1,\ldots,n} |r - r_j| \ .$$

Le problème de l'auto-calibrage peut alors être reformulé de la manière suivante. On cherche $\widehat{\alpha}_v$ avec :

$$\widehat{\alpha}_v = \text{argmin} \sum_{p,i} d(\alpha_v, S_{pi}) \ . \tag{6.12}$$

6.4.5.1 Méthode d'auto-calibrage

Nous proposons un schéma simple et efficace pour la résolution de ce problème. Considérons la figure 6.5 (a) qui montre le graphe de la fonction $d(r, S)$ pour un certain ensemble $S = \{r_1, \ldots, r_n\}$. La fonction est affine par morceaux et change de pente seulement aux abscisses r_j et $\frac{r_j + r_{(j+1)}}{2}$. Afin de trouver le minimum global de $\sum_{p,i} d(r, S_{pi})$, il est donc suffisant d'évaluer cette fonction aux seules abscisses indiquées. L'abscisse qui donne le minimum global est la solution pour α_v. Si un nombre d'équations pair est utilisé, cette abscisse n'est pas unique (voir la figure 6.5 (b), où le minimum global se trouve entre les abscisses 20 et 23). Dans ce cas, on choisit le milieu de l'intervalle comme solution pour α_v.



(a) Le graphe de la distance $d(r, S)$ avec $S = \{20, 40, 80\}$.

(b) Le graphe pour la somme des distances $d(r, S) + d(r, T)$ avec $S = \{20, 40, 80\}$ et $T = \{5, 23, 69\}$.

FIG. 6.5: *Illustration de la détermination de α_v à l'aide de l'équation (6.12). Les lignes verticales dans les graphes indiquent où les fonctions doivent être évaluées pour trouver le minimum global.*

Un exemple de graphes des 3 équations de Kruppa pour 2 vues est montré dans la figure 6.6.

Si beaucoup de vues sont utilisées pour l'auto-calibrage, des méthodes robustes [128] devraient être appliquées pour l'estimation de la racine commune.

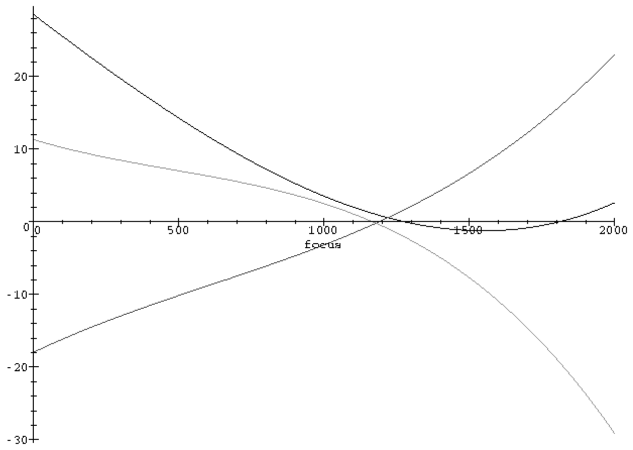


FIG. 6.6: *Exemple de graphes des 3 équations de Kruppa de 2 vues (en fonction de α_v).*

La méthode proposée peut être résumée par le schéma suivant.

Algorithme : Auto-calibrage basé sur un pré-calibrage

0. Pré-calibrage de la caméra : détermination des relations fonctionnelles des paramètres intrinsèques par rapport à α_v.
1. Estimation de matrices fondamentales à partir de correspondances de primitives dans les images.
2. Formulation des équations de Kruppa : calcul des coefficients des monômes de α_v.
3. Calcul des racines pour chaque équation de Kruppa ; considération des solutions réelles et positives.
4. Détermination de la « meilleure » racine commune à toutes les équations de Kruppa utilisées. Ceci est la solution pour α_v.
5. Calcul des autres paramètres intrinsèques par les relations fonctionnelles avec α_v établies à l'étape 0.

6.4.6 Résultats expérimentaux

Nous avons effectué des expériences avec un objectif EIA Servolens, monté sur une caméra de la marque Pulnix TM-6EX. D'abord, nous avons effectué un pré-calibrage : des calibrages utilisant une mire (voir la figure 6.7) ont été effectués pour plusieurs distances focales, avec la méthode de Faugeras et Toscani [65]. Les résultats de ces calibrages sont montrés dans les graphes de la figure 6.8, où u_0, v_0 et α_u sont affichés en fonction de α_v.

Ensuite, la relation fonctionnelle des paramètres intrinsèques a été déterminée. Nous avons trouvé les relations linéaires suivantes :

$$\alpha_u = 1.466\alpha_v \qquad u_0 = 0.060\alpha_v + 184.44 \qquad v_0 = -0.007\alpha_v + 273.19 \ . \tag{6.13}$$

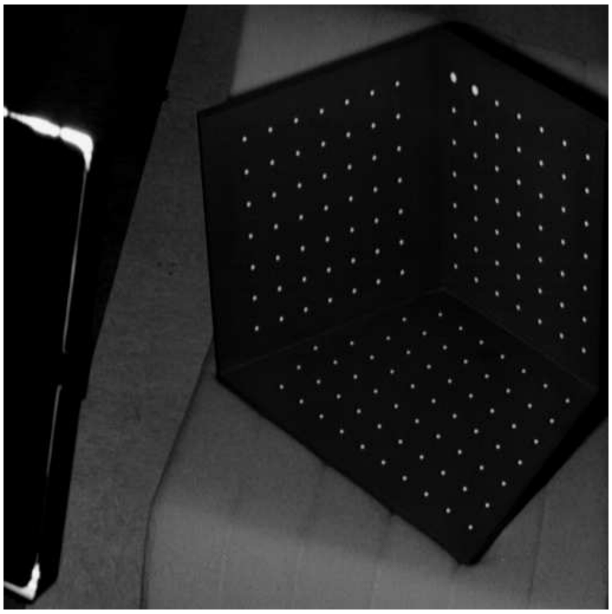
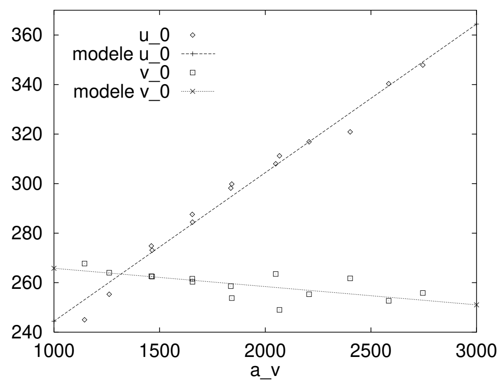
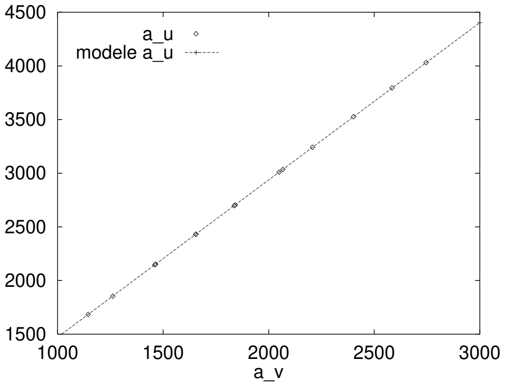


FIG. 6.7: *La mire de calibrage.*



(a) Coordonnées (u_0, v_0) du point principal (en pixels) en fonction de α_v.



(b) Les valeurs de α_u en fonction de α_v.

FIG. 6.8: *Les résultats des calibrages pour différentes distances focales. Pour chaque distance focale, des images de la mire ont été prises de plusieurs positions. Les droites dans les graphes représentent les modèles des interdépendances entre les paramètres intrinsèques.*

L'auto-calibrage a ensuite été effectué basé sur des images de la mire prises de différents points de vue, mais sans utiliser les coordonnées 3D des cibles. Nous comparons notre méthode avec un algorithme effectuant une minimisation non linéaire du critère (6.10) par une méthode du type Levenberg-Marquardt [160]. Contrairement à notre méthode, la minimisation non linéaire requiert une initialisation des paramètres intrinsèques. Nous utilisons 9 jeux d'initialisation différentes, qui sont tous cohérents avec le modèle (6.13) et qui couvrent tout l'intervalle des distances focales considérées.

La minimisation non linéaire est effectuée en 4 modes différents, selon les paramètres intrinsèques qui sont tenus fixes par rapport à leur valeur initiale :

- ▷ le rapport d'échelle τ ainsi que u_0 et v_0 ;
- ▷ le rapport d'échelle τ ;
- ▷ u_0 et v_0 ;
- ▷ aucun des paramètres.

En résumé, une expérience consiste à appliquer une fois notre algorithme (sans initialisation) et d'appliquer les 4 versions de la méthode non linéaire, chacune avec les 9 initialisations différentes. Nous avons effectué 27 expériences de ce type : pour chacune parmi 9 distances focales, 4 images de la mire ont été prises et l'auto-calibrage a été effectué en utilisant 2, 3 ou toutes les vues. Les résultats sont évalués en s'appuyant sur les résultats de calibrages avec la méthode de Faugeras et Toscani, prenant en compte les coordonnées 3D des cibles de la mire.

Dans la figure 6.9 nous résumons les résultats de l'estimation de α_u. Chaque barre représente le nombre d'expériences pour lesquelles l'erreur relative de l'estimation de α_u se trouve dans l'intervalle indiqué sur l'axe horizontal. Les barres en haut des graphes montrent les résultats de notre méthode et celles en bas les résultats de la méthode non linéaire. Nous avons trouvé que la méthode non linéaire atteint les meilleures performances si τ, u_0 et v_0 sont tenus fixes par rapport à leur valeur initiale. Seuls les résultats pour ce mode sont montrés.

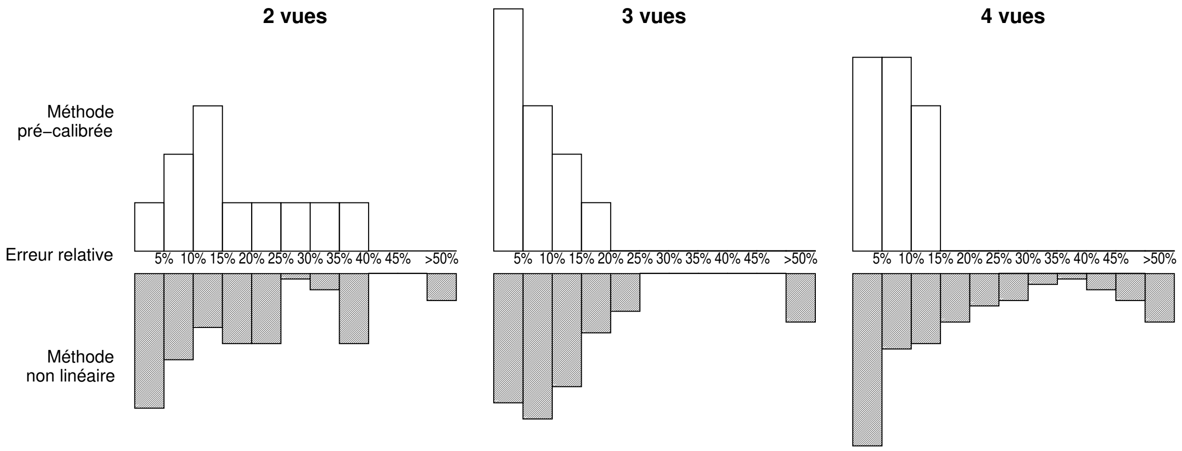


FIG. 6.9: *Résultats de l'auto-calibrage.*

On peut considérer que notre méthode n'a jamais échoué pour plus de 2 vues, tandis que la méthode non linéaire connaît un taux d'échec d'environ 5 %. Tout de même, si l'initialisation de la méthode non linéaire est proche des vraies valeurs, les résultats sont généralement bons.

6.4.7 Discussion

Nous avons démontré comment la prise en compte de l'interdépendance entre les paramètres intrinsèques permet de réduire le nombre de paramètres à estimer lors de l'auto-calibrage. Une méthode non itérative et sans besoin d'initialisation a été proposée et s'est avérée fiable lors d'expériences préliminaires. Le gros inconvénient de la méthode proposée est la nécessité d'un pré-calibrage qui est souvent difficile à réaliser. Il n'est pas clair si le schéma proposé sera efficace pour des systèmes de prise d'images où l'interdépendance des paramètres intrinsèques suit un modèle plus complexe que celui trouvé avec notre équipement.

Nous avons négligé les effets de la mise au point sur les paramètres intrinsèques. Ceux-ci pourraient être pris en compte en effectuant le pré-calibrage pour différentes mises au point. Pour l'auto-calibrage, les différents modèles d'interdépendance peuvent ensuite être évalués. Le meilleur est retenu, la décision pouvant se faire par exemple selon l'erreur de reprojection d'une reconstruction basée sur les paramètres intrinsèques trouvés.

Pour résumer, nous avons examiné si la prise en compte des interdépendances des paramètres intrinsèques peut être utile pour l'auto-calibrage. La réponse a priori est positive, mais une évaluation plus profonde de cette idée reste à faire.

6.5 Conclusion

Il existe une grande variété de méthodes que l'on peut classer sous le sigle de l'auto-calibrage. La plupart des méthodes sont conçues pour des caméras avec des paramètres intrinsèques invariants. Les équations sous-jacentes semblent être relativement bien étudiées et le succès des différentes méthodes dépend certainement largement de la qualité de l'implémentation numérique.

Un autre facteur qui a un impact sur la performance des algorithmes est le choix des points de vue. Beaucoup de séquences d'images sont typiquement prises avec une trajectoire très régulière de la caméra. Nous avons expérimenté la méthode de Triggs (voir 6.2.1.3) avec quelques séquences d'images disponibles sur Internet. Toutes les séquences testées sont dégénérées pour l'auto-calibrage général, c'est-à-dire pour la détermination de tous les 5 paramètres intrinsèques. Dans le chapitre 7 nous donnons une explication, en décrivant des mouvements de caméra critiques pour l'auto-calibrage. D'un côté, des mouvements réguliers sur l'ensemble d'une séquence d'images peuvent prohiber l'auto-calibrage, d'un autre côté des mouvements locaux bien choisis y sont souvent bénéfiques. La possibilité de choix de mouvements particuliers et la stratégie d'évitement de mouvements critiques dépendent finalement de l'application envisagée (voir la discussion en 7.12).

Ce qui est peut-être le plus intéressant pour des applications réelles est l'auto-calibrage de caméras à focale variable. L'auto-calibrage est possible même dans le cas extrême où le seul paramètre intrinsèque invariant est le fait que les pixels sont rectangulaires. En pratique, on pourra toujours faire des hypothèses sur l'invariance de quelques paramètres, comme le rapport d'échelle ou souvent le point principal. Nous avons examiné plus en détail un modèle plus flexible, où la position du point principal est modélisée par rapport à la distance focale. Un tel modèle simplifie l'auto-calibrage et peut le rendre plus fiable, mais ceci pour le prix d'un pré-calibrage.

En conclusion, l'application de l'auto-calibrage demande certaines précautions concernant la prise d'images et de bonnes méthodes numériques.

7. Séquences critiques de mouvements de caméra pour la reconstruction euclidienne non calibrée et l'auto-calibrage

Dans ce chapitre, nous étudions les séquences de mouvements de caméra qui mènent à des ambiguïtés inhérentes pour l'auto-calibrage et la reconstruction euclidienne non calibrée. Deux modèles de caméra sont traités – la caméra planaire habituelle et la caméra linéaire. Notre contribution majeure est une description complète des *séquences de mouvement critiques*. Nous identifions les cas qui ont de l'importance en pratique et dérivons le degré d'ambiguïté causé par les séquences critiques.

Une partie de ces résultats a été publiée dans [187, 188].

7.1 Introduction

Avant de décrire plus précisément notre travail, nous le plaçons dans un contexte général en passant en revue différents types d'ambiguïté qui ont été étudiés en vision tridimensionnelle. Nous distinguons grossièrement deux types d'études – des études portant sur des ambiguïtés intrinsèques ou inhérentes à un problème, qui ne permettent une solution unique avec aucun algorithme ; et des études sur l'instabilité en présence de bruit. Les deux phénomènes sont généralement étroitement liés puisque la présence de bruit pour un scénario à proximité d'une situation intrinsèquement dégénérée est souvent source d'instabilité dans les calculs numériques. Un troisième problème est celui des ambiguïtés qui sont introduites artificiellement par un algorithme utilisé mais qui ne sont pas inhérentes.

Nous mentionnons d'abord plusieurs types d'ambiguïtés inhérentes. Le problème le plus connu est peut-être celui des *surfaces critiques* qui ont été étudiées principalement pour le cas de deux vues et des correspondances de points [24, 92, 99, 100, 107, 113, 119, 122, 137]. Il existe certains types de quadriques tels que, si tous les points 3D observés et les deux centres de projection se trouvent sur la quadrique, la reconstruction des points (et le positionnement des caméra) est ambiguë. Le problème analogue pour des correspondances de droites dans trois vues a aussi été étudié [32, 135, 136]. Ces travaux décrivent aussi des ambiguïtés artificielles introduites par un algorithme spécifique (l'algorithme de Liu-Huang). Récemment, un résultat dual aux surfaces critiques concernant la reconstruction de 6 points vus dans plusieurs images a été obtenu par Maybank [124] : si les 6 points et tous les centres de projection se trouvent sur un certain type de quadrique, la reconstruction est ambiguë. Intuitivement, une configuration proche d'une surface critique provoque des instabilités numériques en présence de bruit, par exemple pour le calcul de la matrice fondamentale [119] ou généralement pour la reconstruction à partir de deux vues [92]. Au lieu de parler de surfaces critiques, on peut alors utiliser la notion de *volume critique* qui décrit l'espace de configurations proches à une surface critique [92]. Un résultat similaire aux surfaces critiques concerne le calibrage d'une caméra à partir d'une vue d'une scène connue. Si tous les points de référence et le centre de projection se trouvent sur une courbe cubique dans l'espace, il existent des solutions ambiguës [31].

Un autre type d'ambiguïté inhérente est l'existence de plusieurs solutions pour des données minimales ; ceci est bien connu pour des tâches diverses d'orientation. Le cas le plus populaire est celui de l'orientation relative entre deux vues calibrées à partir de cinq points [44, 61, 90, 99, 108, 138, 148], où au maximum 10 solutions peuvent exister. Le résultat analogue pour l'orientation relative non calibrée (estimation de la matrice fondamentale) est qu'il y a au maximum 3 solutions avec 7 points [108, 123, 155]. Pour la configuration duale (au sens de la dualité de Carlsson [34]) de 6 points vues dans 3 images, Quan a prouvé que la reconstruction projective est possible, aussi à 3 solutions près [163] (voir aussi [155]). L'ambiguïté de la pose d'une caméra calibrée à partir de 3 points de référence (4 solutions) et à partir de 4 points de référence coplanaires (2 solutions) est aussi bien connue [93, 184, 235]. Des résumés des problèmes d'orientation sont donnés par Wrobel [228] et Strunz [184].

Tandis que l'existence de surfaces critiques dépend et de la configuration de la scène et des positions des caméras, nous connaissons une autre famille d'ambiguïtés qui sont dues uniquement à des mouvements de caméra spéciaux et donc indépendantes de la scène observée. Il s'agit exactement du sujet de ce chapitre, notamment la reconstruction euclidienne non calibrée. Il est bien connu qu'à partir d'une séquence d'images obtenues lors d'un mouvement purement translationnel (même avec changement de direction), uniquement la structure affine peut être trouvée [131, 213]. Avec des mouvements planaires (translations dans un plan et rotations autour d'axes perpendiculaires au plan), la structure métrique de la scène peut être déterminée à une échelle, perpendiculaire au plan du mouvement, près [6, 226]. Une étude sur l'existence de solutions multiples pour l'auto-calibrage à partir de trois vues a été menée par Zeller et Faugeras [232, 234]. Il s'agit en effet d'une étude sur la dégénérescence des équations de Kruppa. Notre travail [187, 188] porte sur une caractérisation globale des mouvements qui créent des ambiguïtés dans la reconstruction euclidienne, affine ou dans l'auto-calibrage, pour un nombre arbitraire de vues et indépendamment de l'algorithme utilisé.

Passons brièvement des ambiguïtés inhérentes au problème d'instabilités dues au bruit dans les positions des points image, qui sont très prononcées pour des configurations spécifiques. Un problème très étudié est celui de l'instabilité de l'estimation du mouvement à partir de deux vues calibrées [2, 19, 42, 55, 56, 231]. Les instabilités sont généralement examinées par rapport à plusieurs facteurs, comme le type de mouvement, le champ de vue, le niveau de bruit dans les images, etc. Szeliski a récemment considéré l'instabilité de l'estimation du mouvement pour des séquences d'images orthographiques et perspectives [199, 200, 201] où il constate que le facteur primordial est l'étendue du mouvement rotationnel de la séquence entière. D'autres travaux concernent par exemple l'instabilité de l'estimation de la matrice fondamentale [119] et de la segmentation d'objets bougeant indépendamment [2].

Dans ce chapitre, nous étudions les ambiguïtés dans la reconstruction euclidienne à partir de séquences d'images non calibrées. Plus concrètement, nous nous intéressons aux ambiguïtés dues aux mouvements de la caméra et qui sont donc présentes indépendamment de la scène observée. Nous dérivons une caractérisation des séquences de mouvements critiques qui est indépendante du nombre de vues et de l'algorithme utilisé. Nous montrons également quels types de mouvements permettent d'obtenir au moins une reconstruction affine.

Organisation de ce chapitre. Les dérivations sont basées sur le concept de la quadrique absolue et son lien avec la reconstruction euclidienne et le calibrage intrinsèque. Dans les sections 7.2 à 7.3, nous passons en revue les principes du calibrage intrinsèque et de la reconstruction euclidienne calibrée. La reconstruction euclidienne non calibrée et ses sous-problèmes, l'auto-calibrage et la reconstruction affine non calibrée, sont traités dans les sections 7.4 et 7.6. Dans la section 7.7, nous introduisons des notations importantes pour ce chapitre et nous présentons une caractérisation des séquences critiques qui est à la base de leur dérivation.

Les séquences critiques pour la caméra linéaire sont dérivées dans la section 7.8 et celles pour la caméra habituelle avec des paramètres intrinsèques constants dans la section 7.9. Dans la section 7.10, nous décrivons quelques ambiguïtés artificielles qui sont dues à l'algorithme utilisé. Des aspects pratiques sur la réduction du degré de l'ambiguïté dans la reconstruction et l'identification automatique de séquences critiques sont discutés dans la section 7.11. Le chapitre est clos par des conclusions.

7.2 Calibrage intrinsèque

Nous rappelons que le calibrage intrinsèque d'une caméra peut être représenté par l'image de la quadrique absolue de l'espace correspondant (cf. la section 2.7.3). Nous illustrons la projection de la quadrique absolue pour les trois modèles de caméra considérés.

Le cas habituel d'une *caméra planaire* est illustré dans la figure 7.1. La quadrique absolue de $\mathcal{P}^3$ est la conique absolue Ω_∞. Sa projection ω_∞ est une conique qui est, tout comme la conique absolue, imaginaire. Puisque le centre de projection est en dehors du plan support de Ω_∞ (le plan à l'infini), la conique ω_∞ est en plus propre.

Le centre de projection d'une *caméra affine* (planaire) se trouve dans le plan à l'infini. L'image de la conique absolue est donc une conique dégénérée. Plus précisément, ω_∞ est une conique de deux points imaginaires conjugués qui se trouvent sur la droite à l'infini du plan image. Ces points sont les

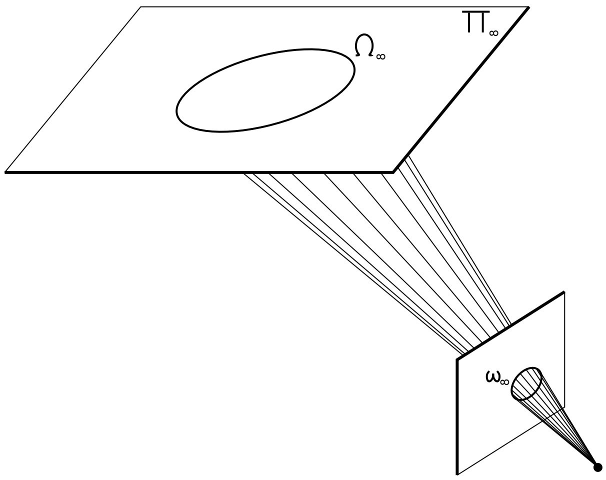


FIG. 7.1: *Projection de la conique absolue pour une caméra planaire.*

projections des points cycliques du plan image (voir figure 7.2, cf. aussi la section 6.2.8). Le centre de projection est un point réel ; les deux points image conjugués ne peuvent donc pas collaber en un seul point réel.

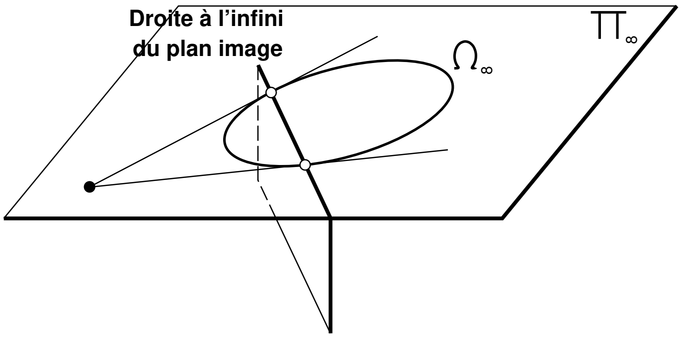


FIG. 7.2: *Projection de la conique absolue pour une caméra affine.*

La quadrique absolue pour la *caméra linéaire* est constituée des deux points cycliques du plan de mouvement. Nous ne considérons que le cas d'une caméra linéaire perspective, c'est-à-dire que le centre de projection se trouve hors la droite à l'infini du plan. La projection de Ω_∞ est une conique ω_∞ constituée de deux points imaginaires conjugués.

7.3 Reconstruction euclidienne calibrée

La reconstruction euclidienne est équivalente à la détermination de la quadrique absolue (cf. la section 2.7.3). Ici, nous considérons le cas de caméras calibrées, i.e. on connaît leur calibrage intrinsèque. Par conséquent, les projections de la quadrique absolue dans les images considérées sont connues. Nous décrivons dans la suite comment on peut déterminer la quadrique absolue en partant d'une reconstruction non calibrée (projective ou affine).

Considérons le cas de deux caméras planaires. Supposons qu'on a établi une reconstruction projective. Pour chaque caméra on peut construire le cône de projection de la conique absolue puisque

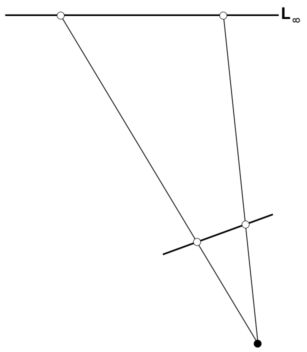


FIG. 7.3: *Projection des points cycliques pour une caméra linéaire.*

son image est connue. Ces deux cônes sont définis dans le repère de la reconstruction projective. L'intersection des cônes consiste en général de deux coniques 3D (voir la figure 7.4), dont une est la conique absolue. La conique absolue est alors déterminée à deux solutions près. Une troisième caméra permet de résoudre cette ambiguïté. Supposons que la vraie conique absolue a été identifiée. Pour obtenir la reconstruction euclidienne il suffit d'effectuer un changement de repère pour que la conique absolue retrouve sa position canonique.

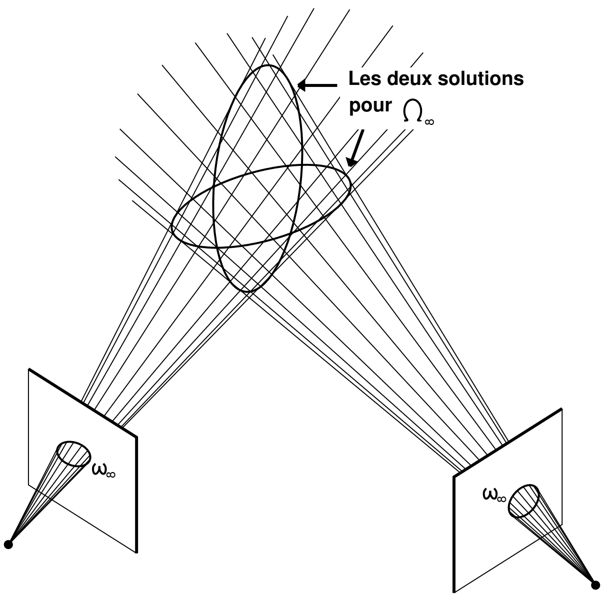


FIG. 7.4: *Détermination de la conique absolue avec 2 caméras planaires calibrées.*

Avec deux ou plusieurs caméras affines, la reconstruction affine est possible sans connaître le calibrage intrinsèque (cf. la section 3.5.2). Pour déterminer la conique absolue, il suffit d'opérer dans le plan à l'infini désormais connu. Chaque caméra affine calibrée donne deux contraintes sur la position de la conique absolue : les rayons de projection des deux points de ω_∞ sont tangents à Ω_∞. Il faut alors un minimum de trois caméras affines afin de pouvoir établir une reconstruction euclidienne (voir la figure 7.5).

