

Quelle information géométrique peut-on obtenir à partir d'une ou plusieurs images prises par projection perspective ?

Patrick Gros

► **To cite this version:**

Patrick Gros. Quelle information géométrique peut-on obtenir à partir d'une ou plusieurs images prises par projection perspective ?. Journées Mathématiques de la Vision, Feb 1993, Rennes, France. 1993. <inria-00590022>

HAL Id: inria-00590022

<https://hal.inria.fr/inria-00590022>

Submitted on 4 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quelle information géométrique peut-on obtenir à partir d'une ou plusieurs images prises par projection perspective ?

Patrick GROS
LIFIA * - INRIA Rhône Alpes
46, avenue Félix Viallet, 38031 Grenoble Cedex 1
Patrick.Gros@imag.fr

Les travaux présentés dans cet article ont été réalisés au sein du projet MOVI du laboratoire LIFIA à Grenoble, par Boubakeur Boufama, Pascal Brand, Patrick Gros, Luce Morin, Long Quan et Françoise Veillon, le tout avec la participation et sous la direction de Roger Mohr. Les contributions de chacun seront précisées dans le fil du texte par les références bibliographiques, auxquelles le lecteur est invité à se reporter pour les détails techniques qui ne seront pas tous donnés ici. L'ensemble du travail a été réalisé dans le cadre du projet ESPRIT - BRA VIVA.

Résumé

En vision par ordinateur, on considère une caméra qui prend des images. En supposant simplement que cette opération de prise de vue est d'un certain type géométrique, et plus précisément que c'est une projection perspective, on peut calculer à partir d'une ou de plusieurs images des quantités géométriques caractéristiques de la scène observée. Après avoir étudié quelques modèles géométriques de caméras, les informations géométriques que l'on peut tirer d'une, deux, trois ou plusieurs images sont étudiées successivement.

1 Introduction

Dans tout cet article, on se place dans le cadre de la vision par ordinateur. On dispose d'une caméra, qui prend des images analogiques. Ces images sont numérisées et stockées dans la mémoire d'un ordinateur. On cherche alors quelle information ces images peuvent fournir à propos de la scène qui a été observée par la caméra. En particulier, on cherche à déterminer les informations numériques que l'on peut calculer, même si on ne connaît pas exactement les paramètres de la caméra : sa focale, sa position, son orientation. . .

Dans ce cas, il est bien entendu impossible de calculer des longueurs, des angles, des surfaces ou même des rapports de ces grandeurs. Le cadre naturel pour une telle étude est celui de la géométrie projective : les grandeurs que l'on peut calculer sont les birapports, et il est même possible de faire de la reconstruction, c'est à dire de retrouver la géométrie tridimensionnelle de la scène observée, ceci à une transformations projective de l'espace près.

L'utilisation de la géométrie projective en vision par ordinateur n'est pas récente [LH 81], mais n'a pris son essor que récemment, notamment par le biais de l'étude du positionnement relatif des objets de la scène [Moh 91] et de celle des invariants [Mun 92, Mun 93]. Ces méthodes sont toutes basées sur le même principe : on suppose que la caméra réalise une projection perspective, et on utilise alors les contraintes fournies par cette projection.

L'autre possibilité pour calculer des grandeurs numériques consiste à étalonner la caméra, c'est à dire à déterminer ses paramètres [Tsa 87]. À partir de plusieurs vues, on peut alors reconstruire exactement la scène observée. L'inconvénient de cette méthode est que l'étalonnage

*Le LIFIA est un laboratoire de l'IMAG, qui est une fédération de laboratoires de l'Institut National Polytechnique de Grenoble et de l'Université Joseph Fourier associés au CNRS

d'une caméra est un processus délicat et instable, qui est à recommencer chaque fois que la caméra est déplacée ou que la mise au point, la focale ou l'ouverture sont modifiées.

Dans cet article, on se place dans le premier cadre ci-dessus. On étudie divers modèles géométriques de caméra, en les comparant au modèle projectif. Puis, supposant ce modèle valide, on étudie successivement les cas où on dispose d'une, deux, trois ou plusieurs images. La contrainte que l'on exploite est l'existence d'une droite reliant tout point de la scène et sa projection dans une image, droite qui passe par un centre de projection qui ne dépend que de l'image considérée et non du point de la scène.

Les informations que l'on peut tirer des images permettent alors soit de reconstruire la scène observée, soit de caractériser ou même de reconnaître les objets qui s'y trouvent.

L'article suit ce plan. Une première section étudie les modèles de caméras, les cas d'une, deux, trois et plusieurs images sont ensuite successivement présentés.

2 Quelques modèles géométriques de caméras

Le but de ce paragraphe est d'étudier quelques modèles géométriques de caméra, principalement pour les comparer au modèle projectif. Cela permet de comprendre ce que ce dernier modèle permet de modéliser et certaines caractéristiques qui lui échappent. Les modèles présentés sont deux modèles optiques, dits des systèmes épais et des systèmes minces, et deux modèles géométriques, de projection perspective et de projection affine. Pour finir, nous montrons comment on peut se ramener au cas projectif lorsqu'il y a des distorsions optiques.

Lorsqu'on cherche le modèle géométrique d'une caméra, on cherche à trouver un modèle qui puisse rendre compte des propriétés géométriques de l'image: par exemple, le fait que des points alignés de la scène donnent des points alignés dans l'image. Ce modèles ne rendent donc pas compte du processus physique de formation de l'image (via l'optique et l'électronique) mais uniquement de la géométrie.

2.1 Approximation des systèmes épais

En optique linéaire, on utilise couramment deux modèles. Celui des systèmes minces, qu'on utilise pour modéliser les lentilles minces, et celui des systèmes épais qui permet de tenir compte de l'épaisseur de la lentille ou du système optique. Commençons donc par étudier ce dernier, qui est a priori le plus complet. Pour plus de détail, on pourra se rapporter à [Gro 93, Sua 74].

