

Une revue des derniers progrès en autocalibration de caméras CCD*

Recent Advances in Camera Auto Calibration

par Thomas SKORDAS

ITMI
61, chemin des Prés – BP 177
F-38244 Meylan cedex

Résumé

Le calibrage d'une caméra CCD consiste à déterminer la relation mathématique existant entre les coordonnées des points tridimensionnels (3D) de la scène et les coordonnées 2D de leurs points images. En considérant le modèle de caméra dit de *sténopé*, la relation entre les coordonnées des points 3D et leurs homologues images est exprimée sous la forme d'une matrice 3×4 appelée *matrice de transformation perspective*. Plusieurs méthodes existent permettant le calcul des éléments de cette matrice. Elles sont toutes basées sur la minimisation d'un critère linéaire (ou non) liant les coordonnées d'un certain nombre de points 3D appartenant à un objet étalon avec les coordonnées 2D de leurs correspondants extraits dans l'image. Ces méthodes sont généralement supervisées, nécessitent l'utilisation de mires spécifiques et supposent que le calibrage soit effectué au début de chaque session de travail et que la matrice de transformation perspective calculée reste stable au cours du temps. Leur emploi est donc mal adapté dans le cas de systèmes de vision active où les paramètres des caméras varient continuellement.

Récemment, plusieurs méthodes ont été proposées permettant d'*autocalibrer* une caméra (ou un système stéréo) : par autocalibrage nous désignons toute méthode permettant de calibrer automatiquement une caméra sans utiliser des objets (mires) spécifiques de calibrage. L'objectif de cet article est alors de présenter ces nouvelles méthodes et de décrire l'état actuel de leur avancement. Les méthodes existantes sont classifiées en trois catégories principales : les méthodes nécessitant d'abord le calcul de la géométrie épipolaire entre deux images successives d'une séquence, celles qui estiment à la fois les paramètres de transformation perspective des caméras et la structure 3D de la scène et enfin celles qui utilisent la connaissance a priori quant au déplacement effectué par une caméra pour estimer les paramètres de celle-ci. Les principes mathématiques de chacune de ces catégories sont présentés et les performances obtenues sont discutées.

Mots clés : Calibrage de caméras, calibrage stéréo, vision active, reconstruction projective.

Abstract

Calibrating a camera involves determining the mathematical relationship between the 3D co-ordinates of a scene point and the corresponding 2D co-ordinates of its image projection. By using a pinhole model of the camera, this relationship can be expressed in terms of a 3×4 matrix called the perspective transformation matrix. Several methods exist for the computation of this matrix. They are based on the minimisation of a linear (or non-linear) criterion relating the co-ordinates of 3D points belonging to a known object (a calibration set-up) with their 2D image projections. These methods are generally supervised in that they necessitate the use of special calibration set-ups and assume that the calibration may be carried out at the beginning of each working session. Their use is therefore not well suited to the case of active vision systems where camera parameters may be continually varying.

Recently, several methods have been proposed for auto calibration of a visual sensor or stereo system : by auto calibration we mean methods allowing one to automatically calibrate a camera without use of special calibration set-ups. The objective of this paper is thus to present these new methods and to describe their current status. These methods may be classified into three categories : methods first requiring the computation of the epipolar geometry between two consecutive images of a sequence; methods that simultaneously estimate the parameters of the perspective transformation matrices and the 3D structure of the scene; and finally methods that use the knowledge of a motion applied to the camera in order to estimate its calibration parameters. The mathematical principles behind each category of method are presented and the results obtained discussed.

Key words : Camera calibration, stereo calibration, active vision, projective reconstruction.

1. Introduction

Le calibrage d'une caméra CCD consiste à déterminer la relation mathématique existant entre les coordonnées des points tridimensionnels (3D) de la scène et les coordonnées 2D de leurs

points images. La détermination de cette relation est une étape nécessaire pour la vision tridimensionnelle, puisque sa connaissance permet d'inférer des informations 3D à partir des données 2D [Faugeras88]. Elle constitue en réalité le point initial pour plusieurs applications de la Vision par Ordinateur où l'on a besoin de la relation entre les images et le monde 3D, comme par exemple la reconnaissance et la localisation d'objets, le contrôle dimen-

* Étude partiellement financée par le Centre National des Études Spatiales – CNES.

sionnel de pièces, la reconstruction de l'environnement pour la navigation d'un robot mobile, etc.

Le modèle mathématique de caméra le plus souvent utilisé en Vision par Ordinateur est le modèle sténopé (« pinhole camera model ») : la caméra est supposée effectuer une transformation perspective parfaite. Selon ce modèle, la relation entre les coordonnées 3D de la scène et leurs homologues images est exprimée sous la forme d'une matrice 3×4 appelée *matrice de transformation perspective*. Cette matrice est composée d'un certain nombre de paramètres optiques internes de la caméra (d'habitude 5 paramètres sont utilisés, communément appelés *paramètres intrinsèques*), de la matrice rotation et du vecteur translation liant le repère associé à la caméra à un repère associé à la scène (communément appelés *paramètres extrinsèques*). En général, un système permettant de calibrer une caméra CCD comprend les éléments suivants :

- Une mire de calibrage (grille ou objet étalon) composée de points dont on connaît précisément les coordonnées 3D dans un repère lié à la mire;
- Un système d'acquisition d'images pour la numérisation et la mémorisation des images de la mire;
- Un algorithme d'extraction des points 2D correspondant aux points 3D de la mire.
- Un algorithme de mise en correspondance des points 2D détectés dans les images avec leurs homologues de la mire.
- Un algorithme de calcul de la matrice de transformation perspective de la caméra liant le repère associé au plan image au repère de coordonnées associé à la mire de calibrage.

Ce processus classique de calibrage est un processus supervisé nécessitant souvent l'intervention de l'opérateur. C'est un processus statique qui s'effectue hors-ligne, avant l'utilisation de la caméra pour une tâche visuelle précise. Une fois la caméra calibrée, ses paramètres ne doivent plus être modifiés en cours d'utilisation. Chaque fois que l'on désire modifier sa mise au point, sa distance focale ou son ouverture du diaphragme, il faudra recalibrer la caméra. Plusieurs travaux de la dernière décennie ont permis d'obtenir des méthodes de calibrage fournissant une très bonne précision et ont démontré que ce processus est bien adapté lorsqu'il s'agit de faire de la métrologie précise dans des conditions bien contrôlées (voir par exemple [ToscaniFaugeras86], [Tsai87], [Tsai89], [Beyer92], mais aussi [MongaHoraud93]).

Ce processus par contre est mal adapté pour être utilisé dans le cas d'un système de vision active (monoculaire ou binoculaire) : un tel système, équipé d'un support robotisé, est sensé être capable d'effectuer un certain nombre de tâches visuelles sans interférence fréquente avec un opérateur humain. Le système doit être capable de se recalibrer rapidement lorsqu'il modifie ses paramètres optiques (zoom, mise au point, ouverture, etc.) pour s'adapter aux changements rapides de l'environnement (dus à la fois aux besoins de la tâche visuelle en cours d'exécution ou aux changements des conditions climatiques ou de température, à la baisse de luminosité, etc.).

Le développement récent de systèmes de vision active d'une part, le fait que la méthode classique de calibrage est assez fastidieuse à mettre en œuvre et demande des dispositifs de laboratoire assez perfectionnés d'autre part, ont motivé le développement de nouvelles méthodes de calibrage nécessitant une intervention minimale de l'opérateur et où le calibrage s'effectue en-ligne. D'où la naissance d'un certain nombre de travaux autour du thème de *l'autocalibrage des caméras* : par autocalibrage nous désignons toute méthode permettant de calibrer automatiquement une caméra (sans aucune intervention manuelle) et sans utiliser des objets spécifiques de calibrage (en particulier sans utilisation de mires). Comme nous allons le voir, afin d'effectuer l'autocalibrage, certaines méthodes s'appuient sur des informations disponibles uniquement dans les images (points de correspondance). D'autres méthodes exploitent une connaissance a priori sur la scène observée : par exemple, la connaissance d'un certain nombre de points scènes coplanaires, ou des points scènes étant très éloignés par rapport au capteur. Enfin, d'autres méthodes exploitent la connaissance (partielle) du mouvement appliqué au capteur.

A l'heure actuelle, trois approches d'autocalibrage se distinguent clairement : elles font toutes l'hypothèse que le problème de correspondance entre des images prises par une caméra en mouvement par rapport à une scène statique est résolu au préalable et qu'un certain nombre de correspondances est alors déjà disponible.

A. La première catégorie de méthodes est celle conçue par Faugeras et son équipe [FaugerasLuongMaybank92], [MaybankFaugeras92], [Luong92], [LuongFaugeras93], [LuongDericheFaugerasPapadopoulo93], [ZhangLuongFaugeras93]. Le calibrage est effectué en deux étapes : d'abord, la géométrie épipolaire entre deux images successives d'une séquence acquise par une caméra en mouvement est déterminée sur la base de correspondances disponibles entre les deux images. Pour cela au minimum huit correspondances images sont nécessaires pour obtenir une solution unique. Utilisant ensuite des techniques de géométrie projective, chaque déplacement de caméra fournit deux équations quadratiques par rapport à cinq coefficients indépendants appelés *coefficients de Kruppa* qui sont directement liés avec les cinq paramètres intrinsèques de la caméra. Trois positions différentes de la caméra sont alors nécessaires afin de calculer ces coefficients. Une fois les coefficients de Kruppa calculés, les paramètres intrinsèques de la caméra peuvent être obtenus. Hartley dans [Hartley92] reprend la même philosophie en ce qui concerne la détermination de la géométrie épipolaire. Compte tenu du fait qu'il est impossible de retrouver l'ensemble des paramètres intrinsèques de la caméra à partir uniquement de la connaissance de la géométrie épipolaire, Hartley a considéré comme connus trois parmi les cinq paramètres intrinsèques de la caméra (fournis par le constructeur). Il calcule alors les deux autres paramètres, ainsi que les paramètres de mouvement de la caméra.

B. Il existe deux inconvénients principaux dans la méthode précédente. D'abord, le problème de calcul de la géométrie épipolaire entre deux vues successives est un processus assez instable numériquement. Deuxièmement, le calcul des paramètres intrinsèques des caméras à partir de trois positions différentes de celles-ci est également difficile à réaliser. Pour pallier ces inconvénients, la seconde catégorie de méthodes considère à la fois les problèmes de reconstruction de scène, de détermination des paramètres intrinsèques d'une caméra et de son mouvement en un seul problème global. Sur la base des correspondances établies entre les différentes images de la séquence, le problème est formulé comme un problème de minimisation globale par rapport à l'ensemble des paramètres. Un grand nombre d'images et de données est alors nécessaire. A l'origine de ce type de méthodes est l'équipe de R. Mohr du LIFIA [MohrVeillonQuan93], [BoufamaMohrVeillon93], [MohrBoufamaBrandt94]. Cette classe de méthodes dépasse largement le simple contexte d'autocalibrage d'une caméra. Toutefois, nous la référençons dans cet article, du fait de son intérêt pour la détermination des paramètres de calibrage.

C. La troisième catégorie est constituée des méthodes utilisées par les systèmes de vision active. La différence essentielle entre un système comportant une caméra mobile et un système actif est que le système actif peut effectuer un certain nombre de mouvements connus avec grande précision. La connaissance de ces mouvements est exploitée par ce troisième type de méthodes : on fait subir à la caméra un certain nombre de mouvements connus et on acquiert des images de points statiques de la scène. Ces points sont poursuivis dans les images et leurs trajectoires observées sont reliées aux paramètres intrinsèques de la caméra. La connaissance de plusieurs points poursuivis ainsi, permet d'obtenir un grand nombre d'équations, dont la résolution permet le calcul des paramètres de la caméra [Basu93], [Dron93], [DuBrady93], [McLauchlanMurray93], [Viéville94].

Cet article présente ces trois types de méthodes, décrit leurs fondations mathématiques et traite de leur précision et de leur complexité de mise en œuvre.

2. Plan de l'article

Le plan de l'article est le suivant :

- La section 3 est une introduction générale sur le problème du calibrage de caméras.
- La section 4 s'intéresse au premier type de méthodes d'autocalibrage qui sont fondées sur le calcul de la géométrie épipolaire. Elle présente les critères utilisés pour le calcul de cette géométrie et discute leur domaine de validité.

- La section 5 présente le second type d'approches que nous appelons *méthodes de reconstruction relative de scènes*. Compte tenu du fait que ces méthodes aboutissent à une reconstruction de scène valable à une transformation projective près, nous présentons également les critères utilisés afin d'aboutir à une reconstruction euclidienne de la scène.
- La section 6 s'intéresse aux méthodes d'autocalibrage appliquées dans le cadre de systèmes de vision active. Plusieurs méthodes sont décrites et leurs résultats sont commentés.
- Enfin, la section 7 récapitule les méthodes d'autocalibrage présentées dans cet article et discute leurs points forts et inconvénients principaux.

Afin de faciliter la lecture de cet article, dans l'Annexe A nous présentons les principaux symboles et notations utilisés tout au long de celui-ci.

3. Calibrage de caméras CCD Quelques rappels

Dans cette section, nous présentons rapidement le problème du calibrage classique d'une caméra CCD et nous donnons quelques définitions d'objets mathématiques que nous utiliserons continuellement dans cet article. Pour plus de détail sur ce problème, le lecteur intéressé peut se référer à [ToscaniFaugeras86], [FaugerasToscani87], [Tsai87], [Tsai89], [PugetSkordas90], [Beyer92] et [MongaHoraud93].

La modélisation de la caméra la plus utilisée est du type modèle sténopé (« pinhole camera ») : tout point P de l'espace se projette en un point p , obtenu par intersection de la droite $[CP]$ (rayon optique) avec le plan rétinien de la caméra. C est un point fixe lié à la caméra, appelé *centre optique* de la caméra. La figure 3.1 illustre cette modélisation :

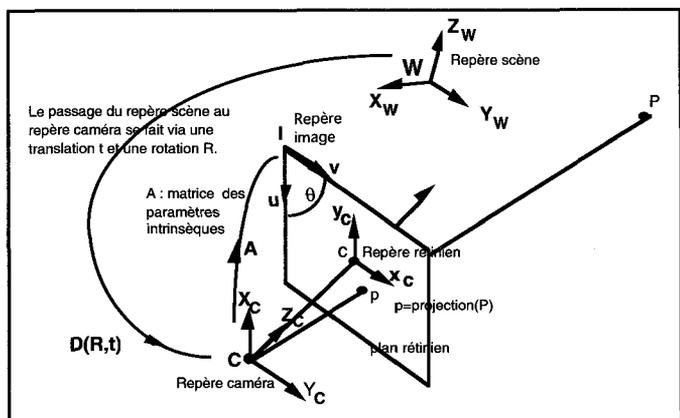


Figure 3.1. – Le modèle sténopé d'une caméra CCD

Soit (C, X_C, Y_C, Z_C) un repère orthonormé associé au centre optique, appelé *repère caméra*. L'axe Z_C est choisi de façon à

être perpendiculaire au plan rétinien et définit l'axe optique de la caméra. Les deux autres axes X_C, Y_C se trouvent par conséquent dans un plan parallèle au plan rétinien et constituent une base orthonormée de ce plan. L'intersection de l'axe Z_C avec le plan rétinien est le point c appelé *point principal*. Le centre optique et le point principal sont distants de la distance focale f de la caméra. En choisissant les unités du repère caméra de sorte que la distance focale soit égale à 1 dans ce repère, nous introduisons un *repère caméra normalisé*. Au point principal on associe un système de coordonnées (c, x_c, y_c) d'unités égales à celles du repère caméra et d'axes parallèles aux axes X_C et Y_C du repère caméra. Ce repère est appelé *repère rétinien* qui est en réalité un repère rétinien normalisé. Soit (W, X_W, Y_W, Z_W) un repère associé à la scène (appelé par la suite *repère scène*) et soit enfin (I, u, v) un système de coordonnées pixel dans lequel sont mesurés les points des images finalement obtenues. Ce repère est appelé *repère image*.

Le modèle sténopé est alors défini par ses paramètres extrinsèques (transformation géométrique entre les repères scène et caméra) et ses paramètres intrinsèques : les coordonnées pixel du point principal dans le repère image, les deux facteurs d'échelle selon les axes u et v du repère image et l'angle θ formé par ces deux axes (voir Annexe C).

Le calibrage d'une caméra CCD consiste à calculer l'ensemble de ses paramètres intrinsèques et extrinsèques. Le calibrage est effectué par le calcul de la matrice de transformation perspective M (c'est à dire une matrice 3×4) liant le repère scène (W, X_W, Y_W, Z_W) au repère image (I, u, v) . Si $P = (X, Y, Z, 1)^t$ est un point observé de l'espace qui se projette dans l'image au point p de coordonnées projectives $(p_u, p_v, p_w)^t = (su, sv, s)^t$ alors on a (rappelons que selon les propriétés présentées dans l'Annexe B, deux triplets de coordonnées (y_1, y_2, y_3) et (u_1, u_2, u_3) spécifient le même point image si et seulement si il existe un facteur d'échelle λ non nul tel que $y_i = \lambda u_i$) :

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = M \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

En introduisant les coordonnées projectives, nous avons implicitement adopté un modèle projectif de la caméra. Nous allons utiliser ce modèle tout au long de cet article. Comme nous le démontrons dans l'Annexe C, son intérêt principal réside dans le fait qu'il rend linéaire la relation entre les coordonnées projectives du repère scène et les coordonnées projectives images. Cette propriété demeure vérifiée quels que soient les repères rétinien et scène choisis. Grâce à cette propriété, nous avons ainsi la possibilité d'utiliser tous les puissants outils offerts par la géométrie projective (dont une introduction est présentée dans l'Annexe B).

Revenons maintenant sur la relation (1). En éliminant le facteur d'échelle s , on obtient :

$$\left. \begin{aligned} u &= \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\ v &= \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \end{aligned} \right\} \quad (2)$$

où m_{ij} sont les éléments de la matrice M .

Chaque point de la scène dont on connaît les coordonnées et les points projetés, fournit 2 relations sur M . Il nous faut donc au moins 6 points pour déterminer M . Ces points doivent être correctement répartis dans la scène et ne pas être coplanaires. Dans ce cas, il est possible de minimiser le critère linéaire suivant ([ToscaniFaugeras86] et [Tsai87]) :

$$\text{nbre points} \sum_{i=1} \left\{ \begin{aligned} &[(m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34})u_i - \\ &(m_{11}X_i + m_{12}Y_i + m_{13}Z_i + m_{14})]^2 + \\ &[(m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34})v_i - \\ &(m_{21}X_i + m_{22}Y_i + m_{23}Z_i + m_{24})]^2 \end{aligned} \right\}$$

La solution $M = 0$ convient. Il faut donc minimiser ce critère sous une contrainte.

La première contrainte étudiée consistait à poser $m_{34} = 1$. Les résultats obtenus étaient plus que médiocres, [ToscaniFaugeras86]. Faugeras et Toscani dans [FaugerasToscani87] ont alors proposé de prendre : $\|m_3\| = 1$ où $m_3 = (m_{31}, m_{32}, m_{33})^t$. Cette méthode conduit à trouver le vecteur propre correspondant à la plus petite valeur propre d'une matrice (3×3) fonction des $(u_i, v_i, X_i, Y_i, Z_i)$.

De la méthode [FaugerasToscani87], il existe un certain nombre de perfectionnements : l'utilisation d'un filtre de Kalman permet par exemple d'obtenir une mesure de l'incertitude des paramètres obtenus [Ayache89]. Luc Robert [Robert93] a proposé une formulation différente du critère en optimisant directement les coefficients de la matrice de projection perspective à partir de valeurs de gradient dans l'image et du modèle 3D de la mire. L'expérience de l'INRIA en calibrage de caméras est résumée dans la thèse de Luong [Luong92] : « concernant les méthodes améliorées, les écarts obtenus sur les paramètres intrinsèques sont à présent inférieurs à 1% sur les facteurs d'échelle et à quelques pixels sur le point principal. Ces méthodes peuvent donc permettre d'obtenir une très bonne précision dans les cas favorables où la mire occupe la plus grande partie de l'image, et est bien positionnée ».

Les méthodes de Beyer constituent aujourd'hui pratiquement la référence dans le domaine [Beyer92]. Beyer a obtenu récemment en laboratoire d'excellentes performances (précision relative atteinte sur un objet 1 : 50000, ce qui signifie que l'on peut déterminer les coordonnées 3D sur un objet de taille 5 m avec une précision de 0.1 mm). Plusieurs images des points scènes sont utilisées. Les points images sont détectés avec une précision maximale de 0.05 pixels (en utilisant une technique de mise en correspondance

aux moindres carrés). L'introduction d'une modélisation des distorsions optiques de la caméra et l'utilisation de techniques de photogrammétrie lui ont permis d'obtenir la précision annoncée précédemment.