Pour ce qui est des caméras linéaires, la reconstruction projective demande un minimum de 3 vues.

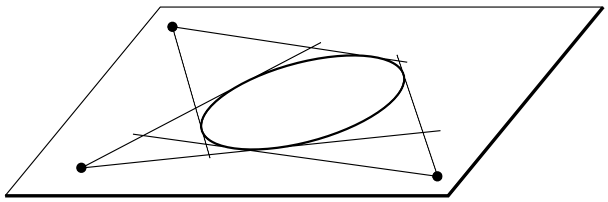


FIG. 7.5: *Détermination de la conique absolue avec 3 caméras affines calibrées.*

Ayant établi la reconstruction projective, les points cycliques peuvent être déterminés en n'utilisant que deux vues. Ils sont trouvés par intersection des rayons de projection de leurs images. Puisqu'il s'agit de deux paires de rayons de projection, il y a 4 points d'intersection et a priori il existent deux possibilités pour l'attribution des points cycliques. Si les deux vues sont obtenues par la même caméra (i.e. par des caméras avec le même calibrage intrinsèque), on peut résoudre cette ambiguïté (cf. la section D.1.1). Sinon, la troisième vue qui est nécessaire pour la reconstruction projective, permet également d'obtenir une solution unique (bien sûr à une transformation euclidienne près).

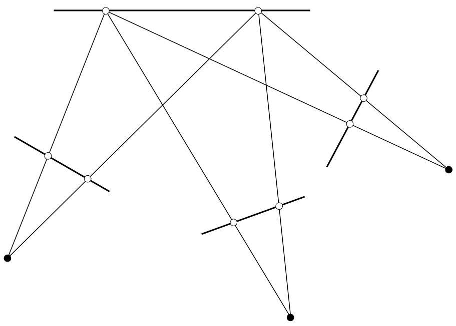


FIG. 7.6: *Détermination des points cycliques avec 3 caméras linéaires calibrées.*

7.4 Reconstruction euclidienne non calibrée

Comme pour le cas calibré, il s'agit de déterminer la quadrique absolue, mais cette fois-ci sans connaître le calibrage intrinsèque des caméras ! Nous excluons aussi des connaissances sur la structure 3D de la scène et sur le mouvement des caméras. Les seules informations a priori concernent la nature de la quadrique absolue et de ses projections :

Caméras perspectives La quadrique absolue est une conique 3D imaginaire et propre, tout comme ses projections.

Caméras affine La quadrique absolue est une conique 3D imaginaire et propre. Son plan support (le plan à l'infini) contient les centres de projection. Ses projections sont constituées de deux points imaginaires conjugués sur la droite à l'infini dans les images.

Caméra linéaire La quadrique absolue consiste de deux points imaginaires conjugués, tout comme ses projections.

La reconstruction euclidienne n'est possible que si des informations supplémentaires sont disponibles. Les seules informations que nous admettons sont la connaissance partielle de paramètres intrinsèques (e.g. le rapport d'échelle ou le fait que les pixels sont rectangulaires) ou le fait que les vues sont prises par la même caméra. Il ne s'agit que de contraintes indirectes sur la position de la quadrique absolue – elles sont effectivement des contraintes sur ses *projections*.

Nous appelons l'ensemble de ces informations sur la nature de la quadrique absolue et sur ses projections les **contraintes** pour la reconstruction euclidienne. Nous pouvons esquisser un procédé pour la reconstruction euclidienne comme suit. Supposons qu'une reconstruction projective (ou affine dans le cas de caméras affines) a été obtenue. La reconstruction euclidienne revient ensuite à la détermination de la quadrique qui satisfait toutes les *contraintes*. S'il y a une solution unique, il s'agit obligatoirement de la quadrique absolue. La reconstruction euclidienne est déterminée en effectuant un changement de repère qui ramène la quadrique absolue à sa position canonique.

S'il existent plusieurs quadriques satisfaisant les contraintes, il n'y a pas de solution unique pour la reconstruction euclidienne parce qu'il y a ambiguïté dans la détermination de la quadrique absolue. Nous appelons toutes ces quadriques des **quadriques absolues potentielles**. Si nous considérons explicitement une telle quadrique qui est *différente* de la vraie quadrique absolue, nous parlons d'une **fausse quadrique absolue**.

7.5 Auto-calibrage

Même si la reconstruction euclidienne n'a pas de solution unique, il est éventuellement possible de déterminer le calibrage intrinsèque d'une caméra. Le meilleur exemple pour cette situation est une caméra planaire effectuant des rotations pures autour du centre de projection (cf. la section 6.2.3.1).

Rappelons que la détermination des paramètres intrinsèques est équivalente à la détermination de la *projection* de la quadrique absolue. S'il existe une seule quadrique absolue potentielle, sa projection donne directement le calibrage intrinsèque. Sinon, il faut distinguer deux cas :

- ▷ les projections de toutes les quadriques absolues potentielles coïncident. Dans ce cas, le calibrage intrinsèque peut être déterminé bien que la reconstruction euclidienne ne soit pas possible ;
- ▷ il existent de fausses quadriques absolues dont des projections sont différentes de la projection de la vraie quadrique absolue. Dans ce cas, ni la quadrique absolue ni sa projection peuvent être identifiées de manière unique et l'auto-calibrage n'est pas possible.

Des exemples pour ces cas sont donnés plus bas.

7.6 Reconstruction affine non calibrée

La reconstruction affine est équivalente à la détermination de l'hyperplan à l'infini du monde. Une caractérisation complète de cet hyperplan est qu'il est le plan support de la quadrique absolue. S'il existe une seule quadrique absolue potentielle, son plan support doit alors être l'hyperplan à l'infini. Sinon, il faut distinguer deux cas :

- ▷ toutes les quadriques absolues potentielles sont « coplanaires ». Dans ce cas, l'hyperplan à l'infini peut être identifié comme étant leur plan support commun, ce qui implique que la reconstruction affine est possible ;
- ▷ il existent de fausses quadriques absolues non coplanaires avec la vraie quadrique absolue. Dans ce cas, il est possible que l'hyperplan à l'infini ne peut pas être identifié de manière unique et

que la reconstruction affine n'est pas réalisable. Nous appelons tous les hyperplans support de quadriques absolues potentielles des **hyperplans à l'infini potentiels**. Nous distinguons aussi les **faux hyperplans à l'infini** du vrai hyperplan à l'infini.

Nous donnons des exemples pour ces cas plus bas.

7.7 Séquences de mouvements critiques

Nous définissons une **séquence d'images** S comme étant un ensemble (qui peut être infini) de matrices de projection P_i :

$$S = \{\mathsf{P}_i | i = 1, ..., m\} \ .$$

Considérons la décomposition des projections en paramètres intrinsèques et extrinsèques :

$$\mathsf{P}_i \sim \mathsf{K}_i \mathsf{R}_i (\quad \mathtt{I} \quad | - \mathbf{t}_i) \ .$$

Nous appelons l'ensemble des orientations extrinsèques, c'est-à-dire l'ensemble $\{(\mathsf{R}_i, \mathbf{t}_i) | i = 1, ..., m\}$, la **séquence de mouvements** sous-jacente à S. L'ensemble des matrices des paramètres intrinsèques K_i est appelé le **calibrage intrinsèque** de la séquence d'images.

Une séquence d'images pour laquelle de fausses quadriques absolues existent est appelée **séquence d'images critique** pour la reconstruction euclidienne. Un repère dans lequel une quadrique absolue potentielle est à la position canonique de la quadrique absolue est appelé **repère euclidien potentiel**. Une reconstruction dans un repère euclidien potentiel est une **reconstruction euclidienne potentielle**.

Dans la suite, nous nous intéressons principalement aux séquences des mouvements sous-jacentes à des séquences d'images critiques. Si une séquence d'images est critique, nous appelons aussi la séquence de mouvements sous-jacente **critique**. Pour une quadrique absolue potentielle Ω, nous disons aussi que la séquence d'images et la séquence de mouvements sont **critiques par rapport à** Ω. Puisque nous considérons le problème de la reconstruction euclidienne, le repère du monde est définie à une transformation euclidienne près. Tout ce qui importe sont alors les mouvements *relatifs* entre les vues. Nous pouvons alors, sans perte de généralité, choisir le repère euclidien dans lequel la séquence de mouvements est définie, de manière arbitraire.

Dans les sections suivantes, nous dérivons les séquences de mouvements critiques pour différents modèles de caméra. Des définitions plus précises des séquences critiques et des quadriques absolues potentielles sont données pour chaque cas traité.

Les modèles de caméra considérés sont les suivants.

- ▷ Caméra planaire perspective avec des paramètres intrinsèques constants. Ceci est le scénario le plus étudié pour la reconstruction euclidienne non calibrée. Le problème est généralement évoqué en termes de l'auto-calibrage de la caméra. Ces deux tâches sont *presque* équivalentes ; la reconstruction euclidienne inclut le (auto-)calibrage mais le réciproque n'est pas toujours vrai. Par exemple, une caméra effectuant des rotations pures peut être calibrée, mais la reconstruction n'est pas possible. Parfois, s'il n'y a pas de possibilité de confusion, nous allons utiliser l'expression auto-calibrage à la place de reconstruction euclidienne non calibrée, afin de simplifier les expressions. Les séquences de mouvements critiques pour ce scénario sont dérivés dans la section 7.9.

- ▷ Caméra linéaire perspective avec des paramètres intrinsèques constants. C'est le cas le plus simple et nous commençons avec lui afin d'introduire notre façon de procéder.

7.8 Séquences critiques de la caméra linéaire

Nous considérons le problème de la reconstruction euclidienne à partir d'une séquence d'images prises par une caméra linéaire avec des paramètres intrinsèques constants. En 6.3, nous proposons une méthode pour le sous-problème de l'auto-calibrage. Ici, nous dérivons les séquences de mouvements critiques pour le problème général de la reconstruction euclidienne.

Rappelons d'abord quelques notions importantes. La quadrique absolue du monde de la caméra linéaire (le plan projectif $\mathcal{P}^2$) est la paire des points cycliques, qui, dans un repère euclidien, ont les coordonnées :

$$\mathbf{I} \sim \begin{pmatrix} 1 \\ +I \\ 0 \end{pmatrix} \qquad \mathbf{J} \sim \begin{pmatrix} 1 \\ -I \\ 0 \end{pmatrix} \quad ,$$

où I est donné par $I = \sqrt{-1}$. Soit Ω_∞ la conique formée de ces deux points. La reconstruction euclidienne revient à déterminer $\mathbf{I}$ et $\mathbf{J}$ ou, puisqu'il s'agit de points imaginaires *conjugués*, à déterminer seulement un des deux points.

Les contraintes pour trouver les points cycliques sont les suivantes.

- ▷ Les points cycliques sont des points imaginaires, tout comme leurs projections (nous ne considérons que des projections avec des coefficients réels).
- ▷ Les paramètres intrinsèques sont constants, donc les projections de chaque point cyclique sont identiques dans toutes les vues (voir la preuve en D.1.1).

Nous soulignons que non seulement les projections de la *conique* Ω_∞ des points $\mathbf{I}$ et $\mathbf{J}$ sont identiques dans toutes les vues, mais aussi les projections des points *individuels*.

Si notre séquence de mouvements consistait de *tous* les mouvements rigides possibles (il y en a ∞^3), les points cycliques seraient les seuls points satisfaisant ces contraintes. Sinon, il y a la possibilité de l'existence d'une ou de plusieurs autres paires de points de $\mathcal{P}^2$ avec les mêmes propriétés. Dans ce cas, les points cycliques ne peuvent pas être identifiés sans ambiguïté et le problème de la reconstruction euclidienne est dégénéré.

Dans la suite, nous définissons quelques notations qui aident à caractériser cette situation.

DÉFINITION 7.1 *Soit S une séquence d'images.*

Un point imaginaire $\mathbf{Q}^I$ de $\mathcal{P}^2$ est un **point cyclique potentiel** *si ses projections dans toutes les vues de S sont des points imaginaires identiques. Son conjugué $\mathbf{Q}^J$ est alors aussi un point cyclique potentiel. Un repère dans lequel un point cyclique potentiel et son conjugué ont les coordonnées $(1, \pm I, 0)^\mathsf{T}$, est un* **repère euclidien potentiel**. *Une reconstruction dans un repère euclidien potentiel est une* **reconstruction euclidienne potentielle**.

La séquence S est **critique** *s'il existe de* **faux points cycliques**, *c'est-à-dire des points cycliques potentiels* différents *des vrais points cycliques. Si $\mathbf{Q}^I$ est un faux point cyclique, on dira que la séquence d'images est* **critique par rapport à $\mathbf{Q}^I$**.

Il est important de noter que la question si un point est un point cyclique potentiel, ne dépend que de la séquence de mouvements sous-jacente à la séquence d'images et non du calibrage intrinsèque. Soit $\mathbf{Q}^I$ un point cyclique potentiel, i.e. :

$$\mathsf{P}_i \mathbf{Q}^I \sim \mathsf{P}_j \mathbf{Q}^I \ , \ 1 \leq i, j \leq m \ .$$

Puisque P_i et P_j ont les mêmes paramètres intrinsèques K, nous pouvons écrire :

$$\mathsf{K}\mathsf{R}_i(\ \mathtt{I}\ \ | -\mathbf{t}_i)\mathbf{Q}^I \sim \mathsf{K}\mathsf{R}_j(\ \mathtt{I}\ \ | -\mathbf{t}_j)\mathbf{Q}^I \ .$$

Il en découle que :

$$\mathsf{R}_i(\ \mathtt{I}\ \ | -\mathbf{t}_i)\mathbf{Q}^I \sim \mathsf{R}_j(\ \mathtt{I}\ \ | -\mathbf{t}_j)\mathbf{Q}^I \ .$$

Donc, il suffit de ne considérer que les mouvements de la caméra pour déterminer les points cycliques potentiels. Nous transférons les définitions 7.1 des séquences d'images directement aux séquences de mouvements sous-jacentes.

Dans la suite nous dérivons les séquences de mouvements critiques. Nous le faisons d'une manière constructive : étant donné un point imaginaire quelconque $\mathbf{Q}^I$ et une position de caméra initiale, nous déterminons toutes les autres positions de caméra qui forment ensemble une séquence critique. Il s'agit donc de trouver les positions de caméra pour lesquelles la projection de $\mathbf{Q}^I$ est la même que dans la vue initiale. Nous dérivons les séquences critiques de deux manières, une plus géométrique, utilisant le birapport de droites et de points, une deuxième algébrique.

Notons d'abord qu'une rotation de $180°$ autour du centre de projection de la caméra linéaire n'affecte pas la projection : effectuer une telle rotation revient à pré-multiplier la matrice de rotation actuelle par la matrice $\mathrm{diag}(-1,-1)$, donc un multiple de l'identité, ce qui ne change pas la matrice de projection au sens des coordonnées homogènes.

7.8.1 Dérivation géométrique des séquences critiques

7.8.1.1 Positions du centre de projection

Soit $\mathbf{Q}^I$ un faux point cyclique et $\mathbf{Q}^J$ son conjugué. Soient $\mathbf{q}_i^I$ et $\mathbf{q}_i^J$ leurs projections dans la i^{e} vue et $\mathbf{i}_i$ et $\mathbf{j}_i$ les projections des vrais points cycliques. Par définition, nous avons :

$$\mathbf{q}_i^I \sim \mathbf{q}_j^I \ , \ \mathbf{q}_i^J \sim \mathbf{q}_j^J \ , \ \mathbf{i}_i \sim \mathbf{i}_j \ , \ \mathbf{j}_i \sim \mathbf{j}_j \ , \ 1 \le i,j \le m \ .$$

Il en découle que le birapport de ces 4 points est le même pour toutes les images :

$$\{\mathbf{i}_i,\mathbf{j}_i;\mathbf{q}_i^I,\mathbf{q}_i^J\} = \{\mathbf{i}_j,\mathbf{j}_j;\mathbf{q}_j^I,\mathbf{q}_j^J\} \ .$$

Par conséquent, le birapport des rayons de projection associés est également constant :

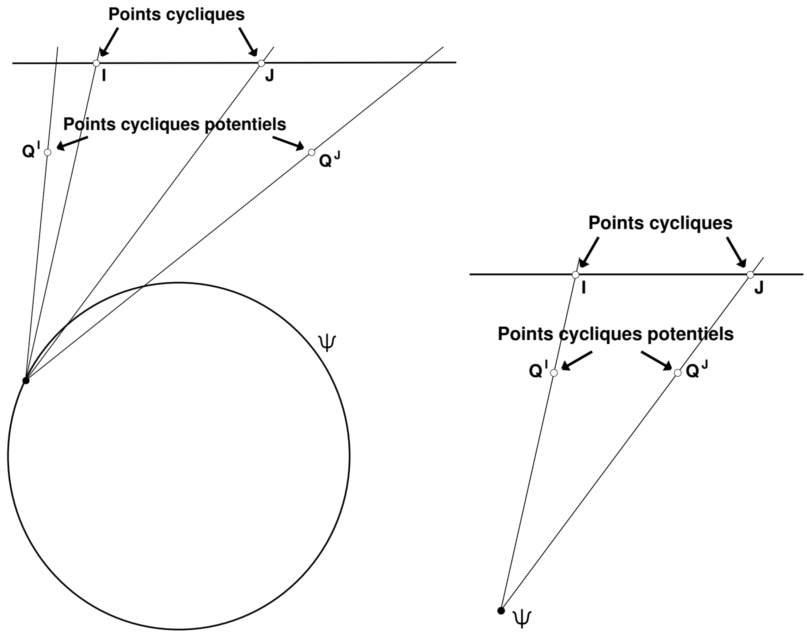
$$\{\langle\mathbf{C}_i,\mathbf{I}\rangle,\langle\mathbf{C}_i,\mathbf{J}\rangle;\langle\mathbf{C}_i,\mathbf{Q}^I\rangle,\langle\mathbf{C}_i,\mathbf{Q}^J\rangle\} = \{\langle\mathbf{C}_j,\mathbf{I}\rangle,\langle\mathbf{C}_j,\mathbf{J}\rangle;\langle\mathbf{C}_j,\mathbf{Q}^I\rangle,\langle\mathbf{C}_j,\mathbf{Q}^J\rangle\} \ ,$$

où $\mathbf{C}_i$ est le centre de projection de la i^{e} vue :

$$\mathbf{C}_i = \begin{pmatrix}\mathbf{t}_i\\1\end{pmatrix} \ .$$

Si nous excluons le cas que $\mathbf{Q}^I$ et $\mathbf{Q}^J$ sont des points à l'infini, nous pouvons conclure que tous les centres de projection doivent se trouver sur une conique[1] Ψ. Cette conique contient également $\mathbf{Q}^I,\mathbf{Q}^J$ et les points cycliques $\mathbf{I}$ et $\mathbf{J}$. Donc, Ψ est un cercle (voir la figure 7.7 (a)). Ce cercle peut dégénérer de manière à avoir un rayon nul, ce qui veut dire qu'il y a un seul centre de projection possible (cf. la figure 7.7 (b)).

1. Quatre points et un birapport définissent une conique de la manière suivante : les points, dont les droites joignant les quatre premiers points ont le birapport donné, forment une conique. Les quatre premiers points se trouvent également sur cette conique.



(a) Cas général. (b) Cercle Ψ de rayon nul.

FIG. 7.7: *Les centres de projection dans une séquence qui est critique par rapport à un point* $\mathbf{Q}^I$ *non à l'infini (et par rapport au point conjugué* $\mathbf{Q}^J$*) se trouvent sur un cercle* Ψ*. Ce cercle contient également* $\mathbf{Q}^I$ *et* $\mathbf{Q}^J$ *(dans le dessin c'est montré autrement, puisqu'il s'agit de points imaginaires).*

Il nous faut aussi regarder le cas que nous avons exclu ci-dessus, i.e. des points $\mathbf{Q}^I$ et $\mathbf{Q}^J$ à l'infini. Puisque les quatre points $\mathbf{Q}^I, \mathbf{Q}^J, \mathbf{I}$ et $\mathbf{J}$ sont maintenant collinéaires, les rayons de projection ont le même birapport, indépendamment de la position du centre de projection. Il n'y a donc aucune contrainte sur la position des centres de projection.

Nous venons de dériver des conditions nécessaires pour des séquences de mouvements critiques par rapport à $\mathbf{Q}^I$ et $\mathbf{Q}^J$, qui concernent uniquement la position des centres de projection. Dans la suite, nous considérons les rotations de la caméra.

7.8.1.2 Rotations de la caméra

Considérons d'abord le cas où $\mathbf{Q}^I$ est à l'infini. Il est facile de voir que les projections d'un point à l'infini sont identiques dans toutes les vues exactement si la caméra n'effectue que des translations pures, suivies éventuellement d'une rotation de 180°. Ce cas est illustré dans la figure 7.8. Dans ce cas, les projections de *tous* les points à l'infini sont identiques dans les vues. L'existence d'un faux point cyclique à l'infini implique alors que tous les points imaginaires à l'infini sont des points cycliques potentiels.

Nous dérivons maintenant des contraintes sur les rotations relatives entre les vues si $\mathbf{Q}^I$ n'est pas à l'infini. Considérons d'abord le cas spécial où le cercle Ψ est de rayon nul, i.e. l'unique centre de projection de la séquence est le point d'intersection des droites $\langle \mathbf{I}, \mathbf{Q}^I \rangle$ et $\langle \mathbf{J}, \mathbf{Q}^J \rangle$ (cf. la figure 7.7 (b)). Les projections de $\mathbf{Q}^I$ et $\mathbf{Q}^J$ coïncident alors avec $\mathbf{i}$ respectivement $\mathbf{j}$. Puisque $\mathbf{i}$ et $\mathbf{j}$ ne sont pas affectés par des rotations de la caméra, c'est aussi le cas pour les projections de $\mathbf{Q}^I$ et $\mathbf{Q}^J$. Nous

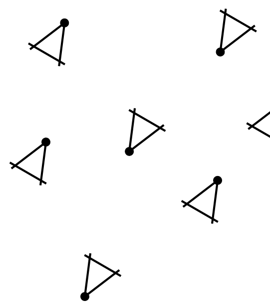


FIG. 7.8: *Séquence de mouvements critique par rapport aux points imaginaires à l'infini : la caméra effectue des translations pures, suivies éventuellement d'une rotation de* $180°$.

concluons que la séquence est critique même si la caméra effectue des rotations quelconques. Ceci n'est pas étonnant : aucune reconstruction, ni projective ni euclidienne, n'est possible avec des images prises du même point de vue, et des rotations arbitraires n'y changent rien.

Considérons maintenant le cas général où le rayon de Ψ est non nul. Soit $\mathbf{Q}$ un point quelconque et $\mathbf{C}$ un centre de projection, tous deux sur Ψ. D'après le théorème de Chasles, le birapport des droites $\langle \mathbf{C}, \mathbf{Q} \rangle$, $\langle \mathbf{C}, \mathbf{Q}^I \rangle$, $\langle \mathbf{C}, \mathbf{I} \rangle$ et $\langle \mathbf{C}, \mathbf{J} \rangle$ est le même pour tous les centres de projection $\mathbf{C}$. Par conséquent, le birapport des projections de $\mathbf{Q}, \mathbf{Q}^I, \mathbf{I}$ et $\mathbf{J}$ est le même dans toutes les vues. Puisque les projections de $\mathbf{Q}^I, \mathbf{I}$ et $\mathbf{J}$ sont identiques dans toutes les vues, il en est de même pour les projections de $\mathbf{Q}$[2]. Comme nous avons pris pour $\mathbf{Q}$ un point quelconque sur Ψ, nous avons alors dérivé une condition nécessaire sur les rotations relatives : la séquence est critique si et seulement si les projections de tous les points de Ψ sont identiques dans toutes les vues. Nous pouvons constater que *tous* les points imaginaires sur Ψ sont des points cycliques potentiels. Dans la figure 7.9 nous montrons un exemple d'une telle séquence critique.

7.8.2 Dérivation algébrique des séquences critiques

Nous redérivons les séquences de mouvements critiques de manière algébrique, afin d'illustrer notre manière de procéder pour les autres modèles de caméra. Les résultats qui sont rapportés dans la suite sont bien sûr identiques aux résultats du paragraphe précédent.

Soit $\mathbf{Q}^I$ un faux point cyclique. Considérons une position de caméra initiale représentée par R_1 et $\mathbf{t}_1$. Nous déterminons les séquences de mouvements critiques par rapport à $\mathbf{Q}^I$ en déterminant toutes les autres positions de caméra pour lesquelles $\mathbf{Q}^I$ a la même projection que dans la vue initiale, i.e. nous cherchons tous les R_2 et $\mathbf{t}_2$ avec :

$$\mathsf{R}_2(\ \mathtt{I}\ | -\mathbf{t}_2)\mathbf{Q}^I \ \sim\ \mathsf{R}_1(\ \mathtt{I}\ | -\mathbf{t}_1)\mathbf{Q}^I \ .$$

Soit $\mathsf{R} = \mathsf{R}_2^\mathsf{T}\mathsf{R}_1$ la rotation relative entre les vues 1 et 2. La condition de la projection identique de $\mathbf{Q}^I$ dans les deux vues s'exprime alors par :

$$(\ \mathtt{I}\ | -\mathbf{t}_2)\mathbf{Q}^I \ \sim\ \mathsf{R}(\ \mathtt{I}\ | -\mathbf{t}_1)\mathbf{Q}^I \ . \tag{7.1}$$

Au lieu de déterminer directement les rotations R_2, nous considérons alors les rotations relatives R entre les vues. Nous séparons les dérivations en deux parties, selon que $\mathbf{Q}^I$ se trouve à l'infini ou non.

2. 3 points et un birapport fixent la position d'un quatrième point.

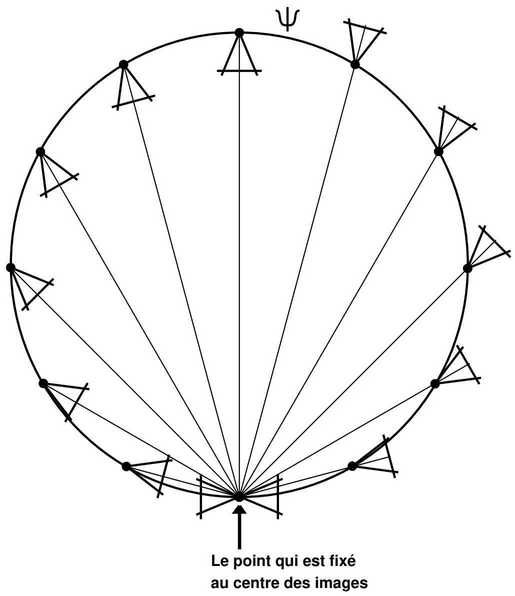


FIG. 7.9: *Nous savons déjà que les centres de projection dans une séquence qui est critique par rapport à un point non à l'infini se trouvent sur un cercle* Ψ. *En plus, les rotations des vues doivent être cohérentes : tous les points du cercle* Ψ *ont la même projection dans toutes les vues. Dans cette figure, nous illustrons ceci pour le cas du point qui se projette sur le centre de l'image. Sur la gauche, nous affichons des vues regardant vers l'intérieur du cercle et sur la droite des vues regardant vers l'extérieur ; ceci pour illustrer que des rotations de* $180°$ *n'affectent pas la transformation de projection d'une caméra linéaire.*

7.8.2.1 Points cycliques potentiels à l'infini

Soit $\mathbf{Q}^I$ donné par :

$$\mathbf{Q}^I \sim \begin{pmatrix} \overline{\mathbf{Q}}^I \\ 0 \end{pmatrix} \quad .$$

L'équation (7.1) s'écrit alors comme :

$$\overline{\mathbf{Q}}^I \quad \sim \quad \mathsf{R}\overline{\mathbf{Q}}^I \quad . \tag{7.2}$$

On constate que les positions $\mathbf{t}_i$ des centres de projection n'ont aucune importance, comme c'est le cas pour les projections de tous les points à l'infini. L'équation (7.2) signifie que $\overline{\mathbf{Q}}^I$ est un vecteur propre de R. Toute matrice de rotation 2×2 a les deux points $(1, \pm I)^{\mathsf{T}}$ comme vecteurs propres. Le cas que $\overline{\mathbf{Q}}^I$ est l'un de ces deux points signifie que $\mathbf{Q}^I$ est un vrai point cyclique. Nous ne nous intéressons qu'à des *faux* points cycliques : les seules rotations avec d'autres vecteurs propres que ceux mentionnés sont les rotations d'angle $0°$ ou $180°$. Par conséquent, pour qu'il existe un faux point cyclique il faut que les rotations relatives entre des paires de vues soient des rotations de $0°$ ou $180°$. Les mouvements de caméra dans la séquence critique sont donc exactement des translations pures, suivies éventuellement de rotations de $180°$.

De telles rotations sont représentées par la matrice identité (à un facteur ± 1 près) ce qui veut dire que *tous* les points à l'infini sont projetés sur des points identiques dans toutes les vues. Par conséquent, s'il existe un seul faux point cyclique, *tous* les points imaginaires sur la droite à l'infini sont des points cycliques potentiels.