Un modèle épais peut être modélisé comme sur la FIG. 1. Le système est caractérisé par deux points principaux H et H' , deux foyers F et F' et l'interstice i . Un point objet B donne un point image B' . Le système d'acquisition de la caméra peut être représenté comme étant un plan P . Si le système optique est cylindrique d'axe Cz , le point B donne comme image sur le plan P un disque de centre B'' et de rayon R_B .

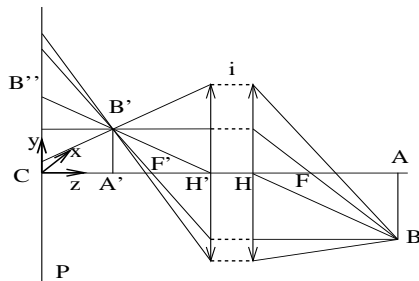


FIG. 1 - : Modèle d'un système épais.

En notant γ la distance CH' et f la distance focale HF , on peut calculer très simplement

R_B et les coordonnées de B' et B'' en fonction des coordonnées (x, y, z) de B . On obtient :

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto B' \begin{pmatrix} xf/(\gamma + f + i - z) \\ yf/(\gamma + f + i - z) \\ \gamma + (z - \gamma - i)f/(\gamma + f + i - z) \end{pmatrix}$$

$$B'' \begin{pmatrix} \gamma x/(\gamma + i - z) \\ \gamma y/(\gamma + i - z) \\ 1 \end{pmatrix} \propto \begin{pmatrix} \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & -1 & \gamma + i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (1)$$

$$R_B = R \left(1 + \frac{\gamma}{f} \frac{z - f - \gamma - i}{z - \gamma - i} \right) \quad (2)$$

On peut donc résumer ce modèle en une projection perspective (équation (1)) et un effet de flou (équation (2)). Avec un tel système, “mettre au point sur le point B ” consiste à choisir γ de telle façon que R_B soit nul. On peut aussi remarquer que l’existence d’un interstice non nul intervient dans les paramètres de la transformation et non dans sa nature.

2.2 Approximation des systèmes minces

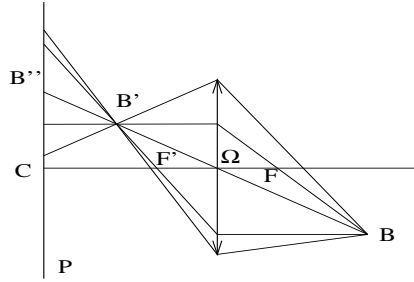


FIG. 2 - : Modèle d’un système mince.

Pour passer des systèmes épais aux systèmes minces, il suffit d’annuler l’interstice du système. On obtient alors le modèle représenté sur la FIG. 2 avec :

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto B' \begin{pmatrix} xf/(\gamma + f - z) \\ yf/(\gamma + f - z) \\ \gamma + f(z - \gamma)/(\gamma + f - z) \end{pmatrix}$$

$$B'' \begin{pmatrix} \gamma x/(\gamma - z) \\ \gamma y/(\gamma - z) \\ 1 \end{pmatrix} \propto \begin{pmatrix} \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & -1 & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3)$$

$$R_B = R \left(1 + \frac{\gamma}{f} \frac{z - f - \gamma}{z - \gamma} \right) \quad (4)$$

Du point de vue de la nature de la transformation, ce cas est donc identique au précédent.

2.3 Modèle de projection perspective

Ce modèle est une simplification du précédent. L’image du point B est l’intersection du plan P et de la droite $B\Omega$. Le modèle peut être représenté comme sur la FIG. 3, et on obtient :

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto B' \begin{pmatrix} xf/(\gamma - z) \\ yf/(\gamma - z) \\ 0 \end{pmatrix}$$

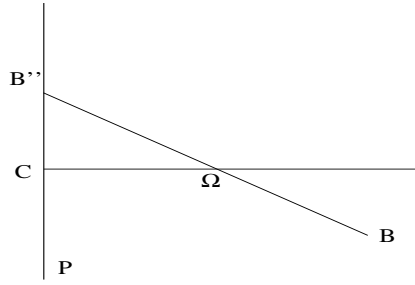


FIG. 3 - : Modèle d'une projection perspective.

$$B'' \begin{pmatrix} \gamma x / (\gamma - z) \\ \gamma y / (\gamma - z) \\ 1 \end{pmatrix} \propto \begin{pmatrix} \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & -1 & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (5)$$

$$R_B = 0 \quad (6)$$

La principale différence par rapport aux modèles précédents est l'absence de flou (cf. équation (6)), due au fait que ce modèle considère un système optique ponctuel, alors que les deux précédents tenaient compte de la surface de celui-ci. Ce système n'a donc pas besoin de mise au point, et on peut le considérer valide pour modéliser une caméra réelle si celle-ci est mise au point pour tous les points d'une scène.

2.4 Modèle de projection affine

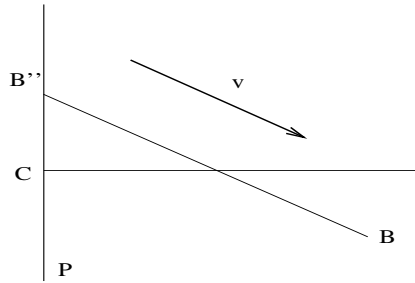


FIG. 4 - : Modèle d'une projection affine.

