



**HAL**  
open science

# Outils géométriques pour la modélisation et la reconnaissance d'objets polyédriques

Patrick Gros

► **To cite this version:**

Patrick Gros. Outils géométriques pour la modélisation et la reconnaissance d'objets polyédriques. Interface homme-machine [cs.HC]. Institut National Polytechnique de Grenoble - INPG, 1993. Français. NNT: . tel-00005127

**HAL Id: tel-00005127**

**<https://theses.hal.science/tel-00005127>**

Submitted on 26 Feb 2004

**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.

**THESE**

*présentée par*

**Patrick GROS**

*pour obtenir le grade de DOCTEUR*

de l'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

(Arrêté ministériel du 30 Mars 1992)

Spécialité : **Informatique**

---

**OUTILS GEOMETRIQUES POUR  
LA MODELISATION ET LA RECONNAISSANCE  
D'OBJETS POLYEDRIQUES**

---

Date de soutenance : 2 Juillet 1993

Composition du jury :

Président : **Claude PUECH**  
Rapporteurs : **Olivier FAUGERAS**  
**Andrew ZISSERMAN**  
Examineurs : **Radu HORAUD**  
**Roger MOHR**

Thèse préparée au sein du laboratoire LIFIA - IMAG - INRIA Rhône - Alpes  
sous la direction de Roger MOHR



# Remerciements

La présente thèse a été effectuée au sein du projet MOVI du LIFIA, Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle à Grenoble. Elle a donc profité d'un environnement scientifique et humain particulièrement riche et du concours de nombreuses personnes que je tiens à remercier ici.

- Je tiens tout d'abord à remercier les membres du jury, Monsieur Puech, professeur à l'Université Joseph Fourier de Grenoble, Monsieur Faugeras, directeur de recherche à l'INRIA Sophia-Antipolis et Monsieur Zisserman, research fellow à l'Université d'Oxford, pour avoir accepté de juger ce travail.
- La région Rhône-Alpes, en lien avec la société ITMI et Monsieur Skordas, a soutenu financièrement mon travail pendant trois ans. Qu'ils trouvent ici la marque de ma gratitude.
- Merci aux personnes du laboratoire, pour la bonne ambiance qui a permis un travail agréable pendant ces trois années. Merci plus particulièrement à l'équipe administrative et à Danièle Herzog d'une part, à l'équipe technique formée de Françoise Nayrolles et d'Henri Saya d'autre part, pour avoir fourni avec beaucoup de bonne volonté le soutien logistique nécessaire à cette thèse.
- Un grand merci à toutes les personnes du projet MOVI pour leurs innombrables contributions : discussions fructueuses, relecture de ma prose, travail en commun, moments de détente, soutien lorsque le moral vacillait..., et en particulier à Emmanuel Arbogast et Luce Morin, collègues thésards, Radu Horaud, Long Quan et Françoise Veillon, les permanents.
- Merci à aux stagiaires qui travaillèrent ou travaillent encore avec moi, Gudrun Socher, Marie Hélène Malissen, Marie Legendre, Edmond Boyer, Sebastien Gelgon et Olivier Bournez, qui, outre leur force de travail, ont apporté nombre d'idées pertinentes pour faire avancer mieux et plus vite le travail présenté dans cette thèse.

- Un merci tout particulier à Roger Mohr, directeur de thèse, ami et maintenant collègue, qui sut donner à cette thèse les bonnes orientations (en direction des invariants par exemple), tout en me laissant le plus libre possible dans mon travail.
- Merci enfin à tous les autres, rencontrés au laboratoire, dans le cadre des projets ORASIS, FIRST, SECOND ou VIVA, ou appartenant au cercle familial, pour leurs contributions dont la diversité n'a eu d'égale que la qualité.

# Sommaire

<b>Introduction</b>	<b>13</b>
<b>1 Géométrie, invariants et vision : théorie</b>	<b>21</b>
1.1 Géométrie projective . . . . .	22
1.1.1 Considérations historiques . . . . .	23
1.1.2 Représentation, espace projectif, coordonnées homogènes . . . . .	25
1.1.3 Dépendance linéaire, sous-espaces, dualité . . . . .	29
1.1.4 Birapport, éléments conjugués . . . . .	31
1.1.5 Transformations projectives . . . . .	35
1.1.6 Coniques . . . . .	37
1.1.7 Géométrie affine . . . . .	41
1.1.8 Géométrie euclidienne . . . . .	42
1.1.9 Autres géométries . . . . .	44
1.2 Théorie des invariants . . . . .	49
1.2.1 Introduction aux invariants . . . . .	50
1.2.2 Méthode infinitésimale de calcul des invariants. . . . .	53
1.2.3 Dénombrement des invariants . . . . .	58
1.2.4 Quelques exemples . . . . .	60
1.2.5 Cas où les transformations ne forment pas un groupe . . . . .	61
1.2.6 Conclusion sur la méthode infinitésimale . . . . .	63
1.2.7 Méthode par généralisation et contrainte . . . . .	63
1.2.8 Invariants de deux points et deux droites . . . . .	65
1.2.9 Invariants de deux coniques . . . . .	67
1.3 Éléments de vision . . . . .	73
1.3.1 L'approximation des systèmes minces . . . . .	73
1.3.2 L'approximation des systèmes épais . . . . .	75
1.3.3 Autres modèles . . . . .	77
1.3.4 De la caméra aux données . . . . .	78

1.4	Conclusion du chapitre . . . . .	81
<b>2</b>	<b>Géométrie, invariants et vision : applications</b>	<b>83</b>
2.1	Géométrie épipolaire et coplanarité . . . . .	84
2.1.1	Introduction à la géométrie épipolaire . . . . .	85
2.1.2	Un test de coplanarité . . . . .	87
2.1.3	Expérimentation . . . . .	93
2.1.4	Résultats obtenus avec des données simulées . . . . .	97
2.1.5	Résultats obtenus avec des images réelles . . . . .	108
2.1.6	Conclusion . . . . .	109
2.2	Calcul d'invariants 3D à partir de plusieurs images . . . . .	113
2.2.1	Etude de différentes configurations . . . . .	114
2.2.2	Expérimentations à partir de données simulées . . . . .	120
2.2.3	Expérimentations à partir de données réelles . . . . .	130
2.2.4	Conclusion . . . . .	132
2.3	Appariement d'images . . . . .	133
2.3.1	Introduction . . . . .	133
2.3.2	Algorithme proposé . . . . .	134
2.3.3	A propos du choix des similitudes . . . . .	137
2.3.4	Expérimentations . . . . .	141
2.3.5	Conclusion . . . . .	145
2.4	Conclusion du chapitre . . . . .	145
<b>3</b>	<b>Modélisation automatique</b>	<b>149</b>
3.1	Introduction . . . . .	149
3.1.1	Etat de l'art . . . . .	150
3.1.2	Méthode proposée . . . . .	152
3.1.3	Plan du chapitre . . . . .	154
3.2	Appariement et correction d'images . . . . .	155
3.2.1	Analyse de quelques sources de bruit . . . . .	156
3.2.2	Correction des images . . . . .	157
3.3	Distance entre images et regroupement . . . . .	159
3.3.1	Evaluation de la ressemblance des images . . . . .	159
3.3.2	Regroupement des images représentant un même aspect . . . . .	162
3.4	Les étapes ultérieures . . . . .	167
3.4.1	La modélisation des aspects . . . . .	167
3.4.2	L'indexation des modèles . . . . .	168
3.4.3	Reconnaissance des images . . . . .	169
3.5	Conclusion du chapitre . . . . .	170
	<b>Conclusion</b>	<b>171</b>
<b>A</b>	<b>Construction de points coplanaires</b>	<b>175</b>
A.1	Première méthode . . . . .	175
A.1.1	Principe de la méthode . . . . .	175
A.1.2	Calculs . . . . .	176
A.2	Deuxième méthode . . . . .	179

---

A.2.1	Principe de la méthode . . . . .	179
A.2.2	Calcul vectoriel . . . . .	181
A.2.3	Calcul analytique . . . . .	182
A.3	Conclusion sur ces méthodes . . . . .	184
<b>Bibliographie de l'auteur</b>		<b>185</b>
<b>Bibliographie</b>		<b>187</b>
<b>Index</b>		<b>199</b>



# Liste des Figures

0.1	Les niveaux de traitement en vision. . . . .	15
0.2	Schéma d'une méthode de modélisation reconnaissance. . . . .	16
0.3	Un ensemble d'images d'un même objet. . . . .	18
1.1	Le parallélisme pour la géométrie affine. . . . .	26
1.2	Les trois espèces de coniques pour la géométrie affine. . . . .	27
1.3	Définition de la droite projective. . . . .	27
1.4	Définition du birapport de 4 droites. . . . .	33
1.5	Définition du birapport de 4 plans. . . . .	34
1.6	Birapports définis par 5 points coplanaires. . . . .	34
1.7	Birapport défini par 5 droites coplanaires. . . . .	35
1.8	Définition de birapports à partir d'une conique. . . . .	38
1.9	Construction de la polaire d'un point. . . . .	39
1.10	Construction d'un triangle auto-polaire. . . . .	40
1.11	Définition de l'angle en géométrie euclidienne 2D. . . . .	43
1.12	Définition de l'angle en géométrie euclidienne 3D. . . . .	45
1.13	Définition de la distance et de l'angle en géométrie non euclidienne plane. . . . .	46
1.14	Par le point $P$ , il passe deux parallèles à la droite $D$ . . . . .	47
1.15	Cas des points du plan et des rotations de centre l'origine. . . . .	51
1.16	Le problème des invariants. . . . .	52
1.17	Cas des points du plan et des translations. . . . .	52
1.18	Cas des couples de points et des isométries. . . . .	53
1.19	Une orbite dans le cas des groupes et dans le cas général . . . . .	63
1.20	Une configuration de 2 points et de 2 droites . . . . .	67
1.21	Interprétation des matrices de coniques en termes de birapports . . . . .	69
1.22	Le modèle des systèmes minces. . . . .	73
1.23	Le modèle des systèmes minces avec capteur. . . . .	74
1.24	Système optique épais avec plan de capteurs. . . . .	76
1.25	Modèle de la projection perspective. . . . .	77