7.8.2.2 Points cycliques potentiels finis

Dans la section 7.7, nous avons déjà mentionné que le repère euclidien du monde peut être choisi de manière arbitraire. Donc, sans perte de généralité, nous pouvons supposer que $\mathbf{Q}^I$ est donné par :

$$\mathbf{Q}^I \sim \begin{pmatrix} aI \\ 0 \\ 1 \end{pmatrix} \quad ,$$

pour un scalaire réel et non nul a. Une des contraintes sur les points cycliques potentiels est que leurs projections sont des points imaginaires. Nous devons alors exclure le cas d'un centre de projection sur la droite engendrée par $\mathbf{Q}^I$ et son point conjugué. Dans notre cas, il s'agit de la droite $Y = 0$, et nous pouvons supposer que :

$$t_{i2} \neq 0 \ , \ \ i = 1, ..., m \ . \tag{7.3}$$

L'équation (7.1) s'écrit comme :

$$\begin{pmatrix} aI - t_{21} \\ -t_{22} \end{pmatrix} \ \sim \ \mathsf{R} \begin{pmatrix} aI - t_{11} \\ -t_{12} \end{pmatrix} \quad . \tag{7.4}$$

Si nous représentons la rotation relative R par :

$$\mathsf{R} = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \quad ,$$

l'équation (7.4) devient :

$$\begin{pmatrix} aI - t_{21} \\ -t_{22} \end{pmatrix} \ \sim \ \begin{pmatrix} aI\cos\beta - t_{11}\cos\beta + t_{12}\sin\beta \\ aI\sin\beta - t_{11}\sin\beta - t_{12}\cos\beta \end{pmatrix} \quad .$$

La multiplication croisée des coefficients de ces vecteurs donne :

$$\begin{aligned} \cos\beta(t_{11}t_{22} - t_{12}t_{21}) + \sin\beta(a^2 - t_{12}t_{22} - t_{11}t_{21}) + & \\ aI(\cos\beta(t_{12} - t_{22}) + \sin\beta(t_{11} + t_{21})) &= 0 \ . \end{aligned}$$

Nous obtenons donc deux équations, dues à l'annulation des parties réelle et imaginaire de cette équation :

$$\cos\beta(t_{11}t_{22} - t_{12}t_{21}) + \sin\beta(a^2 - t_{12}t_{22} - t_{11}t_{21}) = 0 \tag{7.5}$$

$$\cos\beta(t_{12} - t_{22}) + \sin\beta(t_{11} + t_{21}) = 0 \ . \tag{7.6}$$

Positions du centre de projection. Calculer la tangente de β avec les deux équations et les mettre à égalité mène à une équation en termes des seules positions des centres de projection $\mathbf{t}_1$ et $\mathbf{t}_2$:

$$t_{12}t_{21}^2 + t_{12}t_{22}^2 - (t_{12}^2 + t_{11}^2)t_{22} + a^2 = 0 \ .$$

C'est une équation quadratique en les coefficients de $\mathbf{t}_2$, représentée par une matrice symétrique Ψ :

$$(t_{21}, t_{22}, 1) \underbrace{\begin{pmatrix} t_{12} & 0 & 0 \\ 0 & t_{12} & -\frac{1}{2}(a^2 + t_{11}^2 + t_{12}^2) \\ 0 & -\frac{1}{2}(a^2 + t_{11}^2 + t_{12}^2) & a^2 t_{12} \end{pmatrix}}_{\Psi} \begin{pmatrix} t_{21} \\ t_{22} \\ 1 \end{pmatrix} = 0 \ . \qquad (7.7)$$

La matrice Ψ représente un cercle, centré à :

$$\mathbf{C}_\Psi = \begin{pmatrix} 0 \\ \frac{a^2+t_{11}^2+t_{12}^2}{2t_{12}} \\ 1 \end{pmatrix} \ .$$

Nous pouvons conclure que les caméras dans une séquence critique par rapport à $\mathbf{Q}^I$ doivent se déplacer sur un cercle, qui est déterminé par $\mathbf{Q}^I$ et la position $\mathbf{t}_1$ de la vue initiale.

Nous examinons maintenant s'il existent des cas spéciaux où le cercle Ψ dégénère. Le déterminant de Ψ est calculé comme :

$$\det \Psi = -\frac{1}{4} t_{12}((a - t_{12})^2 + t_{11}^2)((a + t_{12})^2 + t_{11}^2) \ .$$

Comme nous avons exclu le cas $t_{12} = 0$ (cf. l'équation (7.3)), il faut donc que les deux conditions suivantes soient remplies pour que Ψ dégénère :

$$t_{11} = 0 \qquad (7.8)$$
$$t_{12} = \pm a \ . \qquad (7.9)$$

Sous ces conditions, Ψ devient :

$$\Psi = \begin{pmatrix} \pm a & 0 & 0 \\ 0 & \pm a & -a^2 \\ 0 & -a^2 & \pm a^3 \end{pmatrix} \ ,$$

ce qui représente une conique de deux droites imaginaires, dont le point d'intersection est le seul point réel[3]. Autrement dit, Ψ est un cercle de rayon nul. Le point réel de Ψ est donné par $(0, \pm a, 1)^\mathsf{T}$, ce qui n'est rien d'autre que la position de caméra initiale (cf. les équations (7.8) et (7.9)). Nous en concluons que, si la vue initiale est prise de cette position, les autres vues doivent également être prises de cette position pour que la séquence soit critique par rapport à $\mathbf{Q}^I$.

Rotations de la caméra. Après avoir traité les positions des centres de projection, il nous reste à examiner les possibilités pour les rotations.

Considérons d'abord le cas spécial où les centres de projection coïncident tous au point $(0, \pm a, 1)^\mathsf{T}$. Évidemment, dans ce cas aucune reconstruction n'est possible puisqu'il n'y a pas d'information de profondeur. On peut vérifier que les équations (7.5) et (7.6) deviennent des tautologies, comme souhaité. Donc, même si la caméra effectue des rotations quelconques autour du centre de projection, la séquence d'images est critique (on verra plus bas que l'auto-calibrage est, en revanche, possible).

Regardons maintenant le cas où la caméra se déplace sur un cercle de rayon non nul. Les équations (7.5) et (7.6) ne dégénèrent pas ce qui veut dire qu'il y a une seule solution pour la tangente de l'angle

3. Les deux droites de la conique sont les droites isotropiques de leur point d'intersection [20], c'est-à-dire les droites qui le relient aux deux points cycliques.

β de la rotation relative entre deux vues. Il y a donc deux rotations relatives possibles, qui diffèrent par une rotation de $180°$. La dérivation d'une formule qui calcule β en fonction de $\mathbf{t}_1, \mathbf{t}_2$ et $\mathbf{Q}^I$ est facile.

Dans l'annexe D.1.2, nous prouvons un fait qui permet de décrire les orientations relatives dans une séquence critique de manière plus illustrative. Notamment, non seulement les projections de $\mathbf{Q}^I$ sont identiques dans toutes les vues, mais aussi les projections de *tous* les points sur le cercle Ψ, soient ils imaginaires ou réels !

7.8.3 Auto-calibrage et reconstruction affine

Avec une séquence critique on ne peut alors pas obtenir une reconstruction euclidienne. Il est tout de même intéressant d'étudier la question si au moins une reconstruction affine ou l'auto-calibrage sont possibles. Dans ce paragraphe, nous donnons des réponses à cette question.

Considérons d'abord le cas d'un seul centre de projection et de rotations quelconques. Évidemment, aucune reconstruction n'est possible. Tout de même, l'auto-calibrage est généralement faisable : si la caméra effectue au moins une rotation de degré différent de $0°$ et $180°$, tous les points cycliques potentiels se trouvent sur les deux droites isotropiques du centre de projection $\mathbf{C}$, i.e. sur les droites reliant $\mathbf{C}$ avec $\mathbf{I}$ et $\mathbf{J}$. Donc, les projections de tous les points cycliques potentiels coïncident et il y a une solution unique pour ω_∞ (la conique des projections des points cycliques) et les paramètres intrinsèques.

Considérons le cas d'un faux point cyclique à l'infini. Nous avons déjà mentionné que l'existence d'un seul tel point implique que tous les points imaginaires de la droite à l'infini sont des points cycliques potentiels. En outre, on peut vérifier qu'il n'y a aucun point cyclique potentiel hors la droite à l'infini. Donc, la droite à l'infini peut être identifiée comme étant la droite supportant tous les points cycliques potentiels ; la reconstruction affine est alors possible. L'auto-calibrage, par contre, n'est pas possible : les projections des points cycliques potentiels couvrent toute la partie imaginaire de la droite image. Il n'est donc pas possible de contraindre la position de ω_∞.

Le dernier cas à considérer concerne un cercle Ψ de rayon non nul. Nous avons déjà pu constater que tous les points imaginaires sur Ψ sont des points cycliques potentiels. Comme pour le cas précédent, les projections des points cycliques potentiels couvrent toute la partie imaginaire de la droite image, ce qui fait que l'auto-calibrage n'est pas possible. En ce qui concerne la reconstruction affine, nous constatons qu'il existe une infinité de droites à l'infini potentielles, notamment toutes les droites reliant des points imaginaires sur Ψ et leurs conjugués. Ces droites sont exactement les droites qui ne coupent Ψ en aucun point réel. Il est claire que la reconstruction affine n'est pas possible.

7.8.4 Résumé

Le tableau 7.1 résume les résultats de cette section.

7.9 Séquences critiques de la caméra planaire avec des paramètres intrinsèques constants

Nous considérons le problème de la reconstruction euclidienne à partir d'une séquence d'images prises par une caméra planaire avec des paramètres intrinsèques constants (voir la section 6.2.1 pour une revue de méthodes d'auto-calibrage).

Points cycliques potentiels	Centres de projection	Rotations	Auto-calibrage	Reconstr. affine
À l'infini	Quelconques	Les rotations relatives sont de 0° ou 180°	Non	Oui
Non à l'infini	Sur un cercle contenant le point cyclique potentiel	Tous les points du cercle sont fixés dans les vues	Non	Non
	Une seule position (cercle de rayon nul)	Quelconques	Oui	Non

TAB. 7.1: *Les séquences critiques de la caméra linéaire. Pour chaque type de séquence critique nous indiquons si l'auto-calibrage ou la reconstruction affine sont possibles.*

La quadrique absolue du monde de la caméra planaire (l'espace projectif $\mathcal{P}^3$) est la conique absolue Ω_∞. La reconstruction euclidienne revient à déterminer Ω_∞. Les contraintes pour trouver la conique absolue sont les suivantes.

- ▷ La conique absolue est une conique imaginaire et propre, tout comme ses projections.
- ▷ Les paramètres intrinsèques sont constants, donc les projections de la conique absolue sont identiques dans toutes les vues.

Contrairement au cas de la caméra linéaire, les projections de la conique absolue ne sont pas fixes point par point ; les coniques image sont seulement globalement identiques.

Si notre séquence de mouvements consistait de *tous* les mouvements rigides possibles (il y en a ∞^6), la conique absolue serait la seule conique 3D ayant ces propriétés. Sinon, une ou plusieurs autres coniques avec les mêmes propriétés peuvent exister. Dans ce cas, la conique absolue ne peut pas être identifiée sans ambiguïté et le problème de la reconstruction euclidienne est dégénéré.

Dans la suite, nous définissons quelques notations qui aident à caractériser cette situation.

DÉFINITION 7.2 *Soit S une séquence d'images.*

Une conique 3D imaginaire et propre Ω est une **conique absolue potentielle** *si ses projections dans toutes les vues de S sont des coniques imaginaires et propres identiques. Un repère dans lequel une conique absolue potentielle a des coordonnées canoniques (cf. la section 2.7.3) est un* **repère euclidien potentiel**. *Une reconstruction dans un repère euclidien potentiel est une* **reconstruction euclidienne potentielle**.

La séquence S est **critique** *s'il existe de* **fausses coniques absolues**, *c'est-à-dire des coniques absolues potentielles* différentes *de la vraie conique absolue. Si Ω est une fausse conique absolue, on dira que la séquence d'images est* **critique par rapport à** Ω.

Il est important de noter que la question si une conique est une conique absolue potentielle, ne dépend que de la séquence de mouvements sous-jacente à la séquence d'images et non du calibrage intrinsèque (voir la preuve dans la section D.2.1). Donc, il suffit de ne considérer que les mouvements de la caméra pour déterminer les coniques absolues potentielles. Notamment, nous pouvons supposer que les matrices de projection prennent la forme (cf. l'annexe D.2.1) :

$$\mathsf{P}_i \sim \mathsf{R}_i(\quad \mathtt{I} \quad | - \mathbf{t}_i) \ .$$

Nous transférons les définitions 7.2 des séquences d'images directement aux séquences de mouvements sous-jacentes.

Dans la suite nous dérivons les séquences de mouvements critiques. Nous le faisons d'une manière constructive : étant donné une conique 3D imaginaire et propre quelconque Ω et une position de caméra initiale, nous déterminons toutes les autres positions de caméra qui forment ensemble une séquence critique. Il s'agit donc de trouver les positions de caméra pour lesquelles la projection de Ω est la même que dans la vue initiale.

7.9.1 Dérivation des séquences critiques

Soit Ω une fausse conique absolue et une position de caméra initiale représentée par R_1 et $\mathbf{t}_1$. Nous déterminons les séquences de mouvements critiques par rapport à Ω en déterminant toutes les autres positions de caméra pour lesquelles Ω a la même projection que dans la vue initiale. Puisque la projection de Ω doit être une conique propre, nous devons exclure que les caméras se trouvent sur le plan support de Ω.

Nous séparons les dérivations en deux parties, selon que Ω se trouve sur le plan à l'infini ou non.

7.9.1.1 Coniques absolues potentielles sur Π_∞

Nous considérons le cas où Ω se trouve sur le plan à l'infini. Soit C sa représentation matricielle dans Π_∞. La projection de Ω dans la i^{e} vue est alors donnée par :

$$\omega_i \sim \mathsf{R}_i \mathsf{C} \mathsf{R}_i{}^{\mathsf{T}} \ .$$

Il est important de noter que la projection de Ω ne dépend que de la rotation de la caméra, non de la position du centre de projection. Ceci est attendu puisque les translations n'affectent pas des entités sur le plan à l'infini.

Dans la suite, nous dérivons des conditions sur les rotations de la caméra qui font que la séquence entière est critique. Considérons les vues 1 et 2. Les projections de Ω sont identiques si :

$$\mathsf{R}_1 \mathsf{C} \mathsf{R}_1^{\mathsf{T}} \sim \mathsf{R}_2 \mathsf{C} \mathsf{R}_2^{\mathsf{T}} \ .$$

Puisque les déterminants de matrices de rotation valent 1, cette équation n'est pas valable seulement à un facteur près, mais elle est exacte :

$$\mathsf{R}_1 \mathsf{C} \mathsf{R}_1^{\mathsf{T}} = \mathsf{R}_2 \mathsf{C} \mathsf{R}_2^{\mathsf{T}} \ .$$

Nous pouvons exprimer cette équation en termes de la rotation relative R entre les vues :

$$\mathsf{R}\mathsf{C} = \mathsf{C}\mathsf{R} \ , \tag{7.10}$$

où $\mathsf{R} = \mathsf{R}_1^{\mathsf{T}} \mathsf{R}_2$.

Soit $\mathbf{Q}$ un vecteur propre de C associé à la valeur propre λ, c'est-à-dire que $\mathsf{C}\mathbf{Q} = \lambda \mathbf{Q}$. De l'équation (7.10), il découle que :

$$\begin{aligned} \mathsf{C}\mathsf{R}\mathbf{Q} &= \mathsf{R}\mathsf{C}\mathbf{Q} \\ &= \lambda \mathsf{R}\mathbf{Q} \ , \end{aligned}$$

c'est-à-dire que $\mathsf{R}\mathbf{Q}$ est également un vecteur propre de C associé à la valeur propre λ. Ceci est valide pour tous les vecteurs propres de C. La rotation relative R doit alors conserver tous les espaces propres de C pour que les projections de Ω soient identiques dans les deux vues. Nous avons considéré seulement la rotation relative entre deux vues. L'extension à plusieurs vues est directe : les projections de Ω sont identiques dans toutes les vues si et seulement si toutes les rotations relatives entre des paires de vues conservent les espaces propres de C.

Selon le type de Ω, plusieurs rotations relatives sont possibles. Nous considérons les différentes possibilités par rapport à la composition du spectre des valeurs propres de C.

CAS 7.1 (UNE VALEUR PROPRE TRIPLE) *La seule conique sur Π_∞ avec une valeur propre triple est la conique absolue, qui n'a pas d'intérêt ici.*

CAS 7.2 (UNE VALEUR PROPRE DOUBLE ET UNE SIMPLE) *L'espace propre associé à la valeur propre double est un plan Π et celui de la simple valeur propre une droite $\mathbf{L}$ perpendiculaire[4] à Π. Une rotation qui conserve globalement Π et $\mathbf{L}$ peut être :*

- *l'identité ;*
- *une rotation autour de $\mathbf{L}$ de degré arbitraire ;*
- *une rotation de $180°$ autour d'une droite de Π incidente avec $\mathbf{L}$.*

Un exemple de séquence critique pour le deuxième cas est montré dans la figure 7.10.

CAS 7.3 (TROIS VALEURS PROPRES DIFFÉRENTES) *Les espaces propres sont trois droites mutuellement perpendiculaires. Les rotations relatives sont donc l'identité ou des rotations de $180°$ autour d'une de ces droites. Ces rotations relatives font partie de celles du cas 7.2.*

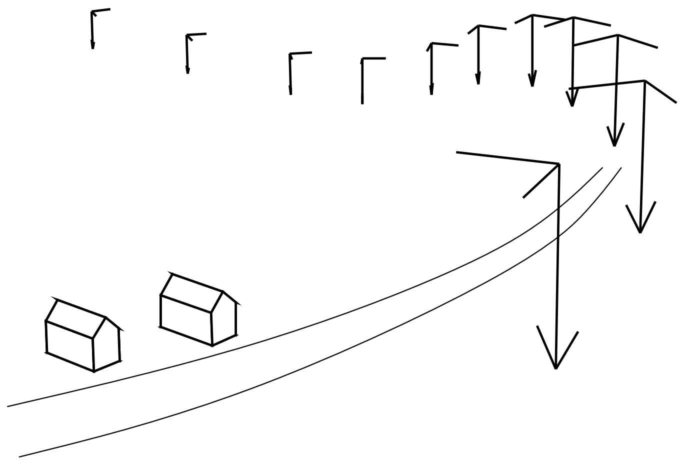


FIG. 7.10: *Séquence de mouvements critique par rapport à une conique sur le plan à l'infini : la caméra effectue des translations quelconques et des rotations autour du même axe (dans le repère local de la caméra).*

Nous résumons les dérivations :

LEMME 7.1 *Les séquences de mouvements critiques par rapport à une conique imaginaire propre sur le plan à l'infini sont exactement les séquences où toutes les rotations relatives entre des paires de vues sont :*

- *l'identité ;*

4. Des vecteurs propres de matrices symétriques associés à des valeurs propres différentes sont orthogonaux.

▷ *une rotation de degré arbitraire autour d'un axe* $\mathbf{L}$ *;*

▷ *ou une rotation de* $180°$ *autour d'un axe perpendiculaire à* $\mathbf{L}$.

Preuve : *Nous avons déjà montré une direction du lemme : si* Ω *est une fausse conique absolue potentielle sur* Π_∞*, les rotations relatives doivent être telles que décrites dans le lemme.*

Il reste à prouver l'existence de fausses coniques absolues si les rotations relatives sont telles qu'elles sont décrites dans le lemme. Soit $\mathbf{Q}^\mathsf{T} = (\overline{\mathbf{Q}}^\mathsf{T}, 0)$ *le point à l'infini de* $\mathbf{L}$*. Toutes les rotations relatives laissent* $\mathbf{Q}$ *invariant. Donc, les projections de* $\mathbf{Q}$ *sont identiques dans toutes les vues :*

$$\mathsf{R}_i\overline{\mathbf{Q}}^\mathsf{T} \sim \mathsf{R}_j\overline{\mathbf{Q}}^\mathsf{T} \quad . \tag{7.11}$$

Définissons la conique Ψ *comme étant la conique sur le plan à l'infini qui est représentée par la matrice* $\overline{\mathbf{Q}}\,\overline{\mathbf{Q}}^\mathsf{T}$*. La projection de* Ψ *dans la* i^e *vue est donnée par :*

$$\omega_i = \mathsf{R}_i\overline{\mathbf{Q}}\,\overline{\mathbf{Q}}^\mathsf{T}\mathsf{R}_i{}^\mathsf{T} \quad .$$

De l'équation (7.11) il découle que les projections de Ψ *sont identiques dans toutes les vues :*

$$\begin{aligned} \omega_i &= \mathsf{R}_i\overline{\mathbf{Q}}\,\overline{\mathbf{Q}}^\mathsf{T}\mathsf{R}_i{}^\mathsf{T} \\ &= \mathsf{R}_j\overline{\mathbf{Q}}\,\overline{\mathbf{Q}}^\mathsf{T}\mathsf{R}_j{}^\mathsf{T} \\ &= \omega_j \quad . \end{aligned}$$

La conique Ψ *n'est tout de même pas une conique absolue potentielle puisqu'elle est dégénérée (par construction, elle est de rang 1).*

Il est facile de prouver que les projections de toutes les coniques dans la famille linéaire engendrée par Ψ *et la conique absolue sont identiques dans toutes les vues. Soit* Ψ' *une conique dans cette famille linéaire. Sa représentation matricielle dans* Π_∞ *est alors donnée par :*

$$\Psi' = \mu \mathtt{I} + \nu\Psi \quad ,$$

pour des scalaires réels μ *et* ν *(rappelons que la représentation matricielle de* Ω_∞ *est l'identité). Sa projection dans la* i^e *vue est alors :*

$$\begin{aligned} \mathsf{R}_i\Psi'\mathsf{R}_i{}^\mathsf{T} &= \mu\mathtt{I} + \nu\omega_i \\ &= \mu\mathtt{I} + \nu\omega_j \\ &= \mathsf{R}_j\Psi'\mathsf{R}_j{}^\mathsf{T} \quad . \end{aligned}$$

La famille linéaire considérée contient des coniques imaginaires et propres différentes de la conique absolue et il existent donc de fausses coniques absolues. □

La preuve du lemme implique deux choses :

▷ l'existence d'une fausse conique absolue sur Π_∞ implique que *toutes* les coniques imaginaires et propres dans la famille linéaire engendrée par cette conique et la vraie conique absolue sont des coniques absolues potentielles ;

▷ une caractéristique commune aux séquences critiques du lemme 7.1 est l'existence d'un point réel dont les projections sont identiques dans toutes les vues de la séquence (le point $\mathbf{Q}$).

7.9.1.2 Coniques absolues potentielles hors de Π_∞

La considération des coniques hors de Π_∞ est plus complexe que de celles sur Π_∞, puisque leur projection dépend non seulement de la rotation de la caméra mais aussi de la position du centre de projection. Avec le lemme 7.1, nous avons donné une condition nécessaire *et* suffisante pour qu'une séquence soit critique par rapport à une conique sur Π_∞. Les résultats pour les coniques absolues potentielles hors de Π_∞, qui seront établis dans la suite de cette section, ne vont pas aussi loin – nous allons uniquement dériver des conditions nécessaires pour qu'une séquence de mouvements soit critique.

Nous reformulons maintenant le problème posé pour pouvoir le résoudre plus facilement. Soit Ω une fausse conique absolue hors de Π_∞ et ω sa projection dans une vue initiale. Nous voulons déterminer toutes les positions de la caméra pour lesquelles Ω se projette également sur ω. La détermination de nouvelles positions est équivalente à la détermination de déplacements rigides de la position initiale. Considérons ce qui se passe lorsqu'on déplace la caméra par une transformation rigide. Les projections de Ω avant et après le mouvement sont désirées être identiques (la conique ω). Donc, les cônes de projection de Ω avant et après le mouvement doivent être égaux à la même transformation rigide près qui a déplacé la caméra.

Considérons le scénario sous un autre angle. Soit Λ le cône de projection de Ω pour la vue initiale. Soit T une transformation rigide qui transforme Λ en un cône Λ' qui contient également Ω. Déplacer la caméra par T résulte en une position pour laquelle Ω a la même image que dans la vue initiale. Le problème de la détermination des séquences critiques par rapport à une conique hors de Π_∞ peut donc être posé ainsi :

PROBLÈME 7.1 *Soit Ω une conique imaginaire propre hors de Π_∞. Supposons qu'une position de caméra initiale est donnée, telle que le centre de projection n'est pas sur le plan support de Ω. Soient ω la projection de Ω dans la vue initiale et Λ le cône de projection associé.*

Déterminons toutes les transformations rigides qui transforment Λ sur des cônes Λ' contenant Ω.

Rappelons qu'un cône est défini par son vertex $\mathbf{V}$ et une section plane qui ne contient pas $\mathbf{V}$. Les cônes Λ' cherchés doivent contenir la conique 3D Ω. Donc, les candidats pour Λ' peuvent être paramétrés par les coordonnées du vertex. On peut donc sous-diviser le problème 7.1 en deux parties :

PROBLÈME 7.2 *Déterminons tous les vertexes $\mathbf{V}$ tels que les cônes engendrés par $\mathbf{V}$ et Ω soient égaux à Λ à une transformation rigide près.*

PROBLÈME 7.3 *Après avoir résolu le problème 7.2, déterminons toutes les transformations rigides entre Λ et les cônes définis par Ω et les vertexes $\mathbf{V}$.*

Qu'est-ce que les solutions de ces problèmes nous disent sur les positions de la caméra cherchées ? Les vertexes qui sont les solutions du problème 7.2, représentent toutes les positions possibles du centre de projection dans la séquence critique par rapport à Ω. Les transformations rigides qui résolvent le problème 7.3 donnent ensuite les rotations relatives entre les vues.

Dans le prochain paragraphe, nous considérons le problème 7.3. Nous ne le résolvons pas complètement, mais nous dérivons le nombre et le type de rotations que la caméra peut effectuer à chaque position du centre de projection. Ensuite, nous déterminons les positions possibles du centre de projection, en résolvant le problème 7.2, et nous faisons la synthèse des résultats sur les positions et les rotations.

Rotations de la caméra. Dans ce paragraphe, nous ne considérons pas directement les rotations relatives entre les vues, mais nous raisonnons sur le nombre de rotations que peut effectuer la caméra à une position précise. Si nous considérons le cône de projection de Ω, nous cherchons alors les rotations autour de son vertex qui laissent le cône globalement invariant. Le nombre et les types de ces rotations dépendent du type du cône.

Un cône isotropique est invariant à toute rotation. Ceci est facile à voir : un cône isotropique contient la conique absolue. Les rotations autour de son vertex laissent donc le vertex et une section plane invariants et par conséquent aussi le cône lui-même. Nous comptons le nombre de rotations possibles par ∞^3.

Un cône circulaire est invariant aux rotations autour de son axe principal et aux rotations de $180°$ dont l'axe est perpendiculaire à l'axe principal. Toute combinaison de ces rotations peut être exprimée par la combinaison d'une rotation autour de l'axe principal et d'une rotation de $0°$ ou $180°$ autour d'un axe $\mathbf{L}$. Nous comptons le nombre de rotations possibles alors par $2 \times \infty$.

Un cône elliptique est invariant aux rotations de $180°$ autour de ses axes. Si l'on tient compte de l'identité, il existent donc 4 rotations différentes qui laissent le cône invariant.

Positions du centre de projection. Dans la section 7.7, nous avons déjà mentionné que le repère euclidien du monde peut être choisi de manière arbitraire. Donc, sans perte de généralité, nous pouvons supposer que Ω se trouve sur le plan $\Pi \equiv (Z = 0)$ et que sa représentation matricielle C dans ce plan est en forme normale euclidienne. Puisque Ω est imaginaire et propre, C est une matrice diagonale :

$$\mathsf{C} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 1 \end{pmatrix} ,$$

avec $a, b > 0$. Soit le centre de projection de la vue initiale donnée par $\mathbf{t}_1$. Le cône de projection de Ω est alors représenté par la matrice Λ suivante (cf. l'annexe D.2.2) :

$$\Lambda_1 \sim \begin{pmatrix} a & 0 & -a\frac{t_{11}}{t_{13}} & 0 \\ 0 & b & -b\frac{t_{12}}{t_{13}} & 0 \\ -a\frac{t_{11}}{t_{13}} & -b\frac{t_{12}}{t_{13}} & \frac{at_{11}^2+bt_{12}^2+1}{t_{13}^2} & -\frac{1}{t_{13}} \\ 0 & 0 & -\frac{1}{t_{13}} & 1 \end{pmatrix} .$$

Soit $\mathbf{t}_2$ le centre de projection d'une deuxième vue. Le cône engendré par $\mathbf{t}_2$ et Ω est donné par :

$$\Lambda_2 \sim \begin{pmatrix} a & 0 & -a\frac{t_{21}}{t_{23}} & 0 \\ 0 & b & -b\frac{t_{22}}{t_{23}} & 0 \\ -a\frac{t_{21}}{t_{23}} & -b\frac{t_{22}}{t_{23}} & \frac{at_{21}^2+bt_{22}^2+1}{t_{23}^2} & -\frac{1}{t_{23}} \\ 0 & 0 & -\frac{1}{t_{23}} & 1 \end{pmatrix} .$$

Nous voulons déterminer les $\mathbf{t}_2$ tels que Λ_2 soit égal à Λ_1 à une transformation rigide près. Ceci est le cas exactement si leurs premières sous-matrices 3×3 ont les mêmes valeurs propres, à un facteur près (voir la preuve dans l'annexe D.2.3).