Dans ce dernier modèle, la projection ne se fait plus par une droite passant par un point fixe Ω , mais selon une direction caractérisée par un vecteur de composantes (v_x, v_y, v_z) . Cela correspond au cas où le point Ω est à l'infini, et revient donc à négliger la profondeur de la scène. On obtient le schéma de la FIG. 4 et les équations :

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto B' \begin{pmatrix} x - z v_x / v_z \\ y - z v_y / v_z \\ 0 \end{pmatrix}$$

$$B'' \begin{pmatrix} k(x - z v_x / v_z) \\ k(y - z v_y / v_z) \\ 1 \end{pmatrix} \propto \begin{pmatrix} 1 & 0 & -v_x / v_z & 0 \\ 0 & 1 & -v_y / v_z & 0 \\ 0 & 0 & 0 & 1/k \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (7)$$

$$R_B = 0 \quad (8)$$

2.5 Cas des non linéarités

De l'étude précédente, on peut retenir que les deux premiers modèles rendent compte de l'effet de flou, alors que le dernier ne tiens pas compte de la profondeur. En dehors de cela, ils sont équivalents au modèle projectif. En particulier, ce sont tous des modèles linéaires, qui projettent une droite sur une droite par exemple. Cela ne permet donc pas de rendre compte des distorsions optiques qui affectent la plupart des objectifs de caméra.

Face à ces distorsions, deux possibilités existent. Soit enrichir le modèle de paramètres supplémentaires qui décrivent ces distorsions, soit transformer l'image pour se ramener au modèle projectif. Dans le premier cas, on étalonne la caméra, en ajoutant des paramètres pour le décentrage et pour la distorsion radiale. On doit alors utiliser une technique de minimisation non-linéaire et cette technique nécessite beaucoup de savoir faire. Cette méthode ne s'applique pas à notre cas où nous cherchons un modèle de la caméra et non ces paramètres.

La deuxième méthode, présentée dans [Peu 92, Bra 94], consiste à effectuer une transformation de l'image, telle que le résultat de cette transformation soit une projection perspective de la scène observée. La principale propriété d'une telle transformation est de projeter une droite en une droite. On prend donc une image d'une grille constituée de droites orthogonales. On calcule alors la correction qu'il faut apporter à la position de chaque intersection de deux images de droites pour que ces images soient elles mêmes droites. Les corrections des points entre les points d'intersection sont calculées par interpolation.

Lorsqu'on prend une nouvelle image, il faut appliquer la correction correspondante à chaque pixel. Une telle méthode permet de corriger toutes les distorsions, et pas seulement la radiale, et elle permet d'arriver à des précisions de localisation de l'ordre d'un cinquantième de pixel. On se ramène ainsi à une image pour laquelle le modèle de projection perspective est valide.

3 Que faire avec une image ?

On suppose, suite à l'étude du paragraphe précédent, qu'on dispose d'une image prise par une caméra, que l'on peut modéliser par une projection perspective. Quelles sont alors les informations numériques relatives à la scène que l'on peut calculer à partir de cette image ?

Il est bien connu que les longueurs ne sont pas significatives. On ne peut pas déterminer la taille de quelqu'un si on ne dispose que d'une photo de cette personne. D'une manière générale, les grandeurs euclidiennes (longueurs, angles, surfaces) et affines (rapports de longueurs ou surfaces) ne se conservent pas. De plus les grandeurs projectives (birapports de longueurs) peuvent ne pas être significatives, car quatre points alignés dans une image ne correspondent pas forcément à quatre points alignés de la scène.

Un célèbre théorème [Bur 90, Cle 90, Mos 92] montre qu'il n'est pas possible de trouver de grandeur projective caractéristique d'une scène 3D à partir d'une seule image. Ce théorème, en même temps qu'il énonce une impossibilité, indique comment il faut tourner cette dernière. Il faut soit savoir que la scène est plane, au moins localement, et c'est le cas qui va être étudié ici, soit disposer de plusieurs images, c'est ce qui va être étudié dans les paragraphes suivants.

3.1 Utilisation du birapport

On se trouve dans le cas géométriquement classique où un plan est projeté projectivement sur un autre. Le lecteur non familier avec la géométrie projective pourra se rapporter à [Com 41] pour une introduction, ou à [Sem 52] pour une présentation plus complète.

Une projection perspective entre deux plans est équivalente à une homographie dont les deux propriétés fondamentales sont de transformer les droites en droites et de conserver les birapports. Des birapports peuvent être définis à partir de quatre points alignés :

$$\{A, B, C, D\} = \frac{\overline{AC}}{\overline{BC}} / \frac{\overline{AD}}{\overline{BD}}$$

mais aussi à partir de quatre droites, de quatre points sur une conique. . . La difficulté d'emploi de ces birapports vient du fait que ce ne sont point des grandeurs euclidiennes. Par exemple, les trois configurations de points de la FIG. 5 sont très proches d'un point de vue géométrique, d'autant plus si la variation de position de B et C est due à du bruit, mais leurs birapports k sont très différents (10^3 et -10^3). Pour comparer des birapports, la différence ou le rapport ne sont donc pas adaptés.

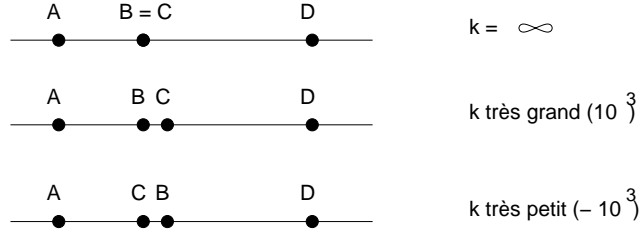


FIG. 5 - : Trois configurations géométriquement proches de birapport très différents.

La méthode à utiliser, développée dans [Mor 93, May 93], consiste à étudier la distribution du birapport pour des configurations aléatoires de points (FIG. 6). Pour comparer deux birapports k_1 et k_2 , on mesure alors la probabilité pour que le birapport d'une configuration aléatoire de points se trouve entre k_1 et k_2 . Il faut aussi tenir compte du fait qu'il y a deux "chemins" entre ces deux valeurs, l'ordinaire, et celui passant par l'infini.