Comme il n'est pas réaliste de déplacer le robot en conservant la mire dans le champ visuel, l'emploi de méthodes de calibrage classique suppose que le calibrage soit effectué au début de chaque session de travail et que les paramètres intrinsèques de la caméra restent stables au cours du temps, or cette hypothèse peut se révéler fautive dans certains cas. En effet, ces paramètres peuvent changer avec les variations de température, les conditions climatiques, les chocs qu'une caméra embarquée subit, etc.; par conséquent, il convient de vérifier souvent leur validité et de les remettre à jour le cas échéant.

De plus, l'emploi de méthodes de calibrage classique est mal adapté au cas d'un système de vision active : un tel système est composé de caméras (une au minimum) dont les paramètres sont changés dynamiquement (en cours de fonctionnement) afin de mieux s'adapter aux besoins d'une tâche visuelle en cours (comme par exemple le changement de distance focale afin de mieux suivre un objet s'éloignant ou se rapprochant à la caméra, etc.). De tels systèmes nécessitent alors une intervention minimale de l'opérateur humain et la possibilité d'effectuer en ligne le calibrage de leurs paramètres.

Ces deux raisons importantes ont alors motivé le développement de travaux d'autocalibrage de caméras.

4. Méthodes d'autocalibrage fondées sur le calcul de la géométrie épipolaire

4.1. INTRODUCTION

Les travaux de cette section sont basés sur le calcul des paramètres intrinsèques d'une caméra en deux étapes successives :

- Dans un premier temps et sur la base d'un certain nombre de correspondances considérées connues entre les images d'une séquence, on calcule la géométrie épipolaire entre deux images successives de la séquence : cette géométrie épipolaire est représentée par une matrice 3×3 appelée *matrice fondamentale*. Cette matrice étant de rang deux et définie à un coefficient multiplicatif près, au minimum 8 correspondances sont nécessaires pour la calculer de façon unique. Le paragraphe 4.2 est entièrement consacré aux méthodes qui ont été développées pour le calcul de la matrice fondamentale.
- Ayant calculé la matrice fondamentale entre deux images successives de la séquence, la relation explicite entre la matrice fondamentale et les paramètres intrinsèques de la caméra est ensuite utilisée pour calculer ces paramètres intrinsèques. Des techniques de la géométrie projective sont alors utilisées pour établir cette relation. Le paragraphe 4.3 présente ces techniques

et discute leur stabilité selon les différents types de mouvement de la caméra.

Connaissant la matrice fondamentale et les paramètres intrinsèques d'une caméra, il est possible de reconstruire la structure euclidienne de la scène. Dans le cas où on ne connaît que la matrice fondamentale liant deux positions successives de cette caméra, il n'est alors possible que de reconstruire la structure projective de la scène. Le paragraphe 4.4 présente la façon dont on peut effectuer cette reconstruction projective et explicite aussi la relation existant entre la matrice fondamentale et la matrice de transformation perspective d'une caméra. Un bon exemple de l'utilisation de l'ensemble des notions et techniques décrites dans cette section est présenté dans le paragraphe 4.5 qui traite du problème de l'autocalibrage d'un système stéréo non calibré en mouvement.

Plusieurs démonstrations de cette section sont appuyées sur des considérations provenant de la géométrie projective. Le lecteur intéressé pourra se référer à [Mohr92], [MohrVeillon92] pour une introduction et à [SempleKneebone52] pour une présentation beaucoup plus détaillée de ce puissant outil. Dans l'Annexe B nous présentons une introduction sommaire de la géométrie projective.

4.2. AUTOCALIBRAGE D'UNE CAMÉRA CALCUL DE LA MATRICE FONDAMENTALE

Considérons le cas d'une caméra en mouvement et deux images prises par celle-ci dans deux positions successives au cours du mouvement. Ces deux images peuvent être considérées comme formant une paire stéréo. A des fins d'une démonstration plus globale, considérons le cas d'un vrai système stéréo composé de deux caméras différentes. Sa configuration épipolaire est présentée Figure 4.1. Le point 3D P et ses deux projections images p et p' définissent un plan qui coupe les plans images gauche et droit respectivement en deux lignes l et l' . Quand P décrit l'espace, toutes ces droites passent par l'intersection de la droite CC' avec le plan image gauche, que nous définissons comme l'épipole gauche e . De manière similaire est défini l'épipole droit e' .

Soient 4 points P_i de l'espace qui donnent 4 couples de droites épipolaires en correspondance ($l_i \leftrightarrow l'_i$). Ces droites épipolaires sont obtenues par intersection d'un faisceau de quatre plans passant par CC' avec les plans images gauche et droit. Selon les propriétés de la géométrie projective présentées dans l'Annexe B, ces deux faisceaux des lignes épipolaires sont en correspondance homographique, car obtenues comme intersection du même faisceau de quatre plans (celui passant par CC') avec deux plans différents de l'espace (les plans images gauche et droit). En effet, lorsqu'un faisceau de quatre plans est coupé par un plan quelconque de l'espace en quatre droites, par définition, le bi-rapport associé à ce faisceau de plans est égal au bi-rapport des quatre droites. Par ailleurs, le bi-rapport associé à un faisceau de plans reste invariant par toute intersection de ce faisceau avec un plan

quelconque de l'espace. Par conséquent, le bi-rapport du faisceau des quatre plans passant par CC' est égal au bi-rapport des quatre droites épipolaires de l'image gauche, mais aussi à celui des quatre droites épipolaires de l'image droite. Toute application projective laissant le bi-rapport d'un faisceau de plans invariant, les deux faisceaux de droites épipolaires sont par définition en correspondance homographique. La relation entre ces droites épipolaires est appelée *transformation épipolaire*.

En conclusion, si nous connaissons 3 droites épipolaires en correspondance et si nous nous donnons une autre droite épipolaire dans l'un des deux plans images, alors nous sommes capables de trouver la droite épipolaire qui lui est associée dans l'autre image : il suffit pour ceci de calculer le bi-rapport.

Le problème de la géométrie épipolaire est un problème à 7 degrés de liberté : 2×2 pour les coordonnées images des épipoles e et e' et trois pour l'homographie entre les faisceaux de droites épipolaires (si τ et τ' sont les directions des droites épipolaires en correspondance, les paramètres (a, b, c, d) de cette homographie relient ces directions : $\tau' = (a\tau + b)/(c\tau + d)$). Parmi les quatre paramètres définissant cette homographie, il n'y a que trois qui sont indépendants, voir [MaybankFaugeras92], [FaugerasLuongMaybank92]).

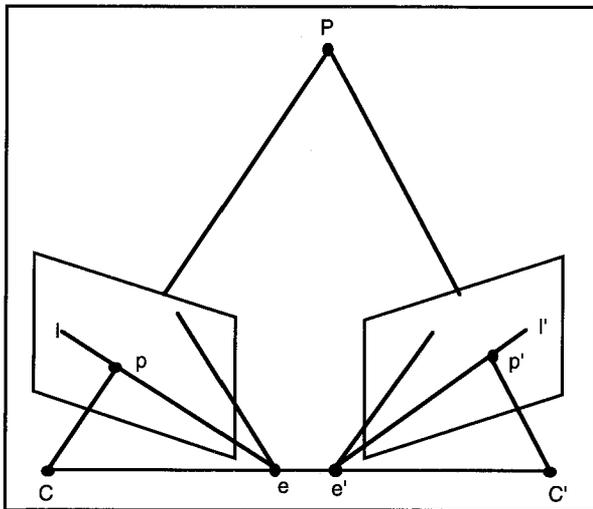


Figure 4.1. – Lignes épipolaires pour un système stéréo

4.2.1. Calcul de la géométrie épipolaire en utilisant la géométrie projective.

Dans ce qui suit, nous présentons une construction de la géométrie épipolaire en utilisant les propriétés de la géométrie projective. Une telle construction est présentée dans [Hartley92], [Mohr92] et [Luong92].

Soit $p = (p_u, p_v, p_w)^t$ un point de l'image gauche et $e = (e_u, e_v, e_w)^t$ l'épipôle gauche, tous deux exprimés en coordonnées projectives images. Les coordonnées projectives (l_u, l_v, l_w) de la droite épipolaire l passant par p et e s'écrivent

$l = e \wedge p = (l_u, l_v, l_w)^t$. L'application $p = (p_u, p_v, p_w)^t \rightarrow e \wedge p = (l_u, l_v, l_w)^t$ est linéaire et peut être représentée par une matrice T_e de rang 2 (matrice antisymétrique) :

$$l = \begin{pmatrix} l_u \\ l_v \\ l_w \end{pmatrix} = \begin{pmatrix} p_w e_v - p_v e_w \\ p_u e_w - p_w e_u \\ p_v e_u - p_u e_v \end{pmatrix} = \begin{pmatrix} 0 & -e_w & e_v \\ e_w & 0 & -e_u \\ -e_v & e_u & 0 \end{pmatrix} \begin{pmatrix} p_u \\ p_v \\ p_w \end{pmatrix} = T_e \begin{pmatrix} p_u \\ p_v \\ p_w \end{pmatrix} \Rightarrow l = T_e p$$

Comme nous avons vu dans le paragraphe précédent, il existe une application projective (homographie) qui transforme une ligne épipolaire de l'image gauche en une ligne épipolaire de l'image droite. Soit Q cette application. On a :

$$l' = Ql$$

et donc :

$$l' = Ql = QT_e p \quad (3)$$

Le point p' (correspondant du point p) appartient à la ligne épipolaire l' et donc $p'^t l' = 0$. On obtient finalement :

$$p'^t Q T_e p = 0$$

ou

$$p'^t F p = 0 \quad (4)$$

où F est la matrice $F = QT_e$, F étant appelée *matrice fondamentale*. Elle est de rang 2, puisque Q est de rang 3 et T_e est de rang 2. Comme l'épipôle est le noyau de la matrice T_e (évidemment on a $T_e e = 0$, e est également le noyau de F et donc $F e = 0$).

F est une matrice 3×3 définie à un coefficient près. Elle a 7 degrés de liberté (2 ddl pour T_e qui est la position de l'épipôle gauche et 5 ddl pour Q dont 2 pour la position de l'épipôle droit et 3 pour l'homographie entre les faisceaux de droites épipolaires que nous avons vue dans le paragraphe précédent).

Finalement observons que si l'on considère l'équation (3), la matrice fondamentale F est une application linéaire qui projette le point p de l'image gauche sur la ligne épipolaire $l' = Fp$ de l'image droite.

4.2.2. Calcul de la géométrie épipolaire en utilisant une méthode classique

Soit de nouveau le système stéréo de la Figure 4.1, repris dans la Figure 4.2. Soit (R, t) la transformation géométrique (rotation, translation) liant les repères associés aux deux caméras. On a : $C' = RC + t$.

Soit P un point de l'espace. Soient p_n et p'_n ses projections images gauche et droite, p_n et p'_n étant respectivement exprimés dans les repères rétiniens gauche et droit normalisés (voir Annexe C pour plus de détail). p_n et p'_n sont en correspondance si et seulement si les trois vecteurs Cp_n , $C'p'_n$ et $t = C'C$ sont coplanaires.

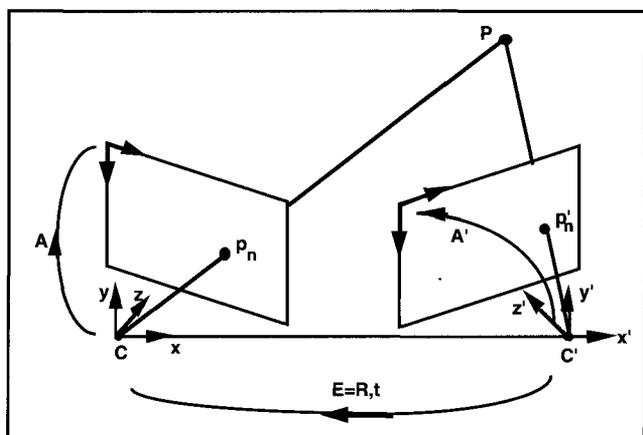


Figure 4.2. – La contrainte de Longuet-Higgins

Exprimons vectoriellement cette remarque dans le repère de la caméra droite : $C'p'_n$ doit appartenir au plan défini par CC' et $R Cp_n$. La contrainte de coplanarité s'écrit donc :

$$C' p'_n \cdot (t \wedge R C p_n) = 0$$

ou de façon équivalente :

$$p'_n \cdot (t \wedge R p_n) = 0 \quad (5)$$

Cette dernière relation est connue sous le nom de Longuet-Higgins [Longuet-Higgins81].

Soit T_t la matrice antisymétrique associée au vecteur translation t , telle que $T_t x = t \wedge x$:

$$T_t = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

En posant $E = T_t R$ l'équation (4.3) se réécrit :

$$p_n'^t E p_n = 0 \quad (6)$$

La matrice E est appelée *matrice essentielle*. Elle est le produit d'une matrice antisymétrique (T_t) par une matrice de rotation (R), donc son rang est 2.

Soient maintenant p et p' les coordonnées projectives images des points p_n et p'_n . Si A et A' sont les matrices des paramètres intrinsèques des caméras gauche et droite respectivement, on a (voir Annexe C) :

$$p = A p_n \text{ et } p' = A' p'_n$$

L'équation de Longuet-Higgins peut alors se réécrire :

$$p'^t F p = 0 \quad (7)$$

où nous avons posé :

$$F = (A'^{-1})^t E A^{-1} \quad (8)$$

Les équations (4) et (8) sont identiques.

La matrice F est la matrice fondamentale. Cette matrice vérifie donc l'équation $p'^t F p = 0$ pour tout couple de points (p, p') appariés dans chaque plan image, et exprimé dans les repères projectifs images. Son rang est celui de la matrice E , les matrices A et A' étant inversibles, donc le rang de F est 2. Notons que F est une matrice (3×3) , définie à un coefficient multiplicatif près, donc dépend de $3 \times 3 - 1 = 8$ paramètres. Le fait que le rang de F soit 2 rajoute une contrainte, donc F ne dépend réellement que de 7 paramètres. Tout appariement de points (p, p') fournit une seule relation sur F : $p'^t F p = 0$ donc il faut connaître au moins 8 couples de points pour la déterminer de manière unique [Luong92].

Comme nous avons vu, la matrice $E = T_t R$ est appelée matrice essentielle du système. F et E représentent donc la même chose, mais dans des bases différentes.

$F = (A'^{-1})^t E A^{-1}$ donc $E = A'^t F A$. Remarquons que : $EE^t = T_t R R^t T_t^t = T_t R R^{-1} (-T_t) = -T_t^2$.

Calcul des épipodes

Nous avons vu précédemment que l'épipole est le noyau de la matrice fondamentale F . En effet, e' appartient à la droite épipolaire $l' = Fp$ et donc vérifie pour tout point p : $e'^t l' = 0$ soit $e'^t F p = 0$. Donc :

$$\forall p \quad e'^t F p = 0$$

ce qui signifie que $e'^t F = 0$ et donc $F^t e' = 0$.

De façon similaire, on peut démontrer que l'épipole gauche vérifie : $F e = 0$.

Du fait d'approximations numériques, il arrive souvent que le déterminant de la matrice F calculée ne soit pas nul. Dans ce cas, les relations précédentes n'ont plus de solution exacte, c'est à dire qu'il n'existe pas de points e et e' vérifiant les équations ci-dessus. Dans ce cas, les épipoles sont calculés en résolvant le problème de minimisation de $\|F e\|^2$ sous la contrainte $\|e\|^2 = 1$. La solution e est le vecteur propre associé à la plus petite valeur propre de $F^t F$. De même, e' est celui associé à la plus petite valeur propre de $F F^t$.

4.2.3. Calcul de la matrice fondamentale

Ce paragraphe s'appuie sur les articles suivants : [Hartley92], [Luong92], [LuongFaugeras93], [LuongDericheFaugerasPapadopoulos93] et [MohrVeillonQuan93].

L'équation vérifiée par F s'écrit pour tout couple de points images appariés (p, p') : $p'^t F p = 0$. De plus, F est de rang 2. Nous avons vu qu'au minimum huit correspondances images sont nécessaires pour la calculer.

4.2.3.1. Critère linéaire

L'équation $\mathbf{p}^t \mathbf{F} \mathbf{p} = \mathbf{0}$ peut se réécrire sous la forme :

$$\mathbf{U}^t \mathbf{a} = \mathbf{0}$$

où :

$$\mathbf{U} = [\mathbf{u}\mathbf{u}', \mathbf{v}\mathbf{v}', \mathbf{u}'\mathbf{u}, \mathbf{v}'\mathbf{v}, \mathbf{u}'\mathbf{v}, \mathbf{v}'\mathbf{u}, \mathbf{1}]^t$$

$$\mathbf{a} = [\mathbf{F}_{11}, \mathbf{F}_{12}, \mathbf{F}_{13}, \mathbf{F}_{21}, \mathbf{F}_{22}, \mathbf{F}_{23}, \mathbf{F}_{31}, \mathbf{F}_{32}, \mathbf{F}_{33}]^t$$

avec $\mathbf{p} = (u, v, 1)^t$ et $\mathbf{p}' = (u', v', 1)^t$. Chaque couple d'appariements $(\mathbf{p}, \mathbf{p}')$ fournit donc une équation de type $\mathbf{U}^t \mathbf{a} = \mathbf{0}$ qui est une équation linéaire et homogène en les neuf éléments de la matrice \mathbf{F} . Si l'on dispose de huit appariements $(\mathbf{p}, \mathbf{p}')$, on peut trouver une solution unique pour \mathbf{F} .

En pratique, on dispose d'un nombre n bien plus grand d'appariements images $(\mathbf{p}_i, \mathbf{p}'_i)$ ($i = 1, \dots, n$). Alors, on peut résoudre les équations $\mathbf{U}^t \mathbf{a} = \mathbf{0}$ au sens des moindres carrés, ce qui revient à minimiser le critère $\min_{\mathbf{F}} [\sum_{i=1}^n (\mathbf{p}_i^t \mathbf{F} \mathbf{p}_i)^2]$, réécrit comme $\min_{\mathbf{a}} \|\tilde{\mathbf{U}} \mathbf{a}\|^2$ sous la contrainte $\|\mathbf{a}\|^2 = 1$, où :

$$\tilde{\mathbf{U}} = \begin{bmatrix} U_1^t \\ \vdots \\ U_n^t \end{bmatrix}$$

La solution \mathbf{a} est le vecteur propre associé à la plus petite valeur propre de la matrice $\tilde{\mathbf{U}}^t \tilde{\mathbf{U}}$.

La méthode utilisée pour résoudre ce critère est non itérative (« closed form ») et le calcul est très sensible au bruit, même avec l'utilisation d'un grand nombre d'appariements. Il a été constaté (voir par exemple [Luong92] et [LuongDericheFaugerasPapadopoulou93]) que la contrainte $\det(\mathbf{F}) = \mathbf{0}$ (rang de \mathbf{F} égal 2) n'est pas satisfaite. Par ailleurs, le critère n'étant pas normalisé, un biais est introduit sur la localisation des épipoles. Pour pallier ce problème, d'autres critères cette fois-ci non linéaires sont introduits.

4.2.3.2. Critères non linéaires

Soit $(\mathbf{p}_i, \mathbf{p}'_i)$, ($i = 1, \dots, n$) un ensemble d'appariements images. L'équation $\mathbf{p}_i^t \mathbf{F} \mathbf{p}_i = 0$ est directement liée avec la distance d du point \mathbf{p}'_i par rapport à sa droite épipolaire $\mathbf{l}'_i = \mathbf{F} \mathbf{p}_i$. La première idée serait alors d'utiliser un critère non linéaire minimisant la somme : $\sum_{i=1}^n d^2(\mathbf{p}'_i, \mathbf{F} \mathbf{p}_i)$ avec :

$$d(\mathbf{p}'_i, \mathbf{F} \mathbf{p}_i) = \frac{|\mathbf{p}'_i{}^t \mathbf{F} \mathbf{p}_i|}{\sqrt{(\mathbf{F} \mathbf{p}_i)_1^2 + (\mathbf{F} \mathbf{p}_i)_2^2}}$$

où le terme $(\mathbf{F} \mathbf{p}_i)_i$ signifie l'élément i du vecteur $\mathbf{F} \mathbf{p}_i$. Pour donner un rôle symétrique aux deux images et pour obtenir une géométrie

épipolaire cohérente, il convient de minimiser le critère suivant :

$$\sum_{i=1}^n \{d^2(\mathbf{p}'_i, \mathbf{F} \mathbf{p}_i) + d^2(\mathbf{p}_i, \mathbf{F}^t \mathbf{p}'_i)\}$$
 où

$$d(\mathbf{p}_i, \mathbf{F}^t \mathbf{p}'_i) = \frac{|\mathbf{p}'_i{}^t \mathbf{F}^t \mathbf{p}_i|}{\sqrt{(\mathbf{F}^t \mathbf{p}'_i)_1^2 + (\mathbf{F}^t \mathbf{p}'_i)_2^2}}$$

est la distance du point \mathbf{p}_i à sa ligne épipolaire $\mathbf{F}^t \mathbf{p}'_i$.