Dans la suite, nous explicitons cette condition. Soient M_1 et M_2 les deux sous-matrices 3×3

(après multiplication avec t_{13}^2 respectivement t_{23}^2) :

$$\begin{aligned}
\mathsf{M}_1 &= \begin{pmatrix} at_{13}^2 & 0 & -at_{11}t_{13} \\ 0 & bt_{13}^2 & -bt_{12}t_{13} \\ -at_{11}t_{13} & -bt_{12}t_{13} & at_{11}^2+bt_{12}^2+1 \end{pmatrix} \\
\mathsf{M}_2 &= \begin{pmatrix} at_{23}^2 & 0 & -at_{21}t_{23} \\ 0 & bt_{23}^2 & -bt_{22}t_{23} \\ -at_{21}t_{23} & -bt_{22}t_{23} & at_{21}^2+bt_{22}^2+1 \end{pmatrix} \quad .
\end{aligned}$$

Supposons que les valeurs propres de M_1 sont données par s_1, s_2 et s_3. Les valeurs propres de M_2 doivent donc être des valeurs $\rho s_1, \rho s_2$ et ρs_3. Les polynômes caractéristiques de M_1 et M_2 doivent alors être de la forme :

$$\begin{aligned}
p_1' &= \lambda^3 - \lambda^2(s_1+s_2+s_3) + \lambda(s_1s_2+s_2s_3+s_3s_1) - s_1s_2s_3 \\
p_2' &= \lambda^3 - \rho\lambda^2(s_1+s_2+s_3) + \rho^2\lambda(s_1s_2+s_2s_3+s_3s_1) - \rho^3 s_1s_2s_3 \quad .
\end{aligned}$$

Inspectons les polynômes caractéristiques de M_1 et M_2 :

$$\begin{aligned}
p_1 &= \lambda^3 - \lambda^2(a^2(t_{11}^2+t_{13}^2)+b^2(t_{12}^2+t_{13}^2)+1) + \lambda t_{13}^2(a^2b^2(t_{11}^2+t_{12}^2+t_{13}^2)+a^2+b^2) - a^2b^2t_{13}^4 \\
p_2 &= \lambda^3 - \lambda^2(a^2(t_{21}^2+t_{23}^2)+b^2(t_{22}^2+t_{23}^2)+1) + \lambda t_{23}^2(a^2b^2(t_{21}^2+t_{22}^2+t_{23}^2)+a^2+b^2) - a^2b^2t_{23}^4 \quad .
\end{aligned}$$

La comparaison des coefficients de p_1, p_1', p_2 et p_2' mène aux équations suivantes :

$$\begin{aligned}
\rho(a^2(t_{11}^2+t_{13}^2)+b^2(t_{12}^2+t_{13}^2)+1) &= a^2(t_{21}^2+t_{23}^2)+b^2(t_{22}^2+t_{23}^2)+1 \\
\rho^2 t_{13}^2(a^2b^2(t_{11}^2+t_{12}^2+t_{13}^2)+a^2+b^2) &= t_{23}^2(a^2b^2(t_{21}^2+t_{22}^2+t_{23}^2)+a^2+b^2) \qquad (7.12) \\
\rho^3 t_{13}^4 a^2b^2 &= t_{23}^4 a^2 b^2 \quad .
\end{aligned}$$

Afin d'obtenir des contraintes sur $\mathbf{t}_2$ en fonction de $\mathbf{t}_1$, il faut éliminer ρ. Une façon de ce faire est de déterminer ρ à l'aide de la première des trois équations et de substituer la solution dans les deux autres équations. Nous obtenons deux équations, une de degré 4 et une de degré 6 dans les coordonnées de $\mathbf{t}_2$. Les coordonnées n'apparaissent qu'avec des exposants pairs. Si l'on considère les carrés des coordonnées, on obtient donc une équation quadratique et une équation cubique. Les points $(t_{21}^2, t_{22}^2, t_{23}^2)^\mathsf{T}$ se trouvent donc sur une courbe de degré 6. Le lieu des centres de projection finalement, est une courbe de degré 12. Dans [188], nous avons rapporté un résultat plus faible, en disant que le lieu des centres de projection consiste au maximum de deux cercles. Les conclusions principales de [188] ne sont pourtant pas affectées.

Dans la suite, nous considérons des cas particuliers selon le type de Ω et de Λ_1. Puisque Ω est une conique imaginaire et propre, elle doit être un cercle ou une ellipse. Le cône Λ_1 peut être elliptique, circulaire ou isotropique. Nous considérons toutes les combinaisons de ces possibilités. Soit Π le plan support de Ω.

CAS 7.4 (Λ_1 : CÔNE ISOTROPIQUE, Ω : ELLIPSE) *Toutes les sections planes d'un cône isotropique sont des cercles. Donc, ce cas n'est pas réalisable.*

CAS 7.5 (Λ_1 : CÔNE ISOTROPIQUE, Ω : CERCLE) *La projection perpendiculaire du vertex $\mathbf{t}_1$ sur Π est le centre de Ω. Le seul autre centre de projection possible est le point $\mathbf{t}_2$ avec :*

$$\mathbf{t}_2 = \begin{pmatrix} t_{11} \\ t_{12} \\ -t_{13} \end{pmatrix} \quad .$$

La position des centres de projection par rapport à Ω est illustrée dans la figure 7.11 (a). Puisque Λ_1 est un cône isotropique, la caméra peut effectuer des rotation arbitraires à toutes les deux positions possibles, sans affecter la projection de Ω.

CAS 7.6 (Λ_1 : CÔNE CIRCULAIRE, Ω : ELLIPSE) *Le vertex $\mathbf{t}_1$ se trouve dans un plan Π_1 qui est perpendiculaire à Π et qui contient un axe de Ω. Il n'y a au total que 4 centres de projection possibles. Ils forment un rectangle dans le plan Π_1, qui est symétrique par rapport à Π (cf. la figure 7.11 (b)). À chaque position, la caméra peut effectuer des rotations comme décrit dans le paragraphe précédent (rotations arbitraires autour de l'axe principal de Λ_1 ou de $180°$ autour d'axes perpendiculaires).*

CAS 7.7 (Λ_1 : CÔNE CIRCULAIRE, Ω : CERCLE) *Comme dans le cas 7.5, la projection perpendiculaire du vertex $\mathbf{t}_1$ sur Π est le centre de Ω. Le seul autre centre de projection possible est le même que dans le cas 7.5. À chaque position, la caméra peut effectuer des rotations comme décrit dans le paragraphe précédent (rotations arbitraires autour de l'axe principal de Λ_1 ou de $180°$ autour d'axes perpendiculaires).*

CAS 7.8 (Λ_1 : CÔNE ELLIPTIQUE, Ω : ELLIPSE) *Ceci donne le cas général, c'est-à-dire que les centres de projection se trouvent sur une courbe de degré 12. Cette courbe est symétrique par rapport à Π et par rapport aux deux plans perpendiculaires à Π et contenant un axe de Ω. Un exemple est montré dans la figure 7.11 (c). À chaque position, la caméra peut effectuer 4 rotations différentes.*

CAS 7.9 (Λ_1 : CÔNE ELLIPTIQUE, Ω : CERCLE) *La caméra peut se déplacer sur les deux cercles esquissés dans la figure 7.11 (d) (cf. l'annexe D.2.4). Cette figure est symétrique par rapport à Π. À chaque position, la caméra peut effectuer 4 rotations différentes.*

7.9.2 Résumé

Le tableau 7.2 résume les résultats des sections précédentes. On peut voir qu'il n'y a aucune séquence critique par rapport à une conique hors de Π_∞ qui contient simultanément « beaucoup » de positions de caméra et « beaucoup » de rotations.

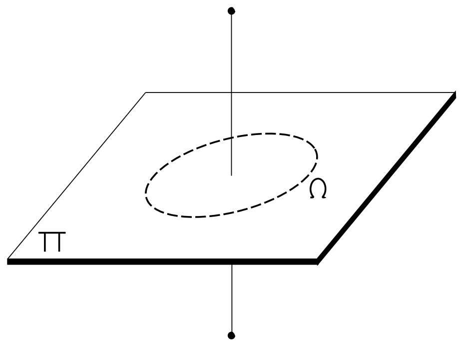
Coniques absolues potentielles	Lieu des centres de projection	# Rotations
À l'infini	Quelconque	$2 \times \infty$
Non à l'infini	Courbe de degré 12	4
	Deux cercles	4
	Coins d'un rectangle	$2 \times \infty$
	Deux points	∞^3

TAB. 7.2: *Les séquences critiques de la caméra planaire. Les chiffres dans la colonne « # Rotations » indiquent le nombre de rotations qui sont possibles à chaque position du centre de projection.*

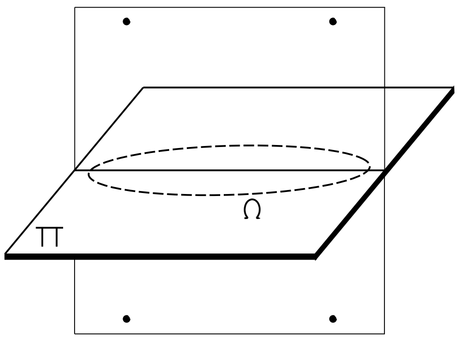
7.9.3 Séquences critiques qui sont courantes en pratique

Quelques-unes des séquences de mouvements critiques sont couramment appliquées en pratique lors de la prise d'images pour la modélisation d'objets, par exemple. Ces séquences sont décrites dans le tableau 7.3 et dans les paragraphes suivants.

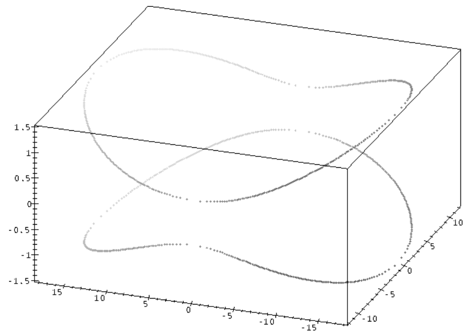
Les mouvements planaires et le cas spécial des mouvements orbitaux sont couramment utilisés pour la modélisation d'objets et dans des applications de navigation. Le premier type dans le tableau



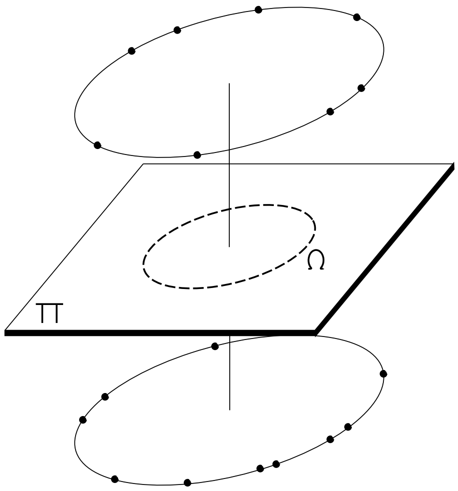
(a) Cas 7.5 et 7.7.



(b) Cas 7.6.



(c) Cas 7.8.



(d) Cas 7.9.

FIG. 7.11: *Le lieu des centres de projection dans une séquence critique.*

7.2 est en quelque sorte une généralisation du mouvement planaire – puisque en général il n'y a que de coniques absolues potentielles sur le plan à l'infini, il n'y a pas de contrainte sur les translations des caméras. Les deux types de mouvements causent donc le même degré d'ambiguïté dans la reconstruction euclidienne. Pourtant, l'auto-calibrage avec ces mouvements ne peut pas forcément être fait avec les mêmes algorithmes. La méthode de Armstrong *et al.* [6] n'est pas adaptée pour le mouvement plus général puisqu'elle est basée sur la coplanarité des centres de projection.

Les translations pures sont également courantes en robotique mobile, du moins pour de courts intervalles de trajectoires. Les rotations pures constituent un bon moyen pour le calibrage intrinsèque d'une caméra, grâce à l'appariement de primitives image plus facile et des équations simples [82].

Description	# $\mathbf{t}$	# R	# Ω
Rotations autour d'axes parallèles et translations quelconques, figure 7.12 (b)	∞^3	∞	∞
Mouvement planaire, figure 7.12 (c)	∞^2	∞	∞
Translations pures	∞^3	1	∞^5
Mouvement orbital, figure 7.12 (a)	$2 \times \infty$	1	∞^2
Rotations pures	1	∞^3	∞^3

TAB. 7.3: *Types de séquences critiques qui sont courants en pratique. Les chiffres dans la colonne « # $\mathbf{t}$ » indiquent le nombre de positions possibles du centre de projection et ceux de la colonne « # R » le nombre de rotations qui sont possibles à chaque position. La dernière colonne contient les nombres de coniques absolues potentielles.*

7.9.3.1 Rotations pures

Soit S une séquence d'images prises en effectuant des rotations pures. Aucune reconstruction 3D n'est possible et S est donc critique pour la reconstruction euclidienne. Il existent toujours des coniques hors du plan à l'infini par rapport auxquelles S est critique : considérons le cône isotropique du centre de projection. Ce cône est invariant aux rotations autour du centre de projection. Donc, toutes les sections planes sont projetées sur la même conique dans toutes les images.

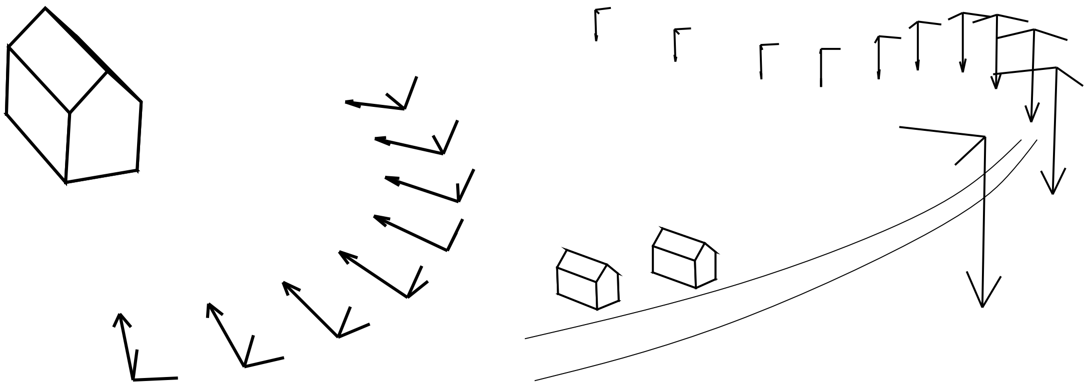
L'auto-calibrage de la caméra est possible exactement s'il n'y a aucune fausse conique absolue sur le plan à l'infini. D'après le lemme 7.1, nous devons alors imposer les contraintes suivantes sur les rotations de la caméra : il faut y avoir au moins deux axes de rotation différents. S'il y a seulement deux axes différents et s'ils sont perpendiculaires, il faut y avoir au moins une rotation d'un angle différent de 180°. Sous ces conditions, l'auto-calibrage est possible (voir la section 6.2.3.1 pour des méthodes pratiques).

7.9.3.2 Translations pures

Nous considérons une séquence de mouvements composée uniquement de translations, éventuellement dans des directions différentes. La séquence est critique par rapport à toutes les coniques imaginaires et propres sur le plan à l'infini. Elle est donc critique pour la reconstruction euclidienne et l'auto-calibrage. Il est connu que la reconstruction affine est possible, même avec un seul mouvement [131].

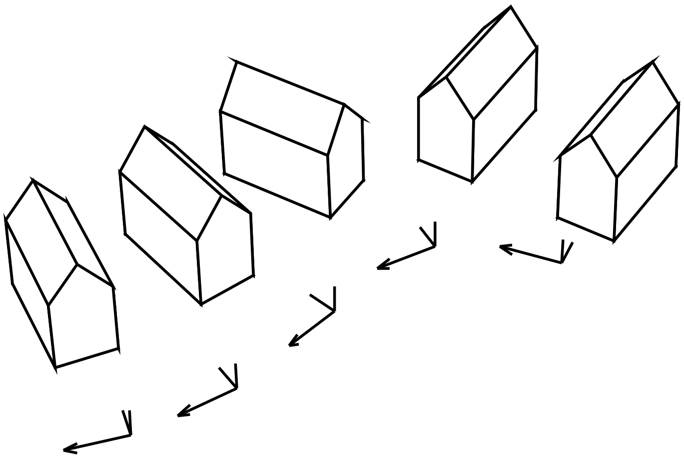
Pour un seul mouvement, il existent pourtant des coniques absolues potentielles hors de Π_∞ : soit $\mathbf{Q}$ le point au milieu de deux centres de projection. Dans l'annexe D.2.5, nous montrons que toutes les coniques propres et imaginaires dont le centre coïncide avec $\mathbf{Q}$, sont des coniques absolues potentielles.

Il y a donc plusieurs plans à l'infini potentiels, le vrai plan à l'infini et tous les plans qui passent par $\mathbf{Q}$. La raison pourquoi la reconstruction affine est quand même possible de manière unique est que *toutes* les coniques imaginaires et propres sur Π_∞ sont des coniques absolues potentielles, ce qui n'est pas le cas pour les autres plans (voir D.2.5).



(a) Mouvement orbital.

(b) Rotations autour d'axes parallèles et translations quelconques.



(c) Mouvement planaire.

FIG. 7.12: *Illustration de quelques types de séquences critiques. Les caméras sont représentées par les axes du repère local, les flèches indiquant les axes optiques.*

7.9.3.3 Mouvements planaires

Les séquences de mouvements planaires (cf. la section 6.2.3.2) sont critiques par rapport à des coniques sur le plan à l'infini puisque toutes les rotation relatives s'effectuent autour du même axe[5]. En général, il y a une famille de coniques absolues potentielles.

A priori, l'utilisation de contraintes sur les paramètres intrinsèques permet de réduire le degré de l'ambiguïté de la reconstruction euclidienne. Armstrong *et al.* proposent par exemple d'utiliser l'information que les pixels sont rectangulaires [6]. En général, cette contrainte réduit le nombre de solutions à deux. Nous montrons dans la suite que la contrainte n'est pas effective si les caméras sont positionnées telles que les plans images sont perpendiculaires au plan du mouvement et qu'un axe de

5. Nous rappelons que par rotation relative nous entendons la partie rotationnelle du mouvement de caméra, exprimée par une rotation autour du centre de projection.

pixels est parallèle à ce plan.

Supposons que le plan du mouvement est le plan $Z = 0$. Les rotations des caméras peuvent alors être représentées par :

$$\mathsf{R}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} \cos\gamma_i & -\sin\gamma_i & 0 \\ \sin\gamma_i & \cos\gamma_i & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos\gamma_i & -\sin\gamma_i & 0 \\ 0 & 0 & 1 \\ -\sin\gamma_i & -\cos\gamma_i & 0 \end{pmatrix} .$$

Considérons la famille des coniques sur le plan à l'infini paramétrée par μ comme :

$$\Omega = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \mu \end{pmatrix} .$$

La projection de ces coniques dans la i^{e} vue est donnée par :

$$\omega_i = \mathsf{R}_i \Omega \mathsf{R}_i^{\mathsf{T}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

Les projections de chaque conique Ω sont donc identiques dans toutes les vues. En outre, les coefficients $\omega_{i_{12}}$ et $\omega_{i_{21}}$ sont nuls, ce qui signifie que les axes de la conique ω_i sont parallèles aux axes du repère image. Si les pixels sont rectangulaires, la transformation de la conique dans le repère des pixels (par la matrice des paramètres intrinsèques) n'affecte pas cette propriété.

Le fait que les pixels sont rectangulaires s'exprime exactement par la contrainte que les axes de l'image de la conique absolue sont parallèles aux axes du repère des pixels. Nous avons montré que, pour la configuration considérée, il existe toute une famille de coniques absolues potentielles qui satisfont cette contrainte. Donc, il n'y a aucune réduction du degré de l'ambiguïté de la reconstruction euclidienne.

7.9.3.4 Mouvements orbitaux

Ceci est un cas spécial des mouvements planaires. Il y a donc les mêmes coniques absolues potentielles. En plus, il existent des coniques absolues potentielles hors de Π_∞ : soit Π le plan du mouvement et $\mathbf{L}$ l'axe de rotation. Tous les cercles imaginaires, qui sont centrés en $\mathbf{L}$ et dont le plan support est parallèle à Π, sont des coniques absolues potentielles. Ces cercles sont esquissés dans la figure 7.13.

Dans chaque plan parallèle à Π et aussi dans le plan à l'infini il existe donc une famille d'une dimension de coniques absolues potentielles[6]. Il n'y a donc a priori aucun moyen de distinguer le plan à l'infini des autres plans ce qui signifie que la reconstruction affine ne peut être obtenue qu'à un paramètre près. Comme pour les mouvements planaires, la structure euclidienne du plan Π et des plans parallèles à Π peut être déterminée : les coniques absolues potentielles sont des cercles sur des plans parallèles. Donc, elles s'intersectent toutes exactement dans les deux points cycliques de ces plans. Les points cycliques peuvent donc être déterminés de manière unique ce qui donne la structure euclidienne dans les plans considérés.

6. Ceci n'est pas tout à fait exact : seules les coniques *imaginaires et propres* dans une famille sont concernées.

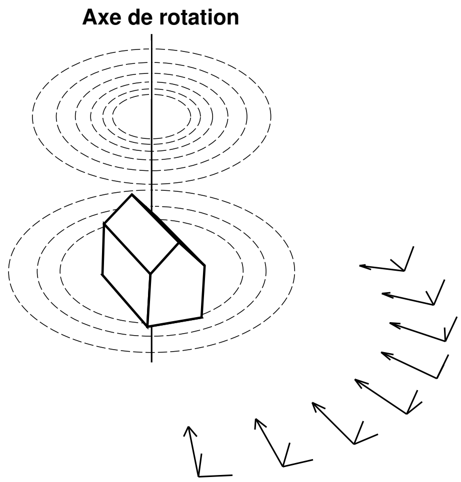


FIG. 7.13: *Coniques absolues potentielles pour un mouvement orbital.*

7.10 Ambiguïtés artificielles

Nous abordons ici le problème d'ambiguïtés artificielles, qui ne sont pas inhérentes à l'auto-calibrage, mais qui sont causées par la méthode utilisée. Pour appliquer une méthode avec succès, il est souhaitable d'être conscient de ses insuffisances. En effet, toutes les méthodes d'auto-calibrage révisées au 6.2.1 sont concernées : toutes les méthodes conçues pour des mouvements généraux ne savent pas traiter le cas de rotations pures de la caméra et toutes les méthodes conçues pour des mouvements spécifiques ne permettent pas l'auto-calibrage si le mouvement ne suit pas le modèle supposé. Ce problème pourrait être résolu en appliquant un procédé de choix de modèle qui consiste à essayer plusieurs méthodes, conçues pour différents types de mouvements ; le modèle offrant le meilleur compromis entre sa taille (paramètres du mouvement spécifique, par exemple) et la qualité du résultat (résidu) est retenu comme la bonne solution (voir [203] pour une revue de critères pour le choix de modèle). Ceci n'est pourtant pas le sujet de ce paragraphe.

Il s'agit plutôt de montrer quelques exemples d'ambiguïtés artificielles, pour illustrer le problème et fournir une explication d'occasionnels mauvais résultats des méthodes concernées. Le but n'est pas de donner un catalogue complet de ces dégénérescences.

7.10.1 Les équations de Kruppa

Nous montrons dans la suite que les équations de Kruppa sont dégénérées si les caméras se trouvent sur une sphère et si tous les axes optiques passent par le centre de la sphère, $\mathbf{V}$. Nous parlons dans la suite de **caméras convergentes** pour décrire cette situation. C'est une configuration assez courante pour la modélisation d'objets de petite ou moyenne taille.

Considérons une sphère Φ quelconque qui est aussi centrée en $\mathbf{V}$. Pour toutes les caméras dans notre configuration, le cône de projection de la sphère (le cône circulaire avec le centre de projection comme vertex et qui englobe la sphère) coupe le plan image en un cercle, qui est centré au point principal. Puisque toutes les caméras ont la même distance de $\mathbf{V}$, ce cercle est le même dans tous les plans image. Les caméras ont le même calibrage intrinsèque et, par conséquent, la représentation du cercle image dans le repère des pixels est identique pour toutes les vues. Si les pixels ne sont pas carrés, la transformation du cercle en coordonnées pixels est une ellipse. Cette ellipse, étant la projection commune de la même quadrique, la sphère Φ, dans toutes les vues, satisfait alors les équations de Kruppa, qui sont des contraintes d'appariement de quadriques !

Il y a une famille linéaire de sphères centrées dans $\mathbf{V}$, qu'on obtient en faisant varier le rayon. Donc, il y a aussi une famille d'ellipses dans l'image satisfaisant les équations de Kruppa. Seules les ellipses imaginaires peuvent donner une matrice des paramètres intrinsèques : celle-ci est obtenue par une décomposition de Cholesky de la matrice de l'ellipse, ce qui n'est possible que si cette dernière est définie positive (ce qui caractérise exactement les coniques imaginaires). Il reste quand même un nombre infini de calibrages intrinsèques potentiels.

Il y a une autre chose intéressante à remarquer. Notamment, toutes les ellipses en question ont le même centre, les mêmes directions principales et surtout, le même rapport des longueurs des axes. Ce rapport est égal au rapport d'échelle de la caméra. Donc, bien qu'il ne soit pas possible de déterminer un calibrage intrinsèque unique, le rapport d'échelle peut être trouvé. Il reste tout de même un grand inconvénient : même si le point principal, l'angle entre les axes pixels et le rapport d'échelle sont connus, il reste impossible de déterminer α_u et α_v individuellement, c'est-à-dire d'estimer la distance focale.

Nous avons donné une explication au phénomène observé par Zeller [232] (voir aussi 6.2.1.5), que les solutions pour α_u et α_v sont distribuées le long d'une droite : pour une configuration de caméras exactement convergentes, tous les points sur la droite sont des solutions valables (chaque point (α_u, α_v) correspond à une des ellipses image mentionnées ci-dessus). Si les caméras sont *presque* convergentes, comme c'est le cas pour les expériences de Zeller, il n'y a plus qu'un nombre fini de solutions, mais qui se trouvent à proximité de la droite.

La raison pourquoi les équations de Kruppa dégénèrent pour cette configuration, malgré l'absence d'ambiguïtés inhérentes, réside en leur signification géométrique : il s'agit de contraintes d'appariement de *quadriques*. L'auto-calibrage consiste à trouver la conique absolue, qui peut être vue comme une quadrique *plane*. C'est cette planarité qui n'est pas exprimable par les équations de Kruppa.

7.10.2 La méthode linéaire de Triggs

Cette méthode (cf. le paragraphe 6.2.1.3) montre les mêmes faiblesses que les équations de Kruppa, c'est-à-dire qu'elle ne converge pas vers la bonne solution, si les caméras sont convergentes. La raison en est la même : la planarité de la conique absolue ou, dans le formalisme de Triggs, la contrainte de rang 3 du cône dual absolu, n'est pas considérée.

7.10.3 La contrainte du module

Nous montrons dans la suite que la contrainte du module (cf. la section 6.2.1.4) est une caractérisation insuffisante du plan à l'infini et qu'elle entraîne des ambiguïtés artificielles pour la reconstruction affine. Nous rappelons que la contrainte du module exprime l'égalité du module des valeurs propres de l'homographie infinie. Cette propriété peut facilement être démontrée comme dans la section 6.2.1.4. Dans l'annexe D.3, nous dérivons ce résultat d'une manière alternative, qui nous permettra de mieux percevoir la situation.

Notamment, les valeurs propres d'une homographie d'un plan Π ont le même module si elle laisse une conique imaginaire et propre globalement invariante. Ceci veut dire qu'il y a une conique imaginaire et propre sur Π dont les projections dans les deux vues considérées sont identiques. Il s'agit là d'une contrainte *locale* pour des paires de vues ; une caractérisation complète du plan à l'infini doit prendre en compte que la conique avec des projections identiques doit être la même pour toutes les paires de vues.

Considérons l'exemple du mouvement planaire (cf. la figure 7.14). Tous les plans qui sont parallèles au plan du mouvement satisfont les contraintes du module pour toutes les paires de vues : les projections des cercles centrés dans l'axe de la rotation liant deux positions de caméra sont identiques dans les deux vues. Si la séquence de mouvements contient des axes différents, les cercles avec des projections identiques dans des paires de vues sont alors différentes. Quant au plan à l'infini, il supporte la conique absolue, dont les projections sont identiques dans *toutes* les vues. La reconstruction affine basée sur la contrainte du module n'aura donc pas de solution unique pour le mouvement planaire, bien qu'il n'y ait pas d'ambiguïtés inhérentes [6].

En conclusion, on peut constater qu'il ne semble pas être possible de complètement isoler la reconstruction affine de la reconstruction euclidienne, sauf bien sûr si des connaissances affine supplémentaires sont disponibles.

