

Géométrie d'images multiples pour des modèles de caméras généraux

Multi-view geometry for general camera models

Peter Sturm¹, Srikumar Ramalingam²

¹INRIA Rhône-Alpes, 655 Avenue de l'Europe, 38330 Montbonnot, France
Peter.Sturm@inrialpes.fr

² University of California, Dept. of Computer Science, Santa Cruz, USA
srikumar@cse.ucsc.edu

Manuscrit reçu le 24 janvier 2005

Résumé et mots clés

Nous considérons le problème de l'estimation de la structure et du mouvement pour un modèle de caméras hautement général, qui représente une caméra par un ensemble de rayons de projection. Ceci permet de décrire la plupart des types de caméras existants (du moins celles qui opèrent dans le domaine visible), y inclus les caméras sténopé, les caméras avec des distorsions radiales ou plus générales, les caméras catadioptriques (à point de vue unique ou non), etc. Nous introduisons une hiérarchie de modèles de caméras généraux : le modèle le plus général peut posséder des rayons de projection quelconques tandis que le modèle le plus contraint que nous considérons ici est le modèle à point de vue unique (tous les rayons passent par un même point). Parmi les modèles intermédiaires, nous identifions ce que nous appelons les caméras axiales (tous les rayons touchent une même ligne) et les caméras connues sous le nom de « cross-slit » (les rayons touchent deux lignes). Les fondements d'une géométrie d'images multiples pour le modèle de caméras le plus général sont donnés. Ils se manifestent par la formulation de tenseurs d'appariement multi-vues, qui sont l'analogie des matrices fondamentales/essentiels, tenseurs trifocaux ou quadrifocaux des caméras perspectives. Ce cadre théorique général est ensuite spécialisé pour les modèles de caméras intermédiaires mentionnés, pour le cas de deux images.

Modèle de caméras, caméra non centrale, caméra omnidirectionnelle, tenseur d'appariement, géométrie épipolaire, géométrie d'images multiples.

Abstract and key words

We consider the structure from motion problem for a previously introduced, highly general imaging model, where cameras are modeled as possibly unconstrained sets of projection rays. This allows to describe most existing camera types (at least for those operating in the visible domain), including pinhole cameras, sensors with radial or more general distortions, catadioptric cameras (central or non-central), etc. We introduce a hierarchy of general camera models: the most general model has unconstrained projection rays whereas the most constrained model dealt with here is the central model, where all rays pass through a single point. Intermediate models are what we call axial cameras (all rays touch a single line), and x-slit cameras (rays touch two lines). The foundations for a multi-view geometry of completely non-central cameras are given, leading to the formulation of multi-view matching tensors, analogous to the fundamental/essential matrices, trifocal and quadrifocal tensors of perspective cameras. This framework is then specialized explicitly for the two-view case, for the intermediate camera types mentioned above.

Camera model, non-central camera, omnidirectional camera, matching tensor, epipolar geometry, multi-view geometry.

1. Introduction

Beaucoup de différents capteurs sont utilisés en vision par ordinateur, dont les caméras perspectives, les systèmes stéréo, les caméras omnidirectionnelles (par exemple, celles catadioptriques), etc. La plupart des modèles utilisés pour ces caméras sont paramétriques et définis par quelques paramètres intrinsèques (distance focale, coefficients de distorsion, etc.) et considèrent surtout des caméras à point de vue unique. De plus, les algorithmes existants de calibrage, de reconstruction 3-D ou d'estimation du mouvement, sont le plus souvent conçus pour un seul modèle de caméras à la fois (voir par exemple [5, 15, 11]).

Le but de notre travail est de relâcher ces contraintes : nous voulons proposer et développer des approches de calibrage, de reconstruction 3-D etc. qui puissent être appliquées quels que soient les types des caméras utilisées, notamment les caméras omnidirectionnelles et/ou n'ayant pas de point de vue unique. Pour ce faire, nous renonçons aux modèles paramétriques classiques et adoptons un modèle très général [13] : une caméra acquiert des images qui consistent en un ensemble de pixels ; chaque pixel capte la lumière qui se propage le long d'un rayon (rayon de projection). Une caméra est alors complètement modélisée par :

- les coordonnées de ces rayons (en 3-D, données par rapport à un repère local de la caméra) ;

- la correspondance entre pixels et rayons.

Ce modèle général permet de décrire la plupart des types de caméra, par exemple (cf. la figure 1) :

- des caméras avec des distorsions optiques quelconques, telles les distorsions radiales ou tangentielles ;

- les caméras catadioptriques, c'est-à-dire des caméras qui perçoivent la scène au travers d'une réflexion dans un miroir, typiquement de forme convexe. De tels systèmes peuvent avoir un point de vue unique [2, 10], mais uniquement si le miroir ainsi que la position relative miroir-caméra sont bien choisis. Si un miroir sphérique est utilisé, ou un miroir dont la surface ne correspond pas à une quadrique [16], le système catadioptrique n'aura pas de point de vue unique ;

- des systèmes stéréo (deux caméras ou plus) : conceptuellement, on peut considérer un système stéréo comme un seul capteur qui consiste de l'ensemble des pixels des caméras et des rayons associés. Il s'agit bien évidemment d'un capteur qui n'a pas de point de vue unique ;

- d'autres systèmes d'acquisition, dont beaucoup n'ont pas de point de vue unique et/ou sont de type omnidirectionnel [3, 4, 14, 6, 7, 18, 23, 24, 28, 33, 34] ;

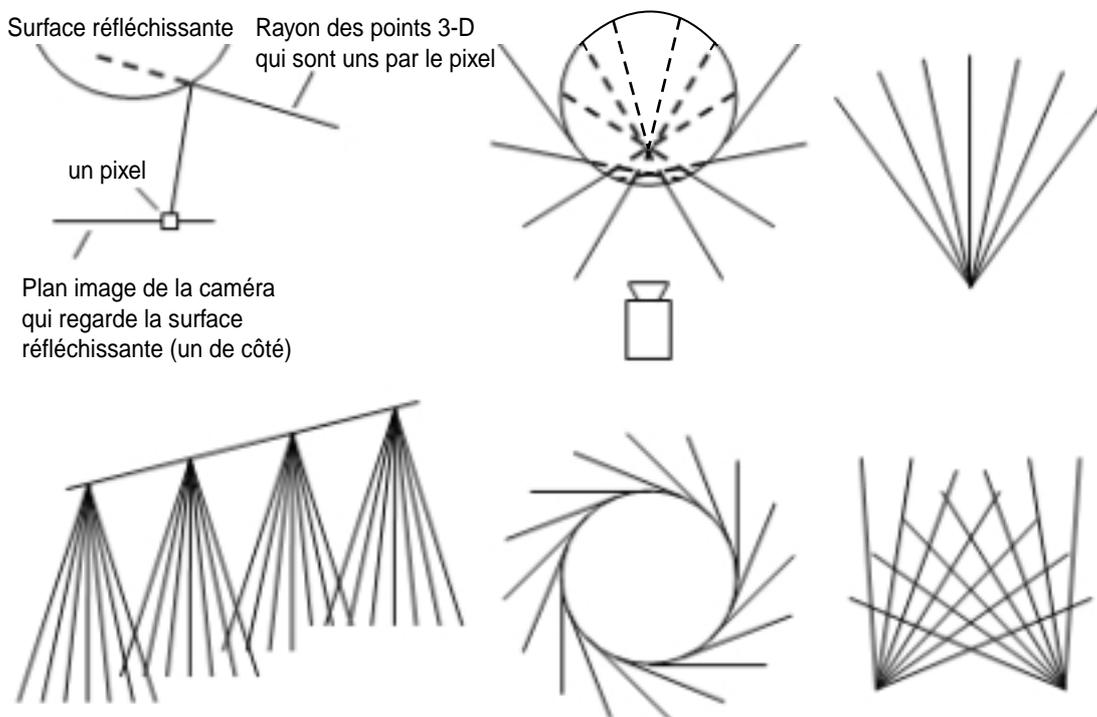


Figure 1. Exemples de types de caméras. Première ligne :

(i) Système catadioptrique (notons que les rayons ne passent pas par les pixels associés).

(ii) Système catadioptrique basé sur un miroir sphérique

(ce système n'a pas de point de vue unique – les rayons ne se coupent pas en un seul point).

(iii) Caméra à point de vue unique (par exemple, caméra perspective, avec ou sans distorsion radiale ou autre)

Deuxième ligne : (i) Caméra de type « push-broom ».

(ii) Système d'acquisition dit « omni-vergent » [24, 28].

(iii) Système stéréo.