Compte tenu du fait que $\mathbf{p}'_i{}^t \mathbf{F}^t \mathbf{p}_i = \mathbf{p}_i^t \mathbf{F} \mathbf{p}_i$, on aboutit finalement à la minimisation du critère suivant (critère **DIST**) qui est un critère normalisé :

Critère DIST

$$\sum_{i=1}^n \left[\frac{(\mathbf{p}_i^t \mathbf{F} \mathbf{p}_i)^2}{(\mathbf{F} \mathbf{p}_i)_1^2 + (\mathbf{F} \mathbf{p}_i)_2^2} + \frac{(\mathbf{p}'_i{}^t \mathbf{F} \mathbf{p}_i)^2}{(\mathbf{F}^t \mathbf{p}'_i)_1^2 + (\mathbf{F}^t \mathbf{p}'_i)_2^2} \right]$$

La minimisation de ce critère est un problème de minimisation non linéaire que l'on peut résoudre par une méthode numérique itérative du type Levenberg–Marquardt (voir dans [Fletscher87] et [GillMurrayWright81]), mais aussi dans [PhongHaraudYassine-Tao95]). La valeur initiale est celle fournie par le critère linéaire présenté ci-dessus.

Une autre façon d'améliorer le critère linéaire est la suivante. Les coordonnées des appariements images $(\mathbf{p}_i, \mathbf{p}'_i)$, ($i = 1, \dots, n$), sont généralement entachées d'un certain bruit. On peut alors multiplier chaque terme sommé $(\mathbf{p}^t \mathbf{F} \mathbf{p})^2$ du critère linéaire par un coefficient de telle sorte que tous les termes aient la même covariance. Si le bruit est de moyenne nulle, isotrope, uniformément réparti et indépendant selon chaque axe image, alors on montre que le nouveau critère à minimiser est un critère normalisé qui s'écrit (critère **GRAD**) (voir Luong92) :

Critère GRAD

$$\sum_{i=1}^n \frac{(\mathbf{p}_i^t \mathbf{F} \mathbf{p}_i)^2}{(\mathbf{F} \mathbf{p}_i)_1^2 + (\mathbf{F} \mathbf{p}_i)_2^2 + (\mathbf{F}^t \mathbf{p}'_i)_1^2 + (\mathbf{F}^t \mathbf{p}'_i)_2^2}$$

Ce critère à minimiser est semblable au précédent. Il peut également être minimisé par une méthode numérique itérative du type Levenberg–Marquardt.

4.2.3.3. Paramétrisation de la matrice fondamentale

Comme nous l'avons déjà mentionné, la matrice \mathbf{F} est de rang 2. Nous pouvons ainsi chercher les minima des critères sous la contrainte $\text{rang}(\mathbf{F}) = 2$ ($\det(\mathbf{F}) = \mathbf{0}$), ce qui revient sans grande perte de généralité à chercher \mathbf{F} sous la forme [Luong92] (forme L) :

Forme L

$$\mathbf{F} = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 a_1 + a_8 a_4 & a_7 a_2 + a_8 a_5 & a_7 a_3 + a_8 a_6 \end{pmatrix}$$

où $(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)$ sont 8 coefficients définis à un multiple près. Cette représentation permet d'assurer que \mathbf{F} est

une matrice singulière du fait que sa troisième ligne est une combinaison linéaire des deux premières.

On peut aussi la rechercher sous la forme [Luong92] (forme T) :

Forme T

$$F = \begin{bmatrix} b & a & -ae_v - be_u \\ -d & -c & ce_v + de_u \\ de'_v - be'_u & ce'_v - ae'_u & -ce_v e'_v - de'_v e_u + ae_v e'_u + be_u e'_u \end{bmatrix}$$

où (e_u, e_v) et (e'_u, e'_v) représentent les coordonnées images des épipoles e et e' et (a, b, c, d) sont les coefficients de l'homographie liant les faisceaux de droites épipolaires que nous avons vue dans le paragraphe 4.2. Cette formulation présente l'avantage de donner immédiatement la position des épipoles.

4.2.3.4. Résultats obtenus

Afin de minimiser les critères non linéaires, une solution initiale pour F est nécessaire. Elle peut être obtenue en utilisant le critère linéaire. Les résultats obtenus par Luong et al. ([LuongDericheFaugerasPapadopoulo93], [Luong92]) montrent que l'utilisation des critères non linéaires améliore toujours la précision de la solution obtenue avec le critère linéaire. Par ailleurs, les auteurs ont constaté que, parmi les quatre méthodes possibles obtenues par combinaison des critères DIST et GRAD avec les paramétrisations de la matrice F sous la forme L et T , la méthode DIST-L (c'est à dire la méthode minimisant le critère DIST avec paramétrisation de F sous la forme L) est légèrement plus performante : elle fournit moins de faux minima et les épipoles trouvés sont plus proches des épipoles corrects.

Luong [Luong92] et Luong et Faugeras [Luong Faugeras93] ont montré que la qualité du résultat obtenu dépendait essentiellement de la position des points dans l'espace 3D. Le meilleur résultat est obtenu lorsque les points sont bien répartis sur toute la scène. Ils ont établi ce résultat après avoir simulé plusieurs scènes en fixant un certain nombre de points et un certain nombre de plans dans lesquels les points se trouvent. Ils ont notamment démontré que lorsque les points scènes appartiennent à un nombre restreint de plans (par exemple points sur une mire bi-plan), alors le calcul de la matrice fondamentale devient très instable. Dans ce dernier cas, une autre méthode tenant compte de la spécificité de la scène peut être utilisée. Cette méthode, appelée *méthode de plans*, est présentée par la suite.

4.2.3.5. La méthode de plans

Dans le cas où les points scène P_i appartiennent tous au même plan (π) , il existe une homographie entre ces points et les projections images p_i . Si nous disposons d'une seconde image et de points p'_i , alors il existe également une relation homographique entre les points p_i et leurs correspondants p'_i . Cette homographie (application projective linéaire) est représentée par une matrice 3×3 inversible H , telle que :

$$p'_i = Hp_i \quad (9)$$

La présence de 4 correspondances dans les deux images (où par trois les points images sont non alignés) suffit pour calculer cette homographie. La présence d'autres correspondances peut être utilisée pour résoudre l'équation (9) au sens de moindres carrés ou en utilisant des critères non linéaires similaires à ceux de DIST et de GRAD présentés précédemment.

Revenons maintenant au problème du calcul de la matrice fondamentale. Compte tenu que les points appariés (p_i, p'_i) vérifient également l'équation $p_i^t F p_i = 0$, nous obtenons alors :

$$\forall p : p^t F p = (Hp)^t F p = p^t H^t F p = 0$$

et donc nous concluons que la matrice $H^t F$ est antisymétrique et on a :

$$H^t F + F^t H = 0 \quad (10)$$

La relation (10) est utilisée par Faugeras et Luong [Luong92], [LuongFaugeras93] pour estimer la matrice fondamentale à partir des homographies (connaissance du fait que des points scène sont coplanaires). En réalité chaque homographie fournit 6 équations qui sont linéaires en les coefficients de F . Deux homographies au minimum sont nécessaires pour déterminer F . Cette méthode est appelée *calcul de la matrice fondamentale à partir de la méthode de plans*.

Faugeras et Luong ont comparé expérimentalement la méthode de plans avec la méthode générale de calcul de la matrice fondamentale. Ils ont observé que :

- Lorsqu'on calcule la matrice fondamentale par la méthode générale en utilisant des points scène appartenant à des structures planes (en nombre fixe), plus le nombre de points augmente, meilleure est la qualité de l'estimation. D'autre part, pour un nombre de points scène donné, plus le nombre de plans augmente, meilleurs sont les résultats (l'amélioration dans ce cas étant beaucoup plus importante que dans le cas précédent). Enfin, ils concluent que « les meilleures estimations sont obtenues avec un faible nombre de points répartis sur un nombre élevé de plans plutôt qu'avec un grand nombre de points répartis sur un faible nombre de plans ». A titre d'illustration, avec un bruit de 0.2 pixels, la matrice F obtenue à partir de 10 points faisant partie de 7 plans est meilleure que celle obtenue à l'aide de 70 points contenus dans 3 plans.
- Dans le cas de la mire bi-plan, généralement utilisée pour le calibrage classique d'une caméra, la méthode des plans fournit de meilleurs résultats que la méthode générale. Dès que les points scène appartiennent à plus de trois plans, la méthode générale fournit de meilleurs résultats. Si le nombre de plans augmente, alors les résultats de la méthode générale s'améliorent nettement, alors que ceux de la méthode des plans se dégradent !

En définitive, il vaut mieux avoir peu de points bien répartis sur plusieurs plans, qu'un grand nombre inclus dans un nombre limité de plans.

Calcul direct des épipoles à partir de 6 points scènes

Il est intéressant d'observer que la matrice homographique H peut être utilisée pour le calcul des épipoles d'un système stéréo. Mohr et al. [Mohr92], [MohrQuanVeillonBoufama92] et Shashua [Shashua93] par exemple proposent de calculer les épipoles de la manière suivante (voir Figure 4.3) : à partir de 4 points scène coplanaires P_i ($i = 1, \dots, 4$) et de leurs images gauches p_i et droites p'_i , il est possible de calculer l'homographie H telle que $p'_i = Hp_i$. Soient P_5 et P_6 deux points non coplanaires par cinq avec les points P_i . Les épipoles sont alors donnés par :

$$e' = (p'_5 \wedge Hp_5) \wedge (p'_6 \wedge Hp_6)$$

$$e = (p_5 \wedge H^{-1}p'_5) \wedge (p_6 \wedge H^{-1}p'_6)$$

En fait Hp_5 et Hp_6 sont les correspondants des points p_5 et p_6 lorsque les points scènes correspondant sont coplanaires avec les points P_i ($i = 1, \dots, 4$). Ainsi les points p'_5 (image droite de P_5) et Hp_5 définissent une ligne épipolaire; p'_6 (image droite de P_6) et Hp_6 définissent une autre ligne épipolaire. Le produit vectoriel de ces deux lignes épipolaires détermine l'épipole droit e' . De façon similaire est déterminé l'épipole gauche e . Luong [Luong92] a expérimenté ce type de méthode en considérant en réalité plusieurs points coplanaires combinés avec un ensemble de points non coplanaires. Il a essayé alors de retrouver l'épipole au sens de moindres carrés (calcul de l'épipole comme intersection de toutes les droites épipolaires générées par la configuration choisie). Il conclut au fait que les résultats obtenus sont extrêmement sensibles au bruit : le point d'intersection de toutes ces droites n'est pas déterminé avec une bonne précision.

Remarque

La méthode précédente peut être utilisée pour déterminer si un point quelconque de la scène appartient au plan défini par 4 autres points scènes. En effet, pour que le point P_5 par exemple appartienne au plan défini par les 4 points P_i ($i = 1, \dots, 4$), il faudrait que les points images Hp_5 et p'_5 soient confondus (cf. Figure 4.3).

4.3. RÉCUPÉRATION DES PARAMÈTRES INTRINSÈQUES

Dans ce paragraphe, nous présentons deux méthodes permettant de calculer les paramètres intrinsèques d'une caméra. Pour ceci, on fait subir à la caméra un certain nombre de déplacements et à l'aide des méthodes présentées précédemment, les matrices fondamentales entre ces positions sont calculées. Utilisant des techniques de géométrie projective, les paramètres intrinsèques de la caméra sont calculés. Les paramètres du déplacement peuvent enfin être récupérés (« camera egomotion »).

Connaissant l'ensemble de ces paramètres, il est possible d'effectuer une reconstruction 3D des points images appariés dans le repère associé à la caméra pour une position donnée de celle-ci.

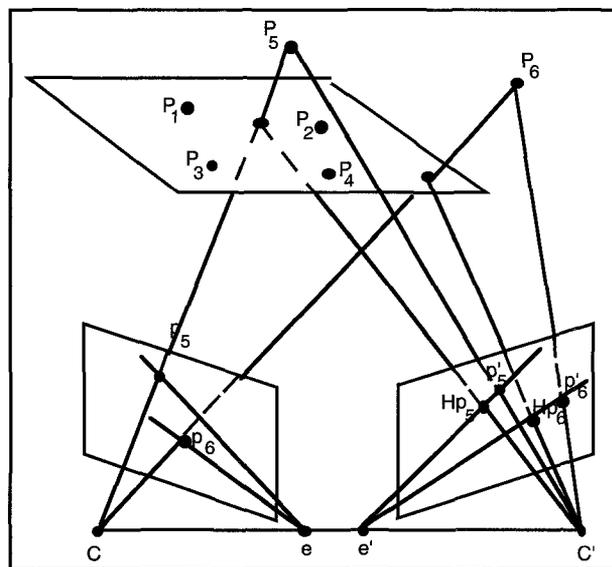


Figure 4.3. – Détermination des épipoles à partir de quatre points scène coplanaires et de deux points non coplanaires avec les précédents.

Par définition, les paramètres intrinsèques sont les coefficients de la matrice A (voir Annexe C). La connaissance de A et de la matrice fondamentale F nous donne E (la matrice essentielle), puisque $E = A^t F A$, donc le déplacement relatif entre deux positions successives de la caméra. Notons qu'il est possible à l'aide de la matrice M (matrice de transformation perspective) obtenue à l'aide d'une mire bi-plan de trouver toutes ces grandeurs.

Nous présentons enfin dans ce paragraphe une troisième méthode originale permettant de calculer les paramètres intrinsèques et le déplacement relatif d'une caméra en mouvement. Elle s'appuie sur des outils de géométrie projective et ne calcule pas les matrices fondamentales liant les différentes positions de la caméra.

4.3.1. La méthode de Faugeras et Luong

Faugeras [FaugerasLuongMaybank92] et Luong [Luong92] proposent de calculer les paramètres intrinsèques d'une caméra en analysant les images obtenues après plusieurs déplacements rigides (translations plus rotations) de celle-ci. Sans connaissances a priori sur la scène, ils cherchent dans un premier temps la matrice fondamentale qui lie la caméra avant et après chaque déplacement et calculent ensuite la géométrie épipolaire liée à chaque déplacement (épipoles et paramètres de l'homographie liant les lignes épipolaires avant et après déplacement).

Utilisant la matrice essentielle, pour chaque déplacement de caméra on a la relation : $E = A^t F A$. La matrice essentielle est sujette à deux conditions ([HuangFaugeras89], [Luong92]) :

$$\det(E) = 0 \quad \text{et} \quad \frac{1}{2} \text{trace}^2(EE^t) - \text{trace}(EE^t)^2 = 0$$

La première condition est automatiquement satisfaite (car $\text{rang}(F)=2$). En remplaçant E , par son expression $A^t F A$, dans

la seconde condition, on obtient une contrainte polynomiale de degré 4 en les coefficients de E . Comme E est quadratique en les coefficients de A , nous nous trouvons alors avec une contrainte de degré 8 en les coefficients de A laquelle est difficilement exploitable.

Luong, Faugeras et Maybank proposent alors une autre méthode pour calculer les paramètres intrinsèques. Elle est dite méthode de la *conique absolue*. Les auteurs montrent qu'il existe une conique Ω de l'espace scène se trouvant au plan de l'infini, dont la projection ω sur le plan image est indépendante par déplacement rigide de la caméra. Cela signifie que cette projection ne dépend que des paramètres intrinsèques de la caméra et non pas de paramètres de son déplacement. Les auteurs démontrent qu'en fait la conique ω est définie par la matrice :

$$B = (A^{-1})^t A^{-1}$$

où A est la matrice des paramètres intrinsèques.

Considérons maintenant deux images consécutives prises au cours du déplacement de la caméra. Étant donné que les lignes épipolaires de ces images sont liées par l'homographie épipolaire, ceci implique que les deux lignes épipolaires *tangentes* à ω dans la première image doivent correspondre aux deux lignes épipolaires *tangentes* ω dans la seconde image, sous la transformation épipolaire définie par la matrice fondamentale F (voir Figure 4.4). Ceci résulte en deux équations indépendantes appelées **équations de Kruppa** qui lient les coefficients de la conique avec les coordonnées des épipoles et des lignes épipolaires considérées. Elles sont de degré 2 en les six éléments de la matrice Δ (matrice de

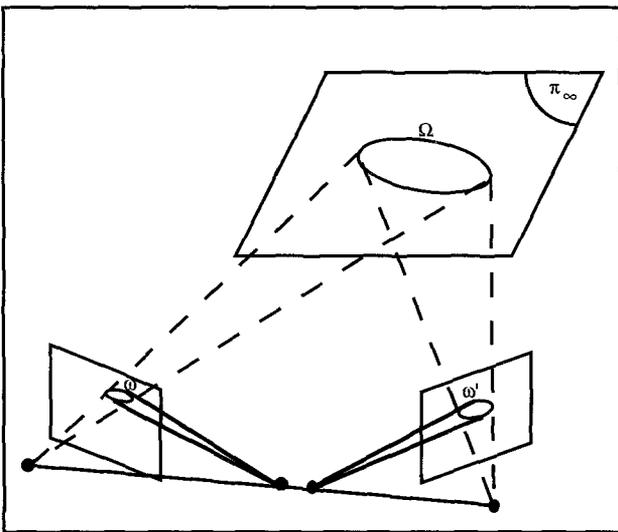


Figure 4.4. – La conique absolue Ω et ses projections gauche et droite ω et ω' , respectivement. Les deux lignes épipolaires tangentes à ω correspondent aux deux lignes épipolaires tangentes à ω' .

Kruppa) définie par :

$$\Delta = \begin{bmatrix} -\delta_{23} & \delta_3 & \delta_2 \\ \delta_3 & -\delta_{13} & \delta_1 \\ \delta_2 & \delta_1 & -\delta_{12} \end{bmatrix}$$

Δ est la matrice qui définit la conique duale de ω , c'est à dire $\Delta = B^* = \det(B)(B^{-1})^t$. Comme Δ est définie à un coefficient multiplicatif près, 5 de ses paramètres sont indépendants (qui sont à relier aux cinq paramètres intrinsèques d'une caméra, comme nous allons le voir ci-dessous). Trois mouvements de la caméra sont alors nécessaires afin de calculer les éléments de la matrice Δ . Luong démontre qu'il n'est pas nécessaire de disposer de trois mouvements indépendants mais que les équations peuvent être résolues même si l'un des trois mouvements est obtenu par composition des deux autres. En particulier, il démontre que dans ce cas, les équations de Kruppa obtenues avec le troisième déplacement sont indépendantes de celles obtenues avec les deux premiers. Ainsi deux déplacements de la caméra sont nécessaires, le troisième étant obtenu comme une composition de deux premiers.

Chaque déplacement de caméra nous fournit deux équations indépendantes en les coefficients de la matrice Δ . Ce déplacement ne doit pas être quelconque. En effet, comme les équations de Kruppa sont liées avec la transformation épipolaire (et donc la matrice fondamentale), les déplacements doivent satisfaire les conditions de stabilité de calcul de cette dernière. Ainsi, si le déplacement est une rotation pure, alors la matrice fondamentale étant nulle, les équations de Kruppa n'ont pas de sens. Si le déplacement est une translation pure, on peut démontrer que dans ce cas les deux équations de Kruppa sont nulles (voir [Luong92], chap. 7). En général, il faut que les déplacements soient relativement grands (par rapport à la distance des objets observés) et qu'ils comportent à la fois une composante rotationnelle et une composante translationnelle importantes. Il faut par ailleurs que la translation ne soit pas parallèle au plan image.

Ces contraintes de déplacement étant respectées, la résolution des équations de Kruppa fournit 32 solutions à explorer et parmi elles une seule est valable. Le calcul des solutions est effectué grâce à la méthode dite *des continuations* (voir [FaugerasLuongMaybank92] et [Luong92] pour plus de détail).