1.26	Calcul d'un déplacement point par point. . . . .	79
1.27	Calcul d'une transformation projective. . . . .	79
2.1	Définition de la géométrie épipolaire. . . . .	85
2.2	Droites épipolaires associées à quelques points de deux images. . . . .	86
2.3	Détermination de la géométrie épipolaire à partir de six appariements de points. . . . .	88
2.4	Test de coplanarité. . . . .	89
2.5	Construction de points coplanaires : première méthode. . . . .	91
2.6	Deuxième méthode de construction de points coplanaires : vue 3D. . . . .	92
2.7	Deuxième méthode de construction de points coplanaires : les deux images. . . . .	92
2.8	Fonction de densité des valeurs du birapport. . . . .	95
2.9	Fonction de répartition des valeurs du birapport. . . . .	95
2.10	Seuils pour le test de coplanarité. . . . .	98
2.11	Histogrammes des valeurs au test de coplanarité pour des quadruplets coplanaires. . . . .	99
2.12	Taux de résultats positifs au test pour des quadruplets coplanaires. . . . .	100
2.13	Taux de résultats positifs au test pour des quadruplets quelconques. . . . .	102
2.14	Histogramme des valeurs au test de coplanarité pour des quadruplets quelconques. . . . .	103
2.15	Corrélation entre les volumes de tétraèdres et les résultats au test pour des bruits de 0 et 0,5 pixels dans les images. . . . .	104
2.16	Corrélation entre les volumes de tétraèdres et les résultats au test pour des bruits de 1 et 2 pixels dans les images. . . . .	105
2.17	Corrélation entre les volumes de tétraèdres et les résultats au test pour des bruits de 3 et 5 pixels dans les images. . . . .	106
2.18	Récapitulatif des résultats du test de coplanarité sur des données simulées : taux de réponses positives des quadruplets coplanaires et quelconques. . . . .	107
2.19	Deux images d'une maison en bois. . . . .	108
2.20	Histogramme des résultats du test de coplanarité pour les images de la maison. . . . .	109
2.21	Numérotation des sommets sélectionnés. . . . .	110
2.22	Résultats obtenus pour quelques configurations. . . . .	110
2.23	Résultats obtenus pour les configurations contenant les points numérotés 5, 8 et 14, et celles contenant les points 21, 23 et 28. . . . .	111
2.24	Cas de six points dans l'espace 3D. . . . .	114
2.25	Cas de quatre points et une droite. . . . .	115
2.26	Cas de trois points et deux droites. . . . .	116
2.27	Cas de deux paires de coplanaires. . . . .	117
2.28	Cas de quatre droites quelconques. . . . .	118
2.29	Cas d'une conique et de trois points. . . . .	119
2.30	Seuils de stabilité des invariants de deux paires de droites coplanaires. . . . .	122
2.31	Histogrammes des distances obtenues lors du calcul de l'invariant de deux paires de droites coplanaires. . . . .	123
2.32	Pourcentage de configurations de deux paires de droites coplanaires donnant des distances inférieures au seuil moyen et à deux et trois fois cette valeur. . . . .	124

2.33	Comparaison entre le calcul à partir des images et celui à partir de la scène 3D. . . . .	125
2.34	Histogrammes des distances entre les invariants calculés à partir de la scène et ceux calculés à partir des images. . . . .	126
2.35	Résultats obtenus avec une méthode de reconstruction. . . . .	128
2.36	Pourcentage d'appariements corrects détectés. . . . .	129
2.37	Pourcentage d'appariements incorrects détectés. . . . .	130
2.38	Trois images d'un dièdre. . . . .	130
2.39	Numérotation des segments du dièdre. . . . .	131
2.40	Distance entre les invariants calculés dans les deux paires d'images. . . . .	132
2.41	Définition des repères liés à une projection perspective. . . . .	138
2.42	Premier exemple d'appariement. . . . .	142
2.43	Deuxième exemple d'appariement. . . . .	143
2.44	Troisième exemple d'appariement. . . . .	144
2.45	Exemple d'appariement d'une scène complexe. . . . .	146
3.1	Le processus de modélisation et reconnaissance. . . . .	154
3.2	Trois images d'un même objet observé du même point sous trois éclairages différents et le résultat de leur traitement. . . . .	155
3.3	Exemple d'arrondi dû à un flou . . . . .	157
3.4	Un exemple d'extrémité libre. . . . .	158
3.5	Deux exemples de corrections possibles des dédoublements (à gauche), et un cas où la correction est impossible (à droite). . . . .	158
3.6	Un exemple d'emploi des techniques de correction. . . . .	159
3.7	Séquence d'images du premier objet . . . . .	161
3.8	Comparaison des deux mesures. . . . .	162
3.9	Les trois aspects d'un premier ensemble d'images. . . . .	165
3.10	Quatre-vingts vues d'un objet. . . . .	165
3.11	Deux représentants pour chacun des sept aspects détectés du deuxième ensemble d'images. . . . .	166
A.1	Construction de points coplanaires : première méthode. . . . .	176
A.2	Deuxième méthode de construction de points coplanaires : vue 3D. . . . .	180
A.3	Deuxième méthode de construction de points coplanaires : les deux images. . . . .	180



# Introduction

## Généralités

La vision artificielle a pour but de permettre à des systèmes automatiques, informatiques ou robotiques, d'acquérir et de traiter des informations visuelles, en s'affranchissant au maximum des capacités visuelles d'un opérateur humain. En cela, elle diffère de l'imagerie, qui utilise aussi des images, mais qui laisse les opérations de traitement de haut niveau et d'interprétation à un opérateur humain.

On peut diviser la vision en deux grands domaines : l'acquisition des images, et le traitement de celles-ci.

L'acquisition des images est un domaine fort ancien. On peut considérer que la peinture en est un lointain ancêtre. Plus proche de nous, la photographie, le cinéma et la vidéo sont, entre autres, des systèmes d'acquisition. Le but est de transformer un signal physique fugitif, composé d'ondes électromagnétiques, en un signal d'une autre nature, électronique analogique ou digital, magnétique, chimique ou informatique. Les domaines concernés par une telle opération sont nombreux : physique, instrumentation, électronique, automatique, traitement du signal, micro-informatique...

Le traitement des images consiste à transformer le signal obtenu pour en tirer des informations de niveau d'abstraction ou de symbolisme croissant. Les informations peuvent être des mesures, des formes, ou une interprétation symbolique complète de la scène observée.

La vision par ordinateur s'inscrit dans ce deuxième grand domaine. On considère des images, représentées de manière interne à l'ordinateur par un tableau d'entiers appelés pixels, chaque entier codant l'intensité lumineuse perçue en un point. Cette information est le résultat d'un long traitement optique et électronique, dont le but est de transformer le signal optique initial en ce tableau. Ce traitement n'est pas sans effets secondaires, et l'information obtenue est très dégradée [Bey 92]. Toute mesure précise à partir du tableau, que nous appellerons désormais image, est délicate ; elle nécessite un calcul d'incertitude rigoureux et des méthodes robustes pour obtenir un résultat utilisable. Cela dit, l'œil humain sait, non sans mal, utiliser de telles images. Celles-ci

doivent donc encore contenir une information utile : le problème consiste à trouver les traitements qui peuvent l'extraire.

Ce problème n'est point trivial. L'homme dispose, en effet, d'un système de vision perfectionné et performant, tant par son capteur stéréoscopique autofocus à ouverture variable, que par les capacités de traitement cérébral, qui lui permettent de passer d'un traitement de bas niveau à une interprétation symbolique. Par contre, l'homme ne possède pas l'expertise de sa propre vision : il est incapable d'expliquer ce qu'il fait pour reconnaître et interpréter ce qu'il voit, et en particulier, il ne sait donner aucune méthode exploitable directement comme algorithme de reconnaissance.

La psychologie cognitive a, entre autres, pour but de rechercher et d'explicitier les mécanismes de la vision humaine ou animale. La vision par ordinateur ne vise pas directement une reproduction sur machine de la vision humaine. Elle vise plutôt à trouver des algorithmes permettant d'arriver à des résultats proches. C'est un domaine neuf, tirant ses outils tant de l'informatique et de l'intelligence artificielle, que du traitement du signal, des mathématiques et de la géométrie ou de la physique.

On peut classer les traitements de la vision par ordinateur en trois niveaux (cf. FIG. 0.1).

1. Un premier niveau, dit bas niveau, concerne le traitement d'une seule image : il s'agit de regrouper les pixels, c'est à dire les entiers de la matrice formant l'image, selon certains critères. L'information présente dans une image est, en effet, considérable, et il convient de la réduire et de la condenser, pour pouvoir l'utiliser plus commodément.

Les critères de regroupement sont variés : on peut regrouper les pixels de valeurs voisines pour former des régions, ou ceux se trouvant sur une même ligne de discontinuité d'intensité lumineuse pour former des contours. On peut aussi extraire des points correspondant à des coins, ou des extremums de courbure. Les contours et régions peuvent à leur tour être traités pour obtenir des segments, des lignes polygonales ou des courbes algébriques, des fonctions splines par exemple.