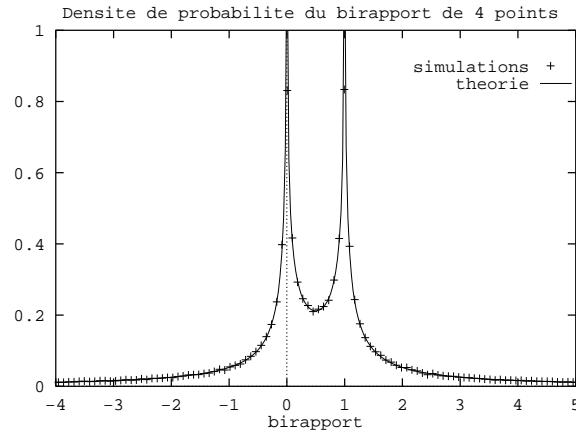


FIG. 6 - : Fonction de distribution du birapport d'une configuration de quatre points alignés aléatoires.

Cette méthode permet ainsi de comparer la ressemblance projective de configurations de points. On trouvera dans les articles cités précédemment, des études complémentaires concernant, par exemple, la stabilité du birapport lorsque les configurations de points sont bruitées.

3.2 Application à l'indexation et à la reconnaissance

Le birapport, outre son intérêt mathématique, peut servir à résoudre des problèmes de vision. Il permet en effet de caractériser une configuration de points ou droites, quelle que soit la transformation projective qu'on fasse subir à cette configuration. Ainsi, une configuration de points d'une scène peut être reconnue dans une image à partir de son birapport quel que soit le point de vue de l'image.

Un exemple est montré sur la FIG. 7. Les deux images représentent deux ensembles de points déduits l'un de l'autre par une application projective. Pour retrouver la correspondance entre les points, on peut, par exemple, calculer les birapports associés à certaines configurations de cinq points et les comparer entre les deux images. Lorsque ces birapports sont égaux, on peut

conclure qu'il existe une transformation projective qui amène une des configurations sur l'autre. C'est ce qui arrive pour les deux configurations de points marqués par des carrés dans la figure.

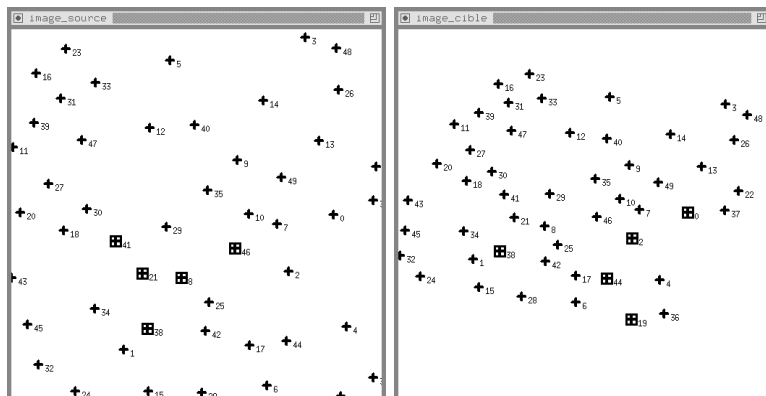


FIG. 7 - : Deux configurations de cinq points définissant les mêmes birapports.

Quelques difficultés. Ce schéma, malgré sa simplicité, est sujet à quelques difficultés de mise en œuvre.

Combinatoire: il y a vingt quatre façons d'ordonner quatre points alignés, mais cela ne donne que six valeurs différentes pour la birapport. Avec cinq points dans le plan, on arrive à cent vingt ordres différents et trente valeurs différentes de birapports. Mais seules deux de ces valeurs sont indépendantes. Comment les choisir ?

Bruit : Lorsque les points ne sont distants que de quelques pixels les uns des autres, une erreur de discrétisation, qui entraîne une erreur d'un pixel sur la localisation d'un point, a un effet catastrophique sur la valeur des birapports.

Ambiguïtés : Outre la transformation qui fait passer globalement d'une image à l'autre, il se peut que d'autres transformations puissent faire transformer des configurations l'une dans l'autre. Le birapport est un invariant pour toutes ces transformations et non pas seulement pour la première, et il ne permet donc pas de distinguer ces deux cas.

Au total, la méthode de calcul employée a une très grande importance. Si on utilise les polynôme symétriques pour extraire deux valeurs de trente birapports de cinq points coplanaires, on nivelle les valeurs et le taux final de discrimination est très faible. Suivant les précautions prises, on arrive à un taux de discrimination de 1/30 à 1/5000, c'est à dire qu'on arrive à reconnaître, à l'aide des seuls birapports, une configuration parmi 30 en moyenne, ou une parmi 5000 avec les techniques les plus stables, ce qui est un résultat remarquable dans le cas de données bruitées.

3.3 D'autres configurations

La méthode présentée ici est applicable à de nombreuses configurations planes, à partir desquelles on peut définir un birapport : quatre points alignés, quatre droites concourantes, cinq points quelconques, cinq droites quelconques, quatre points sur une conique, quatre tangentes à une conique, une conique et deux points, une conique et deux droites, une conique avec un point et une droite, deux coniques, une courbe avec une concavité...

4 Que faire avec deux images ?

Avec une seule image, il n'est pas possible d'obtenir d'information numérique sur une scène tridimensionnelle. Supposons maintenant, que nous disposons de deux images, prises par projection perspective par la même caméra, ou même par deux caméras différentes. On suppose que la mise en correspondance est faite, c'est à dire que l'on sait à quel élément d'une image correspond tout élément de l'autre. Deux éléments se correspondent s'ils sont les projections dans chaque image d'un même élément de la scène observée.

Deux points qui se correspondent ne peuvent localisés n'importe comment. Ils doivent respecter une contrainte, dite contrainte épipolaire. C'est cette contrainte qui va permettre de tirer des informations de la paire d'images.

Pour plus de détails concernant le contenu de ce paragraphe, voir [Gro 93].