Étant donnée Δ , les paramètres intrinsèques peuvent maintenant être calculés. En effet, $\Delta = B^* = AA^t$. Les équations suivantes relient alors les paramètres de Δ avec les paramètres intrinsèques $a_u, a_v, u_0, v_0, \theta$ (à un coefficient multiplicatif λ près) :

$$\begin{aligned} \lambda \delta_1 &= v_0 \\ \lambda \delta_2 &= u_0 \\ \lambda \delta_3 &= u_0 v_0 - a_u a_v \frac{\cot \theta}{\sin \theta} \\ \lambda \delta_{12} &= -1 \\ \lambda \delta_{23} &= -u_0^2 - \frac{a^2}{\sin^2 \theta} \\ \lambda \delta_{13} &= -v_0^2 - \frac{a^2}{\sin^2 \theta} \end{aligned}$$

Il est alors possible à partir de ces relations de déterminer de manière unique les paramètres intrinsèques de la caméra à condition qu'un certain nombre de contraintes soient vérifiées. Ces contraintes peuvent par ailleurs être utilisées pour éliminer un certain nombre de solutions concernant les coefficients δ_{ij} de la matrice Δ qui donnent naissance à des paramètres intrinsèques n'ayant aucune interprétation physique.

Les paramètres intrinsèques ayant été calculés, il est possible de calculer ensuite les paramètres extrinsèques des positions successives de la caméra, c'est à dire son mouvement, dans le repère associé à la position initiale de celle-ci.

La résolution des équations de Kruppa n'est pas une tâche facile. En effet, nous avons vu que leur résolution nécessite l'emploi de méthodes de continuation numérique qui souffrent en général de problèmes de précision et de temps de calcul. Ainsi, Luong dans sa thèse a exploré d'autres méthodes de calcul de ces équations (par exemple une méthode de minimisation globale ou un filtrage récursif de type Kalman étendu), mais les résultats obtenus n'étaient pas vraiment de meilleure qualité. Par ailleurs Luong a essayé d'utiliser des modèles simplifiés des paramètres intrinsèques (par exemple $\theta = \pi/2$, ou u_0, v_0 centrés par rapport à la rétine ou encore un rapport a_u/a_v connu). Les résultats obtenus montrent que la qualité de détermination du point principal est médiocre et qu'il vaut mieux le fixer comme étant au centre de la rétine, alors que le calcul de a_u et a_v est plus stable. Quant à θ , il est en pratique très voisin de 90° .

4.3.2. La méthode de Hartley

La méthode de Hartley ([Hartley92] est assez similaire à celle de Faugeras [FaugerasLuongMaybank92] en ce qui concerne le calcul de la matrice fondamentale. Hartley considère un système stéréo. Sur la base de 8 correspondances minimum gauches-droites et en utilisant un critère linéaire, il calcule d'abord la matrice fondamentale liant les deux caméras de la paire stéréo. Cette matrice à 7 degrés de liberté est utilisée pour le calcul de deux parmi les cinq paramètres intrinsèques des caméras stéréo et de cinq paramètres extrinsèques du système (la matrice essentielle à un multiple près).

Les paramètres intrinsèques considérés sont les deux distances focales des caméras. Hartley détermine d'abord ces deux paramètres en faisant une décomposition en valeurs singulières de la matrice fondamentale et en exploitant des connaissances a priori sur les autres paramètres intrinsèques des caméras (notamment la position image du point principal). Une fois ces deux paramètres déterminés, Hartley calcule ensuite la matrice essentielle ($E = T_t R, R$ étant la matrice de rotation et T_t étant la matrice antisymétrique associée au vecteur translation) reliant le système stéréo. Une fois la matrice essentielle obtenue, il est possible de reconstruire la scène dans le repère de la caméra gauche.

Les résultats expérimentaux obtenus par Hartley sont très sombres et donc difficiles à commenter. Toutefois, les conclusions de ses travaux sont assez proches de celles de Faugeras et Luong :

il semble qu'il vaut mieux fixer les deux paramètres intrinsèques d'une caméra concernant la position image du point principal et ne déterminer que les autres paramètres, plus stables.

4.3.3. La méthode de l'homographie de l'infini

Nous présentons dans ce paragraphe le rôle très important que la connaissance de l'équation du plan à l'infini peut jouer pour le calibrage d'une caméra mobile. Les réflexions que nous présentons sont inspirées notamment des travaux de Viéville [ViévilleZellerRobert95] et Luong et Faugeras [Luong-Faugeras93].

La propriété principale que nous allons exploiter est la suivante : une translation laisse le plan à l'infini invariant (voir Annexe B). Nous allons démontrer qu'ayant les correspondances images, dans deux vues différentes prises par une caméra en mouvement, de 4 points de la scène se trouvant très loin par rapport à la caméra, il est possible de calculer l'homographie du plan de l'infini (de tels points scène sont par exemple les points de l'horizon ou les points se situant très loin par rapport au champ de travail d'une caméra-typiquement une centaine de mètres pour une caméra dont le champ de travail est entre 1 et 10 mètres). Ayant l'homographie du plan de l'infini et deux vues différentes, il est possible de calculer la matrice de rotation reliant ces deux vues et ensuite les paramètres intrinsèques de cette caméra.

Soit donc une caméra en mouvement. Considérons le repère qui lui est associé à l'instant τ . Nous allons utiliser ce repère comme repère de référence.

Exprimons la projection des points scène à l'infini dans le repère de référence. Si A est la matrice des paramètres intrinsèques, alors :

$$p_\infty = [A \ 0] \begin{bmatrix} P_\infty \\ 0 \end{bmatrix} \Rightarrow p_\infty = AP_\infty$$

La caméra se déplace et à l'instant $\tau + d\tau$ elle prend une autre image de la scène. Si $[R, t]$ est son déplacement entre les instants τ et $\tau + d\tau$, alors on a :

$$p'_\infty = M' \begin{bmatrix} P_\infty \\ 0 \end{bmatrix} = [A \ 0] \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_\infty \\ 0 \end{bmatrix} \\ \Rightarrow p'_\infty = [AR \ At] \begin{bmatrix} P_\infty \\ 0 \end{bmatrix} = ARP_\infty$$

et en utilisant l'équation précédente on a :

$$p'_\infty = ARP_\infty = ARA^{-1}p_\infty \Rightarrow p'_\infty = ARA^{-1}p_\infty = H_\infty p_\infty$$

H_∞ est une matrice 3×3 . Elle est l'homographie de l'infini. Elle lie les projections images de tout point scène appartenant au plan de l'infini. Pour la déterminer, il faut au minimum 4 correspondances images (p_i, p'_i) ($i = 1, \dots, n$) provenant de points scène appartenant au plan à l'infini. Pour cela, il suffit de considérer au moins 4 points de l'horizon mis en correspondance dans une séquence d'images (au moins deux vues). Si plusieurs correspondances images sont disponibles (avec $n \geq 4$, n étant le

nombre d'appariements), il est possible de minimiser un critère non linéaire du type :

$$\min_{H_\infty} f(H_\infty) = \sum_{i=1}^n \{(p'_i - H_\infty p_i)^2 + (p_i - H_\infty^t p'_i)^2\}$$

qui peut être minimisé en utilisant la méthode de Levenberg-Marquardt (voir aussi la méthode de plans, paragraphe 4.2.3.5).

On peut remarquer que $H_\infty = ARA^{-1}$ et donc cette homographie ne contient que les composantes rotationnelles du mouvement de la caméra. Elle est indépendante de la translation, comme on pourrait s'y attendre : en effet, il est connu que toute translation laisse les points à l'infini invariants (voir Annexe B).

Comme $R = A^{-1}H_\infty A$ et comme R est une matrice de rotation (donc $RR^t = I$), on a :

$$\begin{aligned} RR^t = I &\Rightarrow (A^{-1}H_\infty A)(A^{-1}H_\infty A)^t = I \\ &\Rightarrow H_\infty AA^t H_\infty^t = AA^t \Rightarrow H_\infty \Delta H_\infty^t = \Delta \end{aligned}$$

où

$$\Delta = AA^t = \begin{bmatrix} u_0^2 + \frac{a_x^2}{\sin^2 \theta} & u_0 v_0 - \frac{a_x a_y \cot \theta}{\sin \theta} & u_0 \\ u_0 v_0 - \frac{a_x a_y \cot \theta}{\sin \theta} & v_0^2 + \frac{a_y^2}{\sin^2 \theta} & v_0 \\ u_0 & v_0 & 1 \end{bmatrix}$$

Par conséquent, si H_∞ est connu, Δ peut être déterminée à l'aide de l'équation $H_\infty \Delta H_\infty^t = \Delta$ et par là, la matrice des paramètres intrinsèques A de la caméra peut être entièrement récupérée. L'équation $H_\infty \Delta H_\infty^t = \Delta$ fournit en réalité neuf équations linéaires en les cinq éléments de la matrice Δ (Δ étant une matrice symétrique définie positive et à un coefficient multiplicatif près). Parmi ces neuf équations, cinq sont seulement indépendantes (voir [ViévilleZellerRobert95]).

Pour déterminer l'homographie H_∞ et donc les éléments de la matrice Δ , deux vues sont en théorie nécessaires. Toutefois, Viéville et al. ont observé que, pour des raisons de stabilité numérique, il est préférable d'utiliser deux déplacements (et donc trois vues), un premier déplacement générant une disparité horizontale et un second, une disparité verticale.

Remarquons que la matrice Δ définit en réalité l'image de la conique absolue de l'espace que nous avons vue au §4.3.1. Dans ce même paragraphe nous avons présenté les formules permettant de calculer les paramètres intrinsèques une fois que les éléments de la matrice Δ ont été calculés.

Une fois la matrice Δ calculée, la rotation de la caméra peut être retrouvée par l'équation :

$$R = A^{-1}H_\infty A$$

Soit P un point scène qui n'appartient pas au plan de l'infini. On peut tester ceci à partir de ses projections (p, p') dans deux images consécutives de la séquence : il suffit que $p' \neq H_\infty p$.

Ce point vérifie alors les équations suivantes :

$$p = [A \ 0] \begin{bmatrix} P \\ 1 \end{bmatrix} \Rightarrow p = AP$$

$$p' = [AR \ At] \begin{bmatrix} P \\ 1 \end{bmatrix} = ARP + At$$

$$\Rightarrow p' = ARA^{-1}p + At \Rightarrow p' = H_\infty p + At$$

Ayant un certain nombre de correspondances (p_i, p'_i) ($i = 1, \dots, n$) entre deux images consécutives de la séquence et ayant calculé l'homographie de l'infini et les paramètres intrinsèques de la caméra, le vecteur translation t peut être récupéré (à un facteur d'échelle près) par résolution aux moindres carrés de l'équation :

$$\sum_{i=1}^n \|H_\infty p_i + At - p'_i\|^2$$

Remarquons enfin que si le point scène considéré est le centre optique de l'image acquise à l'instant τ , alors $P = [0, 0, 0]^t$ et donc sa projection dans l'image acquise à l'instant $\tau + 1$ (qui est l'épipoles associé à cette image) sera telle que :

$$e' = At$$

Par conséquent, il est aussi possible de retrouver la translation t (à un coefficient près bien sûr) connaissant l'épipoles de l'image $\tau + 1$ ainsi que les paramètres intrinsèques de la caméra.

Hartley dans [Hartley94] a présenté une version de cette méthode dans le cas d'une caméra en rotation pure autour de son centre optique. En effet, il est facile de démontrer que lorsqu'une caméra effectue une telle rotation, les points perçus dans deux images différentes, sont liés par une homographie, la même que celle liant les points de l'infini présentée précédemment. Dans ce cas, les homographies liant différentes images acquises par une rotation pure de la caméra sont calculées grâce à des points poursuivis dans ces images. Ces homographies sont ensuite utilisées pour retrouver les paramètres intrinsèques et les matrices de rotation entre les différentes images, utilisant les équations exposées dans ce paragraphe.

4.4. RECONSTRUCTION RELATIVE DE LA SCÈNE

Ayant calculé les paramètres de mouvement d'une caméra (exprimés via la matrice fondamentale ou la matrice essentielle), il est possible de procéder à une reconstruction relative dans le repère associé par exemple à la première position de la caméra. Nous présentons dans ce paragraphe la façon dont on peut utiliser ces informations afin d'obtenir une reconstruction projective de scènes. Une telle méthode a été proposée par Faugeras [Faugeras92] qui démontre qu'il est possible d'effectuer une reconstruction de la scène dans un repère défini à l'aide de cinq points scènes quelconques à condition qu'ils ne soient pas coplanaires par 4 et que

l'on connaisse la matrice fondamentale F . Dans ce qui suit nous présentons succinctement cette méthode.

Soit M une matrice de transformation perspective associée à une caméra, P_i un point scène et p_i son image. Ces points sont liés par la relation bien connue $p_i = MP_i$. Considérons maintenant de nouveau un système stéréo composé de deux caméras. Soient 5 points scène P_i vus dans les deux caméras et soient p_i, p'_i $i = 1, \dots, 5$ leurs images par les caméras gauche et droite.

Les points 3D P_i peuvent être choisis comme la base d'un repère de l'espace projectif P^3 et donc on peut considérer que ces points ont les coordonnées projectives suivantes : $(1, 0, 0, 0)^t, (0, 1, 0, 0)^t, (0, 0, 1, 0)^t, (0, 0, 0, 1)^t, (1, 1, 1, 1)^t$.

Dans le plan image gauche, quatre points p_i (projections des P_i) définissent une autre base projective de l'espace P^2 et donc leurs coordonnées peuvent être écrites comme $(1, 0, 0)^t, (0, 1, 0)^t, (0, 0, 1)^t$ et $(1, 1, 1)^t$. De même dans le plan image droit, 4 projections p'_i définissent un autre repère projectif. En écrivant maintenant pour $i = 1, \dots, 4$ que $a_i p_i = MP_i$, (a_i étant des coefficients), on peut déterminer les coefficients de la matrice M .

$$\begin{aligned} a_1(1, 0, 0)^t &= M(1, 0, 0, 0)^t \\ \text{nous donne } m_{11} &= a_1, m_{21} = 0, \text{ et } m_{31} = 0 \\ a_2(0, 1, 0)^t &= M(0, 1, 0, 0)^t \\ \text{nous donne } m_{12} &= 0, m_{22} = a_2, \text{ et } m_{32} = 0 \\ a_3(0, 0, 1)^t &= M(0, 0, 1, 0)^t \\ \text{nous donne } m_{13} &= 0, m_{23} = 0, \text{ et } m_{33} = a_3 \\ a_4(1, 1, 1)^t &= M(0, 0, 0, 1)^t \\ \text{nous donne } m_{14} &= a_4, m_{24} = a_4, \text{ et } m_{34} = a_4 \end{aligned}$$

Donc la matrice M a la forme :

$$M = \begin{bmatrix} a_1 & 0 & 0 & a_4 \\ 0 & a_2 & 0 & a_4 \\ 0 & 0 & a_3 & a_4 \end{bmatrix}$$

Soit $p_5 = (\alpha, \beta, \gamma)^t$ la projection dans l'image gauche du point $P_5 = (1, 1, 1, 1)^t$.

On a : $a_5(\alpha, \beta, \gamma)^t = M(1, 1, 1, 1)^t$, et donc $a_1 + a_4 = a_5\alpha$, $a_2 + a_4 = a_5\beta$, $a_3 + a_4 = a_5\gamma$, et en posant $\nu = a_4$, $\mu = a_5$ et $x = \frac{\mu}{\nu}$, la matrice M ci-dessus peut se réécrire sous la forme :

$$\begin{aligned} M &= \begin{bmatrix} \mu\alpha - \nu & 0 & 0 & \nu \\ 0 & \mu\beta - \nu & 0 & \nu \\ 0 & 0 & \mu\gamma - \nu & \nu \end{bmatrix} \\ &= \begin{bmatrix} x\alpha - 1 & 0 & 0 & 1 \\ 0 & x\beta - 1 & 0 & 1 \\ 0 & 0 & x\gamma - 1 & 1 \end{bmatrix} \quad (11) \end{aligned}$$

Posons maintenant $M = [\hat{M} \ m]$ avec

$$\hat{M} = \begin{bmatrix} x\alpha - 1 & 0 & 0 \\ 0 & x\beta - 1 & 0 \\ 0 & 0 & x\gamma - 1 \end{bmatrix} \text{ et } m = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Une expression similaire peut être obtenue pour M' en fonction de $x' = \frac{\mu'}{\nu'}$ et du point image $p'_5 = (\alpha', \beta', \gamma')^t$.

On a successivement :

Coordonnées du centre optique de la caméra gauche

$$MC = [0 \ 0 \ 0]^t \Rightarrow C = \left[\frac{1}{1 - \alpha x}, \frac{1}{1 - \beta x}, \frac{1}{1 - \gamma x}, 1 \right]^t \quad (12)$$

Épipoles

On a :

$$\begin{aligned} MC = 0 &\Rightarrow [\hat{M} \ m] \begin{bmatrix} C \\ 1 \end{bmatrix} = 0 \Rightarrow \hat{M}C + m = 0 \\ &\Rightarrow C = \begin{bmatrix} \hat{M}^{-1}m \\ -1 \end{bmatrix} \end{aligned}$$

L'épipole droit est :

$$\begin{aligned} e' = M'C &= [\hat{M}' \ m'] \begin{bmatrix} \hat{M}^{-1}m \\ -1 \end{bmatrix} = \hat{M}'\hat{M}^{-1}m - m' \\ &= \begin{bmatrix} \frac{\alpha'x' - \alpha x}{\alpha x - 1} \\ \frac{\beta'x' - \beta x}{\beta x - 1} \\ \frac{\gamma'x' - \gamma x}{\gamma x - 1} \end{bmatrix} \quad (13) \end{aligned}$$

De façon similaire, on peut trouver que l'épipole gauche est donné par :

$$\begin{aligned} e = MC' &= [\hat{M} \ m] \begin{bmatrix} \hat{M}'^{-1}m' \\ -1 \end{bmatrix} = \hat{M}\hat{M}'^{-1}m' - m \\ &= \begin{bmatrix} \frac{\alpha x - \alpha'x'}{\alpha'x' - 1} \\ \frac{\beta x - \beta'x'}{\beta'x' - 1} \\ \frac{\gamma x - \gamma'x'}{\gamma'x' - 1} \end{bmatrix} \quad (14) \end{aligned}$$

Il est facile de prouver que :

$$e' = -\hat{M}'\hat{M}^{-1}e \text{ et } e = -\hat{M}\hat{M}'^{-1}e'$$

Connaissant la matrice fondamentale F , nous avons vu au § 4.2 comment il est possible de calculer les coordonnées des épipoles gauche et droit. Étant donné qu'il existe une relation liant les coordonnées projectives des épipoles gauche et droit, il est possible de déterminer les paramètres inconnus x et x' ci-dessus et donc de déterminer entièrement les matrices de projection perspective gauche et droite M et M' .

A partir de 5 points scène et de leurs projections dans les images gauche et droite, Faugeras est capable de déterminer entièrement les matrices M et M' . Ceci suppose que la matrice fondamentale ait été auparavant calculée et pour ceci nous avons vu qu'au minimum huit correspondances gauche-droite sont nécessaires.

Ligne épipolaire

On a :

$$p_\infty = M \begin{bmatrix} P_\infty \\ 0 \end{bmatrix} = [\widehat{M} \ m] \begin{bmatrix} P_\infty \\ 0 \end{bmatrix} = \widehat{M}P_\infty$$

$$\Rightarrow P_\infty = \widehat{M}^{-1}p_\infty$$

et donc

$$p'_\infty = M'P_\infty = [\widehat{M}' \ m'] \begin{bmatrix} P_\infty \\ 0 \end{bmatrix} = \widehat{M}'P_\infty = \widehat{M}'\widehat{M}^{-1}p_\infty$$

La ligne épipolaire droite est définie par les points e' et p'_∞ , donc elle est donnée par le produit vectoriel :

$$l' = e' \wedge p'_\infty = T_{e'}p'_\infty = \begin{bmatrix} 0 & -e'_w & e'_v \\ e'_w & 0 & -e'_u \\ -e'_v & e'_u & 0 \end{bmatrix} p'_\infty \quad (15)$$

où $T_{e'}$ est la matrice antisymétrique associée aux coordonnées de l'épipole e' .

Reconstruction projective

Les coordonnées d'un point scène sont données par :

$$P = \mu C + \lambda \begin{bmatrix} P_\infty \\ 0 \end{bmatrix} \quad (16)$$

Par ailleurs, on a $p' = M'P$ et donc en multipliant l'équation (16) par M' on obtient :

$$p' = \mu M'C + \lambda \widehat{M}'\widehat{M}^{-1}p = \mu e' + \lambda \widehat{M}'\widehat{M}^{-1}p$$

A partir de cette dernière équation, on peut trouver les paramètres μ et λ et donc les coordonnées du point 3D vont être calculées à partir de (16) comme :

$$P = \mu \begin{bmatrix} \frac{1}{x\alpha-1} \\ \frac{1}{x\beta-1} \\ \frac{1}{x\gamma-1} \\ -1 \end{bmatrix} + \lambda \begin{bmatrix} \frac{X}{x\alpha-1} \\ \frac{Y}{x\beta-1} \\ \frac{Z}{x\gamma-1} \\ 0 \end{bmatrix} \quad (17)$$

où $P = (X, Y, Z)^t$.