2. Un deuxième niveau concerne le calcul de caractéristiques numériques de la scène observée : il peut s'agir de distances, surfaces, volumes, mouvements, vitesses, ou d'autres mesures comme des birapports. De telles mesures demandent la plupart du temps de disposer de plusieurs images. Un premier problème consiste alors à apparier les images, ou à "poursuivre" des indices, c'est à dire à retrouver la trace dans chaque image d'une même caractéristique de la scène observée. Le deuxième problème important est l'obtention de mesures précises et peu bruitées. Les algorithmes de ce niveau permettent la reconstruction de la scène, c'est à dire le positionnement 3D relatif des divers éléments de la scène, le positionnement de la caméra par rapport à la scène, l'étude des mouvements des différentes composantes de la scène ou la reconnaissance de celles-ci, si cela est possible, à partir des caractéristiques géométriques.
3. Le troisième niveau, dit haut-niveau, concerne l'interprétation de ce qui est observé : il s'agit alors d'identifier le contenu de la scène, et de pouvoir le décrire dans un langage verbal, plutôt que géométrique ou mathématique.

Niveau	Résultat du traitement	Objets obtenus	Niveau de traitement
1	scène ↓	objets	acquisition
	image ↓	pixels	bas niveau
	primitives	contours, régions segments, courbes	niveau intermédiaire
2	↓ grandeurs numériques	vitesse, positions surfaces, volumes	
3	↓ interprétation	noms d'objets	haut niveau

FIG. 0.1 - Les niveaux de traitement en vision.

### Contenu de la thèse

Le travail décrit dans cette thèse se situe au niveau intermédiaire. Il concerne la modélisation automatique d'objets polyédriques pour la reconnaissance. On souhaite pouvoir reconnaître automatiquement des objets dans des images où ils sont représentés. Cette reconnaissance peut être utilisée par un robot mobile pour identifier son environnement, par un convoyeur industriel pour trier des pièces ou vérifier leur qualité.

Pour reconnaître un objet, il faut le connaître au préalable, c'est à dire disposer d'une information suffisante à son propos pour pouvoir le distinguer d'autres objets. L'acquisition de cette connaissance préalable est le but de la modélisation de l'objet. Il s'agit de savoir quelle information est nécessaire pour la reconnaissance, et de déterminer d'où et comment cette information peut être extraite.

Dans cette thèse la modélisation est réalisée à partir d'images. On connaît un objet par un ensemble d'images le représentant. On va alors extraire de ces images l'information nécessaire pour le reconnaître. Ce processus est représenté sur le schéma de la FIG. 0.2. On part d'un objet dont on prend plusieurs images. Ces images sont traitées et le modèle de l'objet est construit à partir d'elles. Le modèle obtenu est rangé dans une base de modèles. Lors de la reconnaissance, on recherche si un des objets dont le modèle est dans la base de donnée n'est pas représenté sur une nouvelle image.

La première difficulté consiste à déterminer le type d'information utile qui peut être extrait des images dont on dispose. On considère deux types d'informations dans ce travail : celles de type numérique et celles de type visuel. Ce dernier vocable concerne la notion d'aspect visuel, notion qui se rapproche de celle d'aspect topologique. Ces deux types d'informations, numérique et visuel, donnent les deux grands axes de cette thèse. Ceux-ci ne sont toutefois pas complètement indépendants : leur utilisation simultanée permet même d'éviter certains problèmes des algorithmes ne relevant que d'un seul de ces domaines : forte combinatoire et faible robustesse.

**Le premier axe** de cette thèse concerne donc l'extraction de grandeurs numériques caractéristiques d'un objet à partir d'un ensemble d'images de celui-ci. La plupart

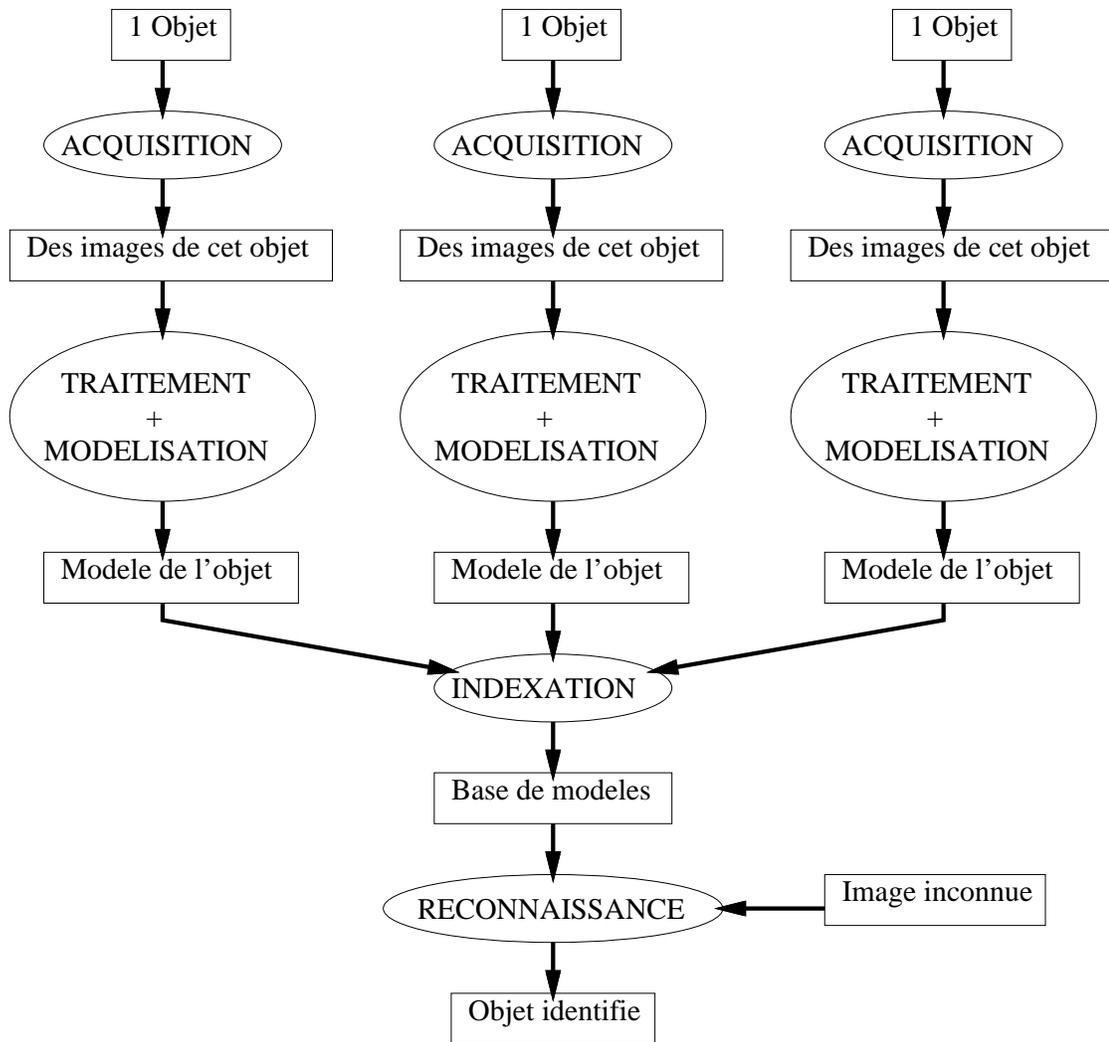


FIG. 0.2 - Schéma d'une méthode de modélisation reconnaissance.

des grandeurs habituelles, longueur, angle ou surface, ne sont pas conservées dans les images. Il s'agit donc de rechercher si on peut trouver d'autres grandeurs qui se conservent, et comment on peut calculer ces grandeurs simplement à partir des images de l'objet dont on dispose.

Le but d'une telle recherche est de caractériser chaque objet par des quantités numériques. Pour les objets du monde 3D habituel, on utilise souvent pour les décrire leurs longueur, largeur, profondeur, circonférence, surface ou volume. Cette information n'est pas calculable à partir d'images. On va donc chercher d'autres grandeurs, mais le but est le même.

Le cadre naturel de cette étude est celui de la géométrie et de la théorie des invariants. Les grandeurs recherchées doivent, en effet, dépendre de l'objet observé et non de l'angle sous lequel il est observé. D'où la recherche d'invariants indépendants des transformations liées au changement de point de vue.

L'utilisation des invariants en vision n'est pas récente, mais ce n'est que récemment

qu'elle a pris son véritable essor. Diverses voies ont été étudiées : invariants 2D ou 3D, calculés à partir d'une ou de plusieurs images. Ces différents cas correspondent à diverses hypothèses : dispose-t-on d'informations préalables, d'une ou de plusieurs images ? Si on a plusieurs images, sont-elles appariées ?

Dans cette thèse, l'étude des invariants commence par un exposé général de la géométrie projective et de la théorie des invariants. Cette étude fournit un cadre conceptuel très riche dans lequel les problèmes de vision peuvent être posés de manière claire et naturelle. Cette étude n'est pas menée dans toute la généralité que peut avoir aujourd'hui la théorie des invariants, mais elle l'est de manière à permettre la réponse à certaines questions concernant l'indépendance et le dénombrement des invariants.

Diverses méthodes de calcul effectif des invariants sont proposées dans la littérature. Pour notre part, nous utilisons la méthode infinitésimale pour exposer la théorie. Pour la pratique, nous proposons une méthode géométrique, basée sur la recherche des invariants canoniques liés à la géométrie dans laquelle on se place. Cette méthode a l'inconvénient, par rapport à certaines méthodes algébriques ou symboliques, de ne pas être systématique et automatique. Elle requiert de l'agilité géométrique pour trouver les invariants de certaines configurations. En revanche, elle se révèle très simple d'emploi et très puissante dans le cas où on ne considère que des éléments d'ordre inférieur ou égal à deux : points, droites, plans et coniques. L'utilisation de cette méthode est aussi possible pour des courbes d'ordre supérieur, en utilisant des éléments différentiels, comme les droites bitangentes ou d'inflexion.

Cette méthode est appliquée à trois problèmes :

1. la recherche des points coplanaires d'un objet ;
2. le calcul d'invariants 3D caractéristiques d'un objet 3D ;
3. l'appariement des images.

Dans ces trois applications, une attention toute particulière est portée à la robustesse et à l'application des algorithmes proposés à des images réelles (par opposition à des images de synthèse).