- un exemple où le modèle énoncé ci-dessus ne s'appliquerait pas est celui d'une caméra qui regarde une scène à travers une interface entre deux matières. Considérons par exemple une caméra qui regarde dans l'eau mais n'y est pas plongée: les rayons de projection sont réfractés et si la caméra se déplace, l'ensemble des rayons ne se déplacera pas de manière rigide.

Bien évidemment, le modèle de caméras que nous utilisons n'est qu'une approximation: en réalité, un pixel capte non pas un seul rayon de lumière, mais plutôt de la lumière qui se propage dans un certain volume. Cette remarque s'applique pourtant à la majorité des modèles existants. D'autres aspects importants, par exemple ceux liés à la photométrie, sont très bien décrits dans [5, 15, 11].

Dans cet article, nous introduisons les fondements pour une géométrie d'images multiples pour le modèle de caméras générique décrit ci-dessus. Ils s'expriment au travers de tenseurs d'appariement, similairement aux matrices fondamentales ou essentielles et aux tenseurs trifocaux ou quadrifocaux des caméras perspectives. Nous rappelons ici simplement que les tenseurs d'appariement servent à donner des contraintes pour l'appariement de primitives géométriques entre images. Ceci sera mieux expliqué dans la suite.

Nous formulons la géométrie d'images multiples pour des caméras calibrées, c'est-à-dire pour lesquelles la relation pixels-rayons est connue. Ainsi, les correspondances entre pixels d'images différentes se traduiront directement en correspondances de rayons de projection en 3-D. Les tenseurs d'appariement que nous allons dériver, agissent alors sur les coordonnées des rayons.

Nous introduisons également une hiérarchie naturelle de modèles de caméras: le modèle le plus général consiste d'un ensemble non contraint de rayons de projection tandis que le modèle le plus contraint considéré ici est celui des caméras à point de vue unique. Dans la suite, nous utilisons l'expression concise de *caméra centrale* pour désigner les caméras à point de vue unique. Des caméras sans point de vue unique sont appelées *caméras non centrales*. Un modèle intermédiaire est ce que nous appelons une *caméra axiale*: une caméra telle qu'il existe une droite en 3-D qui touche tous les rayons de projection. Ce modèle comprend les caméras de type «push-broom» [14] et certaines caméras catadioptriques. Une sous-classe est celle des caméras dites de type «cross-slit»: il existe deux droites en 3-D qui touchent tous les rayons (les caméras de type «push-broom linéaire» sont en effet de ce type).

La géométrie d'images multiples, formulée d'abord pour des caméras non centrales générales, est ensuite spécialisée à ces modèles intermédiaires, pour le cas de base de deux vues (géométrie épipolaire). Il existe plusieurs travaux sur la géométrie épipolaire de caméras omnidirectionnelles, centrales ou non [9, 12, 20, 21, 25, 30, 32]. Le but de la plupart de ces travaux est d'obtenir des contraintes d'appariement pour des caméras non calibrées (ce qui est difficile même pour certaines caméras catadioptriques centrales). Dans cet article, nous abordons le cas de caméras calibrées et donnons un traitement assez complet du problème.

Cet article est structuré comme suit. Dans la section 2, nous rappelons la définition et des propriétés des coordonnées de Plücker pour les droites en 3-D, qui sont utilisées pour paramétrer les rayons de projection. Une hiérarchie de modèles de caméras est proposée dans la section 3. Des paramétrisations de rayons de projection pour différents modèles de caméras, sont proposées en section 4. La géométrie d'images multiples pour le modèle de caméras général est développée dans la section 5. Cette géométrie est ensuite explorée en détail pour le cas de deux vues et différents modèles de caméras, en section 6. Notations utilisées: les matrices sont notées en sans empattement (L, R, \dots), les vecteurs en caractères gras ($\mathbf{a}, \mathbf{b}, \dots$) et les scalaires en caractères italiques (u, v, \dots). Les coefficients de tenseurs, matrices ou vecteurs sont des scalaires, donc notés en italique ($T_{i,j,k}, L_{i,j}, \dots$). Le produit vectoriel de deux vecteurs de longueur 3 est écrit $\mathbf{u} \times \mathbf{v}$. La notation $[\mathbf{u}]_{\times}$ désigne la matrice anti-symétrique de dimension 3×3 définie par le produit vectoriel: $[\mathbf{u}]_{\times} \mathbf{v} = \mathbf{u} \times \mathbf{v}$. La transposée d'une matrice est notée par L^T . Les vecteurs sont parfois interprétés comme des matrices à une colonne; la transposée \mathbf{a}^T d'un vecteur désigne donc une matrice à une ligne.

2. Coordonnées de Plücker

Nous représentons les rayons de projection par des droites en 3-D, en utilisant leurs coordonnées de Plücker. Nous en utilisons la définition suivante.

Soient \mathbf{A} et \mathbf{B} deux points 3-D, donnés en coordonnées homogènes. La droite définie par ces points peut être représentée par la matrice 4×4 anti-symétrique L , dite matrice de Plücker:

$$L = \mathbf{AB}^T - \mathbf{BA}^T = \begin{pmatrix} 0 & A_1B_2 - A_2B_1 & A_1B_3 - A_3B_1 & A_1B_4 - A_4B_1 \\ A_2B_1 - A_1B_2 & 0 & A_2B_3 - A_3B_2 & A_2B_4 - A_4B_2 \\ A_3B_1 - A_1B_3 & A_3B_2 - A_2B_3 & 0 & A_3B_4 - A_4B_3 \\ A_4B_1 - A_1B_4 & A_4B_2 - A_2B_4 & A_4B_3 - A_3B_4 & 0 \end{pmatrix}$$

Notons que la matrice de Plücker d'une droite est indépendante (à l'échelle près) de la paire des points sur cette droite ayant servi à son calcul.

Une représentation alternative de la droite est le vecteur des coordonnées de Plücker, de longueur 6:

$$\mathbf{L} = \begin{pmatrix} A_4B_1 - A_1B_4 \\ A_4B_2 - A_2B_4 \\ A_4B_3 - A_3B_4 \\ A_3B_2 - A_2B_3 \\ A_1B_3 - A_3B_1 \\ A_2B_1 - A_1B_2 \end{pmatrix} \quad (1)$$

Nous identifions deux sous-vecteurs de longueur 3, \mathbf{a} et \mathbf{b} :

$$\mathbf{a} = \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} L_4 \\ L_5 \\ L_6 \end{pmatrix}$$

Tableau 1. Classes de caméras, définies par des configurations de points ou droites en 3-D qui touchent tous les rayons de projection.

Points/droites touchant les rayons	Description de la classe
Aucun	Caméra non centrale
1 point	Caméra centrale
2 points	Caméra ayant un seul rayon de projection
1 droite	Caméra axiale
1 point, 1 droite	Caméra 1-D centrale
2 droites qui ne se coupent pas	Caméra de type cross-slit
2 droites coplanaires	Union d'une caméra centrale et d'une caméra 1-D non centrale
3 droites coplanaires sans point d'intersection commun	Caméra 1-D non centrale

Ces deux vecteurs satisfont la contrainte dite de Plücker: $\mathbf{a}^T \mathbf{b} = 0$. Un vecteur de longueur 6 correspond à des coordonnées de Plücker d'une droite si et seulement si il vérifie cette contrainte. Avec cette définition de \mathbf{a} et \mathbf{b} , la matrice de Plücker peut s'écrire :

$$\mathbf{L} = \begin{pmatrix} [\mathbf{b}]_{\times} & -\mathbf{a} \\ \mathbf{a}^T & 0 \end{pmatrix}$$

Considérons maintenant comment les droites sont transformées par des changements de repère. Soit une transformation euclidienne, définie par une matrice de rotation \mathbf{R} et un vecteur de translation \mathbf{t} , qui agit sur les points 3-D comme suit :

$$\mathbf{C} \rightarrow \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{C}$$

Les coordonnées de Plücker sont alors transformées ainsi :

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \rightarrow \begin{pmatrix} \mathbf{R} & \mathbf{0} \\ -[\mathbf{t}]_{\times} \mathbf{R} & \mathbf{R} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$$

Notons finalement que deux droites \mathbf{L}_1 et \mathbf{L}_2 se coupent exactement si la relation suivante est satisfaite :

$$\mathbf{L}_2^T \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \mathbf{L}_1 = \mathbf{a}_2^T \mathbf{b}_1 + \mathbf{b}_2^T \mathbf{a}_1 = 0 \quad (2)$$