Relation avec la matrice fondamentale

Nous savons que la ligne épipolaire associée au point p est donnée par $l' = Fp$. Par ailleurs, nous avons vu que $l' = T_{e'}p'_\infty$, mais

comme $p'_\infty = \widehat{M}'\widehat{M}^{-1}p_\infty$, on a $l' = T_{e'}\widehat{M}'\widehat{M}^{-1}p_\infty$. Donc $Fp = T_{e'}\widehat{M}'\widehat{M}^{-1}p$, d'où :

$$F = T_{e'}\widehat{M}'\widehat{M}^{-1} = \begin{bmatrix} 0 & -\frac{\beta'x'-1}{\beta x-1} \frac{\gamma'x'-\gamma x}{\gamma x-1} & \frac{\gamma'x'-1}{\gamma x-1} \frac{\beta'x'-\beta x}{\beta x-1} \\ \frac{\alpha'x'-1}{\alpha x-1} \frac{\gamma'x'-\gamma x}{\gamma x-1} & 0 & -\frac{\gamma'x'-1}{\gamma x-1} \frac{\alpha'x'-\alpha x}{\alpha x-1} \\ -\frac{\alpha'x'-1}{\alpha x-1} \frac{\beta'x'-\beta x}{\beta x-1} & \frac{\beta'x'-1}{\beta x-1} \frac{\alpha'x'-\alpha x}{\alpha x-1} & 0 \end{bmatrix} \quad (18)$$

En conclusion, nous observons qu'en choisissant 5 points scène (et connaissant la matrice fondamentale F), ils forment un repère projectif dans lequel il est possible de reconstruire tout autre point apparié dans le système stéréo. Cette reconstruction est bien évidemment définie à une transformation projective près (qui est la transformation d'un repère scène absolu au repère défini par ces 5 points 3D). Si la position des 5 points choisis est connue dans le repère scène absolu, alors les points reconstruits peuvent être exprimés dans ce repère absolu (passage de l'espace projectif à l'espace Euclidien).

4.5. AUTO-CALIBRAGE D'UN SYSTÈME STÉRÉO EN MOUVEMENT

Nous présentons dans ce paragraphe un bon exemple de l'utilisation de l'ensemble des notions et techniques que nous avons décrites dans cette section. Il traite du problème de l'autocalibrage d'un système stéréo non calibré en mouvement, problème abordé par Zhang, Luong et Faugeras dans [ZhangLuongFaugeras93].

Soit un système stéréo non calibré qui est en déplacement. Les caméras stéréo acquièrent des images avant et après le déplacement. On obtient ainsi quatre images d'une même scène et on cherche à trouver à la fois les paramètres intrinsèques et les paramètres extrinsèques de l'ensemble. Un ensemble de points de correspondances est supposé connu à la fois entre les images stéréo (au minimum huit points pour chaque position stéréo) mais aussi temporellement entre les images acquises par la caméra gauche et entre les images acquises par la caméra droite.

Pour résoudre ce problème, Zhang et al. proposent de minimiser un critère non linéaire aux moindres carrés qui comporte 25 inconnues (5 pour chaque matrice de paramètres intrinsèques et 5 pour chaque déplacement entre les images, puisque la translation ne peut être déterminée qu'à un coefficient près). Ils démontrent ensuite que parmi les 5 paramètres intrinsèques de chaque caméra, avec un seul déplacement du système stéréo, deux paramètres uniquement peuvent être déterminés a_u et a_v . La raison de ceci est que chaque déplacement fournit deux contraintes sur les paramètres de Kruppa comme nous l'avons vu dans le paragraphe 4.3, et donc un seul déplacement ne suffit pas pour déterminer l'ensemble de ces paramètres et donc pour calculer tous les paramètres intrinsèques. Pour cette raison, les trois autres paramètres intrinsèques de chaque caméra sont supposés connus. En effet, le paramètre θ est très proche de $\pi/2$ et u_0 et v_0 sont pris au centre de l'image.

Parmi les 15 paramètres des différents déplacements, il n'y en a que 12 qui sont réellement indépendants, donc au total il y a 16 paramètres à déterminer (deux paramètres intrinsèques de la caméra gauche, deux paramètres intrinsèques de la caméra droite, cinq paramètres extrinsèques liant le système stéréo, cinq paramètres extrinsèques liés au déplacement de la caméra gauche et un paramètre extrinsèque lié au déplacement de la caméra droite). Par conséquent, un total de 15 points de correspondance (qui, bien évidemment, ne doivent pas être distribués arbitrairement dans la scène) est nécessaire entre toutes les images afin de déterminer ces paramètres. Le critère formé est alors un critère quadratique par rapport aux paramètres inconnus et est minimisé en utilisant un algorithme de type Gauss-Newton ou Levenberg-Marquardt.

Cet algorithme a besoin d'une bonne valeur initiale des paramètres inconnus. Cette initialisation est fournie par le calcul des matrices fondamentales associées à un couple d'images (temporelles ou stéréo). Pour cela il faut alors disposer d'au minimum huit correspondances entre les différents couples d'images formés, ce qui nécessite au total la connaissance d'au minimum 24 points de correspondance (huit au minimum pour les couples stéréo et huit pour chaque couple temporel d'images). Donc le problème global est toujours surdéterminé. Le calcul des matrices fondamentales est effectué en utilisant un critère de type DIST que nous avons présenté dans le paragraphe 4.2. De ces matrices, les auteurs en déduisent les paramètres intrinsèques a_u et a_v de chaque caméra. Pour cela, ils utilisent les deux contraintes sur la matrice essentielle E utilisées par Luong (la méthode des équations de Kruppa). De là, ils en déduisent les rotations entre couples d'images, et les translations (à un multiple près).

La reconstruction dans un repère caméra devient alors possible (à un coefficient multiplicatif près). Les résultats que Zhang et al. ont obtenu semblent bons : quelques millimètres d'erreur sur la reconstruction des points d'une mire de 300 millimètres.

L'avantage de cette méthode par rapport à celle de Luong [Luong92] réside dans l'exploitation des contraintes de rigidité du système stéréo lors du calcul des paramètres intrinsèques. En effet, Zhang et al. démontrent que la précision d'estimation des paramètres intrinsèques de chaque caméra autocalibrée séparément est considérablement améliorée lorsque l'on tient en compte les stéréo correspondances.

Enfin, remarquons que Zhang et al. démontrent aussi que les erreurs sur les paramètres u_0 et v_0 influent peu sur la qualité et la précision de cette reconstruction. Les résultats présentés montrent que pour une déviation de 35 pixels de la position de u_0 et v_0 par rapport au centre de l'image, la reconstruction 3D obtenue reste de qualité.

5. Méthodes de reconstruction relative de scènes

Les méthodes introduites dans la section précédente présentent deux inconvénients principaux. D'abord, le calcul de la matrice fondamentale entre deux vues successives d'une caméra est comme nous l'avons vu un processus délicat. Certains mouvements ne permettent pas un calcul stable et de plus, les points scènes dont les correspondances images sont utilisées pour calculer la matrice fondamentale doivent être bien répartis dans la scène. Deuxièmement, le calcul des paramètres intrinsèques des caméras à partir de trois positions différentes de celles-ci est généralement difficile à réaliser.

Pour pallier ces inconvénients, un autre type de méthodes a été développé. Ces méthodes ne décomposent pas le problème de calibrage en deux étapes, mais considèrent à la fois le problème de calibrage des paramètres intrinsèques d'une caméra et de son mouvement en un seul problème global. Sur la base des correspondances établies entre les images prises par une caméra en mouvement par rapport à une scène statique, le problème est formulé comme un problème de minimisation globale par rapport à l'ensemble des paramètres.

Un grand nombre d'images et de données est nécessaire afin de résoudre le problème d'optimisation non linéaire à plusieurs paramètres résultant. Ce type de méthodes a été initialement proposé par l'équipe de R. Mohr [MohrVeillonQuan93], [BoufamaMohrVeillon93], [MohrBoufamaBrandt94].

5.1. LA MÉTHODE DE MOHR ET AL.

Le problème posé par Mohr et al est le suivant. Étant donné une caméra qui se déplace et acquiert m images de la scène, comment est-il possible de retrouver la structure de la scène en ayant uniquement un ensemble de correspondances images.

Soient $\{P_i, i = 1, \dots, n\}$ n points inconnus de la scène et m ($m \geq 2$) vues de ces points. Supposons que ces points ont été appariés dans ces m vues. On dispose alors au total de nm points images p_{ij} ($\{i = 1, \dots, n\}$ et $\{j = 1, \dots, m\}$). Par ailleurs, pour chaque position image, on a 11 paramètres inconnus (les 11 paramètres de la matrice de projection M associée à cette position de caméra). Au total on a donc $11m + 3n$ inconnues. Chaque projection d'un point 3D dans l'image nous fournit deux relations sur M (les équations (2)). On dispose alors de $2nm$ équations pour $11m + 3n$ inconnues. Si m et n sont suffisamment grands (ce qui revient à dire que l'on doit posséder plusieurs correspondances images dans plusieurs vues), alors on a des équations redondantes que l'on peut minimiser pour trouver l'ensemble des paramètres inconnus. En observant que les équations sont valables à une application projective près, une base projective quelconque peut être choisie dans l'espace par

rapport à laquelle la scène est reconstruite projectivement. Pour l'espace projectif P^3 , 5 points non coplanaires par 4 forment une base. Leurs coordonnées peuvent être celles de la base canonique : $(1, 0, 0, 0)^t$, $(0, 1, 0, 0)^t$, $(0, 0, 1, 0)^t$, $(1, 1, 1, 1)^t$. Mohr et al. présentent un algorithme qui permet de sélectionner 5 points images de façon à ce que leurs homologues 3D (qui forment la base projective) ne soient pas coplanaires par 4.

En considérant les coordonnées projectives, le problème est formulé comme celui d'une minimisation selon les paramètres :

$$(X_i, Y_i, Z_i, W_i, m_{11}^j, \dots, m_{34}^j)$$

$$\text{pour } \{i = 1, \dots, n\} \text{ et } \{j = 1, \dots, m\}$$

où (X_i, Y_i, Z_i, W_i) sont les coordonnées projectives (homogènes) des points scène et $m_{ik}^{(j)}$ sont les coefficients de la matrice projective $M^{(j)}$. Les coordonnées des points scène sont définies à un multiple près, on peut leur ajouter la contrainte $X_i^2 + Y_i^2 + Z_i^2 + W_i^2 = 1$. Finalement le critère à minimiser est :

$$\sum_{k=1}^{2mn+n} \left(\frac{f_k(u_{ij}, v_{ij}, X_i, Y_i, Z_i, m_{11}^{(j)}, \dots, m_{34}^{(j)})}{\sigma_k} \right)^2$$

avec $f_k()$ une des fonctions suivantes :

$$f_k() = u_{ij} - \frac{m_{11}^{(j)} X_i + m_{12}^{(j)} Y_i + m_{13}^{(j)} Z_i + m_{14}^{(j)} W_i}{m_{31}^{(j)} X_i + m_{32}^{(j)} Y_i + m_{33}^{(j)} Z_i + m_{34}^{(j)} W_i}$$

ou

$$f_k() = v_{ij} - \frac{m_{21}^{(j)} X_i + m_{22}^{(j)} Y_i + m_{23}^{(j)} Z_i + m_{24}^{(j)} W_i}{m_{31}^{(j)} X_i + m_{32}^{(j)} Y_i + m_{33}^{(j)} Z_i + m_{34}^{(j)} W_i}$$

σ_k est l'écart type de chaque mesure image u_{ij} ou v_{ij} . Le critère est minimisé par l'algorithme de Levenberg-Marquardt et l'algorithme semble converger quelles que soient les conditions initiales.

Nous remarquons alors que Mohr établit une reconstruction projective sans établir la correspondance épipolaire. Il a seulement besoin d'un ensemble d'appariements images. Bien évidemment, si la matrice fondamentale est disponible, cet appariement est facilité (car la recherche de correspondances est effectuée uniquement le long des lignes épipolaires).

Comme nous l'avons déjà remarqué, toute la reconstruction est projective. Une forme projective est définie à une application linéaire près, mais aucune information métrique n'est présente. Seulement les propriétés projectives sont préservées, par exemple les points alignés restent alignés et les points coplanaires restent coplanaires. Pour effectuer le passage de l'espace projectif à l'espace affine ou euclidien, des informations supplémentaires sont nécessaires et notamment les coordonnées euclidiennes de 5 points formant la base projective doivent être connues.

Passage de l'espace projectif à l'espace affine et l'espace Euclidien

Afin de résoudre ce passage de l'espace projectif à l'espace affine et l'espace Euclidien, Mohr et al. proposent une démarche intéressante dans [BoufamaMohrVeillon93]. Ils proposent dans un premier temps de munir l'espace d'un repère défini par 5 points de référence (repère projectif, comme présenté ci-dessus). Ensuite, ils cherchent à déterminer une matrice de passage Ω de l'espace projectif à l'espace euclidien. $\Omega = (\omega_{ij})$ est une matrice homogène 4×4 qui a 15 degrés de liberté. Par conséquent au minimum 15 contraintes doivent être établies pour déterminer cette matrice. Ceci est fait en exploitant quelques contraintes sur la scène.

Soient $P' = (X_{P'}, Y_{P'}, Z_{P'}, W_{P'})^t$ les coordonnées projectives d'un point scène et soient $P = \Omega P' = (X_P, Y_P, Z_P, 1)^t$ ses coordonnées euclidiennes.

Pour obtenir une solution euclidienne unique, un repère est fixé dans la scène. Ensuite sont rajoutées les connaissances du plan horizontal (XOY) et de deux plans verticaux perpendiculaires l'un par rapport à l'autre, XOZ et YOZ . Des directions verticales dans l'image sont calculées à partir de points de fuite.

Ces connaissances se traduisent de la manière suivante :

Fixer un point : La connaissance de la position exacte d'un point dans la scène permet d'obtenir trois équations linéaires en les coefficients de Ω . En $-x$ par exemple on a :

$$X_P = \frac{\omega_{11} X_{P'} + \omega_{12} Y_{P'} + \omega_{13} Z_{P'} + \omega_{14} W_{P'}}{\omega_{41} X_{P'} + \omega_{42} Y_{P'} + \omega_{43} Z_{P'} + \omega_{44} W_{P'}}$$

Appartenance au plan horizontal : Si le point P appartient au plan horizontal, alors $Z_P = 0$ et donc on obtient une équation linéaire en les coefficients de Ω du type :

$$\frac{\omega_{31} X_{P'} + \omega_{32} Y_{P'} + \omega_{33} Z_{P'} + \omega_{34} W_{P'}}{\omega_{41} X_{P'} + \omega_{42} Y_{P'} + \omega_{43} Z_{P'} + \omega_{44} W_{P'}} = 0$$

Le même type d'équations est obtenu pour des points appartenant aux plans verticaux.

Alignement vertical : Si A et B sont deux points scènes alignés verticalement, alors on aura $X_A - X_B = 0$ et $Y_A - Y_B = 0$ qui fournissent deux autres contraintes sur Ω :

$$\frac{\omega_{11} X_{A'} + \omega_{12} Y_{A'} + \omega_{13} Z_{A'} + \omega_{14} W_{A'}}{\omega_{41} X_{A'} + \omega_{42} Y_{A'} + \omega_{43} Z_{A'} + \omega_{44} W_{A'}} = \frac{\omega_{11} X_{B'} + \omega_{12} Y_{B'} + \omega_{13} Z_{B'} + \omega_{14} W_{B'}}{\omega_{41} X_{B'} + \omega_{42} Y_{B'} + \omega_{43} Z_{B'} + \omega_{44} W_{B'}}$$

$$\frac{\omega_{21} X_{A'} + \omega_{22} Y_{A'} + \omega_{23} Z_{A'} + \omega_{24} W_{A'}}{\omega_{41} X_{A'} + \omega_{42} Y_{A'} + \omega_{43} Z_{A'} + \omega_{44} W_{A'}} = \frac{\omega_{21} X_{B'} + \omega_{22} Y_{B'} + \omega_{23} Z_{B'} + \omega_{24} W_{B'}}{\omega_{41} X_{B'} + \omega_{42} Y_{B'} + \omega_{43} Z_{B'} + \omega_{44} W_{B'}}$$

Distance entre deux points : Si la distance entre deux points est d , alors $(X_A - X_B)^2 + (Y_A - Y_B)^2 + (Z_A - Z_B)^2 = d^2$, ce qui

donne :

$$\left(\frac{\omega_{11}X_{A'} + \omega_{12}Y_{A'} + \omega_{13}Z_{A'} + \omega_{14}W_{A'}}{\omega_{41}X_{A'} + \omega_{42}Y_{A'} + \omega_{43}Z_{A'} + \omega_{44}W_{A'}} - \frac{\omega_{11}X_{B'} + \omega_{12}Y_{B'} + \omega_{13}Z_{B'} + \omega_{14}W_{B'}}{\omega_{41}X_{B'} + \omega_{42}Y_{B'} + \omega_{43}Z_{B'} + \omega_{44}W_{B'}} \right)^2 + \left(\frac{\omega_{21}X_{A'} + \omega_{22}Y_{A'} + \omega_{23}Z_{A'} + \omega_{24}W_{A'}}{\omega_{41}X_{A'} + \omega_{42}Y_{A'} + \omega_{43}Z_{A'} + \omega_{44}W_{A'}} - \frac{\omega_{21}X_{B'} + \omega_{22}Y_{B'} + \omega_{23}Z_{B'} + \omega_{24}W_{B'}}{\omega_{41}X_{B'} + \omega_{42}Y_{B'} + \omega_{43}Z_{B'} + \omega_{44}W_{B'}} \right)^2 + \left(\frac{\omega_{31}X_{A'} + \omega_{32}Y_{A'} + \omega_{33}Z_{A'} + \omega_{34}W_{A'}}{\omega_{41}X_{A'} + \omega_{42}Y_{A'} + \omega_{43}Z_{A'} + \omega_{44}W_{A'}} - \frac{\omega_{31}X_{B'} + \omega_{32}Y_{B'} + \omega_{33}Z_{B'} + \omega_{34}W_{B'}}{\omega_{41}X_{B'} + \omega_{42}Y_{B'} + \omega_{43}Z_{B'} + \omega_{44}W_{B'}} \right)^2 = d^2$$

Utilisant ces équations, Ω est calculée en utilisant une méthode de moindres carrés non linéaire. Puisque Ω est définie à un facteur d'échelle près, la contrainte suivante est rajoutée : $\sum_{i,j} (\omega_{ij})^2 = 1$.

Le problème est ensuite écrit comme un problème de minimisation de :

$$\sum_{k=i}^n \left(\frac{f_k(\omega_{11}, \dots, \omega_{44})}{\sigma_k} \right)^2$$

où $f_k()$ est une des fonctions définies précédemment et σ_k est l'écart type associé à la k ème mesure. Utilisant Levenberg-Marquardt, le critère précédent est ainsi minimisé.

Il est assez clair que cette démarche est très intéressante si nous travaillons dans un environnement structuré (bureau, hôpital, etc.), voire même à l'extérieur dans des sites où des contraintes sur la scène peuvent facilement être exploitées.

6. Les méthodes d'autocalibrage des systèmes de vision active

6.1. INTRODUCTION

Le deuxième type d'approches, bien qu'évitant le problème difficile d'estimation de la matrice fondamentale, est encore assez lourd à utiliser pour calibrer une caméra en ligne et reconstruire une scène statique. Nous avons en effet vu que pour arriver à la détermination des paramètres inconnus, il faudra intégrer plusieurs images et avoir la correspondance image de plusieurs points de scène. De plus, la résolution des équations quadratiques résultantes est très coûteuse, empêchant pratiquement une utilisation en ligne de ce genre de méthodes.

Par conséquent, ces méthodes répondent de manière non satisfaisante à la contrainte principale d'un système de vision active, à savoir d'être capable d'effectuer un calibrage en-ligne lorsque les paramètres des caméras composant le système changent en

fonction des tâches visuelles à effectuer. Pour cette raison, un troisième type de méthodes est en cours de développement. Le principe est fondé sur le fait que le système actif peut effectuer des déplacements connus avec grande précision. On fait alors subir à la caméra un certain nombre de mouvements connus et on acquiert des images de points statiques de la scène. Ces points sont poursuivis dans les images et leurs trajectoires observées sont reliées aux paramètres intrinsèques de la caméra. La connaissance de plusieurs points poursuivis ainsi, permet d'obtenir un grand nombre d'équations, par résolution desquelles les paramètres de la caméra sont récupérés.