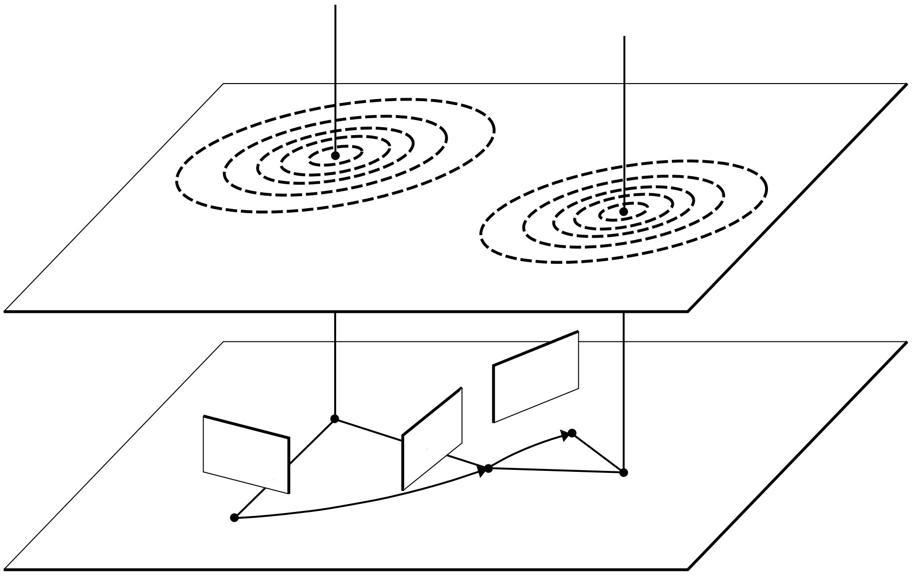


FIG. 7.14: *Illustration de la dégénérescence de la contrainte du module pour le mouvement planaire.*

7.11 Aspects pratiques

Dans la section 7.9.3 nous avons discuté des types de séquences critiques qui sont courants dans des applications. Ils ne sont pas nombreux, mais la plupart des séquences d'images disponibles sont effectivement critiques ou proches d'une séquence critique. Avec les séquences d'images utilisées pour la reconstruction projective dans le chapitre 5, par exemple, l'auto-calibrage n'a pas été possible[7].

Dans la suite, nous discutons quelques aspects pratiques concernant la réduction du degré de l'ambiguïté dans la reconstruction et l'identification automatique de séquences de mouvements critiques.

7. Il y a bien sûr d'autres facteurs qui influencent les résultats de l'auto-calibrage, comme la méthode utilisée (nous avons utilisée la méthode de Triggs [210]) et la qualité de la reconstruction projective.

7.11.1 Réduction du degré de l'ambiguïté dans la reconstruction

Il est possible de réduire le degré de l'ambiguïté dans la reconstruction si des connaissances supplémentaires sur la structure de la scène, le calibrage intrinsèque ou les mouvements de la caméra sont données. En ce qui concerne les paramètres intrinsèques, on peut pratiquement toujours supposer que les pixels sont rectangulaires. En outre, pour beaucoup d'applications, au moins le rapport d'échelle de la caméra est connu. Ces contraintes sont faciles à gérer dans un algorithme du style ajustement de faisceaux. Pour d'autres méthodes il n'est pas toujours évident comment utiliser ces contraintes dès le départ. Armstrong *et al.* par exemple montrent que le fait que les pixels sont rectangulaires réduit l'ambiguïté de la reconstruction euclidienne à partir de mouvements planaires en général de manière à n'avoir plus que deux solutions possibles [6]. Pourtant, la contrainte n'est considérée qu'*après* avoir effectué la reconstruction euclidienne qui souffre d'une ambiguïté d'un degré.

Une autre façon de rejeter de fausses solutions est de vérifier si la reconstruction est physiquement possible, i.e. si tous les points 3D se trouvent devant les caméras, dans lesquelles ils sont vus.

Le meilleur moyen de combattre l'effet des séquences critiques est tout simplement de les éviter si possible, en effectuant des mouvements de caméra qui sont « loins » d'être critiques. Une stratégie exemplaire est appliquée par des photogrammètres professionnels : pour mesurer un objet de petite ou de moyenne taille, des images sont prises tout en contournant l'objet (mouvement orbital), en alternant des rotations latérales (autour de l'axe optique) de $90°$. Sans ces rotations latérales, la séquence de mouvements serait critique. Aussi, d'après le lemme 7.1, des rotations latérales de $180°$ n'y changeraient rien. On peut donc supposer que des rotations de $90°$ font que la séquence est la plus loin possible d'être critique. Cet argument intuitif explique l'observation empirique des photogrammètres qu'avec des rotations de $90°$ les correlations dans l'estimation des paramètres intrinsèques sont les plus petites.

7.11.2 Identification automatique de séquences critiques

Un système général de reconstruction euclidienne non calibrée devrait être capable de gérer des séquences critiques. Au moins la présence de solutions ambiguës devrait être détectée. Mieux serait encore l'identification du type de séquence de mouvements sous-jacente ce qui permettrait d'étiqueter correctement l'ambiguïté, e.g. comme une reconstruction affine. En outre, si la séquence critique est identifiée, un algorithme qui est spécialement conçu pour ce type de séquences critiques peut être appliquée afin d'obtenir éventuellement une reconstruction plus stable et plus précise.

Une méthode qui identifie des mouvements singuliers entre des vues consécutives est proposée par Viéville et Lingrand [216]. La méthode est basée sur un mécanisme de sélection de modèle pour la matrice fondamentale (cf. aussi [204]). Des formes particulières de la matrice fondamentale donnent lieu à des équations moins complexes pour l'auto-calibrage. L'inconvénient de la méthode est que seulement des mouvements entre des vues consécutives sont considérées ; il est donc possible que des *séquences* de mouvements critiques ne sont pas détectées. La méthode est tout de même très bien adaptée pour des applications en robotique où beaucoup de mouvements de caméra sont « réguliers ».

Il serait envisageable de développer une approche similaire de choix de modèle pour des séquences de mouvements : des algorithmes de reconstruction conçus pour différents types de mouvements peuvent être effectués. Ensuite, la reconstruction qui offre le meilleur compromis entre l'erreur résiduelle (l'erreur de reprojection) et la complexité du modèle (nombre de paramètres pour le mouvement et le calibrage intrinsèque) sera retenue (voir [203] pour une revue de critères de choix de modèle). Cette approche peut être raffinée en sélectionnant les algorithmes à effectuer d'une manière appropriée, en s'appuyant sur des « indices » pour la présence de certains types de mouvement. Pour des mouvements planaires, par exemple, tous les épipôles sont collinéaires, et pour les translations pures les matrices fondamentales sont anti-symétriques.

7.12 Conclusion

Dans ce chapitre, nous avons dérivé les séquences de mouvements de caméra qui sont critiques pour la reconstruction euclidienne non calibrée. Deux cas ont été considérés, la caméra linéaire et la caméra planaire, toutes les deux avec des paramètres intrinsèques constants. Des résultats préliminaires pour la caméra affine et la caméra planaire avec des paramètres intrinsèques variables ont été obtenus ; dans l'avenir proche, nous examinerons ces cas plus en profondeur.

Avec notre formalisme, nous avons pu clairement confirmer l'importance des séquences critiques pour la pratique ; en effet, la plupart des séquences d'images existantes sont critiques. Comme règle générale, on peut dire que plus les mouvements sont réguliers, plus la reconstruction euclidienne est affectée par le problème d'ambiguïtés. Il faut tenir compte des séquences de mouvements critiques pour toute application de reconstruction euclidienne non calibrée ; il est préférable de tout simplement éviter ces mouvements.

Nous avons décrit des séquences qui sont exactement critiques. Il serait très intéressant de mener une étude similaire à celle sur les « Gefährliche Räume » [92] (volumes critiques), afin d'évaluer le degré d'instabilité numérique de la reconstruction que causent des séquences de mouvements qui sont proches d'être critiques.

8. Conclusions

Dans cette thèse, nous présentons nos contributions à la recherche dans le domaine de la vision par ordinateur non calibrée, en particulier sur deux problèmes clé de ce domaine, la reconstruction projective et l'auto-calibrage.

Pour la reconstruction projective, nous avons développé une méthode optimale pour le cas de deux images. La méthode est basée sur un critère d'optimalité qui est valable pour tous les niveaux de reconstruction – projectif, affine et euclidien. Son applicabilité n'est donc pas bornée sur la reconstruction projective. Une tendance en vision 3D est la détermination du type du mouvement qui est sous-jacent à une séquence d'images. Dans ce contexte, nous avons étudié des cas particuliers du mouvement entre deux vues, qui permettent de simplifier l'expression mathématique de notre méthode.

Nous avons également développé une méthode élégante et performante pour la reconstruction projective à partir de plusieurs vues. La méthode est basée sur la factorisation d'une seule matrice contenant tous les points dans toutes les images. Le schéma de reconstruction par factorisation a été initialement introduit pour des modèles de caméra affines et notre méthode constitue l'extension au cas des caméras perspectives. Dû à une implémentation numérique soigneuse, la méthode fournit des résultats stables et précis.

Nous avons esquissé une approche pour la reconstruction à partir de longues séquences d'images. La méthode est potentiellement capable de remédier au problème de données incomplètes, c'est-à-dire de points qui ne sont pas visibles dans toutes les images. Des expériences préliminaires démontrent que cette méthode permet d'établir une reconstruction à partir d'une longue séquence d'images, mais il reste à faire des expériences avec des données incomplètes.

D'une manière générale, nous sommes convaincus que la création de modèles 3D à grande échelle ne s'effectuera normalement pas de manière complètement automatique. Une approche plus réaliste est la mise à disposition de puissants outils de reconstruction ou fusion de modèles à un utilisateur, qui décidera de leur utilisation. Sur une petite échelle, notre approche de reconstruction projective pour de longues séquences d'images correspond à cette idée.

La deuxième partie de cette thèse concerne l'auto-calibrage et la reconstruction euclidienne non calibrée. Le cas le plus intéressant et le plus utile est l'auto-calibrage de caméras avec un objectif à focale variable. Nous avons examiné les influences du changement de focale aux autres paramètres intrinsèques de la caméra. Ainsi, nous avons pu démontrer que la prise en compte des interdépendances qui existent entre les paramètres intrinsèques peut être profitable pour l'auto-calibrage.

Notre contribution majeure au problème de l'auto-calibrage est une étude des mouvements de caméra qui causent la dégénérescence. De manière théorique nous avons dérivé tous les types de mouvements critiques. Il s'agit d'un problème important pour la pratique – la plupart des séquences d'images existantes sont dégénérées pour l'auto-calibrage. Notre étude permet de donner des conseils pour la prise d'images afin d'éviter les situations dégénérées.

Un axe de recherche que nous envisageons de poursuivre est l'application de méthodes de sélection de modèle. Ces mécanismes peuvent être utilisés pour la détermination de cas particuliers concernant tous les éléments qui interviennent dans la reconstruction : le mouvement de la caméra, la structure de la scène et les paramètres intrinsèques de la caméra. Il serait par exemple souhaitable d'effectuer l'auto-calibrage d'une caméra en adoptant un modèle affine ou perspectif du calibrage, selon la situation. Des méthodes de choix de modèle peuvent également servir à déterminer le type de dégénérescence pour l'auto-calibrage ce qui permet d'interpréter le résultat comme étant ambiguë, mais aussi d'extraire le sous-ensemble d'informations qui ont pu être déterminées correctement.

Bibliographie

[1] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 7(4):384–401, July 1985.

[2] G. Adiv. Inherent ambiguities in recovering 3D motion and structure from a noisy flow field. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(5):477–489, May 1989.

[3] J. Aisbett. An iterated estimation of the motion parameters of a rigid body from noisy displacement vectors. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 12(11):1092–1098, November 1990.

[4] J.Y. Aloimonos. Perspective approximations. *Image and Vision Computing*, 8(3):179–192, August 1990.

[5] M. Armstrong, A. Zisserman, and P. Beardsley. Euclidean structure from uncalibrated images. In E. Hancock, editor, *Proceedings of the fifth British Machine Vision Conference, York, England*, volume 2, pages 509–518, September 1994.

[6] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1064 of *Lecture Notes in Computer Science*, pages 3–16. Springer-Verlag, April 1996.

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

[8] N. Asada, A. Amano, and M. Baba. Photometric calibration of zoom lens systems. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 186–190. IEEE Computer Society Press, August 1996.

[9] A. Azarbayejani and A.P. Pentland. Recursive estimation of motion, structure, and focal length. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(6):562–575, June 1995.

[10] R. Basri. Paraperspective $\equiv$ affine. *International Journal of Computer Vision*, 19(2):169–179, 1996.

[11] A. Basu. Active calibration. In *Proceedings of* IEEE *International Conference on Robotics and Automation, Atlanta, Georgia, USA*, volume 2, pages 764–769. IEEE Computer Society Press, May 1993.

[12] 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.

[13] 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.

[14] P. Beardsley, A. Zisserman, and D. Murray. Sequential update of projective and affine structure from motion. Technical Report 2012/94, University of Oxford, England, August 1994.

[15] 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.

[16] R. Berthilsson, A. Heyden, and G. Sparr. Recursive structure and motion from image sequences using shape and depth spaces. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 444–449. IEEE Computer Society Press, June 1997.

[17] H.A. Beyer. Some aspects of the geometric calibration of CCD-cameras. In ISPRS *Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland*, pages 68–81, June 1987.

[18] H.A. Beyer. Accurate calibration of CCD cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 96–101, 1992.

[19] S.D. Blostein and T.S. Huang. Error analysis in stereo determination of 3-D point positions. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 9(6):752–765, November 1987.

[20] W. Boehm and H. Prautzsch. *Geometric Concepts for Geometric Design*. A.K. Peters, 1994.

[21] B. Boufama and R. Mohr. Epipole and fundamental matrix estimation using the virtual parallax property. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1030–1036, June 1995.

[22] B. Boufama, R. Mohr, and F. Veillon. Euclidean constraints for uncalibrated reconstruction. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 466–470, May 1993.

[23] 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.
ftp://ftp.imag.fr/pub/MOVI/theses/brand.ps.gz.

[24] A. Brandenberger. *Fehlertheorie der äußeren Orientierung von Steilaufnahmen*. PhD thesis, Eidgenössische Technische Hochschule, Zürich, Switzerland, 1946.

[25] T.J. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(1):90–99, January 1986.

[26] M.J. Brooks, W. Chojnacki, and L. Baumela. Determining the ego-motion of an uncalibrated camera from instantaneous optical flow. *Journal of the Optical Society of America*, 14(10), October 1997.

[27] M.J. Brooks, L. de Agapito, D.Q. Huynh, and L. Baumela. Direct methods for self-calibration of a moving stereo head. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 415–426. Springer-Verlag, April 1996.

[28] D.C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32:444–462, May 1966.

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

[30] A.R. Bruss and B.K.P. Horn. Passive navigation. *Computer Vision, Graphics and Image Processing*, 21(1):3–20, January 1983.

[31] T. Buchanan. The twisted cubic and camera calibration. *Computer Vision, Graphics and Image Processing*, 42(1):130–132, April 1988.

[32] T. Buchanan. Critical sets for 3D reconstruction using lines. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 730–738. Springer-Verlag, May 1992.

[33] A.W. Burner. Zoom lens calibration for wind tunnel measurements. In S.F. El-Hakim, editor, *Proceedings of the* SPIE *Conference on Videometrics IV, Philadelphia, Pennsylvania, USA*, volume 2598 of SPIE *Proceedings Series*, pages 19–33. SPIE - Society of Photo-Optical Instrumentation Engineers, October 1995.

[34] S. Carlsson. Duality of reconstruction and positioning from projective views. In *Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA*, pages 85–92, June 1995.

[35] Y.Q. Cheng, R.T. Collins, A.R. Hanson, and E.M. Riseman. Triangulation without correspondences. In *Proceedings of* ARPA *Image Understanding Workshop, Monterey, California, USA*, pages 993–1000, November 1994.

[36] S. Christy and R. Horaud. Euclidean reconstruction: from paraperspective to perspective. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, pages 129–140. Springer-Verlag, April 1996.

[37] M.A.R. Cooper and S. Robson. Theory of close range photogrammetry. In K.B. Atkinson, editor, *Close Range Photogrammetry and Machine Vision*, chapter 2, pages 9–51. Whittles Publishing, Roseleigh House, Latheronwheel, Caithness, KW5 6DW, Scotland, UK., 1996.

[38] J. Costeira and T. Kanade. A multi-body factorization method for motion analysis. In E. Grimson, editor, *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1071–1076. IEEE, IEEE Computer Society Press, June 1995.

[39] G. Cross and A. Zisserman. Quadric reconstruction from dual-space geometry. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 25–31. IEEE Computer Society Press, January 1998.

[40] J.L. Crowley, P. Bobet, and C. Schmid. Maintaining stereo calibration by tracking image points. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 483–488, June 1993.

[41] K. Daniilidis and J. Ernst. Active intrinsic calibration using vanishing points. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 708–713. IEEE Computer Society Press, June 1996.

[42] K. Daniilidis and H.H. Nagel. Analytical results on error sensitivity of motion estimation from two views. In O. Faugeras, editor, *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, volume 427 of *Lecture Notes in Computer Science*, pages 199–208. Springer-Verlag, April 1990.

[43] C. Debrunner and N. Ahuja. Motion and structure factorization and segmentation of long multiple motion image sequences. In *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 217–221, 1992.

[44] M. Demazure. Sur deux problèmes de reconstruction. Technical report, INRIA, 1988.

[45] D. Dementhon and L.S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15(1/2):123–141, 1995.

[46] R. Deriche, Z. Zhang, Q.T. Luong, and O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, volume 800 of *Lecture Notes in Computer Science*, pages 567–576. Springer-Verlag, May 1994.

[47] F. Devernay and O. Faugeras. From projective to euclidean reconstruction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 264–269, June 1996.

[48] 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.

[49] 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.

[50] H. Ebner. Self-calibrating block adjustment. *Bildmessung und Luftbildwesen*, 44(4):128–139, 1976. Invited paper of Commission III of the XIII. Congress of the ISPRS.

[51] R. Enciso, T. Viéville, and O. Faugeras. Approximation du changement de focale et de mise au point par une transformation à trois paramètres. Technical Report 2071, INRIA, France, October 1993.

[52] R. Enciso, T. Viéville, and O. Faugeras. Approximation du changement de focale et de mise au point par une transformation affine à trois paramètres. *Traitement du Signal*, 11(5), 1994.

[53] R. Enciso, T. Viéville, and O. Faugeras. Comment simplifier le processus de calibration ? In *4èmes Journées Internationales en Informatique, "Interface Mondes Réels et Virtuels"*, Montpellier, France, June 1995.

[54] W. Faig. Calibration of close-range photogrammetric systems: Mathematical formulation. *Photogrammetric Engineering and Remote Sensing*, 41(12):1479–1486, December 1975.

[55] J.Q. Fang and T.S. Huang. Solving three-dimensional small-rotation motion equations: Uniqueness, algorithms, and numerical results. *Computer Vision, Graphics and Image Processing*, 26(2):183–206, May 1984.

[56] J.Q. Fang and T.S. Huang. Some experiments on estimating the 3-D motion parameters of a rigid body from two consecutive image frames. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 6(5):545–554, September 1984.

[57] 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.

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

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

[60] O. Faugeras, S. Laveau, L. Robert, G. Csurka, and C. Zeller. 3D reconstruction of urban scenes from sequences of images. In A. Gruen, O. Kuebler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pages 145–168. Birkhäuser Verlag, 1995.

[61] O. Faugeras and S. Maybank. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 3(4):225–246, 1990.

[62] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 951–956, June 1995.

[63] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self-calibration: Theory and experiments. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 321–334. Springer-Verlag, May 1992.

[64] O.D. Faugeras, F. Lustman, and G. Toscani. Motion and structure from point and line matches. In *Proceedings of the 1st International Conference on Computer Vision, London, England*, pages 25–34. IEEE Computer Society Press, June 1987.

[65] O.D. Faugeras and G. Toscani. Camera calibration for 3D computer vision. In *Proceedings of International Workshop on Machine Vision and Machine Intelligence, Tokyo, Japan*, 1987.

[66] C.S. Fraser, M.R. Shortis, and G. Ganci. Multi-sensor system self-calibration. In S.F. El-Hakim, editor, *Proceedings of the* SPIE *Conference on Videometrics IV, Philadelphia, Pennsylvania, USA*, volume 2598 of SPIE *Proceedings Series*. SPIE - Society of Photo-Optical Instrumentation Engineers, October 1995.

[67] C.S. Fraser and S.A. Veress. Self-calibration of a fixed-frame multiple-camera system. *Photogrammetric Engineering and Remote Sensing*, 46(11):1439–1445, 1980.

[68] G.H. Golub and C.F. van Loan. *Matrix Computation*. The Johns Hopkins University Press, Baltimore, 1989.

[69] R. Hartley. Extraction of focal lengths from the fundamental matrix. Technical report, G.E. CRD, Schenectady, NY, 1993.

[70] R. Hartley. Lines and points in three views - an integrated approach. Technical report, G.E. CRD, 1994.

[71] 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.

[72] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[73] R. Hartley and P. Sturm. Triangulation. In *Proceedings of* ARPA *Image Understanding Workshop, Monterey, California, USA*, pages 957–966, November 1994.

[74] R. Hartley and P. Sturm. Triangulation. In *Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns, Prague, Czech Republic*, pages 190–197, September 1995.

[75] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[76] R. Hartley and A. Zisserman. Private communication, July 1997.

[77] R.I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 579–587. Springer-Verlag, 1992.

[78] R.I. Hartley. Cheirality invariants. In *Proceedings of* DARPA *Image Understanding Workshop*, pages 745–753, 1993.

[79] 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.

[80] R.I. Hartley. An algorithm for self calibration from several views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pages 908–912, 1994.

[81] R.I. Hartley. Lines and points in three views - an integrated approach. In *Proceedings of* ARPA *Image Understanding Workshop, Monterey, California, USA*, pages 1009–1016, November 1994.

[82] 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.

[83] 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.

[84] R.I. Hartley and R. Gupta. Linear pushbroom cameras. In J.-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, volume 800 of *Lecture Notes in Computer Science*, pages 555–566. Springer-Verlag, May 1994.

[85] A. Heyden. Reconstruction from image sequences by means of relative depths. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1058–1063, June 1995.

[86] A. Heyden. Projective structure and motion from image sequences using subspace methods. In M. Frydrych, J. Parkkinen, and A. Visa, editors, *Proceedings of the 10th Scandinavian Conference on Image Analysis, Lappeenranta, Finland*, volume II, pages 963–968, June 1997.

[87] A. Heyden and K. Åström. A canonical framework for sequences of images. In *Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA*, pages 45–52, June 1995.

[88] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 339–343, August 1996.

[89] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 438–443. IEEE Computer Society Press, June 1997.

[90] A. Heyden and G. Sparr. Reconstruction from calibrated cameras - a new proof of the kruppa-demazure theorem. *Discrete & Computational Geometry*, 1996. submitted.

[91] S.D. Hippisley-Cox and J. Porrill. Auto-calibration - kruppa's equations and the intrinsic parameters of a camera. In E. Hancock, editor, *Proceedings of the fifth British Machine Vision Conference, York, England*, pages 771–779. British Machine Vision Association, September 1994.

[92] W. Hofmann. *Das Problem der "Gefährlichen Flächen" in Theorie und Praxis - Ein Beitrag zur Hauptaufgabe der Photogrammetrie*. PhD thesis, Fakultät für Bauwesen, Technische Universität München, Germany, 1953.

[93] R.J. Holt and A.N. Netravali. Camera calibration problem: Some new results. *Computer Vision, Graphics and Image Processing: Image Understanding*, 54(3):368–383, November 1991.

[94] R. Horaud, S. Christy, and F. Dornaika. Object pose: the link between weak perspective, para perspective, and full perspective. Technical report, Inria, September 1994.

[95] R. Horaud, S. Christy, and R. Mohr. Euclidean reconstruction and affine camera calibration using controlled robot motions. In *Proceedings of the* IEEE/RSJ *International Conference on Intelligent Robots and Systems, Grenoble, France*, September 1997.

[96] R. Horaud and G. Csurka. Self-calibration and euclidean reconstruction using motions of a stereo rig. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 96–103, January 1998.

[97] R. Horaud, R. Mohr, F. Dornaika, and B. Boufama. The advantage of mounting a camera onto a robot arm. In *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, Xian, China*, pages 206–213, April 1995.

[98] R. Horaud and O. Monga. *Vision par ordinateur : outils fondamentaux.* Éditions Hermès, Paris, 1993.

[99] B.K. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.

[100] B.K.P. Horn. Motion fields are hardly ever ambiguous. *International Journal of Computer Vision*, 1(3):259–274, 1987.

[101] T.S. Huang and O.D. Faugeras. Some properties of the E matrix in two-view motion estimation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(12):1310–1312, December 1989.

[102] D. Jacobs. Linear fitting with missing data: Applications to structure-from-motion and to characterizing intensity images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 206–212. IEEE Computer Society Press, June 1997.

[103] A. Kara, D.M. Wilkes, and K. Kawamura. 3D structure reconstruction from point correspondences between two perspective projections. *Computer Vision, Graphics and Image Processing: Image Understanding*, 60(3):392–397, November 1994.

[104] J. Koenderink and A. van Doorn. Affine structure from motion. *Journal of the Optical Society of America A*, 8(2):377–385, 1991.

[105] J.J. Koenderink and A.J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, October 1989.

[106] O. Kölbl. Selbstkalibrierung von Aufnahmekammern. *Bildmessung und Luftbildwesen*, (1):31–37, 1972.

[107] J. Krames. Zur Ermittlung eines Objektes aus zwei Perspektiven (Ein Beitrag zur Theorie der „gefährlichen Örter"). *Monatshefte für Mathematik und Physik*, 49:327–354, 1941.

[108] E. Kruppa. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Sitzungsberichte Österreichische Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse, Abteilung II a*, 122:1939–1948, 1913.

[109] S. Laveau. *Géométrie d'un système de N caméras. Théorie, estimation, et applications*. Thèse de doctorat, École Polytechnique, May 1996.

[110] J.M. Lawn and R. Cipolla. Robust egomotion estimation from affine motion parallax. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 205–210. Springer-Verlag, May 1994.

[111] F. Li, M. Brady, and C. Wiles. Fast computation of the fundamental matrix for an active stereo vision system. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1064 of *Lecture Notes in Computer Science*, pages 157–166. Springer-Verlag, April 1996.

[112] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.

[113] H.C. Longuet-Higgins. Multiple interpretations of a pair of images of a surface. In *Proceedings of the Royal Society of London*, volume 418 of *A*, 1988.

[114] Q.T. Luong. *Matrice fondamentale et autocalibration en vision par ordinateur*. Thèse de doctorat, Université de Paris-Sud, Orsay, France, December 1992.

[115] Q.T. Luong and O. Faugeras. Self-calibration of a stereo rig from unknown camera motions and point correspondences. Rapport de recherche, INRIA, July 1993.

[116] Q.T. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *International Journal of Computer Vision*, 17(1):43–76, 1996.

[117] Q.T. Luong and O.D. Faugeras. Camera calibration, scene motion and structure recovery from point correpondences and fundamental matrices. *International Journal of Computer Vision*, 1993. submitted.

[118] Q.T. Luong and O.D. Faugeras. An optimization framework for efficient self-calibration and motion determination. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pages 248–252. IEEE Computer Society Press, October 1994.

[119] Q.T. Luong and O.D. Faugeras. A stability analysis of the fundamental matrix. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, volume 800 of *Lecture Notes in Computer Science*, pages 577–588. Springer-Verlag, May 1994.

[120] Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. Technical report, University of California, Berkeley, EECS, Cory Hall 211-215, University of California, Berkeley, CA 94720, October 1993.

[121] 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.

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

[123] 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.

[124] S.J. Maybank and A. Shashua. Ambiguity in reconstruction from images of six points. Technical Report TR 96-22, Department of Computer Science, Hebrew University, Jerusalem, Israel, November 1996.

[125] C. McGlone. Bundle adjustment with geometric constraints for hypothesis evaluation. In *Proceedings of the XVIII* ISPRS-*Congress, Vienna, Austria*, pages 529–534, July 1996.

[126] P.F. McLauchlan and D.W. Murray. A unifying framework for structure and motion recovery from image sequences. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 314–320, 1995.

[127] P.F. McLauchlan and D.W. Murray. Active camera calibration for a head-eye platform using the variable state-dimension filter. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 18(1):15–21, January 1996.

[128] P. Meer, D. Mintz, A. Rosenfeld, and D.Y. Kim. Robust regression methods for computer vision: a review. *International Journal of Computer Vision*, 6(1):59–70, 1991.

[129] R. Mohr, B. Boufama, and P. Brand. Accurate projective reconstruction. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 203–227, October 1993.

[130] R. Mohr, F. Veillon, and L. Quan. Relative 3D reconstruction using multiple uncalibrated images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 543–548, June 1993.

[131] T. Moons, L. van Gool, M. van Diest, and E. Pauwels. Affine reconstruction from perspective image pairs. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 249–266, October 1993.