Les deux premières applications sont des exemples particulièrement probants de l'utilisation de la "méthode des invariants canoniques" dans des cas complexes. L'avantage de cette méthode est de fournir une interprétation géométrique simple des calculs effectués et du résultat obtenu. Cela permet de détecter facilement les cas instables ou certains calculs mal conditionnés.

La dernière de ces applications fait le lien avec le deuxième axe de la thèse. L'utilisation des invariants se révèle, en effet, un moyen efficace d'apparier deux images : l'algorithme proposé est beaucoup plus stable que les algorithmes d'appariements basés sur les recherches d'isomorphismes de graphes, il nécessite moins d'informations préalables que ceux basés sur l'utilisation de la géométrie épipolaire, et il exige moins de contraintes sur le mouvement de la caméra entre les prises de vue que les algorithmes de "poursuite" (tracking en anglais) ou de corrélation.

**Le deuxième axe** concerne l'apprentissage des aspects d'un objet, appelé encore modélisation de cet objet. Pour reconnaître un objet, il faut disposer d'informations à son sujet. Le problème est donc de déterminer quelles informations sont nécessaires

pour cela, de savoir à partir de quelles données on peut extraire ces informations, et comment cette extraction peut être faite en pratique. C'est cette phase d'extraction des informations et de leur organisation qui est appelée modélisation.

Les grandeurs numériques envisagées au paragraphe précédent ne suffisent généralement pas : elles permettent de décrire les caractéristiques d'un objet, mais pas sa nature. On va donc essayer d'extraire un autre type d'information des images. La difficulté vient du fait que celui-ci n'est pas numérique : cette recherche s'inscrit dans un cadre bien moins formel, et il est difficile de juger des résultats de cette recherche autrement que subjectivement.

Considérons, par exemple, les images de la FIG. 0.3. Elles contiennent, pour un observateur humain, suffisamment d'information pour pouvoir reconnaître l'objet représenté. Cette information est de grande taille (13 images de taille  $512 \times 512$  pixels dans le présent exemple, mais il peut y en avoir une centaine), diffuse (la valeur d'un pixel ne donne qu'une information extrêmement faible sur l'objet) et bruitée. Pour modéliser cet objet, il est nécessaire de déterminer l'information utile contenue dans ces images, et de trouver le moyen d'extraire cette information de l'ensemble d'images.

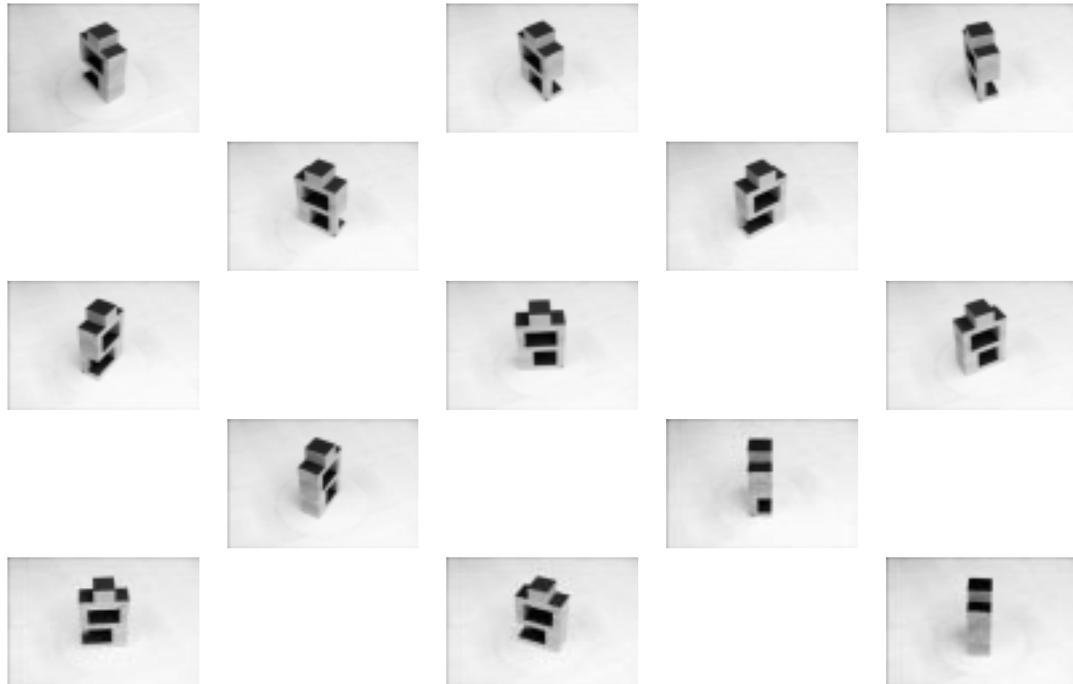


FIG. 0.3 - Un ensemble d'images d'un même objet.

Les systèmes de modélisation proposés dans la littérature sont nombreux et variés. Cependant, peu d'entre eux ont réussi à s'imposer et à être utilisés dans un cercle plus vaste que l'équipe de leurs concepteurs. Cela est d'abord dû à la difficulté du problème : les solutions sont généralement adaptées à diverses hypothèses et circonstances explicites ou implicites, mais n'ont pas de caractère très général. Il ne s'agit pas, bien sûr, de vouloir réaliser un système de modélisation universel, mais, pour un domaine

d'application bien déterminé, de disposer d'un système applicable à un grand nombre d'images, qui ne soit pas lié à un matériel particulier, aux conditions particulières d'un laboratoire ou au savoir-faire d'un expérimentateur donné.

Il est proposé un nouvel algorithme de modélisation dans cette thèse. Celui-ci est basé sur divers principes tirés de systèmes existants, organisés de manière à obtenir une grande robustesse.

- Le bruit dans les images étant très important et difficile à modéliser, les modèles d'objets sont construits à partir de ce qui est perçu, donc à partir de données similaires à celles qui serviront pour la reconnaissance. Ceci est un principe très classique en classification automatique.
- La reconnaissance d'objets représentés sur des images 2D est plus aisée à partir de modèles 2D. La plupart des méthodes pour images 2D et modèles 3D incluent généralement une étape de projection du modèle qui est souvent délicate (par exemple, calcul du graphe d'aspect à partir d'un modèle CAO).
- La robustesse vis-à-vis du bruit est un objectif prioritaire.

A cela, on ajoute une conjecture qui donne son originalité à la méthode : la phase d'appariement est la plus informative. Reconnaître consiste, en effet, à appairer une image et un modèle, modéliser consiste à appairer plusieurs images pour en extraire l'information stable. C'est cette conjecture qui sous-tend l'organisation de l'algorithme de modélisation et la cohérence globale du système de modélisation et reconnaissance.

La méthode proposée est la suivante : on prend de nombreuses images de l'objet à modéliser depuis des points de vue différents. Ces images sont appariées deux par deux puis regroupées selon les aspects qu'elles représentent. Un modèle de chaque aspect est construit : il regroupe les caractéristiques stables de cet aspect. Ces modèles sont destinés à être indexés dans une base de données, qui sera utilisée pour la reconnaissance.

## Contributions majeures

Les principales contributions de cette thèse sont donc les suivantes :

1. la réalisation d'un test de coplanarité et l'étude de stabilité de ce test : à partir de deux images d'un objet, on retrouve les configurations de 4 points coplanaires de l'objet (voir le paragraphe 2.1) ;
2. le calcul des invariants 3D et l'étude de la stabilité de ces invariants : à partir de deux images d'un objet, on calcule des grandeurs numériques caractéristiques de cet objet (voir le paragraphe 2.2) ;
3. l'appariement de deux images par la technique des invariants locaux : on retrouve ainsi automatiquement où sont les projections dans chacune des deux images de chaque arête ou sommet de l'objet observé (voir le paragraphe 2.3) ;
4. la recherche automatique des aspects d'un objet à partir d'un ensemble d'images de celui-ci : cette recherche des aspects est un pas important de l'algorithme de modélisation des objets (voir le chapitre 3).

## Plan de la thèse

Un premier chapitre regroupe les considérations théoriques : introduction à la géométrie projective, théorie et calcul des invariants, introduction à la vision. Il fixe le cadre dans lequel prend place le reste du travail, et expose deux théories très classiques en mathématiques d'une manière plus adaptée à la vision par ordinateur.

Le deuxième chapitre regroupe les applications des invariants à la vision. La première application concerne la recherche de points coplanaires d'une scène. Cette recherche est effectuée à partir de deux images appariées, et suppose connue la géométrie épipolaire de ces deux images. L'utilisation d'images de synthèse permet une étude précise de la stabilité du test de coplanarité proposé, test qui est ensuite validé avec des images réelles. La réciproque de ce test permet, quant à elle, de construire les projections dans les deux images de nouveaux points coplanaires.

La deuxième application est le calcul d'invariants 3D. Ce calcul utilise les résultats précédemment obtenus. Dans ce cas aussi, des résultats sont donnés sur des images de synthèse et des images réelles.

La dernière application concerne l'appariement d'images. Cette partie, illustrée avec des exemples réels, fait le lien avec le chapitre suivant.

Le troisième et dernier chapitre traite de la modélisation automatique des aspects d'un objet. Cette modélisation est basée sur l'appariement des images. Les trois étapes principales en sont :

1. l'appariement initial des images et l'amélioration de cet appariement ;
2. la détection des aspects et le regroupement des images selon l'aspect qu'elles représentent ;
3. la modélisation de chaque aspect, et le calcul de quantités numériques qui complètent ce modèle.

La conclusion dégage les perspectives ouvertes par ce travail.

## Guide de lecture

La rédaction de cette thèse essaye de répondre à deux contraintes opposées. D'une part, la géométrie projective et la théorie des invariants nécessitent une grande rigueur dans l'exposé, pour permettre une compréhension fine des notions. D'autre part, certains lecteurs peu familiers de ces théories ont besoin d'approches plus intuitives pour "faire leur intuition".