Ce type de méthodes est présenté par la suite. Cette section s'appuie alors sur les travaux de [Basu93], [Dron93], [DuBrady93], [McLauchlanMurray93], [Viéville94].

6.2. LA MÉTHODE DE VIÉVILLE

Le principe de la méthode de Viéville [Viéville94] est le suivant : on opère sur la caméra (ou les caméras) une série de rotations de même axe, dont on connaît les angles avec une grande précision (la précision typique demandée est de 0.01° avec une répétabilité meilleure que 0.1°). Viéville a démontré qu'en suivant deux points fixes de la scène et en faisant subir 4 rotations à la caméra, il est possible de retrouver ses paramètres intrinsèques, et une partie de ses paramètres extrinsèques (qui sont dans ce cas la position et la direction de l'axe de rotation. La méthode permet de déterminer entièrement la direction de l'axe de rotation et à un coefficient multiplicatif près, la position du centre de rotation).

La méthode consiste donc à suivre un point fixe P de la scène (voir Figure 6.1). La trajectoire du point image associé au point P en fonction de l'axe de rotation (représenté par le vecteur unitaire \vec{k}) et de l'angle θ de la rotation s'écrit :

$$u = \frac{N_1 \cos(\theta) + S_1 \sin(\theta) + O_1}{N_3 \cos(\theta) + S_3 \sin(\theta) + O_3} = u(G, \theta)$$

$$v = \frac{N_2 \cos(\theta) + S_2 \sin(\theta) + O_2}{N_3 \cos(\theta) + S_3 \sin(\theta) + O_3} = v(G, \theta)$$

(u et v étant les coordonnées du point projeté dans les images) où les paramètres N , S et O sont définis à un coefficient multiplicatif près. Plus précisément, le vecteur $G = [O_1, O_2, O_3, S_1, S_2, S_3, N_1, N_2, N_3]^t$ est fonction des paramètres de calibrage : de l'axe \vec{k} de rotation, des coordonnées de C' (la projection du centre optique C de la caméra sur l'axe de rotation), de la matrice des paramètres intrinsèques A et bien sûr des coordonnées du point scène P qui est suivi.

Avec quatre rotations θ_i dont on connaît les angles avec précision, 8 relations sont tirées. Il est donc possible de déterminer les 9 coefficients N_i , S_i et O_i à un multiple près. Pour les déterminer, Viéville propose de minimiser le critère :

$$\sum_{4 \text{ rotations } \theta_i} \left\{ \left(u_i - \frac{N_1 \cos(\theta_i) + S_1 \sin(\theta_i) + O_1}{N_3 \cos(\theta_i) + S_3 \sin(\theta_i) + O_3} \right)^2 + \right.$$

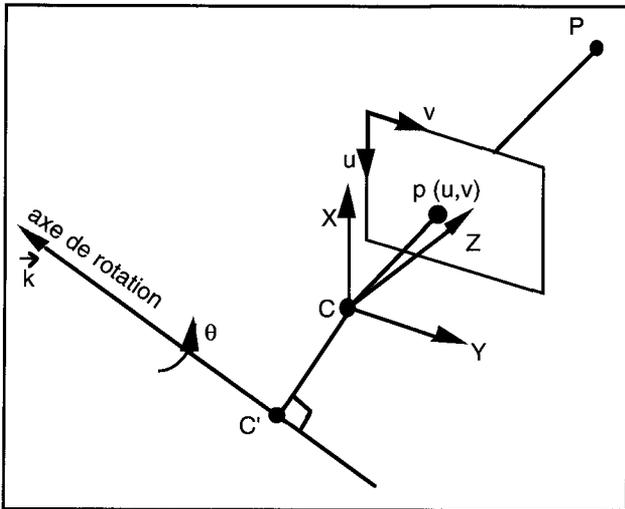


Figure 6.1. – La méthode de Viéville.

$$\left\{ v_i - \frac{N_2 \cos(\theta_i) + S_2 \sin(\theta_i) + O_2}{N_3 \cos(\theta_i) + S_3 \sin(\theta_i) + O_3} \right\}^2$$

à partir d'une valeur initiale obtenue par un développement limité (approximation par un polynôme du 3ème degré de la trajectoire du point scène suivi au moins dans 4 images). Pour chaque point scène ainsi suivi, les paramètres N_i , S_i et O_i qui lui sont associés, sont ainsi obtenus.

Les paramètres intrinsèques

La détermination des paramètres intrinsèques est effectuée de la manière suivante.

On utilise les relations suivantes reliant les coefficients déterminés avec la matrice A , l'axe \vec{k} de rotation et le point observé P :

$$\left. \begin{aligned} A^{-1}S &= \vec{k} \wedge A^{-1}N \\ \vec{k} \wedge A^{-1}O &= \frac{1}{2} \vec{k} \wedge N \\ A^{-1}N + A^{-1}O &= \frac{1}{2} P \end{aligned} \right\} \quad (19)$$

Par ailleurs, Viéville démontre que les vecteurs $(\vec{k}, A^{-1}N, A^{-1}S)$ forment un repère orthonormé et que $\|A^{-1}N\| = \|A^{-1}S\|$.

En utilisant ces dernières relations, on obtient :

$$\left. \begin{aligned} (A^{-1}N) \cdot A^{-1}S &= 0 \Rightarrow (A^{-1}N)^t (A^{-1}S) = 0 \\ \|A^{-1}N\| &= \|A^{-1}S\| \\ \Rightarrow (A^{-1}N)^t (A^{-1}N) &= (A^{-1}S)^t (A^{-1}S) \end{aligned} \right\} \quad (20)$$

d'où :

$$\alpha_1 = \frac{(N_1 - u_0 N_3)(S_1 - u_0 S_3)}{a_u^2} + \frac{(N_2 - v_0 N_3)(S_2 - v_0 S_3)}{a_v^2} + N_3 S_3 = 0$$

$$\alpha_2 = \frac{(N_1 - u_0 N_3)^2 - (S_1 - u_0 S_3)^2}{a_u^2} + \frac{(N_2 - v_0 N_3)^2 - (S_2 - v_0 S_3)^2}{a_v^2} + N_3^2 - S_3^2 = 0$$

où (u_0, v_0, a_u, a_v) sont les paramètres intrinsèques de la caméra. Avec deux points suivis, quatre relations sont obtenues, ce qui est suffisant a priori pour déterminer (u_0, v_0, a_u, a_v) . Il est possible d'obtenir des expressions linéaires si le vecteur à calculer est le vecteur :

$$b = \left(\frac{1}{a_u^2}, \frac{1}{a_v^2}, \frac{u_0}{a_u^2}, \frac{v_0}{a_v^2}, \frac{u_0^2}{a_u^2} + \frac{v_0^2}{a_v^2} \right)^t$$

Par un changement de variables, il est possible d'obtenir des équations linéaires en les paramètres du vecteur b . Compte tenu du fait que b a 5 paramètres, trois points scène sont maintenant nécessaires afin de calculer ces paramètres.

Th. Viéville propose d'améliorer ces équations en calculant la matrice Λ des covariances associées aux grandeurs α_1 et α_2 , puis en minimisant $(\alpha_1, \alpha_2)^t \Lambda^{-1} (\alpha_1, \alpha_2)$. C'est une technique classique destinée à réduire l'importance des paramètres obtenus de manière moins fiable que d'autres. Le critère ainsi modifié présente la particularité de sommer des termes de même covariance (Cf. §4.2 critère GRAD obtenu de manière analogue).

Les paramètres extrinsèques

Connaissant A , il est possible de retrouver le vecteur de rotation \vec{k} . En effet, les vecteurs \vec{k} , $A^{-1}N$ et $A^{-1}S$ forment un repère orthonormé et la direction de \vec{k} est obtenue de la relation suivante :

$$\vec{k} = \frac{(A^{-1}N) \wedge (A^{-1}S)}{\|(A^{-1}N) \wedge (A^{-1}S)\|^2} \quad (21)$$

Il est donc possible, connaissant cet axe, de déterminer la rotation liant la caméra au socle.

De même, il est possible de retrouver à une inconnue près la projection orthogonale du centre optique sur l'axe de rotation, ainsi que la position du point scène P . Si une information supplémentaire est disponible (par exemple, la distance 3D entre les deux points scène suivis), le facteur d'échelle peut être récupéré; il est alors possible de retrouver tous les paramètres extrinsèques. Cette remarque se conçoit aisément : nous ne changeons rien aux résultats (trajectoire+équations) si nous opérons une homothétie d'axe de rotation sur l'ensemble (caméra, points). Cette homothétie représente notre inconnue.

Si C est le centre optique, C' sa projection sur l'axe de rotation, et λ le facteur d'échelle, on a :

$$\vec{C}C' = \lambda \left[\{S^t (A^{-1})^t A^{-1}O\} A^{-1}S + \{N^t (A^{-1})^t A^{-1}O\} A^{-1}N \right] \quad (22)$$

$$\vec{C}P = \frac{\|\vec{C}C'\| \|A^{-1}S\| \|A^{-1}N + A^{-1}O\}}{\sqrt{\{S^t (A^{-1})^t A^{-1}O\}^2 + \{N^t (A^{-1})^t A^{-1}O\}^2}} \quad (23)$$

L'algorithme

Ainsi l'algorithme proposé par Viéville est le suivant :

- a. Poursuivre au moins deux points statiques de la scène pendant au moins quatre images obtenues en faisant une rotation connue autour d'un axe de rotation fixe et enregistrer leurs coordonnées images.
- b. Pour chacun des points scène suivis, identifier ses coefficients $G = (O, S, N)^t$. Pour ceci, il faut d'abord trouver une solution initiale en approchant la trajectoire image observée par un polynôme de degré 3 et ensuite minimiser le critère non linéaire donné précédemment.
- c. Calculer les paramètres intrinsèques avec les relations (20)
- d. Calculer les paramètres extrinsèques en utilisant les équations (21), (22) et (23).

Résultats obtenus

Les résultats obtenus par Viéville semblent prometteurs, et ceci peut se comprendre de manière intuitive : il suppose que les angles sont connus avec une grande précision, et comme la trajectoire d'un point fixe de la scène peut très bien décrire toute l'image, il est possible de récupérer des coefficients N_i, S_i et O_i avec une très grande précision (avec une translation, la trajectoire décrite par le point sera a priori plus courte, et donc l'information recueillie conduira ainsi à une solution moins précise). Connaissant ces coefficients avec une bonne précision, deux relations sont obtenues sur les paramètres intrinsèques, et sont utilisées pour déterminer des valeurs très correctes.

Th. Viéville montre aussi que si la trajectoire des deux points étudiée est horizontale, alors les paramètres intrinsèques relatifs à cet axe (i.e., v_0 et a_v) sont estimés beaucoup plus précisément que les paramètres intrinsèques associés à l'autre axe (i.e., u_0 et a_u). En définitive, si l'on est capable de concevoir un robot doté d'une caméra pouvant librement subir des rotations autour de deux axes orthogonaux, et dont on peut connaître les angles avec précision, alors il sera possible d'obtenir les paramètres intrinsèques de cette caméra, et une partie des paramètres extrinsèques avec de bons résultats.

Bien évidemment, plus le nombre de points scène suivi est grand, meilleur est le résultat et plus le nombre d'images utilisées par point suivi est grand, meilleure est l'estimation des paramètres N_i, S_i et O_i associés à la trajectoire du point scène suivi. Dans ses résultats expérimentaux, Viéville montre qu'une bonne précision sur les paramètres intrinsèques est obtenue avec au minimum 10 points scène suivis dans au moins une dizaine d'images. Quant aux paramètres extrinsèques, le vecteur de rotation k est assez précis, mais le point C' (projection du centre optique sur l'axe de la rotation, cf. Figure 6.1) est très instable. Malheureusement, Viéville n'a pas pour l'instant cherché à obtenir de résultats de reconstruction 3D et donc il n'est pas possible de se prononcer sur la précision atteinte par une telle méthode.

Calcul de la matrice fondamentale utilisant la méthode de Viéville

Nous pouvons observer qu'en utilisant cette méthode, il est possible de déterminer la matrice fondamentale d'un système stéréo. En effet, si nous possédons les paramètres N_i, S_i et O_i associés à la trajectoire d'un point, alors nous devons respecter pour tout angle θ la relation :

$$p^t(\theta) F p(\theta) = [O^t FO + N^t FN] + [N^t FO + O^t FN] \cos(\theta) + [S^t FO + O^t FS] \sin(\theta) + [S^t FS - N^t FC] \sin^2(\theta) + [S^t FN + N^t FS] \cos(\theta) \sin(\theta) = 0.$$

Les fonctions (cosinus, sinus, constante, \sin^2 , et $\cos \sin$) étant indépendantes, nous en déduisons que :

$$\left. \begin{aligned} O^t FO + N^t FN &= 0 \\ N^t FO + O^t FN &= 0 \\ S^t FO + O^t FS &= 0 \\ S^t FS - N^t FC &= 0 \\ S^t FN + N^t FS &= 0 \end{aligned} \right\}$$

Une trajectoire conduit à 5 relations, donc deux trajectoires sont suffisantes a priori. A notre connaissance, une telle méthode n'a pas été implantée à l'heure actuelle.

6.3. LA MÉTHODE DE L'UNIVERSITÉ D'OXFORD

6.3.1. La méthode de McLaughlan and Murray

Dans le cadre du projet ESPRIT 5390 RTGC, l'Université d'Oxford a également développé une méthode de calibrage d'un système de vision active composé d'une caméra montée sur une tourelle [McLaughlanMurray93]. La tourelle est dotée de deux degrés de liberté (rotations en site et azimut) de paramètres connus θ_s, θ_a . Les paramètres recherchés de la caméra sont uniquement les deux facteurs d'échelle a_u et a_v selon les axes u et v . Afin de calculer ces paramètres, l'idée est d'utiliser un certain nombre de points scène stationnaires qui sont poursuivis dans les images acquises lorsque la tourelle de support de caméra effectue des rotations de valeur connue. Pour simplifier les équations, les auteurs supposent qu'il n'y a pas de rotation entre le repère tourelle et le repère associé à la caméra, que la translation du centre optique de la caméra est nulle lorsque la caméra subit des rotations (en réalité ils supposent que l'axe de rotation passe par le centre optique de la caméra). Ces simplifications permettent aux auteurs de travailler directement avec les directions visuelles des points poursuivis sans se préoccuper de leur distance par rapport à la caméra. Ils obtiennent ainsi deux équations reliant les paramètres a_u et a_v recherchés, les coordonnées en X et Y des points scènes, les coordonnées images (u, v) des points scènes poursuivis et les valeurs θ_s, θ_a des angles de rotation appliqués :

$$\left. \begin{aligned} uF_1(X, Y, \theta_s, \theta_a) + a_u F_2(X, Y, \theta_s, \theta_a) &= 0 \\ vF_1(X, Y, \theta_s, \theta_a) - a_v F_3(Y, \theta_s) &= 0 \end{aligned} \right\}$$

La particularité de la méthode présentée réside dans l'utilisation d'un filtre de Kalman linéaire (résultant d'une linéarisation au premier ordre des équations ci-dessus) dont la dimension du vecteur d'état est variable pour estimer au mieux les paramètres recherchés. En effet, le vecteur d'état est composé des deux paramètres recherchés (paramètres globaux) et de paramètres locaux associés aux coordonnées en X et en Y des points scène poursuivis dans les images qui sont estimés en même temps que les paramètres globaux. Le formalisme introduit permet de rajouter ou d'enlever des paramètres locaux du vecteur d'état en fonction du fait que les points scènes sont vus ou non dans les images (sont mis en correspondance ou non). Ceci permet notamment d'intégrer dans le filtre d'une manière naturelle un nombre variable de points scènes poursuivis d'une image à l'autre et de mieux exploiter toute l'information présente à un instant donné avec un coût de calcul linéaire par rapport au nombre des paramètres locaux (rappelons que ce coût est cubique dans le formalisme d'un filtre de Kalman classique).

Les points poursuivis sont des coins. La caméra parcourt une grande partie de l'espace en effectuant des mouvements en zigzag à vitesse angulaire constante. Plusieurs points sont poursuivis, contribuant à une bonne estimation des paramètres recherchés. Les résultats présentés dans l'article sont assez sommaires et démontrent qu'une convergence est atteinte même avec une initialisation éloignée jusqu'à 50% de la vraie valeur. Globalement, la méthode est intéressante surtout à cause de l'introduction d'un nouveau formalisme de Kalman, mais il n'y a malheureusement à l'heure actuelle aucune étude concernant l'influence du bruit image, de l'influence des erreurs sur la mesure des angles de rotation, du nombre des paramètres à utiliser au minimum pour assurer une bonne convergence du filtre et de la qualité des résultats obtenus.

6.3.2. La méthode de Brady

L'idée de base de Brady [DuBrady93] est la même que précédemment : imposer un certain nombre de mouvements connus à la caméra afin de mesurer les paramètres intrinsèques de la caméra par observation des points stationnaires de la scène. Supposons que la caméra subit une rotation $R = (r_x, r_y, r_z)$ connue, le mouvement (\dot{u}, \dot{v}) d'un point stationnaire scène dans le plan image est alors donné par :

$$\begin{bmatrix} a_u \dot{u} \\ a_v \dot{v} \end{bmatrix} = \begin{bmatrix} a_u a_v u v r_x - (a_u^2 u^2 + 1) r_y + a_v v r_z \\ (a_v^2 v^2 + 1) r_x - a_u a_v u v r_y - a_u u r_z \end{bmatrix}$$

où a_u, a_v sont les deux facteurs d'échelle en u et en v respectivement, $u = u_p - u_0, v = v_p - v_0, (u_p, v_p)$ sont les coordonnées images d'un point scène stationnaire, et (u_0, v_0) sont les coordonnées images du point principal.

Si maintenant le mouvement de la caméra et le flot optique sont connus, dans les équations précédentes les seules inconnues sont les 4 paramètres intrinsèques de la caméra. Chaque point stationnaire observé fournit deux équations. Théoriquement deux points

suffisent pour déterminer les inconnues. Si plusieurs points scènes sont observés, alors les paramètres intrinsèques sont trouvés par minimisation du critère :

$$\sum_i K_i K_i^T \rightarrow \min$$

avec

$$K_i = \begin{bmatrix} a_u \dot{u} \\ a_v \dot{v} \end{bmatrix} - \begin{bmatrix} a_u a_v u v r_x - (a_u^2 u^2 + 1) r_y + a_v v r_z \\ (a_v^2 v^2 + 1) r_x - a_u a_v u v r_y - a_u u r_z \end{bmatrix} \quad (A)$$

En effectuant d'abord une rotation autour de l'axe des u , Du et Brady récupèrent le facteur a_v via un processus aux moindres carrés. Une rotation autour de l'axe des v leur permet de récupérer le facteur a_u . Toutefois, le flot optique de points scènes mesuré étant en pratique assez instable, l'algorithme de calcul des paramètres intrinsèques est mal conditionné.

Pour y remédier, les auteurs proposent d'utiliser à la place du flot optique, les trajectoires des points scènes observés. En effectuant des rotations autour de l'axe des u , des points scènes sont poursuivis dans les images et leurs trajectoires sont enregistrées. Ces trajectoires décrivent des hyperboles qui sont symétriques par rapport à l'axe $v_p = v_0$. En extrayant les axes de symétrie de toutes les trajectoires et en les approchant par une représentation unique, on peut récupérer v_0 . u_0 est trouvé par une méthode similaire en effectuant des rotations autour de l'axe des v . Le plus important est que cette méthode est indépendante de la présence d'une distorsion radiale car celle-ci préserve les symétries des lignes passant par le point principal. Quant aux facteurs d'échelle, ils sont récupérés de la façon suivante : les équations (A), dans le cas d'une rotation autour de l'axe des u deviennent :

$$\begin{bmatrix} a_u \Delta u_p \\ a_v \Delta v_p \end{bmatrix} = \begin{bmatrix} a_u a_v (u_p - u_0)(v_p - v_0) \\ 1 + a_v^2 (v_p - v_0)^2 \end{bmatrix} \quad (B)$$

avec $\Delta u_p = \dot{u}_p / r_x, \Delta v_p = \dot{v}_p / r_y$. Si maintenant le point observé est proche de l'axe des u , c'est à dire que $v_p - v_0 \approx 0$, alors nous avons que $a_v = 1 / \Delta v_p$. De façon similaire, a_u est trouvé pour les points observés proches de l'axe des v : $a_u = 1 / \Delta u_p$.

Les résultats expérimentaux présentés sont préliminaires et peu convaincants à l'heure actuelle. Il n'existe aucune comparaison avec des méthodes classiques, donc il est difficile de se prononcer pour la stabilité ou la robustesse de la méthode face à la présence de bruit, des erreurs de poursuite, etc. Les erreurs dues à la précision de mesure des angles de rotation ne sont pas évaluées. De plus, les auteurs font l'hypothèse que les axes de rotations de la caméra coïncident avec ceux du support de la caméra et que ces axes passent par le centre optique, hypothèse qui est difficile à vérifier dans la pratique.