4.1 Présentation de la géométrie épipolaire

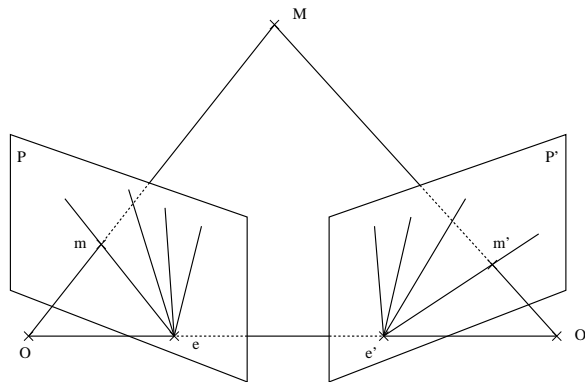


FIG. 8 - : Définition de la géométrie épipolaire

Considérons la FIG. 8. Un point M est projeté sur deux images P et P' , par deux caméras dont les centres de projection sont respectivement O et O' . Le plan (MOO') coupe P et P' en deux droites, dites droites épipolaires, auxquelles doivent appartenir m et m' les projections de M . Lorsque M varie, le plan (MOO') varie, mais les droites d'intersection avec P et P' passent par deux points fixes, e et e' , appelés épipoles des images.

Lorsqu'on connaît un point m d'une image, il définit une droite épipolaire me . Son correspondant ne peut alors se trouver que sur la droite épipolaire correspondante. Cette contrainte peut s'exprimer par le fait qu'il existe une matrice F de dimension 3×3 , de rang 2, appelée matrice fondamentale, telle que pour tout couple de points se correspondant, on a : ${}^t m F m' = 0$.

Cette contrainte permet, par exemple, d'apparier point à point deux droites appariées globalement.

4.2 Un test de coplanarité

Une des applications immédiates de la géométrie épipolaire est la réalisation d'un test de coplanarité. On dispose des projections dans deux images de quatre points. On suppose qu'on connaît la géométrie épipolaire (c'est à dire les épipoles et la correspondance entre les droites épipolaires). On désire savoir si les quatre points de la scène dont on connaît les projections sont coplanaires.

Pour cela, il suffit de calculer les birapports associés, dans chaque image, à la configuration de cinq points formée des quatre projections et de l'épipole. Ces birapports sont les mêmes dans les deux images si et seulement si les points correspondants de la scène observée sont coplanaires.



FIG. 9 - : Deux images d'une maison en bois.

Voici les résultats de ce test sur des images réelles. On part des deux images de la FIG. 9. Divers points de ces images sont numérotés comme indiqué sur la FIG. 10. On cherche alors les points qui sont dans le plan défini par les points 5, 8 et 14. Pour cela, on considère successivement les autres points, on calcule les birapports associés et on les compare, comme indiqué précédemment. Les résultats sont donnés pour ce plan et un autre, dans le tableau de la FIG. 11.

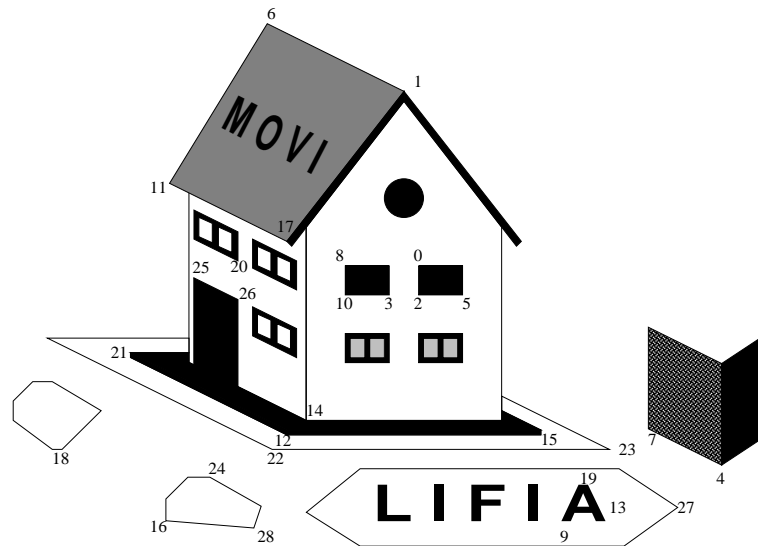


FIG. 10 - : Numérotation des points.

Les points qui sont effectivement coplanaires donnent des différences entre birapports plus petites. A cause du bruit présent dans les images, ces différences ne sont toutefois pas nulles. D'autre part, il n'y a pas de seuil naturel séparant les points coplanaires des points presque coplanaires (comme les points 7 ou 15), et ces derniers des points non coplanaires. Il y a là une réelle difficulté pratique.

4.3 Calcul d'invariants 3D

Reconstruction de points coplanaires. Le test présenté au paragraphe précédent peut être inversé. Étant donné trois points et une droite, on peut calculer le point de la droite coplanaire avec les trois points donnés. Il suffit pour cela de chercher quel est le point de la droite qui vérifie le test de coplanarité.

Mathématiquement, cela s'exprime très simplement. Soient a_1 , b_1 , et c_1 les trois projections des points dans la première image, et a_2 , b_2 et c_2 les projections dans la deuxième image. On

Points 5, 8 et 14		Points 21, 23 et 28	
4ème point	Résultat	4ème point	Résultat
0	0.00123847	18	0.00114258
3	0.00142479	12	0.00336455
17	0.00144425	24	0.00361946
1	0.00161304	16	0.0062559
15	0.00307953	19	0.00675905
18	0.00347562	9	0.00860071
2	0.00444478	14	0.00919245
7	0.00576232	22	0.0108486
23	0.0076242	4	0.0158454
4	0.00886061	7	0.0282956
13	0.0133648	5	0.038735
19	0.0157966	2	0.0424568
27	0.0165858	3	0.0453531
9	0.0214658	0	0.0480963
25	0.0353054	25	0.0503932
26	0.0383085	26	0.0510969
21	0.044384	10	0.0540821
11	0.0544166	20	0.0553584
6	0.0563005	6	0.058392
		8	0.0588166
		1	0.0643018
		11	0.0665695
		17	0.0715031

FIG. 11 - : Résultats obtenus pour les configurations contenant les points numérotés 5, 8 et 14, et celles contenant les points 21, 23 et 28.

note e_1 et e_2 les deux épipoles et l_1 et l_2 les projection de la droite. Alors, si T est l'homographie plane qui transforme les points a_1, b_1, c_1, e_1 en a_2, b_2, c_2, e_2 , on a $d_2 = l_2 \wedge ({}^tT^{-1}.l_1)$, où d_2 est la projection dans la deuxième image du point de la droite coplanaire avec les trois autres.