A destination de ces lecteurs, on a indiqué des conseils pour une "lecture rapide" qui évite les passages les plus délicats mathématiquement et ceux d'intérêt moindre pour l'enchaînement des idées. D'autre part, on a présenté beaucoup de notions de manière intuitive avant de donner leur formulation mathématique rigoureuse.

# Chapitre 1

## Géométrie, invariants et vision : théorie

Dans ce chapitre sont rassemblées les considérations générales et théoriques qui sont la base du reste du travail. Mais au delà de la simple description d'un vocabulaire qui ne serait nécessaire que pour pouvoir formuler les résultats ultérieurs de manière pratique, il s'agit de montrer comment la géométrie fournit un cadre conceptuel puissant dans lequel les problèmes de vision peuvent être posés, interprétés et résolus de manière simple et efficace.

Le problème consiste à trouver un langage dans lequel les problèmes de la vision, tant quantitatifs que qualitatifs, puissent être posés et résolus. Ce langage doit donc permettre de décrire des images et de manipuler de telles descriptions.

Diverses approches ont été proposées. Tout d'abord, on peut utiliser un langage de haut-niveau, qui permet des raisonnements symboliques. Cette option est souvent prise par les chercheurs en apprentissage automatique [Liq 89, Con 90].

Une deuxième approche consiste à considérer plusieurs niveaux de représentations, avec des langages de descriptions différents pour chaque niveau. Une telle approche a été développée par Marr [Mar 82].

Une troisième approche utilise un langage proche du signal image et des concepts tels la réponse impulsionnelle, la bande passante... Une telle description est adaptée aux études de bas niveau [Can 86, Der 87a].

La dernière approche décrit les images en termes géométriques et utilise toute la puissance de la géométrie. Une telle approche n'est pas récente en vision [LH 81].

Ces différentes approches se différencient principalement par leur niveau d'abstrac-

tion. Marr avait une idée très générale de la vision, qui a vieilli techniquement depuis, mais garde tout son intérêt par sa généralité et son originalité. Les approches de haut niveau sont, certes, adaptées aux raisonnements symboliques, mais le passage entre les images et leur description est très difficile, voire impossible pour des images non simplistes en l'état actuel des connaissances. Les descriptions de bas-niveau sont adaptées au traitement effectif des images. La difficulté réside alors dans la pauvreté des raisonnements possibles.

La géométrie permet de concilier une description variée des images, quoique de bas niveau, en termes de points, droites, courbes, et un outillage très complet, tant quantitatif (dérivation, intégration, invariants...), que qualitatifs (théorie de la morphogénèse pour le calcul des aspects d'un objet).

Pour notre part, nous utilisons une description géométrique et topologique. Les images contiennent des segments de droites et des points liés par des relations. C'est l'ensemble de ces données qui constitue la description d'une image.

Ce chapitre contient les fondements théoriques pour l'étude de telles images. Il est divisé en trois parties. La première concerne la géométrie, ou plus précisément quelques géométries utiles en vision : géométries projective, affine et euclidienne. A titre d'information, un premier paragraphe donne quelques indications historiques et le dernier situe, sans les approfondir, d'autres géométries, non-euclidienne, algébrique et riemannienne, et la topologie.

Une deuxième partie expose la théorie des invariants, dans un cadre simple, c'est à dire sans traiter ce problème dans son cadre algébrique le plus large, mais de manière suffisamment précise pour pouvoir en tirer des résultats efficaces en géométrie, comme, par exemple, le dénombrement des invariants fonctionnellement indépendants existant dans un problème géométrique.

La dernière partie concerne la vision, et plus particulièrement l'approximation des transformations réalisées par une caméra par des modèles optiques et géométriques. C'est cette étude qui fonde la pertinence de l'utilisation de la géométrie en vision. Le problème du bruit dans les caméras n'est qu'effleuré et les traitements dits de bas niveau ne sont rappelés que pour mémoire.

## 1.1 Géométrie projective

Dans toute la suite de cette thèse, la géométrie projective sera considérée comme la géométrie étalon. Cela est justifié tout d'abord par la simplicité de sa caractérisation par Möbius : elle transforme, dans un espace réel, un point en un point et une droite en une droite, ce qui est, au moins approximativement, le cas dans les images obtenues avec une caméra électronique. D'autre part, l'étude du paragraphe 1.3 montre que la projection perspective est un modèle pertinent pour une caméra.

Ce paragraphe a pour but de rappeler les principales notions de la géométrie projective. Il ne vise ni l'exhaustivité des notions ni la démonstration rigoureuse des résultats énoncés. Pour une étude plus complète, le lecteur pourra se reporter soit à [Com 41] pour une introduction, soit à [Sem 52] pour une étude plus détaillée.

La présentation de la géométrie faite ici suit le plan suivant : après de brefs rappels historiques, on définit la notion de représentation d'un ensemble, c'est à dire son paramétrage par des coordonnées homogènes, puis on étudie les propriétés et concepts

invariants par changements de paramétrage. Ce sont :

- la dépendance linéaire qui définit les droites, plans et hyperplans ;
- le birapport qui définit la relation de conjugaison ;
- les coniques.

Les géométries affine, euclidienne et non euclidiennes sont alors définies comme des restrictions de la géométrie projective : on n'autorise plus alors qu'un nombre plus restreint de paramétrages, ce qui permet d'obtenir des invariants et des concepts plus riches.

### 1.1.1 Considérations historiques

Connaître l'histoire des concepts géométriques a deux conséquences appréciables. Tout d'abord, cela oblige à une certaine modestie vis à vis de ses propres trouvailles. Pour les sujets nous intéressant, la grande majorité des résultats datent soit de l'antiquité grecque, soit du dix-neuvième siècle. Trouver un résultat tant soit peu nouveau dépasse donc amplement le cadre de cette thèse.

Deuxièmement, cela permet de mieux comprendre la portée des concepts et leur articulation. Ainsi, la géométrie n'apparaît pas comme unifiée, mais comme subissant un grand mouvement d'unification (émergence des structures algébriques sous-jacentes) lui donnant un champ d'application très vaste, qui la renouvelle et la diversifie sans cesse. Ce paragraphe est largement inspiré du livre de Dahan-Dalmédico et al. [DD 86] et de l'article "Géométrie" de l'Encyclopædia Universalis.

**Les origines.** Dans les civilisations babylonienne et égyptienne, la géométrie était pratique et pragmatique. Il s'agissait de répondre à des problèmes de la vie courante : calculs d'aires de champs, constructions d'objets, d'habitations, de monuments. . . Quelques notions géométriques sont utilisées : propriétés des similitudes pour reproduire des dessins, valeur de  $\pi$  estimée à 3, 3 et 1/8 ou 3 et 1/6 pour calculer des aires, ainsi que quelques formules pour calculer le volume des pyramides.

La géométrie grecque opère une rupture nette avec ces origines pragmatiques. La géométrie devient abstraite et déductive. L'espace géométrique est conçu comme étant infini, la différence entre les dessins réalisés et les objets géométriques est perçue, un grand effort est fait pour démontrer les résultats. Les plus brillantes contributions sont dues à Euclide, Apollonius et Archimède.

On doit à Euclide, avec les *Eléments*, la première présentation axiomatique et rigoureuse de la géométrie. Certes, de nombreuses imperfections subsistent, mais l'œuvre d'Euclide reste la référence en géométrie pendant plus de quinze siècles. Le postulat des parallèles est très vite perçu comme ayant un statut particulier. Apollonius, un élève de l'école d'Euclide au musée d'Alexandrie, publie un traité sur les sections coniques. Archimède a pour sa part obtenu des résultats en théorie des aires, et il utilise une méthode d'exhaustion qui est l'ancêtre des techniques d'intégration.

Suite à la fin de la période grecque, la géométrie n'avance plus guère, mais l'héritage d'Euclide est sans cesse retravaillé et élargi par les mathématiciens arabes, puis européens à la fin du Moyen âge. De nombreuses tentatives, infructueuses bien entendu,

sont faites pour démontrer le postulat des parallèles à partir des autres postulats des *Eléments*.

**Le développement des géométries.** Il faut attendre le début du quinzième siècle pour voir apparaître les prémices du renouvellement de la géométrie. L'architecte florentin Brunelleschi (1377-1446) décrit les règles de la perspective, règles qui sont reprises et exposées par Piero della Francesca (1470) et Albert Dürer (1525). Dans son ouvrage publié en 1511 [Alb 11], Alberti pose la question de savoir quelles sont les propriétés géométriques communes à deux vues perspectives d'une même figure. Cette question, posée à l'époque dans le cadre de la peinture et de l'architecture, sera la question à la base de la photogrammétrie lors de l'apparition de la photographie, et de la vision par ordinateur lors de l'apparition des ordinateurs et des caméras électroniques. Alberti devient ainsi un précurseur pour ces domaines modernes.

La personnalité la plus marquante, géométriquement parlant, du dix-septième siècle est sans conteste le lyonnais Girard Desargues, architecte, ingénieur et familier des ouvrages d'Apollonius. Dans son ouvrage principal [Des 39], il décrit les coniques comme des perspectives de cercles, il introduit les points conjugués par rapport à une conique et il traite des pôles et des polaires dans le cas général. Cette dernière notion avait été introduite par Apollonius, qui séparait les différents cas de coniques. La confidentialité de son ouvrage tiré à seulement une cinquantaine d'exemplaires, la difficulté du vocabulaire utilisé employant de nombreuses analogies végétales, expliquent un relatif insuccès des idées de Desargues.

Ces idées seront reprises par Pascal et La Hire, puis oubliées durant le dix-huitième siècle, siècle dominé par les méthodes infinitésimales naissantes et la géométrie des coordonnées de Descartes [Des 37]. Les méthodes de géométrie synthétique (par opposition à la géométrie analytique de Descartes) réapparaissent avec la géométrie descriptive de Monge, qu'il introduit à l'Ecole Centrale des Travaux Publics, qui devient un an plus tard Ecole Polytechnique, et à l'Ecole Normale de l'an III.