Interprétation euclidienne des coordonnées de Plücker. Si les points \mathbf{A} et \mathbf{B} sont donnés en coordonnées affines ($A_4 = B_4 = 1$), les vecteurs \mathbf{a} et \mathbf{b} peuvent s'interpréter comme suit. Notons d'abord $\bar{\mathbf{A}}^T = (A_1, A_2, A_3)$ et $\bar{\mathbf{B}}^T = (B_1, B_2, B_3)$. Nous avons alors $\mathbf{a} = \bar{\mathbf{B}} - \bar{\mathbf{A}}$ et $\mathbf{b} = \bar{\mathbf{B}} \times \bar{\mathbf{A}}$. Le vecteur \mathbf{a} est donc le vecteur directeur de la droite. Quant à \mathbf{b} , il est orthogonal au plan engendré par l'origine et la droite. Finalement, la distance carrée de la droite de l'origine, c'est-à-dire la distance carrée du point sur la droite le plus proche de l'origine, est donnée par :

$$d^2 = \frac{\mathbf{b}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}$$

3. Une hiérarchie de modèles de caméras

Une **caméra non centrale** peut avoir des rayons de projections quelconques, tandis que pour une **caméra centrale**, il existe un point – le **centre optique** – qui se trouve sur tous les rayons de projection. Un cas intermédiaire est ce que nous appelons ici celui des **caméras axiales** : il existe une droite qui touche tous les rayons de projection. Nous l'appelons l'**axe de la caméra** (à ne pas confondre avec l'axe optique du modèle perspectif). Des exemples de caméras qui se trouvent dans cette classe sont : - les caméras dites de type cross-slit [22, 35] (ou bien, *x-slit* ou *two-slit*), et le cas particulier des caméras dites push-broom linéaire [14]. Ces caméras forment en effet une sous-classe des caméras axiales, comme il l'est expliqué plus bas ;

- des systèmes stéréo consistant de deux caméras centrales ou de plusieurs caméras centrales avec des centres optiques collinéaires ;

- certaines caméras catadioptriques non centrales : si le miroir est une surface de révolution est si le centre optique de la caméra centrale qui le regarde (pas nécessairement une caméra perspective) se trouve sur l'axe de révolution, il s'agit d'une caméra axiale. Il est facile de vérifier que dans ce cas, tous les rayons de projection réfléchis par le miroir, coupent l'axe de révolution, qui joue donc le rôle de l'axe de la caméra axiale. Remarquons qu'une caméra catadioptrique avec un miroir sphérique est obligatoirement non centrale, et de type axial. Un autre exemple est celui des systèmes basés sur un miroir conique (avec centre optique sur l'axe du cône) [7]. Le modèle axial peut aussi servir à modéliser des systèmes catadioptriques qui, à cause d'un mauvais alignement entre miroir et caméra, ne produisent pas une projection centrale.

Ces trois classes de modèles de caméras peuvent aussi être définies ainsi : existence d'un espace linéaire de dimension d qui a une intersection non vide avec chacun des rayons de projection.

Avec cette définition, $d = 0$ correspond aux caméras centrales, $d = 1$ aux caméras axiales et $d = 2$ aux caméras complètement non centrales.

Des classes intermédiaires existent. Les caméras de type cross-slit, déjà mentionnées, sont un cas spécial des caméras axiales : il existe deux droites qui toutes deux touchent tous les rayons de la caméra. Similairement, on peut définir des caméras 1-D centrales (caméras ayant une seule ligne de pixels) par un point et une droite en 3-D qui touchent chacun des rayons. Le tableau 1 résume des modèles de caméras définis de cette manière, dont certains n'ont bien sûr aucun intérêt pratique. Une approche similaire pour la définition de classes de caméras a été explorée dans [22] ; le but de cette approche était de trouver *une seule* primitive géométrique qui touche tous les rayons de projection et qui puisse donc servir à la définition d'une classe de caméras. Cette approche est moins générale et moins intuitive que celle adoptée ici.

Il est intéressant de considérer certaines de ces classes de caméras plus en détail, grâce à l'observation suivante. Beaucoup d'algorithmes existants de calibrage ou d'estimation du mouvement procèdent typiquement en deux étapes : (1) estimation d'une matrice ou d'un tenseur en résolvant des systèmes d'équations linéaires (par exemple, la matrice de projection [1], la matrice essentielle [25], les tenseurs trifocaux [15], les tenseurs de calibrage [31], etc.) ; (2) ensuite, les paramètres recherchés (paramètres intrinsèques, matrice de rotation, etc.) sont extraits de ces matrices ou tenseurs. Il y a deux problèmes intrinsèques :

- si un algorithme qui a été développé pour une classe de caméras est appliqué à une classe de caméras différente, l'étape (1) ne donnera pas de solution. Par exemple, l'estimation d'une matrice fondamentale perspective (matrice 3×3) à partir de correspondances de points issues de deux caméras catadioptriques, ne donnera évidemment aucun résultat exploitable ;

- ce premier problème est évident ; un problème plus subtil est le suivant. Si un algorithme qui a été développé pour une classe de caméras est appliqué à une caméra appartenant à une sous-classe, la résolution du système linéaire dans l'étape (1) n'aura pas de solution unique ; il y aura en effet une infinité de solutions. L'étape (2), si effectuée avec une de ces solutions choisie au hasard, donnera un résultat incorrect. Un exemple simple pour illustrer ce problème est l'estimation de la matrice fondamentale perspective, à partir de correspondances de points d'une scène qui ne contient qu'un plan : l'estimation est sous-contrainte. Un exemple plus proche de nos préoccupations est le suivant : si l'on tente d'estimer la matrice essentielle du modèle de caméras *non central* (une matrice 6×6 [25]) à partir de correspondances obtenues de caméras *centrales* ou *axiales*, alors le système d'équations linéaires associé n'aura pas de solution unique.

Ce deuxième problème nous amène à définir la géométrie d'images multiples pour le modèle le plus général d'abord, puis à la spécialiser à des sous-classes. Dans cet article, nous traitons les caméras centrales, axiales, de type cross-slit et le modèle complètement non central.

4. Paramétrisations

La géométrie d'images multiples sera formulée en utilisant les coordonnées de Plücker des rayons de projection. Les rayons de projection de toutes les classes de caméras sauf de la plus générale, appartiennent à des sous-ensembles particuliers des droites en 3-D, cf. la section précédente. Nous pouvons alors tenter de choisir un système de coordonnées local à une caméra tel que les vecteurs de coordonnées de ses rayons aient une forme particulière, menant à des tenseurs d'appariement de forme simplifiée. Comme nous considérons des caméras calibrées, les rayons sont donnés dans un repère métrique et nous pouvons alors appliquer des rotations et translations pour en choisir un de « sympathique ». Les transformations et paramétrisations de rayons de projection appropriées pour différents modèles de caméras sont expliquées dans la suite.

4.1. Caméras centrales

Tous les rayons passent par un même point, le centre optique. Nous distinguons les cas où le centre optique est un point fini ou à l'infini.

Centre optique fini. Nous choisissons un repère local avec comme origine le centre optique. Ainsi, pour tous les rayons de projection le vecteur \mathbf{b} (cf. la section 2) est nul, c'est-à-dire que les rayons ont des coordonnées de la forme :

$$\mathbf{L} = \begin{pmatrix} \mathbf{a} \\ \mathbf{0} \end{pmatrix}$$

Notons que ceci est en accord avec le fait que les tenseurs d'appariement des caméras perspectives ont une « taille de base » de 3 : la matrice fondamentale par exemple, est une matrice 3×3 .

Centre optique infini. Dans ce cas (par exemple, une caméra orthographique), nous ne pouvons pas choisir le centre optique comme origine. À la place, nous orientons le repère tel que le centre optique ait les coordonnées $(0,0,1,0)^T$. Les rayons de projection sont alors de la forme :

$$\mathbf{L} = (0 \quad 0 \quad a_3 \quad b_1 \quad b_2 \quad 0)$$

4.2. Caméras axiales

Tous les rayons touchent une droite particulière, l'axe de la caméra. Nous distinguons les cas d'un axe étant une droite à l'infini ou non.

Axe fini. Nous choisissons un repère où l'axe de la caméra coïncide avec l'axe des Z. Pour les rayons de projection nous avons alors $L_6 = b_3 = 0$:

$$\mathbf{L} = \begin{pmatrix} \mathbf{a} \\ b_1 \\ b_2 \\ 0 \end{pmatrix}$$

Axe infini. Choisissons un repère où l'axe est la droite à l'infini avec les coordonnées $(1,0,0)^T$ (coordonnées d'une droite sur le plan à l'infini). Ceci correspond aux coordonnées de Plücker $(0,0,0,1,0,0)^T$. Les rayons de projection ont donc des coordonnées avec $L_1 = a_1 = 0$ (ceci est basé sur l'équation (2)).

$$\mathbf{L} = \begin{pmatrix} 0 \\ a_2 \\ a_3 \\ \mathbf{b} \end{pmatrix}$$

Pour les deux cas, nous voyons que les rayons de projection d'une caméra axiale peuvent être représentés par 5 coordonnées de Plücker. Les tenseurs d'appariement auront alors une taille de base de 5; par exemple, la matrice essentielle pour des caméras axiales sera de dimension 5×5 (voir plus loin).