Calcul d'invariants 3D. On peut alors appliquer cette technique de "reconstruction de points coplanaires" au calcul d'invariants 3D. En effet, de la même façon que dans le plan, on peut calculer des birapports à partir de configurations de points, droites et coniques dans l'espace. Pour cela, on réalise un certain nombre de constructions. Ainsi, à partir de six points A, B, C, D, E et F , on construit les points F', E' et D' conformément à la FIG. 12. On obtient alors six points dans le plan auxquels on peut associer des birapports.

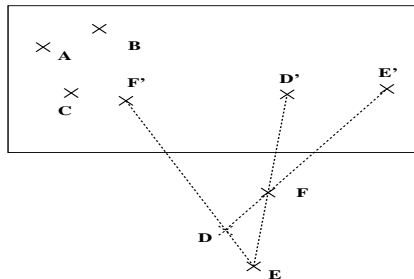


FIG. 12 - : Calcul de birapports associés à six points dans l'espace.

On voudrait pouvoir faire ces constructions à partir des seules projections des points dans

deux images. Ces constructions étant principalement constituées de calculs d'intersections entre plans et droites, elles sont réalisables par la méthode de "reconstruction des points coplanaires" présentée plus haut.

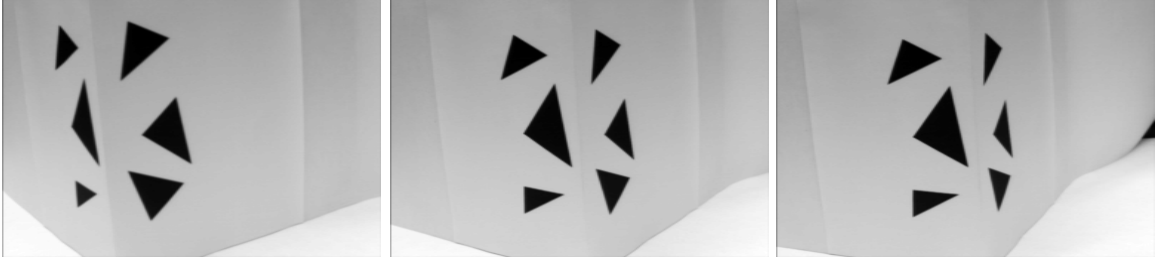


FIG. 13 - : Trois images d'un dièdre.

Voici un exemple d'application sur des données réelles. On part des trois images d'un dièdre montrées à la FIG. 13. A partir des deux premières, on calcule les birapports associées à 7 configurations de 4 droites. Ces mêmes birapports sont calculés à partir des configurations correspondantes dans les deux dernières images. On compare alors tous ces birapports dans le tableau de la FIG. 14. Les lignes correspondent aux calculs à partir des 2 premières images, et les colonnes à ceux des 2 dernières images. Les valeurs sont les distances entre birapports.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1	0.00077	0.03524	0.04605	0.01199	0.08238	0.09704	0.00793
S_2	0.03428	0.00020	0.08110	0.04704	0.11742	0.13208	0.02712
S_3	0.03767	0.07214	0.00915	0.02491	0.04547	0.06013	0.04483
S_4	0.01852	0.05299	0.02831	0.00576	0.06463	0.07929	0.02568
S_5	0.07972	0.11419	0.03290	0.06696	0.00343	0.01808	0.08688
S_6	0.10603	0.14051	0.05921	0.09327	0.02289	0.00823	0.11320
S_7	0.00794	0.02654	0.05476	0.02070	0.09108	0.10557	0.00078

FIG. 14 - : Distance entre les invariants calculés dans les deux paires d'images.

Les birapports des configurations en correspondance donnent les différences les plus petites (sur la diagonale du tableau), mais on retrouve les mêmes limites que précédemment : pas de seuil naturel, bruit, ambiguïtés (qu'en est-il de S_1 et S_7 ?). Ces limites montrent la nécessité d'emploi de techniques robustes pour la détection des primitives dans les images.

5 Que faire avec trois images ?

On se place dans un cadre similaire à celui du paragraphe précédent, mais avec trois images maintenant. Ces images peuvent ne pas avoir été prises avec la même caméra. Ce qui a été décrit au paragraphe précédent, en utilisant la géométrie épipolaire, s'applique donc ici aussi. Ce qui est nouveau avec trois images, c'est que la connaissance de la géométrie épipolaire, dont le calcul est délicat ou parfois impossible [Luo 92], n'est plus un préalable au calcul de birapports dans les images.

5.1 Calcul d'informations 3D

Le schéma de calcul suivant montre comment on obtient une contrainte sans utiliser la géométrie épipolaire. Pour plus de détails, se reporter à [Qua 94]. On considère six points dans l'espace. On peut donner à cinq d'entre eux les coordonnées homogènes canoniques d'une base

projective : $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$ et $(1, 1, 1, 1)$. N'ayant qu'une information projective dans les images, ce choix de coordonnées est possible dès lors que quatre de ces points ne sont pas coplanaires.