Les idées de la géométrie projective de Desargues sont reprises au dix-neuvième siècle par Jean Victor Poncelet, officier napoléonien, fait prisonnier pendant la campagne de Russie, et qui, dans sa geôle russe, écrit une remarquable synthèse des idées projectives, qui sera publiée [Pon 22] et dans laquelle se retrouvent tous les géomètres en réaction contre la trop forte emprise de la géométrie analytique.

Poncelet introduit un certain nombre de principes fondamentaux de la géométrie. Il met pour la première fois l'accent sur la notion de transformation. Transformation entre points par projection ou homologie, et entre points et droites par dualité. Il ajoute même un principe de continuité de ces transformations, mais qu'il ne justifie pas. L'influence de Poncelet fut très grande et elle exacerba l'opposition entre géométrie analytique et géométrie synthétique qui s'interdisait tout recours aux coordonnées, revendiquant ainsi son autonomie.

L'opposition entre ces géométries va être réduite peu à peu. Chasles et Möbius reprennent l'étude des transformations et celles de leurs invariants en particulier : angles, rapports, birapports. Ils désignent les transformations projectives les plus générales par le terme d'homographie, et montrent que ce sont les applications qui transforment les points en points, les droites en droites et conservent le birapport. Möbius montre que cette dernière condition découle des deux autres. Les notions de déplacements,

similitudes, transformations affines, collinéations apparaissent clairement.

Le développement de la géométrie est alors rapide. La plupart des notions courantes de géométrie sont clarifiées. C'est à cette époque que se développent les géométries non euclidiennes (voir le paragraphe. 1.1.9).

**Le programme d'Erlangen.** La synthèse de toute cette évolution de la géométrie est fournie en 1872 par Félix Klein dans son discours d'admission à l'Université d'Erlangen intitulé "*Considérations comparatives sur les recherches géométriques modernes*" et appelé fréquemment programme d'Erlangen [Kle 74].

Au lieu de définir chaque géométrie par les objets sur lesquels elle travaille, il la définit par les transformations qui la constituent. Une géométrie devient l'étude des invariants pour un ensemble de transformations. Mais l'apport principal de Klein tient surtout au rapprochement de la notion de transformation avec celle de groupe.

La notion de groupe, introduite dès 1830 par Galois, ne fut pas vraiment diffusée avant que Jordan le fit en 1870, et elle restait confinée au cas des groupes de permutations. Klein adopte d'emblée le point de vue des groupes, en exhibant une suite de groupes imbriqués les uns dans les autres : le groupe projectif, les similitudes (le groupe affine n'est pas donné par Klein), les isométries, les déplacements, les translations. Lorsqu'on passe d'un groupe à un groupe plus général, on ne garde qu'une partie des propriétés.

Mais Klein considère aussi des transformations plus générales, introduisant la géométrie algébrique avec le groupe des transformations birationnelles et la topologie avec le groupe des transformations continues. Il introduit aussi la géométrie en dimension quelconque (dans une multiplicité). C'est en ces termes que le problème général est posé : "*Etant donné une multiplicité et un groupe de transformations de cette multiplicité, en étudier les êtres au point de vue des propriétés qui ne sont pas altérées par les transformations du groupe.*"

**Qu'en retenir ?** Le programme d'Erlangen nous fournit notre cadre d'étude de la géométrie et de la vision. On modélise la caméra dont nous disposons par des lois optiques, puis par un modèle géométrique de la transformation réalisée. Nous étudions alors les invariants pour le type de la transformation trouvée. Les différentes parties de ce programme d'étude constituent les paragraphes qui suivent.

### 1.1.2 Représentation, espace projectif, coordonnées homogènes

Ce paragraphe fournit des approches de la notion d'espace projectif. Les unes, plus intuitives pourront être préférées par les lecteurs peu familiers avec cette géométrie, qui pourront ne faire que parcourir la dernière, plus mathématique. Cette dernière approche permet une compréhension plus fine de la nature de la géométrie projective et du principe de dualité par exemple.

Dans toute la suite,  $\mathbb{K}$  représente un corps, en pratique  $\mathbb{R}$  ou  $\mathbb{C}$ . Les éléments de ce corps seront appelés scalaires. On note  $\mathbb{K}^n$  l'ensemble des n-uplets d'éléments de  $\mathbb{K}$  et  $\mathbb{K}^{n*}$  l'ensemble des n-uplets à l'exception du n-uplet nul.

### Approches intuitives

**Première approche.** On peut définir un espace projectif comme un sur-ensemble d'un espace affine habituel. Prenons l'exemple du plan : on ajoute à ce plan une droite supplémentaire, appelée droite de l'infini, qu'on peut assimiler à la droite de l'horizon. Deux droites parallèles, qui ne se coupent pas dans l'espace affine, se coupent à l'infini dans l'espace projectif : leur point d'intersection est donc sur cette droite de l'infini  $D_\infty$ . Remarquons que toutes les droites parallèles à une même direction, se coupent en un seul point à l'infini. Chaque point de  $D_\infty$  indique ainsi une direction.

Ce fait rejoint une intuition courante : les deux rails du chemin de fer semblent se rejoindre au loin. C'est aussi la pratique en peinture de représenter la "perspective" en faisant se couper les parallèles sur la ligne d'horizon. Cette particularité projective de l'intersection des parallèles fit écrire à Stendhal : «Au commencement de la géométrie, on dit : " On donne le nom de parallèles à deux lignes qui, prolongées à l'infini, ne se rencontreront jamais." Et, dès le commencement de la statique, cet insigne animal de Louis Monge a écrit à peu près ceci : "Deux parallèles peuvent être considérées comme se rencontrant si on les prolonge à l'infini."»

En géométrie classique, les points à l'infini y restent. On a souvent, dans les débuts de l'étude des coniques, l'impression que la parabole n'est qu'une ellipse allongée à l'infini dans une direction, et que l'hyperbole est une ellipse tellement allongée qu'on la voit réapparaître par "l'autre côté" du plan. Cela dit, rien ne permet de conforter cette intuition. Toutes les applications affines transforment les ellipses en ellipses, les paraboles en paraboles, les hyperboles en hyperboles.

En géométrie projective, on ne fait que considérer la droite de l'infini à l'égale des autres. Deux droites se coupent toujours en un point, et il n'y a qu'une seule sorte de conique.

Le choix de la droite de l'infini étant mathématiquement arbitraire, on dénomme ainsi n'importe quelle droite. C'est ce qui est fait dans les FIG.1.1 et 1.2. On a choisi une droite comme droite de l'infini, et on a indiqué ce que devenaient alors les notions classiques de droites parallèles ou non, et les trois espèces de coniques.

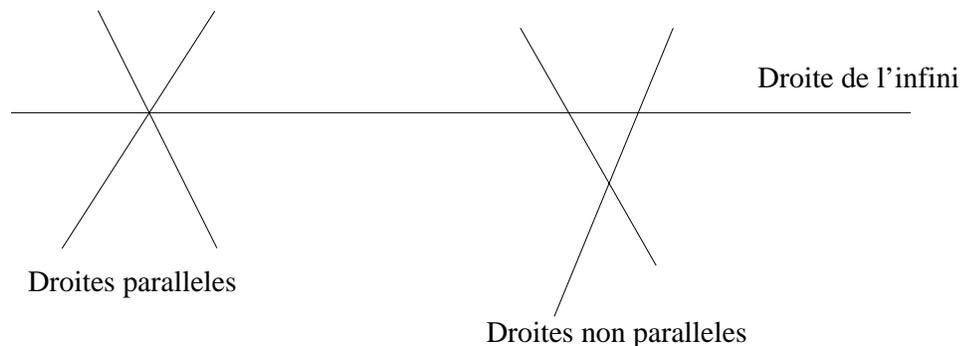


FIG. 1.1 - Le parallélisme pour la géométrie affine.

**Deuxième approche.** Considérons la droite projective. On peut l'envisager comme étant la droite affine à laquelle on a ajouté un point à l'infini. On peut aussi remarquer qu'elle est isomorphe à l'ensemble des droites vectorielles du plan. Considérons dans le

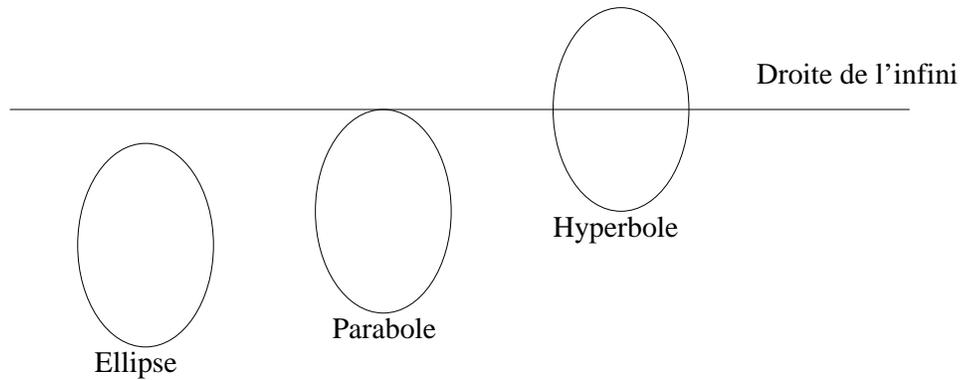


FIG. 1.2 - Les trois espèces de coniques pour la géométrie affine.

plan la droite  $D$ , d'équation  $y = 1$ . (cf. FIG. 1.3). Chaque droite vectorielle, ou chaque droite passant par l'origine, définit un point sur  $D$ . La droite  $Ox$  définit un seul point, dit point à l'infini, qui se trouve aussi bien à "l'infini positif" qu'à "l'infini négatif". Il y a donc une différence entre la droite projective, qui ne contient qu'un seul point à l'infini et la droite numérique achevée  $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$ , qui en contient deux. On peut faire une construction identique pour le plan projectif dans l'espace.