6.4. LA MÉTHODE DE L. DRON

Des idées similaires à celles présentées précédemment sont reprises par Lisa Dron du MIT [Dron93] pour calibrer les

paramètres d'une caméra sans utiliser de mires de calibrage : des caractéristiques sont poursuivies dans une séquence d'images; un algorithme de moindres carrés linéaires permet de considérer des informations provenant de plusieurs images pour estimer les paramètres intrinsèques de la caméra; ces paramètres peuvent être mis à jour dynamiquement quand de nouvelles mesures arrivent.

L'originalité de la méthode par rapport aux précédentes réside dans le fait qu'elle a besoin de connaître uniquement les paramètres de translation de la caméra en mouvement mais pas les paramètres de rotation : selon l'auteur, ceci nécessite un contrôle moins sévère sur les paramètres de contrôle du mouvement de la caméra et rend la méthode moins exigeante vis à vis de la précision d'extraction de caractéristiques images poursuivies comme celle demandée par Brady [DuBrady93], Hartley [Hartley92] ou Faugeras [FaugerasLuongMaybank92]. L'inconvénient est que la méthode nécessite un grand nombre de mouvements de direction différente (pour être stable numériquement, au moins 12 séquences de mouvements différents sont nécessaires). Le deuxième inconvénient de la méthode est que l'on obtient une matrice composée des paramètres intrinsèques multipliés par la matrice de rotation liant la caméra et son support, et non pas directement les paramètres intrinsèques.

6.5. LA MÉTHODE DE BASU

La méthode de Basu [Basu93] consiste à opérer sur la ou les caméras une petite rotation en tangage et une petite rotation en roulis, et à étudier les contours images d'un objet statique. Basu écrit d'abord un certain nombre d'équations qui relient les rotations appliquées avec les paramètres intrinsèques de la caméra recherchés (les deux facteurs d'échelle et la position image du point principal) : les équations sont dérivées en considérant des mouvements en tangage et en roulis très petits et en effectuant ensuite un développement limité d'ordre 1 (stratégie *A*) de ces équations. A partir de ces équations, Basu calcule d'abord (par résolution d'un système de deux équations à deux inconnues) les paramètres u_0 et v_0 de la caméra, puis il obtient les deux facteurs d'échelle. La même procédure est appliquée plusieurs fois jusqu'à obtenir le niveau de précision souhaité. Le problème de cette méthode, bien que théoriquement correcte, réside dans le fait que compte tenu que les rotations appliquées à la caméra sont très petites, les déplacements images des contours observés ne sont pas suffisamment grands pour être fiables. Par conséquent, les erreurs de déplacements observés sont presque du même ordre que les erreurs de localisation des contours.

Pour pallier cet inconvénient, une stratégie alternative est proposée, qui consiste à faire un développement limité d'ordre 2 des équations obtenues et à éliminer les termes négligeables (stratégie *B*). Selon cette stratégie, un seul contour observé suffit pour déterminer le facteur d'échelle a_u , lorsque l'on fait subir à la caméra une petite rotation en tangage. De façon similaire, le facteur a_v est obtenu en observant le même contour après une rotation de

la caméra en roulis. En utilisant ensuite un deuxième contour observé, il est possible de retrouver les coordonnées (u_0, v_0) du point principal. Cette procédure est répétée plusieurs fois jusqu'à l'obtention d'une précision suffisante.

Les résultats expérimentaux que l'auteur présente sont assez intéressants. La stratégie *B* est beaucoup plus fiable et se comporte beaucoup mieux que la stratégie *A* en présence de bruit. Des expérimentations réelles ont démontré que les paramètres intrinsèques de la caméra ont été obtenus avec une erreur relative moyenne de 3% par rapport aux valeurs obtenues avec un calibrage classique.

L'inconvénient majeur de la méthode présentée réside probablement dans le fait que les valeurs obtenues sont très sensibles à la précision sur l'angle des deux rotations : des rotations de 0.03 radians (maximum 0.05 radians) doivent être appliquées de façon très précise afin que la méthode proposée soit valable!

7. Conclusion

Nous avons présenté dans cet article plusieurs méthodes développées récemment concernant le problème d'autocalibrage des caméras CCD, mais aussi le problème de reconstruction relative de scènes.

D'une manière générale, ces nouvelles méthodes peuvent être classifiées en trois catégories différentes : celles qui nécessitent d'abord le calcul de la matrice fondamentale entre des images successives de la séquence, celles qui estiment à la fois les paramètres de transformation perspective de caméras et la structure 3D de la scène et enfin celles qui utilisent la connaissance a priori quant au déplacement effectué par une caméra pour estimer les paramètres intrinsèques de celle-ci.

La première catégorie de méthodes repose sur le calcul de la matrice fondamentale F liant une caméra à l'autre (ou une image en général avec une autre). Toute information susceptible de relier deux plans caméras (position relative, points appariés) est de nature à générer des contraintes sur F . Malheureusement, ce calcul de F s'avère assez instable dans la pratique. Des expériences ont démontré que F dépend de la configuration des points scène dont les images sont utilisées lors du calcul. Qui plus est, le calcul des paramètres intrinsèques résultant est assez délicat. Il dépend du mouvement de la caméra (une seule image est largement insuffisante) et dans le cas général où l'ensemble des paramètres intrinsèques sont considérés comme inconnus, il est difficile de les obtenir. Il vaut mieux considérer dans la pratique que l'angle θ est très proche de $\pi/2$. De plus, comme Zhang l'a montré, les grandeurs u_0 et v_0 influent peu sur la reconstruction. L'hypothèse rencontrée à plusieurs reprises qui consiste à prendre les coordonnées du point principal (u_0, v_0) au centre de la rétine paraît assez raisonnable. Quant aux grandeurs a_u et a_v , elles changent au cours du temps et doivent donc être estimées. Par contre une fois estimés, leur rapport a_u/a_v reste constant.

La deuxième catégorie des méthodes résout plusieurs des inconvénients inhérents à la première catégorie. Pour un système de calibrage en-ligne, le problème principal est la résolution d'un grand nombre d'équations quadratiques par rapport à un grand nombre de paramètres recherchés, donc un coût de calcul important. Des problèmes d'initialisation peuvent apparaître. A l'heure actuelle, elle n'est pas suffisamment expérimentée pour exhiber les conditions de sa convergence et les précisions de reconstruction obtenues en présence des erreurs de correspondances, de bruits images, etc. Toutefois, cette méthode paraît très prometteuse, et peut être utilisée à d'autres problèmes de calibrage (comme par exemple le calibrage entre une caméra et son support « hand-eye calibration » — voir dans [HoraudDornaika-BoufamaMohr94] et [HoraudDornaika95]).

Les méthodes de troisième catégorie, utilisées pour le calibrage de caméras d'un système de vision active, se basent toutes sur l'hypothèse que le mouvement appliqué à la caméra est connu précisément. Dans la plupart des travaux actuels, le mouvement appliqué est une rotation pure dont l'axe est supposé passer par le centre optique de la caméra. A l'heure actuelle, les résultats apparus ne peuvent être considérés que comme préliminaires. Pour la plupart des résultats présentés, les analyses sont encore à un stade quasi-embryonnaire. Notamment aucune étude sur l'influence des hypothèses faites n'existe. Par conséquent, nous attendons très prochainement plusieurs travaux supplémentaires dans ce domaine, devant nous permettre de mieux apprécier ce type de méthodes, leur stabilité et leur robustesse.

Remerciements

L'auteur tient particulièrement à remercier les rapporteurs de cet article qui ont, avec leurs remarques très pertinentes, beaucoup aidé à améliorer la présentation de cet article. Il remercie également M. Bellœil, R. Mohr, R. Horaud, R. Deriche, Th. Viéville, Z. Zang, C. Proy et P. Grandjean pour des discussions passionnantes autour du sujet d'autocalibrage et pour corrections qu'ils ont apportées sur des versions précédentes de cet article. Un grand merci également à toutes les personnes du département Technologie d'ITMI et en particulier à Florence Germain, Patrick Courtney, Fano Ramparany, Serge de Paoli et Patrick Stelmaszyk pour leurs innombrables contributions : discussions fructueuses, relecture de manuscrits, travail en commun, encouragements...

BIBLIOGRAPHIE

[Ayache89] N. Ayache, « Vision stéréoscopique et perception multisensorielle », *InterEditions*, 1989.

[Basu93] A. Basu, « Active calibration : alternative strategy and analysis », In *Procs IEEE Conference on Computer Vision and Pattern Recognition*, New York City, USA, June 1993, pp. 495-500.

[Beyer92] H. Beyer, « Geometric and Radiometric Analysis of a CCD-Camera Based Photogrammetric Close-Range System », *PhD Thesis*, Institut für Geodäsie und Photogrammetrie an der Eifgenössischen Technischen Hochschule Zürich, 1992.

[BoufamaMohrVeillon93] B. Boufama, R. Mohr, F. Veillon, « Euclidean constraints for Uncalibrated Reconstruction », In *Procs. 4th International Conference on Computer Vision (ICCV)*, Berlin, Germany, May 1993, pp. 466-470.

[DuBrady93] F. Du, M. Brady, « Self-Calibration of the Intrinsic Parameters of Cameras for Active Vision Systems », In *Procs. IEEE Conference on Computer Vision and Pattern Recognition, CVPR'93*, New York, June 1993, pp. 477-482.

[Dron93] L. Dron, « Dynamic Camera Self-Calibration for Controlled Motion Sequences », In *Procs. IEEE Conference on Computer Vision and Pattern Recognition, CVPR'93*, New York, June 1993, pp. 501-506.

[FaugerasToscani87] O.D. Faugeras, G. Toscani, « Camera calibration for 3D Computer Vision » In *Proc. International Workshop on Machine Vision and Machine Intelligence*, Tokyo, Japan, February 1987.

[Faugeras88] O.D. Faugeras, « Quelques pas vers la Vision Artificielle en trois dimensions », *Technique et Sciences Informatiques*, 7(6), 1988, p. 547-590.

[Faugeras92] O.D. Faugeras, « What can be seen in three dimensions with an uncalibrated stereo rig », in *Procs. Second European Conference on Computer Vision*, 2nd ECCV, Santa Margherita Ligure, Italy, May 1992.

[FaugerasLuongMaybank92] S. Maybank, O.D. Faugeras, Q.T. Luong, « Camera self-calibration - Theory and experiments », in *Procs. Second European Conference on Computer Vision*, 2nd ECCV, Santa Margherita Ligure, Italy, May 1992.

[Fletcher87] P. Fletcher, « Practical methods of optimization », second edition, *John Wiley and Son Ed.*, 1987.

[GillMurrayWright81] P. Gill, W. Murray, M. Wright, « Practical optimization », *Academic Press, Harcourt Brace Jovanovich, Publishers*.

[Hartley92] R.I. Hartley, « Estimation of relative camera positions for uncalibrated cameras », in *Procs. Second European Conference on Computer Vision*, 2nd ECCV, Santa Margherita Ligure, Italy, May 1992.

[Hartley94] R.I. Hartley, « Self-Calibration from multiple views with a rotating camera », *Third European Conference on Computer Vision*, J. O. Eklundh Ed., Vol. I, Stockholm, Sweden, June 1994, pp. 471-478.

[HoraudDornaika95] R. Horaud, F. Dornaika, « Hand-Eye calibration », in *International Journal of Robotics Research*, 14(3), June 1995.

[HoraudDornaikaBoufamaMohr94] R. Horaud, F. Dornaika, B. Boufama, R. Mohr, « Self Calibration of a Stereo Head Mounted onto a Robot Arm », *Third European Conference on Computer Vision*, J. O. Eklundh Ed., Vol. I, Stockholm, Sweden, June 1994, pp. 455-462.

[HuangFaugeras89] T. Huang, O.D. Faugeras, « Some properties of the E matrix in two-view motion estimation », *IEEE Transactions on PAMI*, vol. 11, 1989, pp. 1310-1312.

[Longuet-Higgins81] C. Longuet-Higgins, « A Computer Algorithm for Reconstructing a Scene from Two Projections », *Nature*, 293 :133-135, 1981.

[Luong92] Q.-T. Luong, « Matrice fondamentale et calibration visuelle sur l'environnement — Vers une plus grande autonomie des systèmes Robotiques », *Thèse de l'université de Paris-Sud*, Centre d'Orsay, Décembre 1992.

[LuongDericheFaugerasPapadopoulo93] Q.T. Luong, R. Deriche, O. D. Faugeras, T. Papadopoulo, « On determining the Fundamental matrix : analysis of different methods and experimental results », *Rapport de Recherche INRIA N° 1894*, Avril 1993.

[LuongFaugeras93] Q.-T. Luong, O.D. Faugeras, « Determining the Fundamental matrix with planes : instability and new algorithms », In *Procs IEEE Conference on Computer Vision and Pattern Recognition*, New York City, USA, June 1993, pp. 489-494.

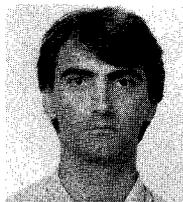
[MaybankFaugeras92] S.J. Maybank, O.D. Faugeras, « A theory of self-calibration of a moving camera », *International Journal of Computer Vision*, 8 :2, 1992, pp. 123-151.

[McLaughlanMurray93] P.F. McLaughlan, D.W. Murray, « Variable state dimension filter applied to active camera calibration », In *Procs SPIE Sensor Fusion VI*, Vol. 2059, September 1993, Boston, MA, USA, pp. 14-25.

- [Mohr92] R. Mohr, « Projective Geometry and Computer Vision », In *Handbook of Pattern Recognition and Computer Vision*, Eds. C. H. Chen, L. F. Pau, and P.S. Wang, World Scientific Publishing Company, 1992.
- [MohrVeillon92] R. Mohr, F. Veillon, « Géométrie projective, une introduction », *Rapport LIFIA*, Octobre 1992, 23 pages.
- [MohrBoufamaBrandt92] R. Mohr, B. Boufama, P. Brandt, « Accurate Preprojective Reconstruction », In *Mundy Zisserman Forsyth Ed.*, Applications of Invariance in Computer Vision, Springer Verlag, Berlin Heidelberg, 1994, pp. 257-276.
- [MohrVeillonQuan93] R. Mohr, F. Veillon, L. Quan, « Relative 3D Reconstruction using multiple uncalibrated images », In *Procs IEEE Conference on Computer Vision and Pattern Recognition*, New York City, June 1993.
- [MongaHeraud93] O. Monga, R. Heraud, « Vision par Ordinateur : Outils fondamentaux », *Editions Hermès*, 1993.
- [PhongHeraudYassineTao95] T.Q. Phong, R. Heraud, A. Yassine, P.D. Tao, « Object Pose from 2-D to 3-D Point and Line Correspondences », to appear in *International Journal of Computer Vision*, 1995.
- [PugetSkordas90] P. Puget, T. Skordas, « Calibrating a mobile camera », in *Journal of Image and Vision Computing*, Vol. 8, N°4, November 1990.
- [Robert93] L. Robert, « Perception stéréoscopique de courbes et de surfaces tridimensionnelles. Application à la Robotique Mobile », *Thèse de l'université de Paris-Sud*, Centre d'Orsay, 1993.
- [Shashua93] A. Shashua, « Projective Depth : A Geometric Invariant for 3D Reconstruction from two Perspective/Orthographic Views for Visual Recognition », In *Procs. International Conference on Computer Vision (ICCV)*, Berlin, Germany, May 1993, pp. 583-590.
- [SempleKneebone52] J.G. Semple, G.T. Kneebone, « Algebraic Projective Geometry », *Oxford Science Publication*, 1952.
- [ToscaniFaugeras86] G. Toscani, O. Faugeras, « The Calibration problem for stereo », In *Procs IEEE Conference on Computer Vision and Pattern Recognition*, CVPR'86, 1986, pp. 15-20.
- [Tsai87] R.Y. Tsai, « A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses », In *IEEE Journal of Robotics and Automation*, 3(4) : 323-344, 1987.
- [Tsai89] R.Y. Tsai, « Synopsis of recent progress on camera calibration for 3D machine vision », in *O. Khatib, J. J. Craig and T. Lozano-Pérez eds.*, The Robotics Review, MIT Press, MA, USA 1989.
- [Viéville94] Th. Viéville, « Auto-calibration of visual sensor parameters on a robotic head », in *International Journal of Image and Vision Computing*, Vol. 12, 1994.
- [ViévilleZellerRobert95] Th. Viéville, C. Zeller, L. Robert, « Using collineations to compute motion and structure in an uncalibrated image sequence », to appear in *International Journal of Computer Vision*, 1995.
- [ZhangLuongFaugeras93] Z. Zhang, Q.-T. Luong, O.D. Faugeras, « Motion of an uncalibrated stereo rig : self-calibration and metric reconstruction », in *Procs. Asiatic Conference on Computer Vision*, November 1993, Osaka, Japan.

LES AUTEURS

Thomas SKORDAS



Thomas Skordas est né à Mylia, Grèce en 1962. Il a reçu son diplôme d'Ingénieur Électricien de l'Université Aristote de Thessalonique en 1984. Il a obtenu le DEA en Automatique et Traitement du Signal de l'École Nationale Supérieure d'Ingénieurs Électriciens de Grenoble en 1985. Il a présenté une thèse de doctorat de l'Institut National Polytechnique de Grenoble, spécialité Informatique, en octobre 1988

(dans le domaine de la Vision par Ordinateur). Pendant sa thèse, il a été soutenu par les fondations Helléniques « Bodossakis » et « Alexandros Onassis ». Depuis Octobre 1988, il travaille à ITMI APTOR S.A., dans le département Technologie. Depuis 1993, il a la responsabilité des activités technologiques en Vision par Ordinateur. Ses principaux domaines de spécialisation sont le calibrage de capteurs visuels, le traitement d'images, la reconnaissance de formes, la modélisation de scènes et l'analyse du mouvement.

Manuscrit reçu le 21 Décembre 1993.

8. Annexes

A.1. NOTATIONS ET SYMBOLES UTILISÉS

Les principaux symboles utilisés dans cet article sont les suivant :

P : Point scène.

Ses coordonnées cartésiennes sont $P = (X, Y, Z)^t$. Ses coordonnées projectives (homogènes) sont $(X, Y, Z, W)^t$.

p : projection dans l'image gauche du point P .

Ses coordonnées projectives exprimées dans le repère image sont $p = (p_u, p_v, p_w)^t = (su, sv, s)^t$. Ses coordonnées en pixels sont $p = (u, v)^t$.

p' : projection dans l'image droite du point P .

Ses coordonnées projectives exprimées dans le repère image sont $p' = (p'_u, p'_v, p'_w)^t = (su', sv', s)^t$. Ses coordonnées en pixels sont $p'(u', v')^t$.

p_n, p'_n : projections dans l'image gauche et droite respectivement du point P . Leurs coordonnées sont exprimées dans les repères rétinien gauche et droit normalisés.

e, e' : épipoles gauche et droit respectivement

l, l' : droites épipolaires gauche et droite respectivement

M, M' : matrices de transformation perspective de la caméra gauche et droite respectivement

$u_0, v_0, a_u, a_v, \theta$: les paramètres intrinsèques d'une caméra

f : la distance focale d'une caméra

A, A' : matrices des paramètres intrinsèques de la caméra gauche et droite respectivement

$D = (R, t)$: matrice de déplacement rigide avec R = matrice de rotation et t = vecteur translation

F = matrice fondamentale liant deux images = matrice de passage entre les coordonnées projectives image

E = matrice essentielle liant les repères associés à deux caméras = matrice de passage entre les coordonnées projectives rétinien normalisées

Δ : matrice de Kruppa

T_x : matrice antisymétrique associée au vecteur x .

A.2. GÉOMÉTRIE PROJECTIVE UNE INTRODUCTION

Comme nous l'avons vu, nous sommes souvent amenés à manipuler des coordonnées « homogènes » :

$$p = \begin{pmatrix} su \\ sv \\ s \end{pmatrix} = MP = M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Le vecteur $(su, sv, s)^t$ représente la droite associée au point $(u, v)^t$ du plan image. La géométrie projective représente une formalisation de cette notion. Les paragraphes qui suivent ont pour objectif de donner quelques définitions et quelques résultats intéressants pour le calibrage de caméras. Pour plus de détail, on peut regarder les articles de [Mohr92], [MohrVeillon92] et surtout le livre [SempleKneebone52].