[132] T. Morita and T. Kanade. A sequential factorization method for recovering shape and motion from image streams. In *Proceedings of* ARPA *Image Understanding Workshop, Monterey, California, USA*, pages 1177–1188, November 1994.

[133] T. Morita and T. Kanade. A sequential factorization method for recovering shape and motion from image streams. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(8):858–867, August 1997.

[134] J.L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. The MIT Press, Cambridge, MA, USA, 1992.

[135] N. Navab and O.D. Faugeras. The critical sets of lines for camera displacement estimation: A mixed euclidean-projective and constructive approach. Technical Report RR-2305, INRIA, July 1994.

[136] N. Navab, O.D. Faugeras, and T. Viéville. The critical sets of lines for camera displacement estimation: A mixed euclidean-projective and constructive approach. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 713–723. IEEE Computer Society Press, May 1993.

[137] S. Negahdaripour. Multiple interpretations of the shape and motion of objects from two perspective images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 12(11):1025–1039, November 1990.

[138] A.N. Netravali, T.S. Huang, A.S. Krishnakumar, and R.J. Holt. Algebraic methods in 3-D motion estimation from two-view point correspondences. *International Journal of Imaging Systems and Technology*, 1:78–99, 1989.

[139] G.N. Newsam, D.Q. Huynh, M.J. Brooks, and H.P. Pan. Recovering unknown focal lengths in self-calibration: An essentially linear algorithm and degenerate configurations. In *Proceedings of the XVIII* ISPRS*-Congress, Vienna, Austria*, volume XXXI, Part B3 of *International Archives of Photogrammetry and Remote Sensing*, pages 575–580, July 1996.

[140] I. Niini. Solution of the interior orientation without control points. *Photogrammetric Journal of Finland*, 13(2):78–84, 1993.

[141] I. Niini. Relative orientation of multiple images using projective singular correlation. Licenciate's thesis, Institute of Photogrammetry and Remote Sensing, University of Technology, Helsinki, Finland, February 1994.

[142] Y. Ohta, K. Maenobu, and T. Sakai. Obtaining surface orientation from texels under perspective projection. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 746–751, 1981.

[143] H. Pan, M.J. Brooks, and G.N. Newsam. Image resituation: Initial theory. In S.F. El-Hakim, editor, *Proceedings of the* SPIE *Conference on Videometrics IV, Philadelphia, Pennsylvania, USA*, volume 2598 of SPIE *Proceedings Series*, pages 162–173. SPIE - Society of Photo-Optical Instrumentation Engineers, October 1995.

[144] H. Pan, D.Q. Huynh, and G.K. Hamlyn. Two-image resituation: Practical algorithm. In S.F. El-Hakim, editor, *Proceedings of the* SPIE *Conference on Videometrics IV, Philadelphia, Pennsylvania, USA*, volume 2598 of SPIE *Proceedings Series*, pages 174–190. SPIE - Society of Photo-Optical Instrumentation Engineers, October 1995.

[145] M. Parent. Praxitèle : gestion de véhicules électriques pour un transport individuel et collectif. In *Proc. of the C-VELEC Conference, Grenoble, France*, February 1995.

[146] M.A. Penna. Camera calibration : a quick and easy way to determine the scale factor. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13(12):1240–1245, December 1991.

[147] J. Philip. Estimation of three-dimensional motion of rigid objects from noisy observations. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13(1):61–66, 1991.

[148] J. Philip. A non-iterative algorithm for determining all essential matrices corresponding to five point pairs. *Photogrammetric Record*, 15(88):589–599, October 1996.

[149] C.J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 97–108. Springer-Verlag, May 1994.

[150] M. Pollefeys and L. Van Gool. A stratified approach to metric self-calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 407–412. IEEE Computer Society Press, June 1997.

[151] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 90–95, January 1998.

[152] M. Pollefeys and L. Van Gool. Self-calibration from the absolute conic on the plane at infinity. In G. Sommer, K. Daniilidis, and J. Pauli, editors, *Proceedings of the 7th International Conference on Computer Analysis of Images and Patterns, Kiel, Germany*, volume 1296 of *Lecture Notes in Computer Science*, pages 175–182. Springer-Verlag, September 1997.

[153] M. Pollefeys, L. van Gool, and A. Oosterlinck. Self-calibration with the modulus constraint. Technical Report KUL/ESAT/MI2/9609, Katholieke Universiteit Leuven, Heverlee, Belgium, May 1996.

[154] M. Pollefeys, L. van Gool, and M. Proesmans. Euclidean 3D reconstruction from image sequences with variable focal lengths. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1064 of *Lecture Notes in Computer Science*, pages 31–42. Springer-Verlag, April 1996.

[155] J. Ponce, T.A. Cass, and D.H. Marimont. Relative stereo and motion reconstruction. Technical Report UIUC-BI-AI-RCV-93-07, Beckman Institute, University of Illinois, Urbana-Champaign, IL, November 1993.

[156] J. Ponce and Y. Genc. Epipolar geometry and linear subspace methods: A new approach to weak calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 776–781. IEEE Computer Society Press, June 1996.

[157] J. Ponce, D.H. Marimont, and T.A. Cass. Analytical methods for uncalibrated stereo and motion reconstruction. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 463–470, May 1994.

[158] K. Prazdny. Egomotion and relative depth map from optical flow. *Biological Cybernetics*, 36:87–102, 1980.

[159] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[160] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.

[161] L. Quan. Affine stereo calibration for relative affine shape reconstruction. In *Proceedings of the fourth British Machine Vision Conference, Surrey, England*, pages 659–668, 1993.

[162] L. Quan. Invariants of 6 points from 3 uncalibrated images. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, volume II, pages 459–470. Springer-Verlag, 1994.

[163] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(1):34–46, January 1995.

[164] L. Quan. Self-calibration of an affine camera from multiple views. *International Journal of Computer Vision*, 19(1):93–105, May 1996.

[165] L. Quan. Uncalibrated 1D projective camera and 3D affine reconstruction of lines. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 60–65, June 1997.

[166] L. Quan and T. Kanade. A factorization method for affine structure from line correspondences. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 803–808, 1996.

[167] L. Quan and R. Mohr. Affine shape representation from motion through reference points. *Journal of Mathematical Imaging and Vision*, 1:145–151, 1992. also in IEEE Workshop on Visual Motion, New Jersey, pages 249–254, 1991.

[168] I.D. Reid and D.W. Murray. Active tracking of foveated feature clusters using affine structure. *International Journal of Computer Vision*, 18(1):41–60, April 1996.

[169] C. Rothwell, G. Csurka, and O. Faugeras. A comparison of projective reconstruction methods for pairs of views. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 932–937, June 1995.

[170] T. Sarjakoski. Bundle block adjustment based on conjugate gradient and direct solution method. In *Proc. of the Symposium of Commission III of the* ISPRS*, Helsinki, Finland*, pages 425–445, June 1982.

[171] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford Science Publication, 1952.

[172] L.S. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *International Journal of Computer Vision*, 16(2):147–182, 1995.

[173] A. Shashua. Projective depth: A geometric invariant for 3D reconstruction from two perspective/orthographic views and for visual recognition. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 583–590. IEEE Computer Society Press, May 1993.

[174] A. Shashua. Projective structure from uncalibrated images: Structure from motion and recognition. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 16(8):778–790, August 1994.

[175] A. Shashua. Algebraic functions for recognition. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, August 1995.

[176] H.Y. Shum, K. Ikeuchi, and R. Reddy. Principal component analysis with missing data and its application to object modeling. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pages 560–565. IEEE, IEEE Computer Society Press, June 1994.

[177] H.Y. Shum, K. Ikeuchi, and R. Reddy. Principal component analysis with missing data and its application to polyhedral object modeling. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(9):854–867, September 1995.

[178] C.C. Slama, editor. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

[179] S. Soatto and P. Perona. Recursive estimation of camera motion from uncalibrated image sequences. In *Proceedings of the 1st* IEEE *International Conference on Image Processing*, pages 58–62. IEEE, 1994.

[180] G. Sparr. Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 328–333. IEEE Computer Society Press, August 1996.

[181] M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. In *Proceedings of DARPA Image Understanding Workshop*, pages 271–283, 1990.

[182] M.E. Spetsakis and Y. Aloimonos. Optimal visual motion estimation: a note. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 14(9):959–964, September 1992.

[183] G.P. Stein. Accurate internal camera calibration using rotation, with analysis of sources of error. In E. Grimson, editor, *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 230–236. IEEE Computer Society Press, June 1995.

[184] G. Strunz. *Bildorientierung und Objektrekonstruktion mit Punkten, Linien und Flächen*. PhD thesis, Fakultät für Bauingenieur- und Vermessungswesen, TU München, Germany, 1993. published in Reihe C, Heft Nr. 408, Deutsche Geodätische Kommission.

[185] P. Sturm. Calibrage dynamique d'une caméra en mouvement par pré-calibrage. In *Journées ORASIS 1996, Clermont-Ferrand, France*, pages 61–66, May 1996.

[186] P. Sturm. Self-calibration of a moving camera by pre-calibration. In R.B. Fisher and E. Trucco, editors, *Proceedings of the seventh British Machine Vision Conference, Edinburgh, Scotland*, volume 2, pages 675–684. British Machine Vision Association, September 1996.

[187] P. Sturm. Critical motion sequences and conjugacy of ambiguous euclidean reconstructions. In M. Frydrych, J. Parkkinen, and A. Visa, editors, *Proceedings of the 10th Scandinavian Conference on Image Analysis, Lappeenranta, Finland*, volume I, pages 439–446, June 1997.

[188] 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.

[189] P. Sturm. Self-calibration of a moving zoom-lens camera by pre-calibration. *Image and Vision Computing*, 15(8):583–589, August 1997.

[190] P. Sturm and L. Quan. Affine stereo calibration. In *Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns, Prague, Czech Republic*, pages 838–843, September 1995.

[191] P. Sturm and L. Quan. Affine stereo calibration. Technical Report LIFIA-29, LIFIA–IMAG, Grenoble, France, June 1995.

[192] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 709–720. Springer-Verlag, April 1996.

[193] P. Sturm and B. Triggs. Structure et mouvement projectifs à partir d'images multiples par factorisation. In *Journées ORASIS 1996, Clermont-Ferrand, France*, pages 175–180, May 1996.

[194] T. Svoboda and T. Pajdla. Efficient motion analysis. RR K335-1995-95, Czech Technical University, Faculty of Electrical Engineering, Department of Control, Computer Vision Laboratory, October 1995.

[195] T. Svoboda and P. Sturm. A badly calibrated camera in ego-motion estimation, propagation of uncertainty. In G. Sommer, K. Daniilidis, and J. Pauli, editors, *Proceedings of the 7th International Conference on Computer Analysis of Images and Patterns, Kiel, Germany*, volume 1296 of *Lecture Notes in Computer Science*, pages 183–190. Springer-Verlag, September 1997.

[196] R. Szeliski and S.B. Kang. Recovering 3D shape and motion from image streams using non-linear least squares. Technical Report CRL 93/3, Digital Equipment Corporation, Cambridge Research Lab, March 1993.

[197] R. Szeliski and S.B. Kang. Recovering 3D shape and motion from image streams using non-linear least squares. *Journal of Visual Communication and Image Representation*, 1(5):10–28, March 1994.

[198] R. Szeliski and S.B. Kang. Direct methods for visual scene reconstruction. In *Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA*, pages 26–33, June 1995.

[199] R. Szeliski and S.B. Kang. Shape ambiguities in structure from motion. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1064 of *Lecture Notes in Computer Science*, pages 709–721. Springer-Verlag, April 1996.

[200] R. Szeliski and S.B. Kang. Shape ambiguities in structure from motion. Technical Report CRL 96/1, Digital Equipment Corporation, Cambridge Research Lab, February 1996.

[201] R. Szeliski and S.B. Kang. Shape ambiguities in structure from motion. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(5):506–512, May 1997.

[202] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

[203] P.H.S. Torr. An assessment of information criteria for motion model selection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 47–52, June 1997.

[204] P.H.S. Torr, A. Zisserman, and S.J. Maybank. Robust detection of degenerate configurations for the fundamental matrix. In E. Grimson, editor, *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1037–1042. IEEE, IEEE Computer Society Press, June 1995.

[205] G. Toscani and O.D. Faugeras. Structure and motion from two noisy perspective views. In *Proceedings of* IEEE *International Conference on Robotics and Automation, Raleigh, North Carolina, USA*, pages 221–227. IEEE Computer Society Press, March 1987.

[206] G. Toscani and O.D. Faugeras. Structure from motion using the reconstruction & reprojection technique. In *Proc. of the* IEEE *Computer Society Workshop on Computer Vision*, pages 345–348. IEEE Computer Society Press, November 1987.

[207] B. Triggs. The geometry of projective reconstruction i: Matching constraints and the joint image. *International Journal of Computer Vision*, 1995. submitted.

[208] B. Triggs. Matching constraints and the joint image. In E. Grimson, editor, *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 338–343. IEEE, IEEE Computer Society Press, June 1995.

[209] B. Triggs. Factorization methods for projective structure and motion. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 845–851, 1996.

[210] 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.

[211] H.P. Trivedi. Estimation of stereo and motion parameters using a variational principle. *Image and Vision Computing*, 5(2):181–183, May 1987.

[212] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 6(1):13–27, January 1984.

[213] L. van Gool, T. Moons, M. Proesmans, and M. van Diest. Affine reconstruction from perspective image pairs obtained by a translating camera. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, volume I, pages 290–294. Proceedings of IAPR Workshop on Computer Vision, IEEE Computer Society Press, October 1994.

[214] T. Viéville. Auto-calibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 12, 1994.

[215] T. Viéville and D. Lingrand. Using singular displacements for uncalibrated monocular visual systems. Rapport de recherche 2678, INRIA, October 1995. also published in ECCV 96.

[216] T. Viéville and D. Lingrand. Using singular displacements for uncalibrated monocular visual systems. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 207–216. Springer-Verlag, April 1996.

[217] T. Viéville and Q.T. Luong. Computing motion and structure in image sequences without calibration. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pages 420–425, Jerusalem, October 1994.

[218] T. Viéville and Q.T. Luong. Motion of points and lines in the uncalibrated case. Technical report, INRIA, 1994.

[219] D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 675–682. IEEE, May 1993.

[220] J. Weng, N. Ahuja, and T.S. Huang. Closed-form solution + maximum likelihood: A robust approach to motion and structure estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, California, USA*, pages 381–386, June 1988.

[221] J. Weng, N. Ahuja, and T.S. Huang. Optimal motion and structure estimation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 15(9):864–884, September 1993.

[222] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accurate evaluation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, October 1992.

[223] J. Weng, T.S. Huang, and N. Ahuja. Error analysis of motion parameter estimation from image sequences. In *Proceedings of the 1st International Conference on Computer Vision, London, England*, pages 703–707. IEEE Computer Society Press, June 1987.

[224] J. Weng, T.S. Huang, and N. Ahuja. Motion and structure from two perspective views: Algorithms, error analysis and error estimation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(5):451–476, May 1989.

[225] T. Wiberg. Computation of principal components when data are missing. In J. Gordesch and P. Naeve, editors, *Proc. 2nd Symposium on Computational Statistics, Berlin, Germany*, pages 229–236, 1976.

[226] C. Wiles and M. Brady. Ground plane motion camera models. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 238–247. Springer-Verlag, April 1996.

[227] R.G. Willson and S.A. Shafer. A perspective projection camera model for zoom lenses. In *Proc. Second Conference on Optical 3D Measurement Techniques, Zürich, Switzerland*, October 1993.

[228] B.P. Wrobel. Minimum solutions for orientation. In *Proc. of the Workshop on Calibration and Orientation of Cameras in Computer Vision, Washington D.C., USA*. Springer-Verlag, August 1992.

[229] Y. Yakimovsky and R. Cunningham. A system for extracting three-dimensional measurements from a stereo pair of TV cameras. *Computer Graphics and Image Processing*, 7(2):195–210, 1978.

[230] Y. Yasumoto and G. Medioni. Robust estimation of three-dimensional motion parameters from a sequence of image frames using regularization. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(4):464–471, July 1986.

[231] G.S.J. Young and R. Chellappa. Statistical analysis of inherent ambiguities in recovering 3D motion from a noisy flow field. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 14(10):995–1013, October 1992.

[232] C. Zeller. *Calibration projective, affine et euclidienne en vision par ordinateur et application à la perception tridimensionnelle*. Thèse de doctorat, École Polytechnique, February 1996.

[233] C. Zeller and O. Faugeras. Projective, affine and metric measurements from video sequences. In *Proc. of the International Symposium on Optical Science, Engineering and Instrumentation*, San Diego, CA, July 1995.

[234] C. Zeller and O. Faugeras. Camera self-calibration from video sequences: the kruppa equations revisited. Rapport de recherche 2793, INRIA, February 1996.

[235] Z. Zeng and X. Wang. A general solution of a closed-form space resection. *Photogrammetric Engineering and Remote Sensing*, 58(3):327–338, March 1992.

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

[237] Z. Zhang. On the epipolar geometry between two images with lens distortion. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 407–411. IEEE Computer Society Press, August 1996.

[238] Z. Zhang, Q.T. Luong, and O. Faugeras. Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction. Rapport de recherche, INRIA, October 1993.

[239] Z. Zhang, Q.T. Luong, and O. Faugeras. Self-calibration of an uncalibrated stereo rig from one unknown motion. In E. Hancock, editor, *Proceedings of the fifth British Machine Vision Conference, York, England*, pages 499–508. British Machine Vision Association, September 1994.

[240] X. Zhuang, T.S. Huang, and R.M. Haralick. Two-view motion analysis: A unified algorithm. *Journal of the Optical Society of America A*, 3(9):1492–1500, September 1986.

[241] H. Ziemann and S.F. El-Hakim. System calibration versus self-calibration. *International Archives of Photogrammetry*, 24(1):117–122, 1982.

[242] 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.

A.

Triangulation

A.1 Triangulation pour des mouvements particuliers

A.1.1 Translation pure

Comme Armstrong l'a observé [7], le polynôme $r_1(t)$ (défini par l'équation 4.7 sur la page 52) est réduit à un polynôme de degré 2, une de ses deux racines étant le minimum global de la fonction de coût $s_1(t)$. Nous redéveloppons ce résultat, mais en distinguant les cas d'épipôles finis et infinis.

A.1.1.1 Épipôles finis

Comme dans la méthode de base décrite dans la section 4.5.2, nous effectuons des transformations rigides dans les images pour simplifier le problème. La différence est qu'ici, nous effectuons la même transformation dans les deux images : une translation qui projette $\mathbf{q}_1$ à l'origine, suivie d'une rotation autour de l'origine qui ramène les épipôles (qui sont les mêmes) sur les axes des u. Le point $\mathbf{q}_2$ est transformé par ces transformations sur un point $\mathbf{q}'$ qui n'est généralement pas l'origine. La matrice fondamentale transformée a la forme simple suivante :

$$\mathsf{F} \sim \begin{pmatrix} 0 & f & 0 \\ -f & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

et les épipôles ont les coordonnées $(1, 0, f)^\mathsf{T}$.

Nous paramétrons les droites épipolaires comme décrit dans la section 4.5.2 par les points sur l'axe des v, $(0, t, 1)^\mathsf{T}$. Dans la suite nous regardons les fonctions de coût pour les deux méthodes POLY et POLY-ABS.

Poly. La fonction de coût est :

$$s_1(t) = \frac{t^2 + (t(fq'_1 - 1) + q'_2)^2}{1 + f^2t^2} \quad .$$

Le numérateur de la dérivée première de $s_1(t)$, dont nous cherchons les racines, est un polynôme quadratique :

$$r_1(t) = f^2q'_2(1 - fq'_1)t^2 + (2 - 2fq'_1 + f^2({q'_1}^2 - {q'_2}^2))t + q'_2(fq'_1 - 1) \quad .$$

Donc, au lieu de déterminer les racines d'un polynôme de degré 6, il suffit de résoudre une équation quadratique.

Poly-Abs. La fonction de coût est :

$$s_2(t) = \frac{|t| + |t(fq'_1 - 1) + q'_2|}{\sqrt{1 + f^2t^2}}$$

Pour $r_2(t)$, nous obtenons le polynôme quadratique :

$$r_2(t) = -f^4{q'_2}^2t^2 + 2f^2q'_2(fq'_1 - 1)t + fq'_1(2 - fq'_1)$$

qui nous donne les deux solutions simples $\frac{q'_1}{fq'_2}$ et $\frac{fq'_1-2}{f^2q'_2}$.

A.1.1.2 Épipôles à l'infini

Ce cas correspond à une translation latérale, c'est-à-dire parallèle au plan image. Dans ce cas, f s'annule et la matrice fondamentale après les transformations rigides est réduite à la forme suivante :

$$\mathsf{F} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \quad .$$

Poly. La fonction de coût et le numérateur de sa dérivée première sont alors :

$$s_1(t) = t^2 + (q'_2 - t)^2 \qquad r_1(t) = 2t - q'_2$$

ce qui nous fournit directement le minimum global de $s_1(t)$: $q'_2/2$ (la dérivée seconde est constante et positive ce qui garantie qu'il s'agit bien d'un minimum).

Poly-Abs. La fonction de coût est :

$$s_2(t) = |t| + |t - q'_2| \quad .$$

Les deux termes de la fonction de coût sont des valeurs absolues de fonctions linéaires en t. Chacun des deux termes a donc une seule racine et le minimum global de la fonction de coût est déterminé ainsi : soient m_1 et m_2 les racines des deux fonctions linéaires. Le minimum global de $s_2(t)$ est celle des racines m_1 et m_2 qui appartient à la fonction linéaire de plus grande pente. Si les deux pentes sont égales, toutes les valeurs entre m_1 et m_2 sont des minima globaux. Pour le cas présent, les deux racines sont $t = 0$ et $t = q'_2$. Les pentes absolues des deux fonctions linéaires sont toutes deux égales à 1, donc toutes les valeurs entre 0 et q'_2 sont des minima globaux de la fonction de coût.

A.1.2 Zoom

Nous allons traiter quelques-uns des différents cas évoqués dans la section 2.3.4.3. Si la translation du centre optique s'effectue dans le plan focal, la matrice fondamentale a la même forme que celle d'un couple de caméras affines. Les simplifications pour l'algorithme de triangulation sont donc les mêmes que pour les caméras affines (voir la section 4.6.1). Sinon, si l'on considère des pixels non rectangulaires, les formes particulières de la matrice fondamentale ne donnent pas lieu à une réduction du degré des fonctions de coût $s_1(t)$ et $s_2(t)$.

Un cas très simple est celui où le centre optique se déplace le long de l'axe optique, le point principal reste fixe et l'angle entre les axes des pixels est de 90° : la matrice fondamentale a la même forme que celle due à une translation pure d'une caméra avec des paramètres intrinsèques constants. L'algorithme de triangulation se simplifie de la manière décrite dans la section A.1.1.

Deux autres cas sont traités dans la suite.

A.1.2.1 Pixels rectangulaires

Après l'application des transformations rigides qui amènent les points image correspondants à l'origine et les épipôles sur les axes des u, F a la forme :

$$\begin{pmatrix} b & c & -be \\ -c & b & ce \\ bd & cd & -bde \end{pmatrix} \quad .$$

Poly. La fonction de coût est :

$$s_1(t) = \frac{t^2e^2}{t^2+e^2} + \frac{d^2(ct-be)^2}{(ct-be)^2+(bt+ce)^2}$$

et le polynôme $r_1(t)$ devient :

$$r_1(t) = e(t^2+e^2)^2(b^2+c^2)(bcd^2t^2 + (c^2d^2e + c^2e^3 + b^2e^3 - b^2d^2e)t - bcd^2e^2) \quad .$$

C'est un polynôme de degré 6, mais sa factorisation montre que 4 des racines sont complexes ($b^2 + c^2 \neq 0$, sinon F serait la matrice $0_{3\times3}$; et $e \neq 0$, sinon l'épipôle serait à l'origine ce qui peut être exclu puisqu'il doit être différent du point image $\mathbf{q}_1$).

Le problème se réduit alors à la résolution d'une équation quadratique :

$$bcd^2t^2 + (c^2d^2e + c^2e^3 + b^2e^3 - b^2d^2e)t - bcd^2e^2 = 0 \quad .$$

Poly-Abs. La fonction de coût est :

$$s_2(t) = |\frac{et}{\sqrt{e^2+t^2}}| + |\frac{d(ct-be)}{\sqrt{(b^2+c^2)(t^2+e^2)}}| \quad .$$

Le polynôme $r_2(t)$ est de degré 8, mais comme $r_1(t)$ il peut être factorisé, ainsi permettant de voir que seules deux racines réelles peuvent exister :

$$r_2(t) = -e^2(t^2+e^2)^3(b^2+c^2)^2(b^2d^2t^2 + 2bcd^2et + c^2d^2e - b^2e^4 - c^2e^4) \quad .$$

Le problème se réduit alors à la résolution d'une équation quadratique :

$$b^2d^2t^2 + 2bcd^2et + c^2d^2e - b^2e^4 - c^2e^4 = 0 \quad .$$

A.1.2.2 Pixels rectangulaires et translation du centre optique dans le plan focal

Comme pour le cas d'une translation pure, les épipôles sont les mêmes dans les deux images. Ceci nous permet d'effectuer les mêmes transformations rigides dans les deux images (voir la section A.1.1). La matrice fondamentale transformée a la forme simple :

$$\mathsf{F} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 0 & b & c \end{pmatrix}$$

et les épipôles sont les points à l'infini des axes des u : $(1,0,0)^\mathsf{T}$.

Poly. La fonction de coût est :

$$s_1(t) = t^2 + \frac{aq_2' + bt + c}{a^2}$$

et le polynôme $r_1(t)$ est une fonction linéaire en t :

$$r_1(t) = (a^2 + b^2)t + b(aq_2' + c) \quad .$$

Ceci nous fournit directement le minimum global de $s_1(t)$: $-b\frac{aq_2'+c}{a^2+b^2}$ (la dérivée seconde est constante et positive ce qui garantit qu'il s'agit bien d'un minimum).

Poly-Abs. La fonction de coût est :

$$s_2(t) = |t| + |\frac{bt + aq_2' + c}{a}| \quad .$$

Les deux termes de la fonction de coût sont des valeurs absolues de fonctions linéaires en t. La racine de l'un des deux termes est le minimum global de la somme (voir la section A.1.1.2 pour plus de détails comment choisir la bonne solution).

A.2 Triangulation pour la caméra linéaire

Dans cette section, nous considérons la reconstruction à partir de caméras linéaires et essayons de trouver un résultat similaire à celui exprimé par la méthode POLY pour le cas de deux vues planaires. Pour deux vues linéaires, rien n'est à faire : les rayons de projection de deux points dans les deux vues se coupent toujours, donc la triangulation devient triviale. Cela est aussi une explication simple pour le fait qu'il n'existe pas de contrainte d'appariement pour deux vues linéaires.

Nous allons donc considérer le cas de trois vues. La formulation du problème est analogue à celle pour deux vues planaires : soient $\mathbf{q}_1 = (u_1, 1)^\mathsf{T}, \mathbf{q}_2 = (u_2, 1)^\mathsf{T}$ et $\mathbf{q}_3 = (u_3, 1)^\mathsf{T}$ des points correspondants mesurés dans les trois vues. Nous cherchons des points $\widehat{\mathbf{q}}_1'$, $\widehat{\mathbf{q}}_2'$ et $\widehat{\mathbf{q}}_3'$, tels que :

$$\begin{aligned} \{\widehat{\mathbf{q}}_1', \widehat{\mathbf{q}}_2', \widehat{\mathbf{q}}_3'\} &= \text{argmin}(d(\mathbf{q}_1, \mathbf{q}_1')^2 + d(\mathbf{q}_2, \mathbf{q}_2')^2) + d(\mathbf{q}_3, \mathbf{q}_3')^2) \\ \text{sous la contrainte} \quad & \textstyle\sum_{i,j,k=1}^2 \mathsf{G}_{ijk} q_{1_i}' q_{2_j}' q_{3_k}' = 0 \quad , \end{aligned} \tag{A.1}$$

où G est le tenseur trifocal adéquat (voir 2.4.1).

Sous l'hypothèse d'un bruit gaussien dans la position des points image, ces points sont le plus vraisemblablement les « vraies » correspondances des points image. Après avoir déterminé $\widehat{\mathbf{q}}'_1, \widehat{\mathbf{q}}'_2$ et $\widehat{\mathbf{q}}'_3$, le point 3D $\widehat{\mathbf{Q}}$ peut être calculé par n'importe quelle méthode de triangulation puisque les rayons de projection considérés s'intersectent précisément.