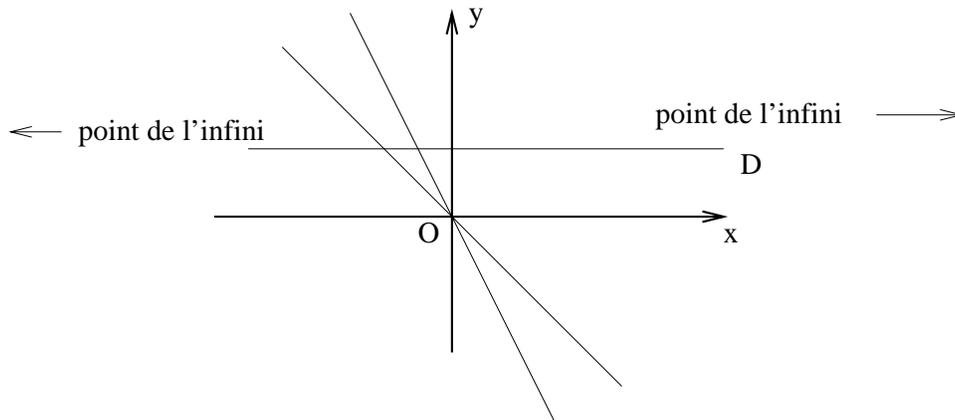


FIG. 1.3 - Définition de la droite projective.

**Troisième approche.** Le problème posé par les deux approches précédentes est celui du paramétrage. Pour la droite projective, on pourrait utiliser le symbole  $\infty$  pour désigner le point de l'infini, mais cela ne résoudrait pas le problème dans le plan ou un espace à 3 ou 4 dimensions. On définit plutôt un nouveau système de coordonnées, dites coordonnées homogènes. Considérons la droite. Dans une base affine, définie par deux points distincts  $A$  et  $B$ , chaque point  $C$  est repéré par un réel  $x$  qui vaut :

$$x = \frac{\overline{AC}}{\overline{AB}}$$

Dans un système de coordonnées homogènes, on va définir ce même point par tout couple de réels  $(u, v)$ , tel que  $v \neq 0$  et  $x = u/v$ . Ainsi, chaque point possède une infinité

de couples de coordonnées homogènes. Le point de l'infini aura lui pour coordonnées, tout couple  $(u, v)$ , tel que  $v = 0$  et  $u \neq 0$ . La droite projective est ainsi isomorphe au quotient de  $\mathbb{R}^{2*}$  par la relation de proportionnalité entre les couples de réels.

Dans le plan, la même construction est possible. Un point de coordonnées affines  $(x, y)$  admet comme coordonnées homogènes tout triplet non nul de réels  $(u, v, w)$ , tels que  $x = u/w$  et  $y = v/w$ . La droite de l'infini contient les points de coordonnées homogènes  $(u, v, 0)$ , où  $u$  et  $v$  ne sont pas simultanément nuls. Même à un facteur de proportionnalité près, on peut donner des coordonnées différentes à l'infini de points qui sont à l'infini.

**Lien avec l'approche mathématique.** Comme souvent, l'approche mathématique procède à l'inverse de l'intuition. Au lieu de définir d'abord les espaces projectifs, puis ensuite les coordonnées homogènes, elle commence par définir ce qu'est un paramétrage par des coordonnées homogènes, puis définit les espaces projectifs comme étant les espaces qui admettent un paramétrage par des coordonnées homogènes. Dans la suite, pour "faire son intuition", on pourra rapporter les définitions qui suivent à un des exemples précédents, soit au premier exemple, fondamental, donné dans ce qui suit.

### Approche mathématique

Commençons par définir la notion de représentation d'un ensemble, c'est à dire de paramétrisation de cet ensemble par des coordonnées homogènes. On associe à tout élément d'un ensemble tous les  $n + 1$ -uplets qui sont proportionnels à un  $n + 1$ -uplet donné non nul. Ces  $n + 1$ -uplets sont appelés coordonnées homogènes de l'élément.

**Définition 1** On dit qu'un ensemble  $E$  d'éléments quelconques admet une **représentation**  $n$ -dimensionnelle par des coordonnées homogènes appartenant au corps  $\mathbb{K}$ , s'il existe une application surjective de  $\mathbb{K}^{n+1*}$  dans  $E$  telle que deux vecteurs de  $\mathbb{K}^{n+1*}$  aient la même image si et seulement s'ils sont proportionnels. Un antécédent d'un point de  $E$  est appelé vecteur de **coordonnées homogènes** de ce point. Ces coordonnées sont donc définies à un facteur de proportionnalité près.

Un espace projectif est alors un ensemble admettant des paramétrisations par coordonnées homogènes se déduisant les unes des autres par des transformations linéaires.

**Définition 2** Un ensemble  $E$  est un **espace projectif** de dimension  $n$  sur le corps  $\mathbb{K}$  s'il admet une classe  $(\mathcal{R})$  de représentations admissibles  $n$ -dimensionnelles par des coordonnées homogènes de  $\mathbb{K}$ , telle que, si  $\mathcal{R}_0$  est une de ces représentations admissibles, la classe  $(\mathcal{R})$  contient toutes les représentations qui se déduisent de  $\mathcal{R}_0$  par une transformation linéaire non singulière :

$$x'_i = \sum_{k=1}^{n+1} a_{ik} x_k \quad (i = 1, \dots, n + 1)$$

**Notation** Tous ces espaces étant isomorphes, on les note  $\mathbb{P}^n(\mathbb{K})$ , ou  $\mathbb{P}^n$  s'il n'y a pas d'ambiguïté sur le corps  $\mathbb{K}$ . Les éléments d'un tel ensemble sont généralement appelés points.

**Premier exemple** Soit  $\sim$  la relation d'équivalence sur  $\mathbb{K}^{n+1*}$  suivante :

$$(x_1, \dots, x_{n+1}) \sim (x'_1, \dots, x'_{n+1}) \text{ ssi } \exists \lambda \in \mathbb{K}^* \quad (x_1, \dots, x_{n+1}) = \lambda(x'_1, \dots, x'_{n+1})$$

Alors l'ensemble quotient  $\mathbb{K}^{n+1*} / \sim$  est un espace projectif de dimension  $n$  sur  $\mathbb{K}$ . Lorsqu'on parle d'espace projectif sans autre précision, c'est généralement à cet espace que l'on fait allusion.

Il existe une injection de  $\mathbb{K}^n$  dans  $\mathbb{P}^n(\mathbb{K})$ . A chaque  $n$ -uplet  $(x_1, \dots, x_n)$  de  $\mathbb{K}^n$ , on fait correspondre le  $(n+1)$ -uplet  $(x_1, \dots, x_n, 1)$ . Cette injection n'est pas surjective : elle n'atteint pas les éléments de  $\mathbb{P}^n(\mathbb{K})$  tels que  $x_{n+1} = 0$ . Ces éléments sont appelés éléments de l'infini. Cette injection n'étant pas unique, l'équation de l'ensemble des éléments de l'infini peut donc varier.

Dans la définition ci-dessus, les représentations admissibles  $\mathcal{R}$  de  $\mathbb{P}^n(\mathbb{K})$  sont reliées par un groupe de transformations linéaires non singulières. Ce groupe est appelé le **groupe projectif** de dimension  $n$  sur  $\mathbb{K}$ . Il est noté  $GP(n, \mathbb{K})$ . La géométrie projective est l'étude des invariants de ce groupe. Dans toute la suite, et sauf mention explicite du contraire, on prendra  $\mathbb{C}$  comme corps de base.

### 1.1.3 Dépendance linéaire, sous-espaces, dualité

Ayant défini les espaces projectifs, la notion primordiale de la géométrie projective est celle d'indépendance linéaire, qui permet de définir droites, plans et hyperplans. De manière intuitive, ces ensembles sont identiques à ce qu'il sont dans les espaces affines, à condition de leur ajouter leurs points à l'infini. Leurs équations, qui sont linéaires dans les coordonnées affines, le sont aussi en coordonnées homogènes. Il n'y a donc pas de concept nouveau dans ce paragraphe, mais plutôt une actualisation du vocabulaire à la géométrie projective.

Le principe de dualité est toutefois digne d'intérêt. Il stipule simplement que l'ensemble des hyperplans d'un espace projectif est aussi un espace projectif. On peut donc manier les hyperplans comme les points : tout ce qui est valable pour les uns l'est pour les autres, il suffit d'adapter le vocabulaire.

**Définition 3** *Un ensemble de points  $P_1, \dots, P_m$  de  $\mathbb{P}^n$  est **linéairement dépendant** pour une représentation admissible  $\mathcal{R}$ , si les coordonnées homogènes de ces points  $x^{(1)}, \dots, x^{(m)}$  sont liées par une relation linéaire :*

$$\exists (\lambda_1, \dots, \lambda_m) \in \mathbb{C}^{m*} \quad \sum_{k=1}^m \lambda_k x^{(k)} = 0$$

*Dans la cas contraire, les points sont **linéairement indépendants**.*

On peut démontrer que cette définition ne dépend pas de la représentation  $\mathcal{R}$  choisie : l'indépendance linéaire est donc une notion projective. Avec cette notion, on peut définir les bases projectives.

**Théorème 1** *Dans  $\mathbb{P}^n$ , soient  $P_1, \dots, P_{n+2}$   $n+2$  points tels que  $n+1$  de ces points ne soient pas linéairement dépendants. Si les coordonnées homogènes de ces points sont fixées, la représentation  $\mathcal{R}$  est entièrement déterminée. Les coordonnées de ces points peuvent être fixées arbitrairement, pour autant que la condition d'indépendance linéaire est vérifiée. Un tel ensemble de point forme une **base projective** de  $\mathbb{P}^n$ .*

La base canonique de  $\mathbb{P}^n$  est formée des points dont les coordonnées homogènes sont  $(1,0,\dots,0)$ ,  $(0,1,0,\dots,0)$ ,  $\dots$ ,  $(0,\dots,0,1)$  et  $(1,\dots,1)$ . Les  $n+1$  premiers points de cette base sont appelés **points de référence**, et le dernier **point unité**. La donnée des points de référence et du point unité est la manière la plus naturelle de définir une représentation  $\mathcal{R}$ .