Par contre le sixième point a des coordonnées qu'on ne peut fixer arbitrairement, et qu'on cherche donc à déterminer. Soient (X, Y, Z, T) ces coordonnées. Ces six points sont projetés dans une des images. quatre de ces projections peuvent être prises comme base projective du plan de projection, $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ et $(1, 1, 1)$, les deux autres ayant des coordonnées connues (mesurées dans l'image) (u_5, v_5, w_5) et (u_6, v_6, w_6) .

Une projection perspective générale a onze paramètres, qui sont inconnus dans notre cas. Les six points nous fournissent chacun deux mesures (leur abscisse et leur ordonnée), soit douze relations en tout. On peut donc éliminer les onze paramètres de la projection d'une des relations. On obtient alors :

$$\begin{aligned} w_6(u_5 - v_5)XY + v_6(w_5 - u_5)XZ + u_5(v_6 - w_6)XT &+ \\ u_6(v_5 - w_5)YZ + v_5(w_6 - u_6)YT + w_5(u_6 - v_6)ZT &= 0 \end{aligned}$$

Comme il y a trois images, on obtient ainsi trois relations, dans lesquelles il y a trois inconnues, les rapports X/T , Y/T et Z/T . On peut résoudre et obtenir ainsi la position dans l'espace du sixième point par rapport aux cinq premiers. Accessoirement, on peut alors résoudre aussi les onze relations restantes pour chaque image, qui nous donnent les paramètres de la projection.

5.2 Application

Le calcul précédent qui permet de "reconstruire projectivement" des points (voir le paragraphe suivant), permet alors de calculer toute grandeur projective sur la scène reconstruite. On peut aussi visualiser simplement le résultat de cette reconstruction. La FIG. 15 montre une image d'une maison en bois. A partir de trois images de cette maison, on peut la reconstruire, c'est à dire retrouver sa géométrie tridimensionnelle.

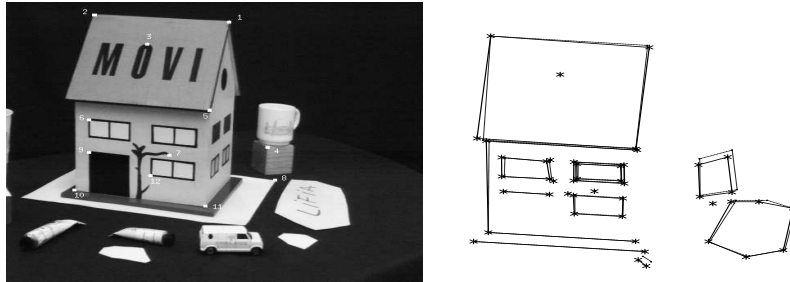


FIG. 15 - : Reconstruction projective à partir de trois images.

On a appliqué la méthode précédente, en choisissant les points 2, 5, 8, 10 et 11 comme base projective de l'espace, et en positionnant les autres points par rapport à ceux-là. Une projection du résultat est montrée dans la partie gauche de la FIG. 15, et est comparé au résultat obtenu avec la méthode du paragraphe suivant.

Cette méthode qui ne nécessite pas le calcul de la géométrie épipolaire s'avère plus stable que les calculs sur deux images.

6 Que faire avec plusieurs images ?

Pour terminer, on peut poser le problème général. On dispose de n images, dans lesquelles on observe p points. Est-il alors possible de calculer la position de ces points dans l'espace, autrement dit de reconstruire la scène ?

Il est clair que sans autre information, il n'est pas possible de retrouver la position exacte des points sans connaître la localisation des caméras. On ne pourra reconstruire qu'à un déplacement près. De même, la taille de la scène ne peut être retrouvée (elle peut être plus grande et observée de plus loin), et il faut aussi reconstruire à une homothétie près. D'autres transformations sont possibles, et cela exclut toute unicité de la solution.

6.1 Reconstruction projective

Reprenons l'équation (5) de la projection perspective. Dans le cas d'une caméra quelconque, elle s'écrit :

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

On peut l'abrégé en : $\lambda m = P.M$, où M est un point de l'espace, P une matrice de projection, m la projection de m dans l'image, et λ un coefficient multiplicatif, dû au fait que les coordonnées projectives sont définies à un coefficient multiplicatif près. Pour mieux faire ressortir les mesures, on peut aussi l'écrire :

$$\begin{aligned} x &= \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\ y &= \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \end{aligned}$$

Dans ces équations, x et y sont les mesures dans l'image, et les coefficients de P et de M sont les inconnues. On dispose de deux telles équations pour chaque point d'une image. Il y en a donc $2 \times n \times p$. Chaque point de l'espace donne trois inconnues (sa position dans l'espace), et chaque caméra a onze paramètres inconnus.

Il faut remarquer que si on trouve une solution P, M à l'équation $\lambda m = P.M$, alors quelle que soit la matrice A de dimension 4×4 inversible, on a $\lambda m = (PA).(A^{-1}M)$ et PA et $A^{-1}M$ sont aussi des solutions. L'espace des solutions est de dimension 15.

Cette matrice correspond à une transformation projective de l'espace quelconque. On ne peut donc trouver la solution qu'à une transformation projective près. Cela avait déjà été signalé en disant qu'on ne pouvait espérer mieux qu'une solution à un déplacement et une homothétie près. En fait, c'est donc à une transformation projective près qu'on travaille. Cela est dû au fait qu'on ne dispose que d'une information projective.

Pour fixer cette matrice A et avoir une solution unique, on peut fixer cinq points de l'espace, en leur donnant les coordonnées homogènes canoniques de la base projective. Cela supprime quinze inconnues. Il reste donc $3p - 15 + 11n$ inconnues. Dès que $2np > 3p - 15 + 11n$, il est possible de résoudre. C'est le cas pour deux images et sept points, trois images et six points... On peut remarquer que l'équation exclut le cas d'une seule image, et que six points sont toujours nécessaires, quel que soit le nombre d'images.

Pour la résolution du système obtenu, on peut utiliser une méthode d'optimisation non linéaire (Levenberg-Marquart) [Moh 93].