4.3. Caméras de type cross-slit

Comme il a été mentionné, les caméras de type cross-slit sont définies par l'existence de deux droites qui coupent tous les rayons de projection. Le cas où ces deux axes se coupent, c'est-à-dire sont coplanaires, n'a pas d'intérêt ici (voir le tableau 1). Deux cas sont alors possibles: (i) tous les deux axes sont des droites finies ou (ii) exactement un des deux axes est une droite à l'infini. Il y a forcément au moins un axe fini; nous adoptons un repère comme il a été décrit ci-dessus pour les caméras axiales avec axe fini. Ceci nous laisse encore des degrés de liberté dans le choix du repère, ce qui sera exploité pour obtenir des coordonnées simples pour le deuxième axe.

Deux axes finis. Ayant fixé le premier axe, nous avons toujours la possibilité d'effectuer des rotations autour, ou des translations le long de cet axe. Nous pouvons alors choisir un repère où le deuxième axe de la caméra coupe l'axe des Y et est parallèle au plan $X - Z$. Ainsi, il contient deux points avec des coordonnées de la forme:

$$\mathbf{A} = \begin{pmatrix} 0 \\ Y \\ 0 \\ 1 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} X \\ 0 \\ Z \\ 0 \end{pmatrix}$$

Ses coordonnées de Plücker sont alors données par:

$$\mathbf{L}_2^T = (X \quad 0 \quad Z \quad -YZ \quad 0 \quad YZ)$$

Les rayons de projection coupent les deux axes et doivent alors être de la forme:

$$\mathbf{L}^T = \left(a_1 \quad a_2 \quad a_3 \quad \left(\frac{YZ}{X} a_1 - Y a_3 \right) \quad b_2 \quad 0 \right)$$

Nous divisons par X , ce qui est permis ici puisque $X \neq 0$ (sinon, le deuxième axe serait parallèle au premier, donc coplanaire, ce qui est exclu ici). Remplaçons ensuite $\frac{YZ}{X}$ par W . Les rayons de projection peuvent alors être paramétrés par les 4 coefficients a_1, a_2, a_3, b_2 (qui sont définis à une échelle près):

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ W & 0 & -Y & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ b_2 \end{pmatrix}$$

Les coefficients W et Y sont connus et identiques pour tous les rayons de projection (ils représentent la position relative des deux axes de caméra, qui est connue puisque la caméra est supposée être calibrée).

Un axe fini et un axe infini. Comme il a été dit ci-dessus, nous fixons d'abord l'axe fini comme pour les caméras axiales, puis pouvons encore effectuer des rotations autour ou des translations le long de cet axe pour obtenir des coordonnées particulières pour l'axe infini. Les translations n'ont pas d'effet sur les coordonnées de l'axe infini; quant à la rotation, nous pouvons la choisir telle que l'axe infini ait les coordonnées $(0, \cos \Theta, \sin \Theta)^T$ (coordonnées homogènes d'une droite à l'infini), pour un certain Θ . Ceci correspond aux coordonnées de Plücker:

$$\mathbf{L}_2^T = (0 \quad 0 \quad 0 \quad 0 \quad \cos \Theta \quad \sin \Theta)$$

Les rayons de projection doivent couper les deux axes, donc sont de la forme:

$$\mathbf{L}^T = (a_1 \quad -a_3 \tan \Theta \quad a_3 \quad b_1 \quad b_2 \quad 0)$$

Définissons $W = -\tan \Theta$. Les rayons de projection peuvent alors être représentés par 4 coefficients a_1, a_3, b_1, b_2 (définis à l'échelle près):

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & W & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_3 \\ b_1 \\ b_2 \end{pmatrix}$$

Les tenseurs d'appariement pour des caméras de type cross-slit auront alors une taille de base de 4.

4.4. Caméras non centrales générales

Aucune simplification des coordonnées des rayons de projection n'est possible; ce seront alors des coordonnées de Plücker générales. Les tenseurs d'appariement auront une taille de base de 6.

5. Géométrie d'images multiples

Nous établissons les fondements d'une géométrie d'images multiples pour des caméras générales (non centrales). Elle sera incarnée, comme avec les caméras perspectives, par des tenseurs d'appariement. Nous montrons comment les obtenir, de manière analogue au cas perspectif.

Ici, nous ne traitons que du cas calibré ; le cas non calibré n'est bien géré que pour les caméras perspectives, puisque ces caméras, calibrées ou non, sont liées par des transformations projectives. Pourtant, quant aux caméras non centrales, un tel lien n'existe pas en général : dans le cas le plus général, chaque paire pixel+rayon de projection peut être complètement indépendante de toutes les autres paires.

Dans la suite, nous rappelons d'abord la notion de tenseur d'appariement, puis les principes d'une approche de géométrie d'image multiples pour les caméras perspectives. Cette approche travaille avec des coordonnées de points dans les images. Nous appliquons ensuite les mêmes idées, tout en travaillant avec des coordonnées de droites en 3-D, pour dériver notre géométrie des caméras non centrales.

5.1. Tenseurs d'appariement

Il n'existe pas de véritable définition de la notion de tenseur d'appariement dans la littérature. Néanmoins, il est entendu qu'il s'agit, pour les caméras perspectives, de tenseurs dont les coefficients dépendent des matrices de projection d'un ensemble de vues considérées et qui permettent de formuler des contraintes d'appariement pour des primitives géométriques dans ces vues. Il est généralement sous-entendu que ces contraintes sont de forme multi-linéaire, c'est-à-dire que les contraintes sont linéaires en les coordonnées de chaque primitive. L'exemple le plus connu est le tenseur d'appariement bifocal, ou bien la matrice fondamentale F , qui donne la contrainte d'appariement bi-linéaire classique :

$$\mathbf{q}_2^T \mathbf{F} \mathbf{q}_1 = 0$$

Plus généralement, pour n vues et des points $\mathbf{q}_1 \cdots \mathbf{q}_n$ dans ces vues, on espère trouver des tenseurs d'appariement T de dimension $3 \times \cdots \times 3$ qui permettent de formuler des contraintes d'appariement multi-linéaires :

$$\sum_{i_1=1}^3 \sum_{i_2=1}^3 \cdots \sum_{i_n=1}^3 q_{1,i_1} q_{2,i_2} \cdots q_{n,i_n} T_{i_1,i_2,\dots,i_n} = 0 \quad (3)$$

Ceci est décrit plus en détail dans la section suivante.

Ici, nous ne traitons que des tenseurs d'appariement pour n points, mais il en existe également pour l'appariement de droites ou d'un mélange de droites et de points [15].

Notons aussi que les tenseurs d'appariement ont deux applications principales :

1. ils permettent d'établir des contraintes d'appariement, voir ci-dessus ;
2. réciproquement, des appariements entre images permettent de calculer les tenseurs. Puisqu'ils dépendent des matrices de projection, on peut espérer de remonter à ces dernières, donc faire de l'estimation du mouvement de caméra, de l'auto-calibrage, de la reconstruction 3-D etc. Il existe un éventail assez large de telles applications et nous renvoyons à [15] pour un tour d'horizon assez complet.

Le but principal de cet article est de montrer l'existence et la forme de tenseurs d'appariement pour nos modèles de caméras généraux. Plus concrètement, nous recherchons des contraintes d'appariement qui seraient de la forme :

$$\sum_{i_1=1}^6 \sum_{i_2=1}^6 \cdots \sum_{i_n=1}^6 L_{1,i_1} L_{2,i_2} \cdots L_{n,i_n} T_{i_1,i_2,\dots,i_n} = 0 \quad (4)$$

où les L sont les coordonnées de Plücker de rayons de projection, ou bien des vecteurs de coordonnées tels que définis section 4. Dans le cas le plus général, les tenseurs d'appariement seraient alors de dimension $6 \times \cdots \times 6$. La contrainte (4) exprimerait que les rayons de projection considérés peuvent se correspondre, c'est-à-dire que les pixels associés à ces rayons peuvent être un appariement potentiel. Ces concepts seront introduits dans la section 5.3.