Contrairement au cas de deux vues planaires, les points cherchés ne peuvent pas être paramétrés par un seul paramètre ; il en faut deux. Nous choisissons deux paramètres v_1 et v_2 tels que :

$$\widehat{\mathbf{q}}'_1 = \begin{pmatrix} v_1 \\ 1 \end{pmatrix} \qquad \widehat{\mathbf{q}}'_2 = \begin{pmatrix} v_2 \\ 1 \end{pmatrix} \quad .$$

Ces deux points fixent la position du troisième par la relation trifocale. Concrètement, le troisième point est donné par :

$$\widehat{\mathbf{q}}'_3 = \begin{pmatrix} -\mathsf{G}_{112}v_1v_2 - \mathsf{G}_{122}v_1 - \mathsf{G}_{212}v_2 - \mathsf{G}_{222} \\ \mathsf{G}_{111}v_1v_2 + \mathsf{G}_{121}v_1 + \mathsf{G}_{211}v_2 + \mathsf{G}_{221} \end{pmatrix} \quad .$$

Comme fonction de coût nous obtenons :

$$s_1(v_1, v_2) = (u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 + \frac{\mathsf{G}_{112}v_1v_2 + \mathsf{G}_{122}v_1 + \mathsf{G}_{212}v_2 + \mathsf{G}_{222}}{\mathsf{G}_{111}v_1v_2 + \mathsf{G}_{121}v_1 + \mathsf{G}_{211}v_2 + \mathsf{G}_{221}})^2 \quad .$$

Afin de trouver le minimum global de s_1, nous calculons ses dérivées partielles par rapport à v_1 et v_2 et annulons leurs numérateurs. Ceci mène à deux équations de la forme :

$$\begin{aligned}
r_1(v_1, v_2) \;=\;& \mathsf{G}_{111}^3 v_1^4 v_2^3 + 3\mathsf{G}_{111}^2\mathsf{G}_{121} v_1^4 v_2^2 + 3\mathsf{G}_{111}\mathsf{G}_{121}^2 v_1^4 v_2 + \mathsf{G}_{121}^3 v_1^4 - \\
& \mathsf{G}_{111}^2(\mathsf{G}_{111}u_1 - 3\mathsf{G}_{211})v_1^3v_2^3 - 3\mathsf{G}_{111}(\mathsf{G}_{111}\mathsf{G}_{121}u_1 - \mathsf{G}_{111}\mathsf{G}_{221} - 2\mathsf{G}_{121}\mathsf{G}_{211})v_1^3v_2^2 - \\
& 3\mathsf{G}_{121}(\mathsf{G}_{111}\mathsf{G}_{121}u_1 - \mathsf{G}_{121}\mathsf{G}_{211} - 2\mathsf{G}_{111}\mathsf{G}_{221})v_1^3v_2 - \mathsf{G}_{121}^2(\mathsf{G}_{121}u_1 - 3\mathsf{G}_{221})v_1^3 - \\
& 3\mathsf{G}_{111}\mathsf{G}_{211}(\mathsf{G}_{111}u_1 - \mathsf{G}_{211})v_1^2v_2^3 + \\
& (3\mathsf{G}_{211}(\mathsf{G}_{111}\mathsf{G}_{221} + 2\mathsf{G}_{121}\mathsf{G}_{211}) - 3\mathsf{G}_{111}u_1(\mathsf{G}_{111}\mathsf{G}_{221} + 2\mathsf{G}_{121}\mathsf{G}_{211}))v_1^2v_2^2 + \\
& (3\mathsf{G}_{221}(\mathsf{G}_{111}\mathsf{G}_{221} + 2\mathsf{G}_{121}\mathsf{G}_{211}) - 3\mathsf{G}_{121}u_1(\mathsf{G}_{121}\mathsf{G}_{211} + 2\mathsf{G}_{111}\mathsf{G}_{221}))v_1^2v_2 - \\
& 3\mathsf{G}_{121}\mathsf{G}_{221}(\mathsf{G}_{121}u_1 - \mathsf{G}_{221})v_1^2 + \ldots \\
\;=\;& 0 \\
r_2(v_1, v_2) \;=\;& \mathsf{G}_{111}^3 v_2^4 v_1^3 + 3\mathsf{G}_{111}^2\mathsf{G}_{211} v_2^4 v_1^2 + 3\mathsf{G}_{111}\mathsf{G}_{211}^2 v_2^4 v_1 + \mathsf{G}_{211}^3 v_2^4 - \\
& \mathsf{G}_{111}^2(\mathsf{G}_{111}u_2 - 3\mathsf{G}_{121})v_2^3v_1^3 - 3\mathsf{G}_{111}(\mathsf{G}_{111}\mathsf{G}_{211}u_2 - \mathsf{G}_{111}\mathsf{G}_{221} - 2\mathsf{G}_{121}\mathsf{G}_{211})v_2^3v_1^2 - \\
& 3\mathsf{G}_{211}(\mathsf{G}_{111}\mathsf{G}_{211}u_2 - \mathsf{G}_{121}\mathsf{G}_{211} - 2\mathsf{G}_{111}\mathsf{G}_{221})v_2^3v_1 - \mathsf{G}_{211}^2(\mathsf{G}_{211}u_2 - 3\mathsf{G}_{221})v_2^3 - \\
& 3\mathsf{G}_{111}\mathsf{G}_{121}(\mathsf{G}_{111}u_2 - \mathsf{G}_{121})v_2^2v_1^3 + \\
& (3\mathsf{G}_{121}(\mathsf{G}_{121}\mathsf{G}_{211} + 2\mathsf{G}_{111}\mathsf{G}_{221}) - 3\mathsf{G}_{111}u_2(\mathsf{G}_{111}\mathsf{G}_{221} + 2\mathsf{G}_{121}\mathsf{G}_{211}))v_2^2v_1^2 + \\
& (3\mathsf{G}_{221}(\mathsf{G}_{111}\mathsf{G}_{221} + 2\mathsf{G}_{121}\mathsf{G}_{211}) - 3\mathsf{G}_{211}u_2(\mathsf{G}_{121}\mathsf{G}_{211} + 2\mathsf{G}_{111}\mathsf{G}_{221}))v_2^2v_1 - \\
& 3\mathsf{G}_{211}\mathsf{G}_{221}(\mathsf{G}_{211}u_2 - \mathsf{G}_{221})v_2^2 + \ldots \\
\;=\;& 0 \quad .
\end{aligned}$$

Former le résultant de ces deux polynômes aboutit en général à un polynôme de degré 49. Il est pourtant possible de simplifier la complexité du problème en appliquant certaines transformation rigides dans les images, qui seront dérivées dans la suite. L'application de transformations rigides ne change en rien le problème de minimisation, qui est basé sur des distances dans les images. Nous disposons en tout de trois degrés de liberté pour ces transformations rigides : les seules transformations rigides dans $\mathcal{P}^1$ sont les translations. Soient $\mathsf{T}_1, \mathsf{T}_2$ et T_3 des translations dans les images, avec :

$$\mathsf{T}_i = \begin{pmatrix} 1 & t_i \\ 0 & 1 \end{pmatrix} \quad .$$

Le tenseur trifocal après les translations est G' avec :

$$\begin{array}{rcl}
\mathsf{G}'_{111} & = & \mathsf{G}_{111} \\
\mathsf{G}'_{112} & = & \mathsf{G}_{112} - \mathsf{G}_{111}t_3 \\
\mathsf{G}'_{121} & = & \mathsf{G}_{121} - \mathsf{G}_{111}t_2 \\
\mathsf{G}'_{122} & = & \mathsf{G}_{122} - \mathsf{G}_{112}t_2 - (\mathsf{G}_{121} - \mathsf{G}_{111}t_2)t_3 \\
\mathsf{G}'_{211} & = & \mathsf{G}_{211} - \mathsf{G}_{111}t_1 \\
\mathsf{G}'_{212} & = & \mathsf{G}_{212} - \mathsf{G}_{112}t_1 - (\mathsf{G}_{211} - \mathsf{G}_{111}t_1)t_3 \\
\mathsf{G}'_{221} & = & \mathsf{G}_{221} - \mathsf{G}_{121}t_1 - (\mathsf{G}_{211} - \mathsf{G}_{111}t_1)t_2 \\
\mathsf{G}'_{222} & = & \mathsf{G}_{222} - \mathsf{G}_{122}t_1 - (\mathsf{G}_{212} - \mathsf{G}_{112}t_2)t_2 - (\mathsf{G}_{221} - \mathsf{G}_{121}t_1 - (\mathsf{G}_{211} - \mathsf{G}_{111}t_1)t_2)t_3 \quad .
\end{array}$$

Il est facile de montrer que jusqu'à trois coefficients du tenseur peuvent être annulés simultanément en appliquant des translations appropriées. Notons que le coefficient G_{111}, qui apparaît dans les termes d'ordre supérieur des polynômes r_1 et r_2, n'est pas influencé par les translations. À l'aide de MAPLE, nous avons tenté de réduire le degré du résultant de r_1 et r_2 en annulant simultanément trois coefficients du tenseur G. Le résultat est un résultant de degré 25, pour lequel nous connaissons directement une factorisation en un polynôme de degré 15 et un polynôme quadratique, élevé à la puissance 5. Ce résultat peut aussi être obtenu en paramétrant le problème directement par les coordonnées du point reconstruit.

L'opération la plus complexe pour la triangulation à partir de trois vues linéaires est donc la résolution d'un polynôme de degré 15. Il est douteux que ce résultat soit applicable, le degré 15 étant peut-être trop élevé pour une résolution fiable. Contrairement au cas de deux vues planaires, la minimisation de la somme des distances absolues peut être formulée par des équations de degré inférieur par rapport à la minimisation de la somme des distances carrées. Ceci est montré dans le paragraphe suivant.

A.2.1 Minimisation de la distance absolue

Comme pour le cas de deux vues planaires, nous adaptons notre méthode de manière à minimiser la somme des distances absolues au lieu de la somme des distances carrées. La fonction de coût est donc :

$$s_2(v_1, v_2) = |u_1 - v_1| + |u_2 - v_2| + |u_3 + \frac{\mathsf{G}_{112}v_1v_2 + \mathsf{G}_{122}v_1 + \mathsf{G}_{212}v_2 + \mathsf{G}_{222}}{\mathsf{G}_{111}v_1v_2 + \mathsf{G}_{121}v_1 + \mathsf{G}_{211}v_2 + \mathsf{G}_{221}}| \quad .$$

Les dérivées partielles sont :

$$\begin{array}{rcl}
\dfrac{\partial s_2}{\partial v_1}(v_1, v_2) & = & \omega_1 + \omega_2 \dfrac{(\mathsf{G}_{212}\mathsf{G}_{111} - \mathsf{G}_{112}\mathsf{G}_{211})v_2^2 + (\ldots)v_2 + \mathsf{G}_{222}\mathsf{G}_{121} - \mathsf{G}_{122}\mathsf{G}_{221}}{(\mathsf{G}_{111}v_1v_2 + \mathsf{G}_{121}v_1 + \mathsf{G}_{211}v_2 + \mathsf{G}_{221})^2} \\
\dfrac{\partial s_2}{\partial v_2}(v_1, v_2) & = & \omega_3 + \omega_4 \dfrac{(\mathsf{G}_{122}\mathsf{G}_{111} - \mathsf{G}_{112}\mathsf{G}_{121})v_1^2 + (\ldots)v_1 + \mathsf{G}_{222}\mathsf{G}_{211} - \mathsf{G}_{212}\mathsf{G}_{221}}{(\mathsf{G}_{111}v_1v_2 + \mathsf{G}_{121}v_1 + \mathsf{G}_{211}v_2 + \mathsf{G}_{221})^2}
\end{array}$$

où $\omega_1, \omega_2, \omega_3$ et ω_4 sont égaux à -1 ou 1, en fonction des signes des termes de la fonction s_2. Nous éliminons les ω_i en annulant les dérivées, en séparant les deux termes respectifs et en les élevant au carré. Ceci mène aux équations suivantes :

$$\begin{aligned} r_1(v_1,v_2) &= (\mathsf{G}_{111}^2 v_1^2 v_2^2 + 2\mathsf{G}_{111}\mathsf{G}_{121} v_1^2 v_2 + \mathsf{G}_{121}^2 v_1^2 + 2\mathsf{G}_{111}\mathsf{G}_{211} v_1 v_2^2 + \dots) \\ &\quad (\mathsf{G}_{111}^2 v_1^2 v_2^2 + 2\mathsf{G}_{111}\mathsf{G}_{121} v_1^2 v_2 + \mathsf{G}_{121}^2 v_1^2 + 2\mathsf{G}_{111}\mathsf{G}_{211} v_1 v_2^2 + \dots) \end{aligned}$$

$$\begin{aligned} r_2(v_1,v_2) &= (\mathsf{G}_{111}^2 v_1^2 v_2^2 + 2\mathsf{G}_{111}\mathsf{G}_{121} v_1^2 v_2 + (\mathsf{G}_{111}\mathsf{G}_{122} - \mathsf{G}_{112}\mathsf{G}_{121} + \mathsf{G}_{121}^2) v_1^2 + \dots) \\ &\quad (\mathsf{G}_{111}^2 v_1^2 v_2^2 + 2\mathsf{G}_{111}\mathsf{G}_{121} v_1^2 v_2 + (\mathsf{G}_{112}\mathsf{G}_{121} - \mathsf{G}_{111}\mathsf{G}_{122} + \mathsf{G}_{121}^2) v_1^2 + \dots) \end{aligned}$$

Afin de trouver les racines communes de r_1 et r_2, nous calculons les résultants des 4 combinaisons des termes dans r_1 et r_2. Chaque résultant est un polynôme de degré 8 en une variable dont nous obtenons directement une factorisation en un polynôme de degré 2 et un de degré 6. Le minimum global est déterminé en évaluant la fonction de coût à toutes les racines réelles de ces polynômes. L'opération la plus complexe pour trouver le minimum global de $s_2(v_1, v_2)$ est donc la recherche des racines d'un polynôme de degré 6 ce qui est plutôt favorable et ce qui rend l'application de cette méthode possible.

B.

Recherche des meilleurs points correspondants par homographie

B.1 Introduction

Nous traitons un problème similaire à celui de la triangulation (voir la section 4.1) : considérons les projections d'un plan Π dans deux images. Soient H l'homographie entre les deux images associée à Π et $\mathbf{q}_1 = (u_1, v_1, 1)^\mathsf{T}$ et $\mathbf{q}_2 = (u_2, v_2, 1)^\mathsf{T}$ les deux projections d'un point $\mathbf{Q}$ de Π. En présence de bruit, $\mathbf{q}_1$ et $\mathbf{q}_2$ ne vont pas se correspondre exactement par H, c'est-à-dire que $\mathsf{H}\mathbf{q}_1 \not\sim \mathbf{q}_2$. Le problème est maintenant, de trouver deux points $\widehat{\mathbf{q}}'_1$ et $\widehat{\mathbf{q}}'_2$ qui soient en correspondance et qui soient le plus proche possible des points originaux $\mathbf{q}_1$ et $\mathbf{q}_2$:

$$\begin{gathered}\{\widehat{\mathbf{q}}'_1, \widehat{\mathbf{q}}'_2\} = \text{argmin}(d(\mathbf{q}_1, \mathbf{q}'_1)^2 + d(\mathbf{q}_2, \mathbf{q}'_2)^2) \\ \text{sous la contrainte} \quad \mathsf{H}\mathbf{q}'_1 \sim \mathbf{q}'_2 \ .\end{gathered} \tag{B.1}$$

Sous l'hypothèse d'un bruit gaussien, ces points sont le plus vraisemblablement les « vraies » correspondances (non bruitées).

Ce problème a été évoqué par Richard Hartley et Andrew Zisserman [76] et nous proposons ici une solution.

B.2 Une solution du problème

Les deux points cherchés ont 4 paramètres, mais grâce à la correspondance par H, nous pouvons les paramétrer par seulement 2 coordonnées u et v : $\widehat{\mathbf{q}}'_1 = (u, v, 1)^\mathsf{T}$ et $\widehat{\mathbf{q}}'_2 = \mathsf{H}\widehat{\mathbf{q}}'_1 = \mathsf{H}(u, v, 1)^\mathsf{T}$. Le critère de minimisation est :

$$
\begin{aligned}
s_1(u,v) &= d(\mathbf{q}_1, \widehat{\mathbf{q}}'_1)^2 + d(\mathbf{q}_2, \widehat{\mathbf{q}}'_2)^2 \\
&= (u_1 - u)^2 + (v_1 - v)^2 && \text{(B.2)} \\
&\quad + (u_2 - \frac{H_{11}u + H_{12}v + H_{13}}{H_{31}u + H_{32}v + H_{33}})^2 + (v_2 - \frac{H_{21}u + H_{22}v + H_{23}}{H_{31}u + H_{32}v + H_{33}})^2 \quad . && \text{(B.3)}
\end{aligned}
$$

Nous cherchons maintenant le minimum absolu de s_1. Une condition nécessaire pour le minimum est que les dérivées partielles de s_1 s'annulent :

$$
\begin{aligned}
\frac{\partial s_1}{\partial u}(u,v) &= H_{31}^3 u^4 + \\
&\quad [3H_{32}v - u_1 H_{31} + 3H_{33}]H_{31}^2 u^3 + \\
&\quad [3H_{32}^2 v^2 + (6H_{32}H_{33} - 3u_1 H_{31} H_{32})v - 3u_1 H_{31} H_{33} + 3H_{33}^2]H_{31} u^2 + \\
&\quad [H_{32}^3 v^3 + (-3u_1 H_{31} H_{32}^2 + 3H_{32}^2 H_{33})v^2 + \ldots]u - \\
&\quad u_1 H_{32}^3 v^3 + \ldots \\
&= 0 \\
\frac{\partial s_1}{\partial v}(u,v) &= H_{32}^3 v^4 + \\
&\quad [3H_{31}u - v_1 H_{32} + 3H_{33}]H_{32}^2 v^3 + \\
&\quad [3H_{31}^2 u^2 + (6H_{31}H_{33} - 3v_1 H_{31} H_{32})u - 3v_1 H_{32} H_{33} + 3H_{33}^2]H_{32} v^2 + \\
&\quad [H_{31}^3 u^3 + (-3v_1 H_{31}^2 H_{32} + 3H_{31}^2 H_{33})u^2 + \ldots]v - \\
&\quad v_1 H_{31}^3 u^3 + \ldots \\
&= 0 \ .
\end{aligned}
$$

Le résultant de ces polynômes est un polynôme de degré 30 et avec des coefficients très complexes. Nous cherchons maintenant à réduire son degré. La latitude que nous avons dans la formulation du problème est que nous pouvons appliquer des transformations rigides dans les deux images – ceci ne change pas le critère de minimisation qui est basé sur des distances. Considérons les coefficients des termes d'ordre élevé des dérivées partielles. Les degrés des polynômes se réduiraient considérablement si l'on pouvait annuler H_{31} et H_{32}. L'application des transformations rigides T_1 et T_2 aux images donne lieu à une nouvelle homographie entre les images :

$$
\mathsf{H}' = \mathsf{T}_2 \mathsf{H} \mathsf{T}_1{}^{-1} \quad .
$$

Avec T_1 et T_2 de la forme :

$$
\mathsf{T}_1 = \begin{pmatrix} \cos\alpha_1 & -\sin\alpha_1 & t_{1_1} \\ \sin\alpha_1 & \cos\alpha_1 & t_{1_2} \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathsf{T}_2 = \begin{pmatrix} \cos\alpha_2 & -\sin\alpha_2 & t_{2_1} \\ \sin\alpha_2 & \cos\alpha_2 & t_{2_2} \\ 0 & 0 & 1 \end{pmatrix}
$$

H' devient :

$$
\mathsf{H}' = \begin{pmatrix} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cos\alpha_1 H_{31} - \sin\alpha_1 H_{32} & \sin\alpha_1 H_{31} + \cos\alpha_1 H_{32} & \cdots \end{pmatrix} \quad .
$$

Évidemment, H'_{31} ou H'_{32} peuvent être annulés en choisissant une valeur appropriée pour l'angle α_1 de la rotation dans la première image. Pourtant, en général H'_{31} et H'_{32} ne peuvent pas s'annuler simultanément. Si c'était possible, H' serait une transformation affine, et ceci est possible si et seulement si l'homographie originale H est aussi une transformation affine. Ce cas particulier est traité dans le paragraphe B.4.

Si nous nous décidons d'annuler H'_{31} en mettant $\alpha_1 = \arctan \frac{H_{31}}{H_{32}}$, la dérivée partielle par u se simplifie considérablement, et le résultant des deux dérivées partielles par rapport à la variable u est un polynôme de la forme suivante :

$$r_1(v) = (H'_{33} + H'_{32}v)^2 \ \ ({H'}^7_{32}v^8 + {H'}^6_{32}(\dots)v^7 + {H'}^5_{32}(\dots)v^6 + \dots) \ .$$

Afin de trouver tous les candidats pour le minimum global de la fonction de coût $s_1(u,v)$ il suffit donc de poursuivre la procédure suivante : on détermine les racines du polynôme $r_1(v)$ et on substitue dans les deux dérivées partielles tour après tour v par les racines de r_1 pour ensuite déterminer les racines en u. Pour chaque valeur de v, on retient les u qui sont racine commune des deux dérivées. Ensuite, le minimum global est déterminé en évaluant $s_1(u,v)$ aux valeurs retenues et en choisissant celles donnant le résultat minimal. L'opération la plus complexe est la détermination des racines d'un polynôme de degré 8.

Nous voyons que la complexité du problème a été réduite grâce à une simple rotation dans la première image. Des réductions supplémentaires sont éventuellement possibles en profitant des degrés de liberté dont on dispose avec les paramètres des transformations rigides restants : $\alpha_2, \mathbf{t}_1, \mathbf{t}_2$. Pourtant, ceci nécessiterait des transformations spécifiques pour chaque paire de points traitée et les calculs pour déterminer ces transformations deviendraient rapidement très complexes.

B.3 Minimisation de la somme des distances absolues

Comme pour le problème de la triangulation, nous examinons aussi la forme du problème lorsque nous tentons à minimiser la somme des distances absolues entre les points originaux et les points corrigés, au lieu de la somme des carrés des distances :

$$\begin{aligned} \{\widehat{\mathbf{q}}'_1, \widehat{\mathbf{q}}'_2\} &= \mathrm{argmin}(d(\mathbf{q}_1, \mathbf{q}'_1) + d(\mathbf{q}_2, \mathbf{q}'_2)) \\ \text{sous la contrainte} \quad & \mathsf{H}\mathbf{q}'_1 \sim \mathbf{q}'_2 \ . \end{aligned} \tag{B.4}$$

En appliquant la même paramétrisation que dans le paragraphe précédent, nous obtenons la fonction de coût :

$$\begin{aligned} s_2(u,v) &= \sqrt{(u_1-u)^2 + (v_1-v)^2} \\ &+ \sqrt{(u_2 - \frac{H_{11}u + H_{12}v + H_{13}}{H_{31}u + H_{32}v + H_{33}})^2 + (v_2 - \frac{H_{21}u + H_{22}v + H_{23}}{H_{31}u + H_{32}v + H_{33}})^2} \ . \end{aligned}$$

Le même développement que pour la méthode POLY-ABS dans la section 4.5.4 peut être appliqué ici. Malgré la simplification $H_{31} = 0$, nous nous retrouvons avec des dérivées partielles de degré 6 (respectivement 8) ce qui semble être trop compliqué pour que la poursuite du développement soit raisonnable.

B.4 Homographies affines

B.4.1 Minimisation de la somme des carrés des distances

Si l'homographie H est une transformation affine, c'est-à-dire si $H_{31} = H_{32} = 0$, les inconnues u et v disparaissent des dénominateurs dans la fonction de coût (B.3) et ses dérivées partielles sont linéaires en u et v :

$$\begin{aligned}
\frac{\partial s_1}{\partial u}(u,v) &= (H_{11}^2 + H_{21}^2 + H_{33}^2)u + (H_{11}H_{12} + H_{21}H_{22})v + \\
&\quad H_{11}H_{13} + H_{21}H_{23} - H_{33}(H_{33}u_1 + H_{11}u_2 + H_{21}v_2) \\
\frac{\partial s_1}{\partial v}(u,v) &= (H_{12}^2 + H_{22}^2 + H_{33}^2)v + (H_{11}H_{12} + H_{21}H_{22})u + \\
&\quad H_{12}H_{13} + H_{22}H_{23} - H_{33}(H_{33}v_1 + H_{12}u_2 + H_{22}v_2) \ .
\end{aligned}$$

Nous obtenons donc une solution directe pour $\frac{\partial s_1}{\partial u}(u,v) = \frac{\partial s_1}{\partial v}(u,v) = 0$:

$$\begin{aligned}
\hat{u} &= \frac{1}{D}(u_1H_{33}^4 + u_1H_{12}^2H_{33}^2 + u_1H_{22}^2H_{33}^2 - v_1H_{11}H_{12}H_{33}^2 - v_1H_{21}H_{22}H_{33}^2 \\
&\quad +u_2H_{11}H_{33}^3 + u_2H_{11}H_{22}^2H_{33} - u_2H_{12}H_{21}H_{22}H_{33} + v_2H_{21}H_{33}^3 + v_2H_{12}^2H_{21}H_{33} - v_2H_{11}H_{12}H_{22}H_{33} \\
&\quad +H_{12}H_{13}H_{21}H_{22} + H_{11}H_{12}H_{22}H_{23} - H_{11}H_{13}H_{22}^2 - H_{11}H_{13}H_{33}^2 - H_{21}H_{23}H_{33}^2 - H_{12}^2H_{21}H_{23}) \\
\hat{v} &= \frac{1}{D}(v_1H_{33}^4 + v_1H_{11}^2H_{33}^2 + v_1H_{21}^2H_{33}^2 - u_1H_{11}H_{12}H_{33}^2 - u_1H_{21}H_{22}H_{33}^2 \\
&\quad +v_2H_{22}H_{33}^3 + v_2H_{11}^2H_{22}H_{33} - v_2H_{11}H_{12}H_{21}H_{33} + u_2H_{12}H_{33}^3 + u_2H_{12}H_{21}^2H_{33} - u_2H_{11}H_{21}H_{22}H_{33} \\
&\quad +H_{11}H_{12}H_{21}H_{23} + H_{11}H_{13}H_{21}H_{22} - H_{12}H_{13}H_{21}^2 - H_{12}H_{13}H_{33}^2 - H_{22}H_{23}H_{33}^2 - H_{11}^2H_{22}H_{23}) \ ,
\end{aligned}$$

où le dénominateur D est donné par :

$$D = H_{33}^4 - 2H_{11}H_{12}H_{21}H_{22} + H_{11}^2H_{22}^2 + H_{11}^2H_{33}^2 + H_{12}^2H_{33}^2 + H_{12}^2H_{21}^2 + H_{21}^2H_{33}^2 + H_{22}^2H_{33}^2 \ .$$

B.4.2 Simplification

Nous pouvons simplifier davantage la solution du problème en appliquant des transformations rigides spécifiques aux images. Soit l'homographie H la composée d'une matrice A et d'un vecteur $\mathbf{b}$:

$$\mathsf{H} = \begin{pmatrix} \mathsf{A} & \mathbf{b} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \ .$$

Soit $\mathsf{A} = \mathsf{UDV}^\mathsf{T}$ une décomposition en valeurs singulières de la sous-matrice A. L'application des transformations rigides :

$$\mathsf{T}_1 = \begin{pmatrix} \mathsf{V}^\mathsf{T} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \qquad \mathsf{T}_2 = \begin{pmatrix} \mathsf{U}^\mathsf{T} & -\mathsf{U}^\mathsf{T}\mathbf{b} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}$$

aux images transforme l'homographie en la matrice diagonale H' :

$$\begin{aligned}
\mathsf{H}' &= \mathsf{T}_2\mathsf{H}\mathsf{T}_1{}^{-1} \\
&= \begin{pmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ ,
\end{aligned}$$

où D_1 et D_2 sont les valeurs singulières de A.

Dû à ces transformations, la fonction de coût et ses dérivées prennent une forme encore plus simple, et la solution pour $\hat{u}$ et $\hat{v}$ (dans les images transformées) est donnée par :

$$\hat{u} = \frac{u_1 + D_1 u_2}{D_1^2} \qquad \hat{v} = \frac{v_1 + D_2 v_2}{D_2^2} \ .$$

B.4.3 Minimisation de la somme des distances absolues

Nous traitons le problème après avoir appliqué les transformations rigides du paragraphe précédent. La détermination du minimum global revient finalement à la recherche des racines d'un polynôme de degré 16, mais qui possède une factorisation avantageuse. Notamment, il est de la forme :

$$\begin{aligned} r_2(u) \quad = \quad & (u_2 - D_1 u)^4 (u_1 - u)^4 \\ & (u^2 + \ldots u + \ldots)(u^2 + \ldots u + \ldots)(u^2 + \ldots u + \ldots)(u^2 + \ldots u + \ldots) \ . \end{aligned}$$

Le plus haut degré des polynômes à résoudre est donc de 2.

C.