La définition de l'espace projectif est la suivante :

Définition

Soit (x_1, \dots, x_{n+1}) un point de \mathbb{R}^{n+1} . On définit l'espace projectif \mathbb{P}^n comme l'espace quotient de $\mathbb{R}^{n+1} - \{0, 0, \dots, 0\}$ par l'équivalence :

$$X \sim Y \iff \text{il existe } \lambda \neq 0, X = \lambda Y$$

où

$$X = (x_1, \dots, x_{n+1}) \text{ et } Y = (y_1, \dots, y_{n+1})$$

A un point de \mathbb{P}^n on associe les coordonnées (x_1, \dots, x_{n+1}) qui sont donc définies à un facteur d'échelle près. On peut aussi avoir \mathbb{P}^n comme l'ensemble des droites vectorielles de \mathbb{R}^{n+1} .

Le dessin de la Figure B.1 illustre cette notion en dimension $n = 2$.

Prolongement canonique de \mathbb{P}^n dans \mathbb{R}^n

On identifie habituellement un point de $\mathbb{P}^n - \{x_1, \dots, x_n, 0\}$ à un point de \mathbb{R}^n par l'application :

$$(x_1, \dots, x_{n+1}) \rightarrow \left(\frac{x_1}{x_{n+1}}, \dots, \frac{x_n}{x_{n+1}} \right)$$

On note que si x_{n+1} tend vers 0, ce point de \mathbb{R}^n tend vers l'infini dans la direction (x_1, \dots, x_n) , ce qui justifie l'appellation d'hyperplan de l'infini pour le plan d'équation $x_{n+1} = 0$. Le point $(x_1, \dots, x_n, 0)$, peut être vu comme une direction de l'espace \mathbb{R}^n .

A toute droite de l'espace projectif correspond donc un point de l'espace affine : il suffit de prendre son intersection avec le plan $x_{n+1} = 1$. Seules les droites du plan $x_{n+1} = 0$ n'ont pas de correspondant. En fait, elles correspondent à des points situés à l'infini, ou plus simplement à des directions asymptotiques. Nous voyons bien sur la Figure B.1 que si M' partait à l'infini, sa droite associée aurait tendance à tendre vers une droite incluse dans le plan (O, x, y) de coordonnées $(x, y, 0)$.

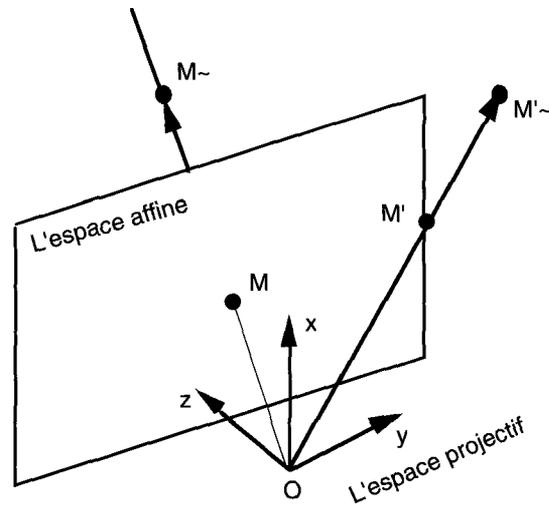


Figure B.1. – L'espace affine de dimension 2 et son espace projectif de dimension 3. A M correspond la droite $M\sim$ et inversement.

Les applications projectives et leurs propriétés

Une *transformation projective* ou *homographie* (en anglais « collineation ») est une application de \mathbb{P}^n dans \mathbb{P}^n définie par une matrice Q de dimension $(n+1) \times (n+1)$ telle que l'image de (x_1, \dots, x_{n+1}) soit définie par :

$$\begin{pmatrix} y_1 \\ \dots \\ y_{n+1} \end{pmatrix} = Q \begin{pmatrix} x_1 \\ \dots \\ x_{n+1} \end{pmatrix}$$

Comme les deux vecteurs colonnes ci-dessus sont définis à un facteur d'échelle près, il en est de même de la matrice Q .

Donc une transformation projective est une application linéaire en coordonnées homogènes. Toutes les transformations affines standards (translations, rotations, changements d'échelles et leur combinaison) peuvent facilement être représentées par des transformations projectives. Par exemple, un déplacement rigide de \mathbb{R}^3 est défini par une translation $t = (a, b, c)^t$ et une matrice de rotation (3×3) Rot . L'application linéaire (projective) associée à ce déplacement rigide (Rot, t) est :

$$\begin{bmatrix} Rot & \vec{t} \\ \vec{0}_3 & 1 \end{bmatrix}$$

On peut facilement démontrer que les applications projectives ont les propriétés suivantes :

Propriétés

- Une translation de \mathbb{R}^n est une application projective qui laisse les points à l'infini invariants.
- Les transformations affines sont les applications projectives qui laissent l'hyperplan de l'infini globalement invariant.

Bases projectives : Une base de l'espace projectif P^n est formée de $n + 2$ points tels que $n + 1$ parmi eux n'appartiennent pas au même hyperplan. Une base canonique est habituellement choisie par $(1, 0, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, 0, \dots, 0, 1)$ auxquels est ajouté le point unitaire $(1, 1, \dots, 1)$. Pour l'espace P^3 , 5 points sont donc nécessaires pour définir une base projective : les quatre premiers définissent les 3 axes de l'espace P^3 et l'origine. Le cinquième est le point unitaire.

Les invariants projectifs

Il existe une notion importante de l'espace projectif : il s'agit des bi-rapports. On définit le bi-rapport de 4 points alignés, de 4 droites coplanaires ayant un point commun et de 4 plans concourants ayant une droite commune.

L'invariant projectif de base : Le bi-rapport est l'invariant projectif de base.

Bi-rapport de 4 points alignés

Soient deux points M et N de P^1 . Tout point A de la droite $[M, N]$ peut s'écrire $A = \kappa M + \mu N$ (cf. Figure B.2), avec (κ, μ) définis à un coefficient de proportionnalité près, définissant ainsi un espace projectif de dimension 1 sur la droite $[M, N]$. Soient $A_i, \{i = 1, \dots, 4\}$ quatre points (κ_i, μ_i) de cette droite. On définit le bi-rapport de ces 4 points alignés A_1, A_2, A_3, A_4 par :

$$\{A_1, A_2 ; A_3, A_4\} = \frac{\frac{\kappa_1 - \kappa_3}{\mu_1 - \mu_3} / \frac{\kappa_2 - \kappa_3}{\mu_2 - \mu_3}}{\frac{\kappa_1 - \kappa_4}{\mu_1 - \mu_4} / \frac{\kappa_2 - \kappa_4}{\mu_2 - \mu_4}} = \frac{\lambda_1 - \lambda_3}{\lambda_1 - \lambda_4} / \frac{\lambda_2 - \lambda_3}{\lambda_2 - \lambda_4}$$

On peut vérifier que ce bi-rapport est indépendant du choix des points M et N .

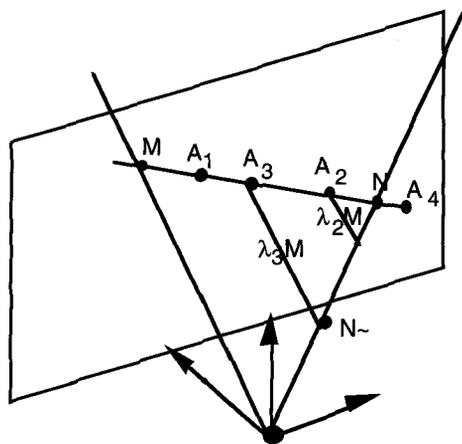


Figure B.2. – La définition du bi-rapport de 4 points alignés. λ s'obtient via une projection parallèlement à un axe.

Propriété : Le bi-rapport de 4 points alignés reste invariant par toute application projective. Pour toute application projective f , on a donc :

$$\{f(A_1), f(A_2) ; f(A_3), f(A_4)\} = \{A_1, A_2 ; A_3, A_4\}$$

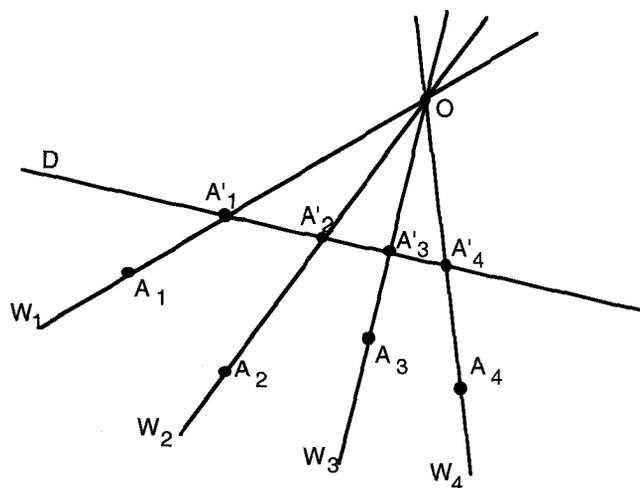


Figure B.3. – Le bi-rapport de 4 droites concourantes est identique à celui défini par les 4 points A'_1, A'_2, A'_3, A'_4 .

Il est possible de démontrer que le bi-rapport s'écrit plus simplement sous la forme :

$$\{A_1, A_2 ; A_3, A_4\} = \frac{\lambda_1 - \lambda_3}{\lambda_1 - \lambda_4} / \frac{\lambda_2 - \lambda_3}{\lambda_2 - \lambda_4} = \frac{\overline{A_1 A_3}}{\overline{A_1 A_4}} / \frac{\overline{A_2 A_3}}{\overline{A_2 A_4}}$$

La notation $\overline{A_1 A_3}$ est la longueur du segment $A_1 A_3$. Cela veut dire qu'étant donné 3 points sur une droite et un bi-rapport, le quatrième point peut parfaitement être positionné sur la droite. 3 points sur une droite définissent une base projective dans P^1 . Tout autre point est positionné de manière unique sur cette droite par rapport à ces trois points avec un bi-rapport. Le bi-rapport est par conséquent la coordonnée projective de tout point par rapport à la base définie par ces 3 points alignés.

Bi-rapport de 4 droites coplanaires concourantes

Pour quatre droites coplanaires W_1, W_2, W_3, W_4 ayant un point O en commun, on définit le bi-rapport par la relation (cf. Figure B.3) :

$$\{W_1, W_2 ; W_3, W_4\} = \frac{\frac{\det(O, A_1, A_3)}{\det(O, A_1, A_4)}}{\frac{\det(O, A_2, A_3)}{\det(O, A_2, A_4)}}$$

où A_i est un point quelconque de W_i . On montre que le bi-rapport ainsi défini est indépendant des points choisis. Lorsque les points sont alignés, on peut vérifier que la définition du bi-rapport de 4 droites coïncide avec la définition du bi-rapport des quatre points alignés :

$$\{W_1, W_2 ; W_3, W_4\} = \frac{\frac{\det(O, A_1, A_3)}{\det(O, A_1, A_4)}}{\frac{\det(O, A_2, A_3)}{\det(O, A_2, A_4)}} = \{A_1, A_2 ; A_3, A_4\}$$

si $(A_1, A_2 ; A_3, A_4)$ sont alignés.

Propriété : De même qu'avec le bi-rapport de 4 points alignés, l'application projective qui transforme une droite en une autre

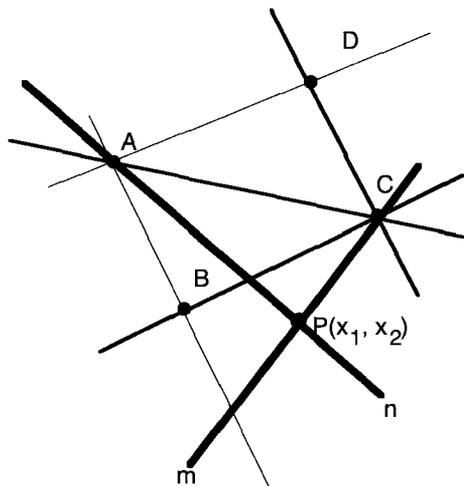


Figure B.4. – Ayant donné 4 points (non alignés trois par trois), tout autre point de l'espace P^2 peut être défini de manière unique par rapport à la base projective définie par ces 4 points. En effet, $P(x_1, x_2)$ est le point défini par l'intersection des droites C_m et A_n où C_m est la droite définie par le bi-rapport $x_1 = \{[CA], [CB], [CD], [CP]\}$ et A_n par le bi-rapport $x_2 = \{[AB], [AC], [AD], [AP]\}$.

droite laisse invariant le bi-rapport de 4 droites coplanaires concourantes W_1, W_2, W_3, W_4 .

En utilisant maintenant ce qui a été décrit ci-dessus, nous pouvons remarquer que dans le cas du plan projectif P^2 , un point P est défini de manière unique par l'intersection de deux droites. Par conséquent, si l'on se donne 4 points non alignés, alors n'importe quel point du plan est défini par deux bi-rapports définissant chacun une droite (voir Figure B.4) :

$$x_1 = \{[CA], [CB], [CD], [CP]\}$$

droites ayant pour point commun le point C

$$x_2 = \{[AB], [AC], [AD], [AP]\}$$

droites ayant pour point commun le point A

En fait, cette construction consiste à doter l'espace projectif d'un repère. Comme celui-ci est de dimension 3, il est parfaitement logique d'utiliser 4 points. L'avantage de cette construction provient du fait que si on connaît les images des 4 points A, B, C , et D par une application projective, alors les coordonnées de l'image du point P restent inchangées dans le nouveau repère. On est donc capable de le positionner.

Bi-rapport de faisceaux de plans

Un faisceau de plans est un ensemble de 4 plans ayant une droite en commun. Le bi-rapport d'un faisceau de plans est égal à celui d'un faisceau de droites résultant de l'intersection du faisceau de plans par un plan sécant.

A.3. Calibrage d'une caméra La matrice de transformation perspective

Considérons le modèle sténopé d'une caméra présenté dans la Figure 3.1. Avec ce modèle, le repère caméra (C, X_C, Y_C, Z_C) et le repère rétinien (c, x_c, y_c) sont liés par la transformation perspective :

$$\frac{f}{Z_C} = \frac{x_c}{X_C} = \frac{y_c}{Y_C}$$

En utilisant les coordonnées projectives pour la caméra et la rétine, cette relation peut se réécrire sous la forme :

$$\begin{bmatrix} \alpha x_c \\ \alpha y_c \\ \alpha \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = U_f \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix}$$

Ainsi, la relation entre les coordonnées rétinienne et les coordonnées caméras (exprimée par la matrice $3 \times 4 U_f$) est linéaire (voir Annexe B). La matrice U_f est une application projective permettant le passage de l'espace projectif P^3 dans le plan projectif P^2 et reflète la projection perspective.

Si l'on passe du repère rétinien (c, x_c, y_c) au repère image (I, u, v) (par le biais d'un changement d'unités et d'une translation de l'origine du repère), nous avons la relation suivante :

$$x_c = \frac{u - u_0}{k_u} \quad \text{et} \quad y_c = \frac{v - v_0}{k_v}$$

ou écrite sous forme de coordonnées projectives :

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = U_{im} \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix}$$

La matrice $3 \times 3 U_{im}$ est une transformation affine représentant un changement d'échelle et une translation. (u_0, v_0) sont les coordonnées du point principal c en pixels et k_u, k_v deux facteurs d'échelle selon les axes u et v .

A partir des équations précédentes, en posant $a_u = k_u f$ et $a_v = k_v f$, nous pouvons écrire la relation suivante liant les coordonnées projectives image aux coordonnées projectives caméra :

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = U_{im} U_f = \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} f k_u & 0 & u_0 & 0 \\ 0 & f k_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_u & 0 & u_0 & 0 \\ 0 & a_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = U \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix}$$

La matrice $3 \times 3 U$ (avec $U = U_{im} U_f$) est une application projective de l'espace projectif P^3 dans le plan projectif P^2 . Elle est constituée des paramètres intrinsèques de la caméra. Un modèle amélioré peut être utilisé pour exprimer la transformation entre le repère rétinien et le repère image en introduisant un paramètre supplémentaire θ qui représente l'angle que font les deux vecteurs u et v du repère image entre eux. Ainsi on obtient la relation suivante entre les coordonnées projectives images et caméra :

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} a_u & -a_u \cot \theta & u_0 & 0 \\ 0 & a_v / \sin \theta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix}$$

En pratique, sur un capteur CCD, $\theta \approx \frac{\pi}{2}$.

Les repères scène (W, X_W, Y_W, Z_W) et caméra (C, X_C, Y_C, Z_C) sont reliés par une transformation rigide $D(R, t)$ composée d'une rotation R et d'une translation t . Ainsi, le passage d'un repère à l'autre s'écrit (en coordonnées projectives) :

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = D \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \text{ avec } D = \begin{bmatrix} R & \vec{t} \\ \vec{0}_3^t & 1 \end{bmatrix}$$

La matrice de calibrage M est la matrice de passage du repère scène (W, X_W, Y_W, Z_W) au repère image (I, u, v).

Elle s'écrit comme :

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} = UD$$

$$= \begin{bmatrix} a_u & -a_u \cot \theta & u_0 & 0 \\ 0 & a_v / \sin \theta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \vec{t} \\ \vec{0}_3^t & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} a_u & -a_u \cot \theta & u_0 & 0 \\ 0 & a_v / \sin \theta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \vec{t} \\ \vec{0}_3^t & 1 \end{bmatrix}$$

cette dernière relation peut s'écrire sous la forme :

$$M = A \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} D$$

A est une matrice 3×3 appelée *matrice des paramètres intrinsèques*. $D(R, t)$ est la matrice 4×4 de déplacement rigide entre

les repères scène et caméra et est appelée *matrice des paramètres extrinsèques*.

$M = (m_{ij})$ dépend donc de 12 paramètres, mais elle est définie à un coefficient multiplicatif près. Elle est appelée *matrice de transformation perspective*. M dépend donc de $3 \times 4 - 1 = 11$ coefficients qui sont à relier avec les 5 paramètres intrinsèques dont dépend la matrice A et les 6 paramètres extrinsèques de la matrice D .

Si un point de l'espace observé, de coordonnées projectives $P = (X, Y, Z, 1)^t$, se projette dans l'image au point de coordonnées projectives $p = (su, sv, s)^t$, il existe la relation suivante :

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = M \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$M = A \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} D = AVD$$

où V est la matrice 3×4 :

$$V = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

La matrice $U = AV$ permet de passer du repère caméra au repère image. Avec le modèle sténopé adopté précédemment, il est impossible de calculer explicitement la distance focale f . On peut alors introduire des coordonnées caméras sans dimension : les unités du repère de la caméra sont choisies de sorte que la distance focale soit égale à 1 dans ce repère. On appelle ce nouveau repère *repère caméra normalisé*. Si maintenant au point principal c de la rétine on associe un système de coordonnées (c, x_c, y_c) d'unités égales à celles du repère caméra et d'axes parallèles aux axes X_C et Y_C du repère caméra, alors ce repère est le *repère rétinien normalisé*. La relation entre les coordonnées caméra (X_C, Y_C, Z_C) , les coordonnées rétiniennes (x_c, y_c) et les coordonnées rétiniennes normalisées (x_c^{norm}, y_c^{norm}) est donnée par la projection perspective :

$$\left. \begin{aligned} x_c = f \frac{X_C}{Z_C} &\Rightarrow \frac{x_c}{f} = \frac{X_C}{Z_C} \\ &\Rightarrow x_c^{norm} = \frac{X_C}{Z_C} \\ y_c = f \frac{Y_C}{Z_C} &\Rightarrow \frac{y_c}{f} = \frac{Y_C}{Z_C} \\ &\Rightarrow y_c^{norm} = \frac{Y_C}{Z_C} \end{aligned} \right\} \Rightarrow \frac{1}{Z_C} = \frac{x_c^{norm}}{X_C} = \frac{y_c^{norm}}{Y_C}$$

Cette dernière relation peut se réécrire sous la forme (en utilisant les coordonnées projectives) :

$$\begin{bmatrix} sx_c^{norm} \\ sy_c^{norm} \\ s \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = V \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix}$$

La matrice 3×4 V lie le repère caméra avec le repère rétinien normalisé. Elle est une application projective permettant le passage de l'espace projectif P^3 dans le plan projectif P^2 et reflète la projection perspective. La relation entre les coordonnées rétinienes normalisées et les coordonnées caméras est ici linéaire.

Enfin, il est facile de voir que la relation entre les coordonnées images et rétinienes normalisées est également linéaire. Elle est donnée par la matrice des paramètres intrinsèques A , qui est également une application projective.

Ainsi, la matrice de projection perspective $M = AVD$ liant les coordonnées scène avec les coordonnées image est le produit d'un ensemble d'applications projectives (applications linéaires en les coordonnées homogènes). Par conséquent, nous avons adopté un modèle de caméra pour lequel la relation entre les coordonnées scène et les coordonnées image est linéaire projective.