5.2. Rappels sur la géométrie d'images multiples des caméras perspectives

Nous rappelons des principes de l'une des approches pour dériver les relations entre images multiples, développée pour les caméras perspectives [8]. Soient P_i des matrices de projection et \mathbf{q}_i des points image. Un ensemble de points image peut constituer une correspondance uniquement s'il existe un point 3-D \mathbf{Q} et des facteurs d'échelle scalaires λ_i tels que, pour tout i :

$$\lambda_i \mathbf{q}_i = P_i \mathbf{Q}$$

Ceci peut être écrit sous la forme d'une équation matricielle :

$$\underbrace{\begin{pmatrix} P_1 & \mathbf{q}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ P_2 & \mathbf{0} & \mathbf{q}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{q}_n \end{pmatrix}}_M \begin{pmatrix} \mathbf{Q} \\ -\lambda_1 \\ -\lambda_2 \\ \vdots \\ -\lambda_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Cette équation exprime que la matrice M , de dimension $3n \times (4+n)$, possède un vecteur nul, c'est-à-dire qu'elle est de rang inférieur à $4+n$. Par conséquent, les déterminants de toutes ses sous-matrices de dimension $(4+n) \times (4+n)$ valent zéro. Ces déterminants sont des expressions multi-linéaires en termes des coordonnées des points image \mathbf{q}_i .

Il faut les développer pour toute sous-matrice possible de la bonne dimension. Notons que seule une sous-matrice contenant deux lignes ou plus associées à chacune des vues, peut donner une contrainte qui lie toutes les matrices de projection. Ainsi, des contraintes peuvent être obtenues pour n vues avec $2n \leq 4 + n$, ce qui implique que des contraintes multi-vues (et multi-linéaires) n'existent que jusqu'à 4 vues.

Les contraintes pour n vues sont de la forme (3), où T désigne donc un tenseur de dimension $3 \times \dots \times 3$, appelé tenseur d'appariement. Les tenseurs dépendent uniquement des matrices de projection P_i et constituent en effet une représentation compacte de celles-ci (représentation qui permet d'extraire les P_i à une transformation projective près).

Notons que dès que l'on considère des caméras calibrées, cette théorie peut s'appliquer à n'importe quelle caméra *centrale* (en plus des caméras perspectives) : des contraintes comme ci-dessus peuvent par exemple être écrites pour des caméras avec distorsion radiale, en termes des coordonnées de points image corrigés.

5.3. Géométrie d'images multiples pour les caméras non centrales

Ici, nous traitons avec des matrices de pose au lieu de matrices de projection (qui, elles, dépendent et de la pose et du calibrage) :

$$P_i = \begin{pmatrix} R_i & \mathbf{t}_i \\ \mathbf{0}^T & 1 \end{pmatrix}$$

Ces transformations euclidiennes représentent des déplacements d'une caméra, ou bien le changement de repère d'un repère global vers le repère local de la caméra. Les points image du paragraphe précédent sont maintenant remplacés par des rayons de projection. Soit le i^{e} rayon représenté par deux points 3-D \mathbf{A}_i et \mathbf{B}_i . Ultérieurement, nous voulons aboutir à des expressions en termes des coordonnées de Plücker des rayons, c'est-à-dire des tenseurs T et des contraintes d'appariement de la même forme que (3), mais avec des tenseurs de dimension $6 \times \dots \times 6$, qui agissent sur des coordonnées de Plücker, voir l'équation (4) ci-dessus. Dans la suite, nous expliquons comment obtenir ces contraintes.

Considérons un ensemble de n rayons de projection, un pour chaque vue, qui chacun est représenté par deux points \mathbf{A}_i et \mathbf{B}_i . Le choix des points sur une droite n'a aucune importance, comme plus tard nous arriverons à des expressions en termes des coordonnées de Plücker.

Les n rayons peuvent être une correspondance s'il existe un point 3-D \mathbf{Q} et des facteurs d'échelle λ_i et μ_i tels que, pour chaque vue $i = 1 \dots n$:

$$\lambda_i \mathbf{A}_i + \mu_i \mathbf{B}_i = P_i \mathbf{Q}$$

Cette équation exprime en effet que le point \mathbf{Q} se trouve sur le i^{e} rayon de projection (le tout exprimé dans le repère local de la i^{e} caméra).

Comme pour les caméras perspectives, nous regroupons les équations de toutes les vues en une équation matricielle :

$$\underbrace{\begin{pmatrix} P_1 & \mathbf{A}_1 & \mathbf{B}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ P_2 & \mathbf{0} & \mathbf{0} & \mathbf{A}_2 & \mathbf{B}_2 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ P_n & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_n & \mathbf{B}_n \end{pmatrix}}_M \begin{pmatrix} \mathbf{Q} \\ -\lambda_1 \\ -\mu_1 \\ -\lambda_2 \\ -\mu_2 \\ \vdots \\ -\lambda_n \\ -\mu_n \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

Cette équation implique que M , de dimension $4n \times (4 + 2n)$, n'est pas de rang plein. Par conséquent, toutes ses sous-matrices de dimension $(4 + 2n) \times (4 + 2n)$ doivent être singulières, c'est-à-dire avoir un déterminant nul. Dans la suite, nous montrons que seules les sous-matrices qui contiennent au moins trois lignes associées à chacune des vues, mènent à des contraintes entre toutes les matrices de pose P_i . Ceci veut dire que nous aurons des contraintes pour n vues si $3n \leq 4 + 2n$, donc $n \leq 4$ comme pour les caméras perspectives.

Regardons pour commencer le cas d'une sous-matrice de M qui contient, pour une des vues, une seule ligne associée. Sans perte de généralité, supposons que ce soit le cas pour la première vue et que seule la première ligne associée à cette vue soit présente dans la sous-matrice de M . Cette sous-matrice sera donc de la forme :

$$N = \begin{pmatrix} \mathbf{P}_{11}^T & A_{11} & B_{11} & 0 & 0 & \dots & 0 & 0 \\ P_2 & \mathbf{0} & \mathbf{0} & \mathbf{A}_2 & \mathbf{B}_2 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ P_n & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_n & \mathbf{B}_n \end{pmatrix}$$

Ici, \mathbf{P}_{11}^T désigne la première ligne de la matrice P_1 . Les deux colonnes contenant les scalaires A_{11} et B_{11} ne contiennent sinon que des zéros. L'une est donc le multiple de l'autre, ce qui implique que le déterminant de N vaut *toujours* zéro. Il n'y a donc pas de contrainte exploitable.

Dans la suite, nous excluons ce cas, c'est-à-dire que nous ne considérons que des sous-matrices de M avec au moins deux lignes associées à chacune des vues. Soit N une telle matrice. Sans perte de généralité, nous commençons le développement de son déterminant avec la colonne contenant \mathbf{A}_1 et \mathbf{B}_1 . Le déterminant sera alors donné comme une somme de termes de la forme :

$$(A_{1,j} B_{1,k} - A_{1,k} B_{1,j}) \det \tilde{N}_{jk}$$

où $j, k \in \{1..4\}$, $j \neq k$ et \tilde{N}_{jk} est obtenue de N en omettant les colonnes contenant \mathbf{A}_1 et \mathbf{B}_1 ainsi que les lignes contenant $A_{1,j}$ et $A_{1,k}$.

Nous observons différentes choses :

- Le terme $(A_{1,j} B_{1,k} - A_{1,k} B_{1,j})$ représente en effet une des coordonnées de Plücker du rayon de la première vue, engendré

par \mathbf{A}_1 et \mathbf{B}_1 (cf. la section 2). En continuant avec le développement du déterminant de \tilde{N}_{jk} et ainsi de suite, le déterminant de N pourra s'écrire sous la forme :

$$\sum_{i_1=1}^6 \sum_{i_2=1}^6 \cdots \sum_{i_n=1}^6 L_{1,i_1} L_{2,i_2} \cdots L_{n,i_n} T_{i_1,i_2,\dots,i_n} = 0$$

où les $L_{i,j}$ sont des coordonnées de Plücker des rayons considérés, obtenues à partir des \mathbf{A}_i et \mathbf{B}_i . Les coefficients des matrices de pose P_i sont, eux, regroupés dans un tenseur T , de dimension $6 \times \cdots \times 6$.

- Si N ne contient que deux lignes associées à la première vue, alors son déterminant sera de la forme :

$$L_{1,x} \left(\sum_{i_2=1}^6 \cdots \sum_{i_n=1}^6 L_{2,i_2} \cdots L_{n,i_n} T_{i_2,\dots,i_n} \right) = 0$$

c'est-à-dire qu'une seule coordonnée du premier rayon sera présente dans l'expression. Cette contrainte ne lie pas toutes les vues entre elles : elle est vérifiée si $L_{1,x} = 0$, ce qui est une condition indépendante des autres vues ou bien si l'expression entre parenthèses vaut zéro, ce qui ne dépend que des vues 2 à n . Ceci explique ce que nous avons constaté plus haut : pour obtenir des contraintes entre toutes les vues, seules des sous-matrices contenant au moins trois lignes pour chacune des vues sont utiles.