Calibrage intrinsèque à partir de l'homographie infinie

Soit H_∞ l'homographie entre deux vues avec le même calibrage intrinsèque K. Sans perte de généralité, nous pouvons supposer que $\det \mathsf{H}_\infty = 1$. H_∞ se calcule à partir de K et de la rotation relative des vues, R, d'après $\mathsf{H}_\infty = \mathsf{K}\mathsf{R}\mathsf{K}^{-1}$. Il en découle que $\mathsf{R} = \mathsf{K}^{-1}\mathsf{H}_\infty\mathsf{K}$ et, grâce à l'orthogonalité de R :

$$\mathsf{R}\mathsf{R}^\mathsf{T} = \mathsf{K}^{-1}\mathsf{H}_\infty\mathsf{K}\mathsf{K}^\mathsf{T}\mathsf{H}_\infty^\mathsf{T}\mathsf{K}^{-\mathsf{T}} = \mathtt{I} \ ,$$

ce que nous pouvons transformer en :

$$\mathsf{H}_\infty\mathsf{K}\mathsf{K}^\mathsf{T}\mathsf{H}_\infty^\mathsf{T} = \mathsf{K}\mathsf{K}^\mathsf{T} \ . \tag{C.1}$$

Cette équation est linéaire en les éléments de la matrice symétrique $\mathsf{X} = \mathsf{K}\mathsf{K}^\mathsf{T}$, qui représente le dual de l'image de la conique absolue (cf. 2.7.3). Pour déterminer X, il faut au moins deux homographies infinies. Ayant trouvé X, K est calculé par une décomposition de Cholesky de X. Le fait que $K_{33} = 1$ nous donne une solution unique pour K. La décomposition de Cholesky n'est possible que si X est définie positive.

Il y a une interprétation géométrique directe à l'équation (C.1). La matrice $\mathsf{K}\mathsf{K}^\mathsf{T}$ représente le dual de l'image de la conique absolue, qui, nous le rappelons, se trouve sur le plan à l'infini. L'image de la conique absolue est fixe pour des vues possédant le même calibrage intrinsèque. Donc, le transfert de la conique $\mathsf{K}\mathsf{K}^\mathsf{T}$ d'une vue à une autre, via l'homographie infinie, résulte en $\mathsf{K}\mathsf{K}^\mathsf{T}$.

D.

Preuves pour la dérivation des séquences critiques

D.1 Caméra linéaire

D.1.1 Projection des points cycliques

PROPOSITION D.1 *La projection d'un point cyclique par une caméra linéaire reste fixe lors de mouvements rigides de la caméra.*

Preuve : *Il suffit de montrer que tout mouvement rigide dans $\mathcal{P}^2$ laisse chacun des deux points cycliques fixe. Considérons le point cyclique $(1, I, 0)^\mathsf{T}$. Il est transformé par un mouvement rigide constitué d'une translation $\mathbf{t}$ et d'une rotation d'angle α d'après*

$$\begin{aligned}
\begin{pmatrix} \cos\alpha & -\sin\alpha & t_1 \\ \sin\alpha & \cos\alpha & t_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ I \\ 0 \end{pmatrix} &= \begin{pmatrix} \cos\alpha - I\sin\alpha \\ \sin\alpha + I\cos\alpha \\ 0 \end{pmatrix} \\
&\sim \begin{pmatrix} 1 \\ \frac{\sin\alpha + I\cos\alpha}{\cos\alpha - I\sin\alpha} \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ I \\ 0 \end{pmatrix} \quad .
\end{aligned}$$

Le résultat analogue peut être obtenu pour le deuxième point cyclique $(1, -I, 0)^\mathsf{T}$. $\square$

D.1.2 Séquences critiques par rapport à un point non à l'infini

PROPOSITION D.2 *Considérons une caméra linéaire avec des paramètres intrinsèques constants se déplaçant sur un cercle Ψ. S'il existe un seul point $\mathbf{Q}_0$ sur Ψ, différent des points cycliques, dont les projections dans toutes les vues sont identiques, ceci est aussi le cas pour tous les autres points sur Ψ.*

Preuve : *Les projections des points cycliques sont fixes dans toutes les vues d'une caméra linéaire avec des paramètres intrinsèques constants. Donc, il existe trois points sur Ψ avec cette propriété – $\mathbf{Q}_0, \mathbf{I}$ et $\mathbf{J}$.*

Considérons un quatrième point $\mathbf{Q}_1$ quelconque sur Ψ. Les centres de projection de la caméra se trouvent également sur le cercle. Donc, les birapports des droites reliant les centres de projection avec les points $\mathbf{Q}_0, \mathbf{I}, \mathbf{J}$ et $\mathbf{Q}_1$ sont tous identiques (théorème de Chasles [171]). Les droites en question sont les rayons de projection de ces 4 points. Donc, le birapport des droites est égale au birapport des points image correspondants. Trois parmi les quatre points ayant des projections fixes et le birapport des quatre points image étant constant, les projections du quatrième point sont aussi identiques dans toutes les vues. □

COROLLAIRE D.1 *La condition suivante est nécessaire et suffisante pour qu'une séquence de mouvements soit critique par rapport à un point non à l'infini : la caméra se déplace sur un cercle et les projections d'un point sur le cercle qui est différent des points cycliques sont identiques dans toutes les vues.*

Preuve : *Le fait qu'une telle séquence de mouvements est critique, se prouve facilement à l'aide de la proposition D.2 : les projections de chaque point sur le cercle sont identiques, donc tous les points imaginaires sont des points cycliques potentiels. Chaque cercle contient des points imaginaires hors de la droite à l'infini. Il existent alors de faux points cycliques et la séquence est critique.*

Le réciproque a été prouvé au cours de la section 7.8. □

D.2 Caméra planaire

D.2.1 Insignifiance des paramètres intrinsèques

Dans la section 7.9 nous prétendons que la question si une séquence de mouvements donnée est critique est indépendante des paramètres intrinsèques effectifs de la caméra. Ceci peut être prouvé facilement : soit ω_i la projection d'une conique 3D Ω dans la vue i. Si Ω est une conique absolue potentielle, nous avons :

$$\omega_i \sim \omega_j \ , \ 1 \leq i,j \leq m \ .$$

Un changement de paramètres intrinsèques correspond à la pré-multiplication des matrices de projection par une matrice $\mathsf{M}_{3\times 3}$ non singulière. Les coniques dans les images sont transformées de la manière suivante :

$$\omega_i' \sim \mathsf{M}^{-\mathsf{T}} \omega_i \mathsf{M}^{-1} \ .$$

Évidemment, les coniques transformées sont également identiques :

$$\omega_i' \sim \omega_j' \ , \ 1 \leq i,j \leq m \ .$$

Donc, les paramètres intrinsèques de la caméra n'ont pas d'importance pour la question si une conique est une conique absolue potentielle. En outre, on peut choisir la transformation M de manière à obtenir un calibrage intrinsèque représenté par la matrice identité. Dans ce cas, les matrices de projection ont la forme :

$$\mathsf{P}_i \sim \mathsf{R}_i(\ \mathtt{I} \ | - \mathbf{t}_i) \ .$$

D.2.2 La matrice d'un cône à partir du vertex et d'une conique

PROPOSITION D.3 *Soit Φ une conique sur le plan support $\Pi \equiv (Z = 0)$, représentée par la matrice :*

$$\begin{pmatrix} a & b & d \\ b & c & e \\ d & e & f \end{pmatrix} \ .$$

Un cône K avec vertex $\mathbf{V} = (X, Y, Z, 1)^\mathsf{T}$ non sur Π, et qui contient Φ, peut être représenté par la matrice :

$$\mathsf{K} = \begin{pmatrix} a & b & -\frac{aX+bY+d}{Z} & d \\ b & c & -\frac{bX+cY+e}{Z} & e \\ -\frac{aX+bY+d}{Z} & -\frac{bX+cY+e}{Z} & \frac{aX^2+cY^2+2bXY+2dX+2eY+f}{Z^2} & -\frac{dX+eY+f}{Z} \\ d & e & -\frac{dX+eY+f}{Z} & f \end{pmatrix} \ .$$

Preuve : *Nous devons prouver que* (i) $\mathbf{V}$ *est un point singulier de* K *et que* (ii) K *contient* Φ *(un cône est défini par son vertex et une conique dont le plan support ne contient pas le vertex).*

(i) *La multiplication* $\mathsf{K}\mathbf{V}$ *résulte en le vecteur nul, donc* $\mathbf{V}$ *est un point singulier de* K.

(ii) *L'intersection de* K *avec le plan* $Z = 0$ *contient exactement les points* $\mathbf{Q} = (X', Y', 0, 1)^\mathsf{T}$ *avec* $\mathbf{Q}^\mathsf{T}\mathsf{K}\mathbf{Q} = 0$. *Cette dernière équation devient alors :*

$$aX'^2 + cY'^2 + 2bX'Y' + 2dX' + 2eY' + f = 0$$

ce qui n'est rien d'autre que l'équation de la conique Φ *dans le plan* $Z = 0$. *Alors,* K *contient* Φ.

□

D.2.3 Transformation rigide de cônes

PROPOSITION D.4 *Deux cônes Λ_1 et Λ_2 sont égaux à une transformation rigide près, si et seulement si leurs premières sous-matrices 3×3 ont les mêmes valeurs propres, à un facteur près.*

Preuve : *Soient les cônes données par :*

$$\Lambda_1 \sim \begin{pmatrix} \mathsf{M}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^\mathsf{T} & c_1 \end{pmatrix}$$

$$\Lambda_2 \sim \begin{pmatrix} \mathsf{M}_2 & \mathbf{b}_2 \\ \mathbf{b}_2^\mathsf{T} & c_2 \end{pmatrix} \ .$$

Nous pouvons déplacer les cônes par des translations sans changer le problème. Nous choisissons d'effectuer des translations telles que tous les deux cônes ont comme vertex l'origine. Les translations n'affectent pas les premières sous-matrices M_1 *et* M_2 *et la condition que le point* $(0,0,0,1)^{\mathsf{T}}$ *est un point singulier fait que les cônes transformés sont représentés par :*

$$\begin{aligned} \Lambda_1' &= \begin{pmatrix} \mathsf{M}_1 & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{pmatrix} \\ \Lambda_2' &= \begin{pmatrix} \mathsf{M}_2 & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{pmatrix} \quad . \end{aligned}$$

Ces cônes sont égaux à une transformation rigide près, si et seulement s'il existe une matrice orthogonale R *telle que :*

$$\begin{pmatrix} \mathsf{R}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix} \Lambda_1' \begin{pmatrix} \mathsf{R} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix} \sim \Lambda_2' \ .$$

Il en découle que :

$$\mathsf{R}^{\mathsf{T}}\mathsf{M}_1\mathsf{R} \sim \mathsf{M}_2 \ .$$

Ceci veut dire que M_1 *et* M_2 *sont conjuguées, à un facteur près, et qu'elles ont les mêmes valeurs propres, à un facteur près. Réciproquement, l'égalité (à un facteur près) des valeurs propres de deux matrices symétriques implique l'existence d'une matrice orthogonale* R *telle que l'équation ci-dessus soit satisfaite.* □

D.2.4 Sur le cas 7.9

Soient Ω un cercle et Λ_1 un cône elliptique comme dans la section 7.9.1.2. Dans la suite, nous déterminons tous les vertexes $\mathbf{t}_2$ tels que les cônes engendrés par $\mathbf{t}_2$ et Ω soient égaux à Λ_1 à une transformation rigide près.

Nous dérivons d'abord une contrainte sur la distance de $\mathbf{t}_2$ du plan support Π de Ω. Soit Λ_2 le cône engendré par un vertex $\mathbf{t}_2$ et Ω. Supposons qu'il existe une transformation rigide T entre Λ_1 et Λ_2. Ceci veut dire que si l'on applique l'inverse de T au plan support Π, sans transformer Λ_1, on obtient un plan Π_2 qui coupe Λ_1 en un cercle de la même dimension que Ω. Ceci n'est possible que si Π_2 a la même distance du vertex de Λ_1 que Π. L'application de T sur Λ_1 résulte en Λ_2 dont le vertex $\mathbf{t}_2$ est alors à la même distance de Π que le vertex de Λ_1.

Avec les notations de la section 7.9.1.2, ceci veut dire que $t_{23} = \pm t_{13}$. Les équations (7.12) se simplifient alors :

$$\begin{aligned} \rho(a^2(t_{11}^2+t_{13}^2)+b^2(t_{12}^2+t_{13}^2)+1) &= a^2(t_{21}^2+t_{13}^2)+b^2(t_{22}^2+t_{13}^2)+1 \\ \rho^2 t_{13}^2(a^2b^2(t_{11}^2+t_{12}^2+t_{13}^2)+a^2+b^2) &= t_{13}^2(a^2b^2(t_{21}^2+t_{22}^2+t_{13}^2)+a^2+b^2) \\ \rho^3 &= 1 \ . \end{aligned}$$

Puisque $\rho = 1$, les deux premières équations deviennent :

$$\begin{array}{rcl} a^2(t_{11}^2 + t_{13}^2) + b^2(t_{12}^2 + t_{13}^2) & = & a^2(t_{21}^2 + t_{13}^2) + b^2(t_{22}^2 + t_{13}^2) \\ t_{11}^2 + t_{12}^2 & = & t_{21}^2 + t_{22}^2 \quad . \end{array}$$

Puisque Ω est un cercle, nous avons $a = b$, et les deux équations sont identiques :

$$t_{11}^2 + t_{12}^2 = t_{21}^2 + t_{22}^2 \quad .$$

Ceci est l'équation d'un cercle en les coordonnées X et Y. Il en découle que $\mathbf{t}_2$ se trouve sur deux cercles dont les plans support sont parallèles et équidistants à Π. Les cercles ont la même dimension, mais elle est différente de celle de Ω (Ω est un cercle imaginaire) et la projection orthogonale de leurs centres sur Π coïncide avec le centre de Ω (cf. la figure 7.11).

D.2.5 Une translation pure

PROPOSITION D.5 *Soit Ω une conique 3D propre et imaginaire hors de Π_∞. Considérons une caméra se déplaçant par une translation pure : soient $\mathbf{t}_1$ et $\mathbf{t}_2$ les positions du centre de projection, avant et après la translation, toutes deux hors du plan support de Ω.*

Les projections de Ω dans les deux vues sont égales si et seulement si le point milieu de $\mathbf{t}_1$ et $\mathbf{t}_2$ coïncide avec le centre de Ω.

Preuve : *Sans perte de généralité, nous pouvons suppose que Ω se trouve sur le plan $Z = 0$ et qu'elle est représentée, dans ce plan, par la matrice :*

$$\Omega \sim \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad .$$

Soit R *la partie rotationnelle de l'orientation de la caméra. Les matrices de projection avant et après la translation sont alors (nous ignorons les paramètres intrinsèques, conformément au paragraphe D.2.1) :*

$$\begin{array}{rcl} \mathsf{P}_1 & \sim & \mathsf{R}(\ \mathtt{I} \ \mid -\mathbf{t}_1) \\ \mathsf{P}_2 & \sim & \mathsf{R}(\ \mathtt{I} \ \mid -\mathbf{t}_2) \quad . \end{array}$$

Leur restriction sur les points du plan $Z = 0$ est :

$$\begin{array}{rcl} \mathsf{P}'_1 & \sim & \mathsf{R} \begin{pmatrix} 1 & 0 & -t_{11} \\ 0 & 1 & -t_{12} \\ 0 & 0 & -t_{13} \end{pmatrix} \\ \mathsf{P}'_2 & \sim & \mathsf{R} \begin{pmatrix} 1 & 0 & -t_{21} \\ 0 & 1 & -t_{22} \\ 0 & 0 & -t_{23} \end{pmatrix} \quad . \end{array}$$

Les projections de Ω dans les deux vues sont données par :

$$\begin{array}{rcl} \omega_1 & \sim & (\mathsf{P}'_1)^{-\mathsf{T}}\, \Omega\, (\mathsf{P}'_1)^{-1} \\ \omega_2 & \sim & (\mathsf{P}'_2)^{-\mathsf{T}}\, \Omega\, (\mathsf{P}'_2)^{-1} \quad . \end{array}$$

L'égalité de ω_1 et ω_2 revient à la contrainte suivante :

$$\begin{pmatrix} a & 0 & a(t_{11}\frac{t_{23}}{t_{13}} - t_{21}) \\ 0 & b & b(t_{12}\frac{t_{23}}{t_{13}} - t_{22}) \\ a(t_{11}\frac{t_{23}}{t_{13}} - t_{21}) & b(t_{12}\frac{t_{23}}{t_{13}} - t_{22}) & a(t_{11}\frac{t_{23}}{t_{13}} - t_{21})^2 + b(t_{12}\frac{t_{23}}{t_{13}} - t_{22})^2 + \frac{t_{23}^2}{t_{13}^2} \end{pmatrix} \sim \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 1 \end{pmatrix} \; .$$

Puisque a et b sont non nuls (sinon, Ω ne serait pas propre), on obtient des conditions nécessaires et suffisantes pour l'égalité de ω_1 et ω_2 :

$$\begin{aligned} t_{11}\frac{t_{23}}{t_{13}} - t_{21} &= 0 \\ t_{12}\frac{t_{23}}{t_{13}} - t_{22} &= 0 \\ \frac{t_{23}^2}{t_{13}^2} &= 1 \; . \end{aligned}$$

Comme nous pouvons exclure que $t_{13} \neq 0$ et $t_{23} \neq 0$ ($\mathbf{t}_1$ et $\mathbf{t}_2$ se trouvent en dehors du plan $Z = 0$), ces conditions se traduisent finalement par :

$$\mathbf{t}_2 = -\mathbf{t}_1 \; .$$

Ceci ne signifie rien d'autre que le point milieu de $\mathbf{t}_1$ et $\mathbf{t}_2$ est, en coordonnées homogènes, le point $(0,0,0,1)^\mathsf{T}$. La preuve s'achève en remarquant que ce point est le centre de Ω. □

COROLLAIRE D.2 *Considérons deux vues d'une caméra se déplaçant par une translation pure. Les coniques absolues potentielles pour ce mouvement sont exactement :*

- ▷ *toutes les coniques propres et imaginaires du plan à l'infini ;*
- ▷ *et toutes les coniques propres et imaginaires dont le centre se trouve au milieu des deux centres de projection.*

Il en découle qu'il existe un seul plan dont toutes les coniques propres et imaginaires sont des coniques absolues potentielles : le plan à l'infini. Sur tous les plans passant par le point milieu des centres de projection, il existent une famille de coniques de trois dimensions, dont les coniques propres et imaginaires sont des coniques absolues potentielles.

D.3 Homographies avec des coniques fixes et contrainte du module pour la reconstruction affine

PROPOSITION D.6 *Une homographie H associée à un plan, qui laisse une conique propre globalement invariante, a des valeurs propres $\{\lambda_1, \lambda_1\lambda_2, \frac{\lambda_1}{\lambda_2}\}$, pour un λ_1 réel. Si deux des valeurs propres sont des nombres complexes, les valeurs propres sont alors $\{\lambda_1, \lambda_1 e^{I\Theta}, \lambda_1 e^{-I\Theta}\}$, pour un Θ réel.*

Preuve : *Soit Φ une conique qui est globalement invariante à H, c'est-à-dire que $\Phi \sim \mathsf{H}^{-\mathsf{T}}\Phi\mathsf{H}^{-1}$. Ceci est équivalent à :*

$$\Phi \sim \mathsf{H}^\mathsf{T}\Phi\mathsf{H} \; . \tag{D.1}$$

Sans perte de généralité, nous pouvons supposer que $\det \mathsf{H} = 1$ *ce qui nous permet d'enlever le facteur multiplicatif inconnu de l'équation (D.1) :*

$$\Phi = \mathsf{H}^\mathsf{T} \Phi \mathsf{H} \ .$$

Soient τ_i *les valeurs propres de* H *et* $\mathbf{v}_i$ *des vecteurs propres correspondants, c'est-à-dire que* $\mathsf{H}\mathbf{v}_i = \tau_i \mathbf{v}_i$. *Alors :*

$$\begin{array}{rcl} \Phi \mathbf{v}_i & = & \mathsf{H}^\mathsf{T} \Phi \mathsf{H} \mathbf{v}_i \\ \Leftrightarrow \ \dfrac{1}{\tau_i} \Phi \mathbf{v}_i & = & \mathsf{H}^\mathsf{T} \Phi \mathbf{v}_i \ , \end{array}$$

c'est-à-dire que $\Phi \mathbf{v}_i$ *est un vecteur propre de* H^T *associé à la valeur propre* $\frac{1}{\tau_i}$. *Puisque* H *et* H^T *doivent avoir les mêmes valeurs propres, nous avons l'égalité suivante entre les spectres de* H *et* H^T *:*

$$\{\tau_1, \tau_2, \tau_3\} = \{\frac{1}{\tau_1}, \frac{1}{\tau_2}, \frac{1}{\tau_3}\}$$

(il s'agit d'une égalité d'ensembles non ordonnés). Il est facile de voir que l'une des valeurs propres doit être égale à 1, *disons* τ_1. *Quant aux deux autres, elles doivent être le réciproque l'une de l'autre :* $\tau_3 = \frac{1}{\tau_2}$. *Si nous omettons la condition que* $\det \mathsf{H} = 1$, *les valeurs propres de* H *deviennent* $\{\lambda_1, \lambda_1\lambda_2, \frac{\lambda_1}{\lambda_2}\}$, *où* $\lambda_1 = \sqrt[3]{\det \mathsf{H}}$ *est un nombre réel et* $\lambda_2 = \tau_2$.

Si τ_2 *(et donc aussi* τ_3*) ne sont pas des nombres réels, ils sont complexes conjugués et d'après* $\tau_2 = \frac{1}{\tau_3} = \frac{1}{\bar{\tau}_2}$ *on peut les écrire comme* $\tau_2 = e^{I\Theta}$ *et* $\tau_3 = e^{-I\Theta}$, *pour un* Θ *réel.* □

PROPOSITION D.7 *Une homographie* H *associée à un plan, qui laisse une conique propre et* imaginaire *globalement invariante, a des valeurs propres du même module, c'est-à-dire des valeurs propres* $\{\lambda, \lambda e^{I\Theta}, \lambda e^{-I\Theta}\}$, *pour des* λ *et* Θ *réels.*

Preuve : *Soit* Φ *une conique imaginaire propre qui est globalement invariante à l'homographie* H, *c'est-à-dire que* $\mathsf{H}^\mathsf{T}\Phi\mathsf{H} = \Phi$. *Sans perte de généralité, nous pouvons supposer que* $\det \mathsf{H} = 1$. *De la proposition D.6 il découle que* H *a des valeurs propres* $\{1, \tau, \frac{1}{\tau}\}$.

Si $\tau = 1$, *alors* 1 *est une triple valeur propre de* H *et la proposition est correct. Maintenant, nous allons prouver par contradiction que si* $\tau \neq 1$ *alors* τ *ne peut pas être réel. Supposons que* τ *est réel et que* $\tau \neq 1$. *Dans ce cas, toutes les valeurs propres de* H *sont différentes et il existe une décomposition de* H *de la forme* $\mathsf{H} = \mathsf{V}\Sigma\mathsf{V}^{-1}$, *avec* $\Sigma = \mathrm{diag}(1, \tau, \frac{1}{\tau})$ *la matrice diagonale des valeurs propres de* H *et* V *une matrice de vecteurs propres correspondant. Nous développons cette relation :*

$$\begin{array}{rcl} \mathsf{H}^\mathsf{T}\Phi\mathsf{H} & = & \Phi \\ \mathsf{V}^{-\mathsf{T}}\Sigma\mathsf{V}^\mathsf{T}\Phi\mathsf{V}\Sigma\mathsf{V}^{-1} & = & \Phi \\ \Sigma\mathsf{V}^\mathsf{T}\Phi\mathsf{V} & = & \mathsf{V}^\mathsf{T}\Phi\mathsf{V}\Sigma^{-1} \ . \end{array}$$

Soit $\mathsf{A} = (a)_{ij} = \mathsf{V}^\mathsf{T}\Phi\mathsf{V}$ *:*

$$\begin{array}{rcl} \Sigma\mathsf{A} & = & \mathsf{A}\Sigma^{-1} \\ \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ \tau a_{21} & \tau a_{22} & \tau a_{23} \\ \frac{a_{31}}{\tau} & \frac{a_{32}}{\tau} & \frac{a_{33}}{\tau} \end{pmatrix} & = & \begin{pmatrix} a_{11} & \frac{a_{12}}{\tau} & \tau a_{13} \\ a_{21} & \frac{a_{22}}{\tau} & \tau a_{23} \\ a_{31} & \frac{a_{32}}{\tau} & \tau a_{33} \end{pmatrix} \ . \end{array}$$

Puisque $\tau \neq 1$, *cette équation implique que :*

$$a_{12} = a_{13} = a_{21} = a_{22} = a_{31} = a_{33} = 0 \ .$$

Alors, $\mathsf{A} = \mathsf{V}^\mathsf{T}\Phi\mathsf{V}$ *est de la forme :*

$$\mathsf{V}^\mathsf{T}\Phi\mathsf{V} = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{23} & 0 \end{pmatrix}$$

où a_{11} *et* a_{23} *sont réels (*$\mathsf{V}^\mathsf{T}\Phi\mathsf{V}$ *est symétrique). Donc, la matrice* $\mathsf{V}^\mathsf{T}\Phi\mathsf{V}$ *représente une parabole ce qui signifie qu'ils existe un point réel* $\mathbf{x}$ *avec :*

$$\mathbf{x}^\mathsf{T}\mathsf{V}^\mathsf{T}\Phi\mathsf{V}\mathbf{x} = 0 \ .$$

$\mathsf{V}\mathbf{x}$ *est aussi un point réel et se trouve sur la conique* Φ*. Le fait que* Φ *contient un point réel est en contradiction avec l'hypothèse que* Φ *est une conique imaginaire. Par conséquent, la supposition que* $\tau \neq 1$ *est réel, est fausse.*

De la proposition D.6 il découle finalement que les valeurs propres sont $\{\lambda, \lambda e^{I\Theta}, \lambda e^{-I\Theta}\}$*, pour des* λ *et* Θ *réels.* □

REMARQUE D.1 *Les réciproques des propositions D.6 et D.7 sont faciles à prouver. Puisque nous ne les utilisons pas, nous en omettons les preuves.*

COROLLAIRE D.3 (CONTRAINTE DU MODULE) *L'homographie infinie pour deux vues avec les mêmes paramètres intrinsèques a des valeurs propres du même module.*

Preuve : *Grâce à l'égalité des paramètres intrinsèques, les projections de la conique absolue dans les deux vues sont identiques. Par conséquent, l'homographie infinie laisse une conique imaginaire et propre invariante et on peut appliquer la proposition D.7 pour compléter la preuve.* □

Index des auteurs cités

Index des mots clés

Vision 3D non calibrée : contributions à la reconstruction projective et étude des mouvements critiques pour l'auto-calibrage

Cette thèse concerne la reconstruction tridimensionnelle d'objets à partir d'images prises par des caméras. Le schéma classique s'appuie sur un calibrage hors ligne des caméras. Ce calibrage n'est pas toujours possible et des applications dynamiques requièrent en général fréquemment une mise à jour du calibrage au vol. Il est donc clairement souhaitable de pouvoir s'affranchir du besoin d'un calibrage hors ligne.

Dans cette thèse, nous nous concentrons d'une part sur des méthodes de reconstruction non calibrée et d'autre part sur le concept de calibrage en ligne ou auto-calibrage. Nous tâchons d'obtenir des mesures tridimensionnelles même avec des images prises par des caméras non calibrées. Les informations obtenues ne sont pas de nature métrique, mais elles sont néanmoins exploitables pour des tâches de reconnaissance d'objets ou d'asservissement de robots. Nous avons développé plusieurs méthodes pratiques pour la reconstruction tridimensionnelle non calibrée.

L'auto-calibrage ou calibrage en ligne consiste à calibrer une caméra uniquement à partir d'images d'objets inconnus. Les images prises au cours d'une application peuvent donc être utilisées simultanément pour l'application elle-même et pour le calibrage de la caméra. Notre contribution majeure dans ce domaine est une étude des conditions de dégénérescence qui s'avèrent apparaître fréquemment en pratique. Plus précisément, nous identifions tous les types de mouvements de caméra qui prohibent l'auto-calibrage.

Mots clés : vision par ordinateur, calibrage, reconstruction tridimensionnelle, reconstruction projective, auto-calibrage, séquence de mouvements critique.

Uncalibrated 3D Vision: Contributions to Projective Reconstruction and Study of the Critical Motions for Self-Calibration

This thesis is about threedimensional reconstruction of objects from images taken by cameras. The classical scheme is based on an off-line calibration of the cameras. Very often, this is not available and, in general, dynamic applications need a frequent update of the calibration. Thus it is clearly desirable to get rid off the requirement of performing an off-line calibration.

In this thesis, we concentrate on the one hand on methods for uncalibrated reconstruction and on the other hand on the concept of on-line or self-calibration. We try to get threedimensional measurements even from images taken with uncalibrated cameras. The obtained information is not of a metric nature, but still useful for tasks such as object recognition or the visually guided control of robot motions. We have developed several methods for uncalibrated threedimensional reconstruction.

Self-calibration consists of calibrating a camera only from images of unknown objects. Images taken in course of an application can thus be used simultaneously for the application itself and for camera calibration. Our major contribution in this domain is a study of degeneracy conditions that turn out to be a very common problem in practice. In concrete terms, we identify all types of camera motion that prevent self-calibration.

Keywords: computer vision, calibration, threedimensional reconstruction, projective reconstruction, self-calibration, critical motion sequence.

Discipline : Informatique
Laboratoire de rattachement : GRAVIR - IMAG - INRIA Rhône - Alpes
655, Avenue de l'Europe, 38330 Montbonnot St Martin