Un exemple. La FIG. 16 montre un exemple d'application de cet algorithme. On part d'une séquence de neuf images de la maison, on choisit cinq des points comme base (en leur donnant des coordonnées euclidiennes telles que la reconstruction soit presque euclidienne), ces points sont marqués par des croix, et on résout. Les points sont entourés de leur ellipsoïde d'incertitude.

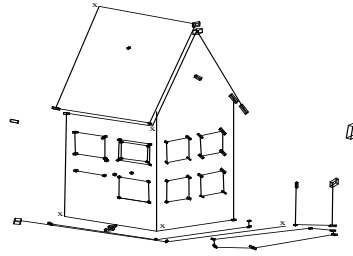


FIG. 16 - : Reconstruction projective de la maison à partir de neuf images.

6.2 Reconstruction euclidienne

Lorsqu'on ne connaît pas les coordonnées euclidiennes des points de référence, il est toujours possible de donner des coordonnées arbitraires. L'algorithme marche, mais la reconstruction est difficile à apprécier à l'œil, car elle est purement projective : les parallèles ne sont plus parallèles, les rapports ne sont pas respectés. . . Seuls les alignements et les birapports sont respectés.

On peut alors remplacer le fait de fixer les coordonnées de cinq points par des contraintes de type affine ou euclidien (parallélisme, verticalité, horizontalité, distances. . .) que l'on ajoute au moment de l'optimisation non linéaire [Bou 93].

La FIG. 17 montre les résultats sur la même séquence d'image de la maison.

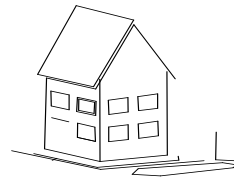


FIG. 17 - : Reconstruction euclidienne de la maison à partir de neuf images.

7 Conclusion

Lorsque la transformation opérée par une caméra est une projection perspective, il est possible de calculer des birapports caractéristiques de la scène observée, de faire de la reconnaissance ou de reconstruire projectivement la scène, cela sans connaître les paramètres de la caméra.

Les avantages des méthodes qui ont été exposées ici sont la simplicité, la généralité et l'absence de calibration des caméras. Par contre, ces méthodes nécessitent une grande précision dans la détection des points, droites et autres primitives de l'image. Les méthodes actuelles qui permettent des précisions allant du dixième au cinquantième de pixel ne sont pour cela pas superflues.

La géométrie projective apparaît ainsi comme un cadre naturel très puissant pour interpréter le contenu des images et pouvoir en tirer des informations numériques précises.

Références

- [Bou 93] B. Boufama, R. Mohr, et F. Veillon. Euclidian constraints for uncalibrated reconstruction. *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 466–470, Mai 1993.
- [Bra 94] P. Brand, R. Mohr, et Ph. Bobet. Distorsion optique : correction dans un modèle projectif. *Actes du 9ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Paris, France*, pages 87–98, Paris, Janvier 1994.
- [Bur 90] J.B. Burns, R. Weiss, et E.M. Riseman. View variation of point set and line segment features. *Proceedings of DARPA Image Understanding Workshop, Pittsburgh, Pennsylvania, USA*, pages 650–659, 1990.
- [Cle 90] D.J. Clemens et D.W. Jacobs. Model-group indexing for recognition. *Proceedings of DARPA Image Understanding Workshop, Pittsburgh, Pennsylvania, USA*, pages 604–613, Septembre 1990.
- [Com 41] H. Commissaire et G. Cagnac. *Cours de Mathématiques Spéciales*. Masson, 1941.
- [Gro 93] P. Gros. *Outils Géométriques pour le modélisation et la reconnaissance d'objets polyédriques*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, Juillet 1993.
- [LH 81] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, volume 293, pages 133–135. XX, Septembre 1981.
- [Luo 92] Q.T. Luong. *Matrice Fondamentale et Autocalibration en Vision par Ordinateur*. Thèse de doctorat, Université de Paris-Sud, Orsay, France, Décembre 1992.
- [May 93] S.J. Maybank. Classification based on the Cross Ratio. *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 113–132, Octobre 1993.
- [Moh 91] R. Mohr et L. Morin. Relative positioning from geometric invariants. *Proceedings of the Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, USA*, pages 139–144, Juin 1991.
- [Moh 93] R. Mohr, F. Veillon, et L. Quan. Relative 3D reconstruction using multiple uncalibrated images. *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 543–548, Juin 1993.
- [Mor 93] L. Morin. *Quelques Contributions des Invariants Projectifs à la Vision par Ordinateur*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, Janvier 1993.
- [Mos 92] Y. Moses et S. Ullman. Limitations of non model-based recognition. *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 820–828, Mai 1992.
- [Mun 92] J.L. Mundy et A. Zisserman, éditeurs. *Geometric Invariance in Computer Vision*. MIT Press, Cambridge, Massachusetts, USA, 1992.
- [Mun 93] J.L. Mundy et A. Zisserman, éditeurs. *Proceedings of the second ESPRIT - ARPA workshop on Applications of Invariance on Computer Vision, Ponta Delagada, Azores, Portugal*, Octobre 1993.
- [Peu 92] B. Peuchot. Camera virtual equivalent model - 0.01 pixel detectors. *14th Annual International Conference IEEE EMBS. Satellite Symposium on 3D Advanced Image Processing in Medecine, Rennes.*, pages 41–45, Novembre 1992.

- [Qua 94] L. Quan. Invariants of 6 points from 3 uncalibrated images. J.O. Eklundh, éditeur, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, volume II, pages 459–470. Springer-Verlag, 1994.
- [Sem 52] J.G. Semple et G.T. Kneebone. *Algebraic Projective Geometry*. Oxford Science Publication, 1952.
- [Sua 74] R. Suardet. *Optique*. Collection de Sciences Physiques. Editions J.-B. Baillière, Paris, France, 1974.
- [Tsa 87] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.