Nous sommes maintenant prêts à établir les différents cas qui mènent à des contraintes entre images. Comme il a été dit, aucune contrainte (multi-linéaire) n'existe qui lierait plus de quatre vues à la fois. Nous résumons alors les cas utiles, de deux à quatre vues, dans le tableau 2, pour les caméras centrales (basé sur la théorie développée pour les caméras perspectives) et non centrales. Le tableau donne les dimensions des sous-matrices de M , dont le déterminant donne des contraintes d'appariement, représentées par des matrices essentielles (pour deux vues), des tenseurs trifocaux ou quadrifocaux. L'écriture détaillée des tenseurs devient alors une tâche plutôt « mécanique ».

Tableau 2. Les cas utiles de contraintes d'appariement pour des caméras centrales et non centrales. Les colonnes surtitrées par « cas utiles » contiennent des entrées de la forme $x - y - z$ etc. Ceci désigne la constitution des sous-matrices de M qui donnent lieu à des contraintes entre toutes les vues : $x - y - z$ par exemple veut dire que les sous-matrices contiennent x lignes associées à une vue, y à une autre, etc.

nombre de vues	central		non-central	
	M	cas utiles	M	cas utiles
2	6×6	3-3	8×8	4-4
3	9×7	3-2-2	12×10	4-3-3
4	12×8	2-2-2-2	16×12	3-3-3-3

6. Le cas de deux vues

Nous avons jusqu'alors expliqué comment formuler des contraintes d'appariement entre deux, trois ou quatre caméras non centrales, représentées par des tenseurs de dimension 6×6 à $6 \times 6 \times 6 \times 6$. Afin de rendre ces résultats plus concrets, nous explorons maintenant en détails le cas de deux vues. Nous montrons comment le tenseur bifocal, ou matrice essentielle, dépend des paramètres de pose (ou bien, du mouvement). Ceci est d'abord fait pour les caméras non centrales générales, puis spécialisé aux caméras axiales, de type cross-slit et finalement aux caméras centrales. À la fin de cette section, nous donnons quelques commentaires sur l'estimation des matrices essentielles, introduites dans la suite.

6.1. Caméras non centrales

Pour plus de simplicité, et sans perte de généralité, nous supposons ici que le repère global coïncide avec le repère de la première caméra. La matrice de pose P_1 est donc l'identité. Quant à la pose de la deuxième caméra, nous omettons alors les indices ; elle sera donc représentée par une matrice de rotation R et un vecteur de translation \mathbf{t} . La matrice M devient alors :

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & A_{1,1} & B_{1,1} & 0 & 0 \\ 0 & 1 & 0 & 0 & A_{1,2} & B_{1,2} & 0 & 0 \\ 0 & 0 & 1 & 0 & A_{1,3} & B_{1,3} & 0 & 0 \\ 0 & 0 & 0 & 1 & A_{1,4} & B_{1,4} & 0 & 0 \\ R_{11} & R_{12} & R_{13} & t_1 & 0 & 0 & A_{2,1} & B_{2,1} \\ R_{21} & R_{22} & R_{23} & t_2 & 0 & 0 & A_{2,2} & B_{2,2} \\ R_{31} & R_{32} & R_{33} & t_3 & 0 & 0 & A_{2,3} & B_{2,3} \\ 0 & 0 & 0 & 1 & 0 & 0 & A_{2,4} & B_{2,4} \end{pmatrix}$$

Pour toute paire de rayons correspondants, M doit être singulière. Comme dans le cas présent de deux vues M est carrée, ceci implique que son déterminant vaut zéro. En le développant d'après les indications de la section précédente, nous obtenons la contrainte :

$$\mathbf{L}_2^T \begin{pmatrix} -[\mathbf{t}] \times \mathbf{R} & \mathbf{R} \\ \mathbf{R} & 0 \end{pmatrix} \mathbf{L}_1 = 0 \quad (5)$$

où les coordonnées de Plücker \mathbf{L}_1 et \mathbf{L}_2 sont définies selon (1). Nous pouvons identifier la matrice essentielle, donnée dans [25] :

$$E_n = \begin{pmatrix} -[\mathbf{t}] \times \mathbf{R} & \mathbf{R} \\ \mathbf{R} & 0 \end{pmatrix} \quad (6)$$

L'équation (5) représente en effet une contrainte épipolaire, tout à fait similaire à celle des caméras perspectives.

6.2. Caméras axiales

Axe fini. Comme il a été expliqué dans la section 3, nous adoptons des repères locaux pour nos caméras où les rayons ont une coordonnée $L_6 = 0$. Ainsi, la contrainte épipolaire (5) peut être exprimée par une matrice essentielle amputée, de dimension 5×5 , qui agit sur des vecteurs de coordonnées de Plücker amputés de leur sixième coordonnée :

$$E_{af} = \begin{pmatrix} -[\mathbf{t}]_{\times} \mathbf{R} & \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \\ R_{31} & R_{32} \end{pmatrix} \\ \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \end{pmatrix} & 0_{2 \times 2} \end{pmatrix} \quad (7)$$

Cette matrice essentielle est obtenue à partir de E_n (cf. (6)) en omettant les sixièmes ligne et colonne.

Remarquons que E_{af} est en général de rang plein, c'est-à-dire de rang 5. Elle est singulière exactement si les axes des deux caméras se coupent. Dans ce cas-ci, les vecteurs nuls gauche et droit de E_{af} donnent les coordonnées des deux axes, exprimées dans le repère de l'autre caméra respectivement (pour obtenir les coordonnées de Plücker, il faut juste rajouter un zéro comme sixième coefficient).

Axe infini. Comme ci-dessus, la contrainte épipolaire (5) se simplifie, et est représentée par une matrice essentielle de dimension 5×5 :

$$\begin{pmatrix} \begin{pmatrix} -t_3 & 0 & t_1 \\ t_2 & -t_1 & 0 \end{pmatrix} \begin{pmatrix} R_{12} & R_{13} \\ R_{22} & R_{23} \\ R_{32} & R_{33} \end{pmatrix} & \begin{pmatrix} R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \\ \begin{pmatrix} R_{12} & R_{13} \\ R_{22} & R_{23} \\ R_{32} & R_{33} \end{pmatrix} & 0 \end{pmatrix}$$

En détail, la matrice essentielle s'écrit :

$$E_{ai} = \begin{pmatrix} t_1 R_{32} - t_3 R_{12} & t_1 R_{33} - t_3 R_{13} & R_{21} & R_{22} & R_{23} \\ t_2 R_{12} - t_1 R_{22} & t_2 R_{13} - t_1 R_{23} & R_{31} & R_{32} & R_{33} \\ R_{12} & R_{13} & 0 & 0 & 0 \\ R_{22} & R_{23} & 0 & 0 & 0 \\ R_{32} & R_{33} & 0 & 0 & 0 \end{pmatrix}$$

Elle est toujours singulière (les axes des deux caméras sont des droites à l'infini, donc se coupent forcément). Son vecteur nul de droite est $(0,0,R_{11},R_{12},R_{13})^T$. Il représente l'axe de la deuxième caméra, exprimé par rapport au repère de la première caméra. Réciproquement, le vecteur nul de gauche est $(0,0,R_{11},R_{21},R_{31})^T$, ce qui représente l'axe de la première caméra, dans le repère de la deuxième.

6.3. Caméras de type cross-slit

Nous considérons les deux cas expliqués dans la section 4.3.

Deux axes finis. La contrainte épipolaire (5) se simplifie et peut s'écrire basée sur une matrice essentielle de dimension 4×4 ,

qui agit sur des vecteurs de coordonnées de Plücker amputés de la forme $(a_1, a_2, a_3, b_2)^T$ (cf. la section 4.3) :

$$E_{xff} = \begin{pmatrix} 1 & 0 & 0 & W_2 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -Y_2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} E_{af} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ W_1 & 0 & -Y_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

où E_{af} est donnée dans (7). Contrairement aux cas précédents, la matrice essentielle contient maintenant non seulement des paramètres de pose, mais aussi des « paramètres intrinsèques » (les coefficients W_i et Y_i des deuxièmes axes des caméras). En détail, elle peut s'écrire :

$$E_{xff} = \begin{pmatrix} -[\mathbf{t}]_{\times} \mathbf{R} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} + \begin{pmatrix} R_{11}(W_1 + W_2) & R_{12}W_2 & R_{13}W_2 - R_{11}Y_1 & R_{12} \\ R_{21}W_1 & 0 & -R_{21}Y_1 & R_{22} \\ R_{31}W_1 - R_{11}Y_2 & -R_{12}Y_2 & -R_{13}Y_2 - R_{31}Y_1 & R_{32} \\ R_{21} & R_{22} & R_{23} & 0 \end{pmatrix}$$

Un axe fini et un axe infini. Similairement, la contrainte épipolaire se simplifie et nous obtenons une matrice essentielle de dimension 4×4 :

$$E_{xfi} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & W_2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} E_{af} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & W_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

6.4. Caméras centrales

Centre optique fini. Les rayons de projection sont ici de la forme $(L_1, L_2, L_3, 0, 0, 0)^T$ (voir la section 3). La contrainte épipolaire (5) se réduit donc à l'expression :

$$(L_{2,1} \quad L_{2,2} \quad L_{2,3}) (-[\mathbf{t}]_{\times} \mathbf{R}) \begin{pmatrix} L_{1,1} \\ L_{1,2} \\ L_{1,3} \end{pmatrix} = 0$$

où nous retrouvons la matrice essentielle « classique » de dimension 3×3 : $E_{cf} = -[\mathbf{t}]_{\times} \mathbf{R}$ [15, 17].

Centre optique infini. Dans ce cas, la matrice essentielle est donnée par :

$$E_{ci} = \begin{pmatrix} t_2 R_{13} - t_1 R_{23} & R_{31} & R_{32} \\ R_{13} & 0 & 0 \\ R_{23} & 0 & 0 \end{pmatrix}$$

Ceci ressemble à la matrice fondamentale affine [27], mais ne lui correspond pas exactement : ici, la matrice essentielle agit sur des droites en 3-D, tandis que la matrice fondamentale agit sur des points image. Par exemple, le vecteur nul de droite de E_{ci} est $(0, R_{32}, -R_{31})$, ce qui représente la droite en 3-D avec les

coordonnées de Plücker $(0,0,0,R_{32},-R_{31},0)^T$. C'est la droite engendrée par les centres optiques des deux caméras (une droite à l'infini), exprimée par rapport au repère de la première caméra.

6.5. Estimation des matrices essentielles

Cet article a une vocation théorique, mais nous voulons néanmoins donner quelques éléments utiles pour l'estimation numérique des matrices essentielles introduites ainsi que pour l'extraction des paramètres de mouvement à partir de celles-ci. Pour les caméras perspectives, deux types majeurs d'algorithmes pour l'estimation de la matrice essentielle et du mouvement ont été développés. La première approche consiste à utiliser le moins de correspondances possible. Pour la matrice essentielle perspective, il s'agit de 5 correspondances de points, et seulement très récemment, un algorithme vraiment minimal a été trouvé [19]. Ce type d'algorithme correspond à la résolution d'équations non linéaires, donnant un nombre fini de solutions. L'autre type d'approche consiste en un premier lieu à ignorer la structure de la matrice essentielle (c'est-à-dire comment elle est construite à partir de R et \mathbf{t}) et de la traiter comme une matrice

de dimension 3×3 quelconque. La contrainte épipolaire permet alors de l'estimer en résolvant un système d'équations linéaires cette fois-ci [17]. Les prix à payer sont un nombre plus important de correspondances requises (8 au minimum) et que la matrice ainsi estimée ne correspond pas en général (en présence de bruit) à une matrice essentielle exacte.

Il existe aussi des algorithmes intermédiaires, mais ceci n'est pas important pour notre propos. Les deux types d'approche extraient finalement les paramètres de mouvement R et \mathbf{t} de la matrice essentielle estimée. Pour la deuxième approche, ceci requiert une solution approximative, puisque la matrice estimée ne respecte pas la structure d'une vraie matrice essentielle.

Dans la suite, nous esquissons la deuxième approche pour deux exemples de matrices essentielles introduites dans cette section ; le développement d'approches minimales (par exemple pour la matrice essentielle 6×6 des caméras non centrales) est un problème ouvert et probablement assez difficile.

Caméras non centrales. La matrice essentielle a 36 coefficients, mais 9 d'entre eux sont toujours zéro et 9 autres apparaissent en double (les coefficients de R). L'estimation linéaire doit alors porter sur 18 coefficients, et comme elle se fait à partir d'équation linéaires et homogènes, 17 correspondances de rayons de projection sont suffisantes.

Tableau 3. Résumé des matrices essentielles pour différents modèles de caméras. La dernière colonne donne le nombre minimum de correspondances de rayons de projections qui sont requises pour une estimation linéaire des matrices essentielles.

Modèle de caméras	Matrice essentielle	Nb.
Caméra non centrale	$E_n = \begin{pmatrix} -[\mathbf{t}] \times R & R \\ R & O_{3 \times 3} \end{pmatrix}$	17
Caméra axiale avec axe fini	$E_{af} = \begin{pmatrix} -[\mathbf{t}] \times R & \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \\ R_{31} & R_{32} \end{pmatrix} \\ \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \end{pmatrix} & O_{2 \times 2} \end{pmatrix}$	16
Caméra axiale avec axe infini	$E_{ai} = \begin{pmatrix} t_1 R_{32} - t_3 R_{12} & t_1 R_{33} - t_3 R_{13} & R_{21} & R_{22} & R_{23} \\ t_2 R_{12} - t_1 R_{22} & t_2 R_{13} - t_1 R_{23} & R_{31} & R_{32} & R_{33} \\ R_{12} & R_{13} & 0 & 0 & 0 \\ R_{22} & R_{23} & 0 & 0 & 0 \\ R_{32} & R_{33} & 0 & 0 & 0 \end{pmatrix}$	11
Caméra de type cross-slit avec deux axes finis	$E_{xff} = \begin{pmatrix} 1 & 0 & 0 & W_2 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -Y_2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} E_{af} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ W_1 & 0 & -Y_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	13
Caméra de type cross-slit avec un axe fini et un axe infini	$E_{xfi} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & W_2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} E_{af} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & W_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	10
Caméra centrale avec centre optique fini	$E_{cf} = -[\mathbf{t}] \times R$	8
Caméra centrale avec centre optique infini	$E_{ci} = \begin{pmatrix} t_2 R_{13} - t_1 R_{23} & R_{31} & R_{32} \\ R_{13} & 0 & 0 \\ R_{23} & 0 & 0 \end{pmatrix}$	4

Comme il a été mentionné ci-dessus, l'estimation linéaire ne donnera pas une matrice essentielle parfaite. Pour extraire les paramètres de mouvement R et \mathbf{t} , nous devons alors en tenir compte. Nous donnons l'esquisse d'un algorithme :

1. Soit A la sous-matrice 3×3 en bas à gauche de E_n . Sans bruit, elle serait égale (à un facteur d'échelle près) à la matrice de rotation R . En présence de bruit, nous pouvons obtenir une matrice de rotation R via la SVD (décomposition en valeurs singulières [26]) de A : si $A = U\Sigma V$ est la SVD de A , alors $R = UV$ est la matrice de rotation qui est la plus proche de A (au sens de la norme de Frobenius) [29]. Si le déterminant de R ainsi calculée vaut -1 , il faut encore multiplier la matrice avec -1 .

2. Déterminer le scalaire λ qui minimise

$$\|\lambda A - R\|_F$$

où $\|\cdot\|_F$ désigne la norme de Frobenius d'une matrice. C'est un problème de moindres carrés assez simple. Multiplier ensuite E_n avec λ .

3. Soit B la sous-matrice 3×3 en haut à gauche de E_n . Calculer \mathbf{t} en minimisant

$$\|B + [\mathbf{t}]_{\times} R\|_F$$

ce qui revient à la résolution d'un problème de moindres carrés.

D'autres algorithmes sont bien sûr possibles.

Caméras axiales avec axe fini. La matrice essentielle E_{af} contient 17 coefficients différents et peut alors être estimée en résolvant un système linéaire, à partir de 16 correspondances. L'algorithme d'extraction de R et \mathbf{t} est assez similaire au cas précédent. Soit A la sous-matrice 3×2 en haut à droite de E_{af} . Nous pouvons estimer R à partir de A basé sur sa SVD (voir les détails dans [29]). Le reste de l'algorithme est analogue à celui des caméras non centrales.

Autres cas. La plupart des autres cas (à l'exception de la caméra centrale avec centre optique infini) peuvent être résolus de manière assez similaire et nous en omettons les détails. La forme des matrices essentielles des différents modèles de caméras ainsi que le nombre minimum de correspondances requises pour une estimation linéaire, sont donnés dans le tableau 3.

7. Conclusions

Nous avons proposé une géométrie d'images multiples pour des caméras non centrales générales, la première d'après notre connaissance. Une hiérarchie naturelle de modèles de caméras a été introduite, regroupant les caméras en classes selon la répartition spatiale de leurs rayons de projection. La géométrie de deux vues a été spécialisée et décrite en détail pour différents modèles de caméras. Nous espérons que ce travail théorique

permet de définir un terrain commun pour de récents efforts dans la caractérisation de la géométrie de caméras non classiques.

Nous donnons quelques perspectives de travail assez immédiates. Les relations géométriques entre des caméras de différents types devraient être simples à dériver selon le schéma utilisé dans ce travail. Par exemple, la matrice essentielle entre une caméra centrale et une caméra axiale sera de taille 3×4 . Aussi, une traduction de nos résultats en notation tensorielle est aisément envisageable. Comme c'est le cas pour les caméras perspectives, des contraintes d'appariement pour des images de droites au lieu de points peuvent probablement être développées.

Dans cet article, nous avons principalement considéré la théorie de la géométrie d'images multiples ; quant à l'estimation numérique des tenseurs et l'extraction des paramètres de mouvement, nous avons esquissé des méthodes pour le cas de deux vues, mais un traitement complet nécessite plus de travail. Des expériences pratiques avec l'estimation de la matrice essentielle pour des caméras non centrales ont pourtant été achevées avec succès, comme ça l'a été le cas pour d'autres auteurs [25]. Sinon, nous travaillons actuellement sur le développement de méthodes de calibrage spécifiques aux caméras axiales et de type cross-slit, dans l'esprit de [31].

Références

- [1] Y.I. ABDEL-AZIZ et H.M. KARARA, «Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry», *ASP/UI Symposium on Close-Range Photogrammetry, Urbana, Illinois*, 1-18, 1971.
- [2] S. BAKER et S.K. NAYAR, «A theory of single-viewpoint catadioptric image formation», *International Journal of Computer Vision*, 35(2): 1-22, 1999.
- [3] H. BAKSTEIN, «Non-central cameras for 3D reconstruction, Rapport Technique CTU-CMP-2001-21», Center for Machine Perception, Czech Technical University, Prague, 2001.
- [4] H. BAKSTEIN et T. PAJDLA, «An overview of non-central cameras», *Computer Vision Winter Workshop, Ljubljana, Slovénie*, 223-233, 2001.
- [5] J.-P. BARRETO et H. ARAUJO, «Paracatadioptric camera calibration using lines», *ICCV*, 1359-1365, 2003.
- [6] R. BENOSMAN et S.B. KANG (éditeurs), *Panoramic Vision Sensors, Theory, and Applications*, Springer Verlag, 2001.
- [7] E. BRASSART, L. DELAHOUCHE, C. CAUCHOIS, C. DROCOURT, C. PEGARD et E. MOUADDIB, «Experimental Results Got With the Omnidirectional Vision Sensor: SYCLOP», *IEEE Workshop on Omnidirectional Vision, Hilton Head, Caroline du Sud*, 145-152, 2000.
- [8] O. FAUGERAS et B. MOURRAIN, «On the geometry and algebra of the point and line correspondences between n images», *ICCV*, 951-956, 1995.
- [9] D. FELDMAN, T. PAJDLA et D. WEINSHALL, «On the epipolar geometry of the crossed-slits projection», *ICCV*, 988-995, 2003.

- [10] C. GEYER et K. DANILIDIS, «A unifying theory of central panoramic systems and practical applications», *ECCV*, 445-461, 2000.
- [11] C. GEYER et K. DANILIDIS, «Paracatadioptric camera calibration», *PAMI*, 24(5): 687-695, 2002.
- [12] C. GEYER et K. DANILIDIS, «Mirrors in Motion: Epipolar geometry and motion estimation», *ICCV*, 2: 766-773, 2003.
- [13] M.D. GROSSBERG et S.K. NAYAR, «A general imaging model and a method for finding its parameters», *ICCV*, 2: 108-115, 2001.
- [14] R.I. HARTLEY et R. GUPTA, «Linear pushbroom cameras», *ECCV*, 555-566, 1994.
- [15] R.I. HARTLEY et A. ZISSERMAN, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.
- [16] R.A. HICKS et R. BAJCSY, «Catadioptric sensors that approximate wide-angle perspective projections», *CVPR*, 545-551, 2000.
- [17] H.C. LONGUET-HIGGINS, «A computer program for reconstructing a scene from two projections», *Nature*, 293: 133-135, 1981.
- [18] J. NEUMANN, C. FERMÜLLER et Y. ALOIMONOS, «Polydioptric camera design and 3D motion estimation», *CVPR*, II: 294-301, 2003.
- [19] D. NISTÉR, «An efficient solution to the five-point relative pose problem», *CVPR*, II: 195-202, 2003.
- [20] T. PAJDLA, «Epipolar Geometry of Some Non-classical Cameras», *Computer Vision Winter Workshop, Bled, Slovenie*, 159-180, 2001.
- [21] T. PAJDLA, T. SVOBODA et V. HLAVAC, «Epipolar Geometry of Central Panoramic Cameras», Dans *Panoramic Vision: Sensors, Theory, and Applications*, R. Benosman et S.B. Kang (éditeurs), 85-114, Springer Verlag, 2001.
- [22] T. PAJDLA, «Geometry of two-slit camera», Rapport Technique CTU-CMP-2002-02, Center for Machine Perception, Czech Technical University, Prague, 2002.
- [23] T. PAJDLA, «Stereo with oblique cameras», *IJCV*, 47(1-3): 161-170, 2002.
- [24] S. PELEG, M. BEN-EZRA et Y. PRITCH, «Omnistereo: Panoramic stereo imaging», *PAMI*, 23(3): 279-290, 2001.
- [25] R. PLESS, «Using many cameras as one», *CVPR*, II: 587-593, 2003.
- [26] W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING et B.P. FLANNERY, *Numerical Recipes in C – The Art of Scientific Computing*, Cambridge University Press, 1992.
- [27] L.S. SHAPIRO, A. ZISSERMAN et M. BRADY, «3D Motion Recovery via Affine Epipolar Geometry», *IJCV*, 16(2): 147-182, 1995.
- [28] H.-Y. SHUM, A. KALAI et S.M. SEITZ, «Omnivergent stereo», *ICCV*, 22-29, 1999.
- [29] P. STURM, «Algorithms for Plane-Based Pose Estimation», *CVPR*, 1010-1017, 2000.
- [30] P. STURM, «Mixing catadioptric and perspective cameras», *Workshop on Omnidirectional Vision, Copenhagen*, 37-44, 2002.
- [31] P. STURM et S. RAMALINGAM, «A generic concept for camera calibration», *ECCV*, 1-13, 2004.
- [32] T. SVOBODA, *Central Panoramic Cameras: Design, Geometry, Egomotion*. Thèse de doctorat, Faculty of Electrical Engineering, Czech Technical University, Prague, 1999.
- [33] R. SWAMINATHAN, M.D. GROSSBERG et S.K. NAYAR, «A perspective on distortions», *CVPR*, II: 594-601, 2003.
- [34] J. YU et L. MCMILLAN, «General linear cameras», *ECCV*, 14-27, 2004.
- [35] A. ZOMET, D. FELDMAN, S. PELEG et D. WEINSHALL, «Mosaicing new views: The crossed-slit projection», *PAMI*, 25(6): 741-754, 2003.



Peter Sturm

Peter Sturm (ingénieur de l'Université de Karlsruhe 1994, DEA de l'INPG 1994, docteur de l'INPG 1997) est Chargé de Recherche à l'INRIA Rhône-Alpes depuis 1999. Pour sa thèse il a reçu le premier Prix de Thèse SPECIF. De 1997 à 1999, il a été Research Fellow à l'Université de Reading (Royaume-Uni).

P. Sturm est régulièrement membre des comités de programme des congrès majeurs de la vision par ordinateur et du traitement d'images et sera Area Chair pour ECCV 2006. Il est relecteur pour environ 20 revues et a co-organisé deux ateliers internationaux (ECCV-OMNIVIS en 2004 et ICCV-BenCOS en 2005). Il est Co-Chairman du groupe de travail III/1, Automatic Calibration and Orientation of Optical Cameras, de l'ISPRS (International Society for Photogrammetry and Remote Sensing).

P. Sturm est l'auteur de plus de 75 publications scientifiques apparues dans des revues ou actes de congrès avec comité de lecture.

Ses compétences scientifiques concernent principalement le calibrage et l'auto-calibrage de (systèmes de) caméras, la reconstruction 3-D à partir d'images ou de séquences vidéo et l'application de caméras omnidirectionnelles. Il travaille également sur la reconstruction de surfaces spéculaires à partir d'images et sur le suivi d'objets dans des séquences vidéo.



Srikumar Ramalingam

Srikumar Ramalingam (ingénieur d'Anna University, Inde, 1994, Master of Science de l'Université de Californie 2004) effectue actuellement un doctorat sous une convention de co-tutelle internationale entre l'Université de Californie à Santa Cruz et l'INP de Grenoble.

Ses domaines d'intérêt scientifiques sont l'infographie et la vision par ordinateur et il travaille actuellement sur les problèmes de calibrage et reconstruction 3-D pour des modèles de caméras généraux.