Un changement de base est équivalent à un changement de représentation. Les coordonnées d'un point dans une des bases se déduisent donc de ses coordonnées dans l'autre base par une transformation du groupe projectif qui ne dépend que des bases. Cette transformation peut être caractérisée par une matrice  $(n+1) \times (n+1)$  de déterminant non nul et définie à un facteur de proportionnalité près. Comme annoncé précédemment, l'indépendance linéaire permet de définir les sous espaces de l'espace projectif.

**Théorème 2** Soient, dans  $\mathbb{P}^n$ ,  $m$  points linéairement indépendants, avec  $m \leq n$ . Alors l'ensemble des points qui sont linéairement dépendants de ces  $m$  points forment un espace projectif de dimension  $m-1$ . Quand cette dimension vaut respectivement 1, 2 et  $n-1$ , cet espace est généralement appelé **droite**, **plan** et **hyperplan**. L'ensemble des sous espaces de  $\mathbb{P}^n$  de même dimension est aussi un espace projectif.

L'ensemble des hyperplans d'un espace projectif est un espace projectif de même dimension. Toute définition, propriété ou théorème pour les points d'un espace projectif est donc aussi valable pour ses hyperplans. La seule différence est liée au vocabulaire employé. Ainsi points et droites de  $\mathbb{P}^2$  d'une part, et points et plans de  $\mathbb{P}^3$  d'autre part ont les mêmes propriétés. Ceci est appelé **principe de dualité**.

Un exemple classique d'emploi de la dualité est le théorème de Desargues. La réciproque de ce théorème est en effet sa forme duale. En le démontrant, on démontre du même coup sa réciproque, par dualité. Voici ce théorème et sa réciproque, les expressions duales étant placées sur une même ligne.

### Théorème 3 (Desargues)

<i>Théorème de Desargues</i>	<i>Réciproque du théorème</i>
<i>Si deux triangles sont tels que les droites joignant leurs sommets correspondants sont concourantes, alors les points d'intersection de leurs cotés correspondants sont colinéaires.</i>	<i>Si deux triangles sont tels que les points d'intersection de leurs côtés correspondants sont colinéaires, alors les droites joignant leurs sommets correspondants sont concourantes.</i>

Examinons quelques cas particuliers de sous espaces.

**Droites dans  $\mathbb{P}^2$ .** Elles sont définies par 2 points linéairement indépendants. Soient  $Q$  et  $R$  deux points et respectivement  $y$  et  $z$  leur vecteur de coordonnées homogènes. Alors leurs coordonnées vérifient une et une seule relation :  $u_1x_1 + u_2x_2 + u_3x_3 = 0$ , où  $(u_1, u_2, u_3)$  est un triplet de scalaires définis à un facteur de proportionnalité près. L'ensemble de tous ces triplets forment une représentation de l'espace projectif des droites de  $\mathbb{P}^2$ . Un point  $P$  de coordonnées  $x$  appartient à cette droite

si ses coordonnées vérifient la même équation, ou, de manière équivalente :

$$\begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix} = 0$$

Il existe, d'autre part, deux scalaires  $\lambda$  et  $\mu$  dont le rapport est déterminé de manière unique, et tels que  $x = \lambda y + \mu z$ . Les couples  $(\lambda, \mu)$ , pour tous les points de la droite, forment une représentation de l'espace projectif que constitue cette droite.

Dans le vocabulaire de l'espace des points, on dit que deux points définissent une droite, que trois points sont colinéaires, qu'une droite est la réunion de tous les points colinéaires à deux points donnés. Dans le vocabulaire de l'espace des droites, on dit, de manière duale, que deux droites définissent un point, que trois droites sont concourantes, qu'un point est l'intersection de toutes les droites concourantes avec deux droites données. Le vocabulaire change, mais les concepts sont les mêmes.

**Plans dans  $\mathbb{P}^3$ .** Ils sont définis de manière analogue par trois points  $Q$ ,  $R$ , et  $S$  de coordonnées  $y$ ,  $z$  et  $t$ . Celles-ci vérifient une équation linéaire  $u_1x_1 + u_2x_2 + u_3x_3 + u_4x_4 = 0$ , où  $(u_1, u_2, u_3, u_4)$  est un vecteur de coordonnées homogènes du plan.

Un point  $P$  de coordonnées  $x$  appartient à ce plan si ses coordonnées vérifient la même équation, ou bien si  $\det(x, y, z, t) = 0$ , ou s'il existe trois scalaires  $\lambda$ ,  $\mu$ ,  $\nu$  tels que  $x = \lambda y + \mu z + \nu t$ .

**Droites dans  $\mathbb{P}^3$ .** Elles sont définies par deux points et forment un espace de dimension 4. Elles n'ont pas de représentation simple par 5 coordonnées homogènes. On peut par contre les caractériser par un vecteur de 6 coordonnées homogènes vérifiant une relation de compatibilité. Voir [Sem 52, chapitre XV].

#### 1.1.4 Birapport, éléments conjugués

Après celle d'indépendance linéaire, la deuxième notion fondamentale est celle de birapport. C'est l'invariant numérique de base, canonique même, de la géométrie projective, l'équivalent projectif du rapport affine ou de la distance euclidienne. Ce birapport est la mesure la plus simple qui reste invariante par changement de coordonnées homogènes. Après une brève présentation, quelques propriétés simples en sont rappelées, dont la conjugaison.

**Définition 4** On appelle **birapport** la fonction rationnelle suivante :

$$\mathbb{C}^4 \longrightarrow \mathbb{C}$$

$$(x_1, x_2, x_3, x_4) \longmapsto \{x_1, x_2; x_3, x_4\} = \frac{x_1 - x_3}{x_2 - x_3} \bigg/ \frac{x_1 - x_4}{x_2 - x_4}$$

Cette fonction peut être formellement étendue avec les conventions suivantes :

$$\frac{a}{\infty} = 0, \frac{0}{0} = \frac{\infty}{\infty} = 1, \frac{a}{0} = \infty, a \cdot \infty = \infty \text{ avec } a \in \mathbb{C}^*$$

Avec ces conventions, on obtient :

$$\begin{aligned} \{\infty, x_2; x_3, x_4\} &= \frac{x_2 - x_4}{x_2 - x_3} ; \{x_1, \infty; x_3, x_4\} = \frac{x_1 - x_3}{x_1 - x_4} ; \\ \{x_1, x_2; \infty, x_4\} &= \frac{x_2 - x_4}{x_1 - x_4} ; \{x_1, x_2; x_3, \infty\} = \frac{x_1 - x_3}{x_2 - x_3} \end{aligned}$$

Par permutation de l'ordre des scalaires, on peut former 24 quadruplets. Cela ne définit toutefois que 6 valeurs différentes du birapport. Si  $\rho$  est une de ces valeurs, les cinq autres sont :

$$\frac{1}{\rho}, 1 - \rho, \frac{1}{1 - \rho}, \frac{\rho}{\rho - 1}, \frac{\rho - 1}{\rho}$$

Il existe trois cas particuliers où ces six valeurs ne sont pas distinctes :

1. si deux des quatre valeurs sont confondues, les 6 birapports obtenus valent :  $1, 1, 0, 0, \infty, \infty$ .
2. dans un cas particulier important appelé **conjugaison harmonique**, les birapports obtenus sont :  $-1, -1, 0.5, 0.5, 2, 2$  ;
3. dans le dernier cas, appelé quadruplet équi-anharmonique, les six birapports sont  $-\omega, -\omega, -\omega, -\omega^2, -\omega^2, -\omega^2$ , où  $\omega = e^{2i\pi/3}$ .

On peut alors définir le birapport de divers éléments géométriques.

- Quatre points d'une droite projective définissent un birapport. Soient  $(P_i)_{i \in \{1,2,3,4\}}$  quatre points de coordonnées homogènes respectives  $(\lambda_i, \mu_i)$ . On définit  $\theta_i = \lambda_i/\mu_i$ , ce nombre valant éventuellement  $\infty$  si  $\mu_i$  est nul. Alors, la quantité :

$$\{P_1, P_2; P_3, P_4\} = \frac{(\theta_1 - \theta_3)(\theta_2 - \theta_4)}{(\theta_2 - \theta_3)(\theta_1 - \theta_4)}$$

est indépendante de la représentation de la droite choisie. On l'appelle birapport des quatre points.

- Un faisceau de droites est un ensemble de droites concourantes. Dans  $\mathbb{P}^2$ , un tel faisceau de quatre droites se coupant en un point  $P_0$  définit un birapport (cf. FIG. 1.4). En effet, le birapport des quatre points d'intersection de ces droites avec une droite quelconque ne passant pas par  $P_0$  ne dépend pas de cette dernière droite. On l'appelle birapport des quatre droites du faisceau. Lorsque ces droites sont définies par leur point d'intersection  $P_0$  et par un point quelconque de chacune des droites  $P_1, P_2, P_3, P_4$ , ces points devant être différents de  $P_0$  mais pouvant ne pas être alignés, on peut utiliser la formule de Möbius [Möb 85] pour calculer le birapport :

$$\{P_0; P_1, P_2, P_3, P_4\} = \frac{|P_0 P_1 P_3| |P_0 P_2 P_4|}{|P_0 P_1 P_4| |P_0 P_2 P_3|} \quad (1.1)$$

où  $|ABC|$  est le déterminant  $3 \times 3$  des coordonnées des points  $A, B$  et  $C$  :

$$|ABC| = \begin{vmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ t_A & t_B & t_C \end{vmatrix}$$

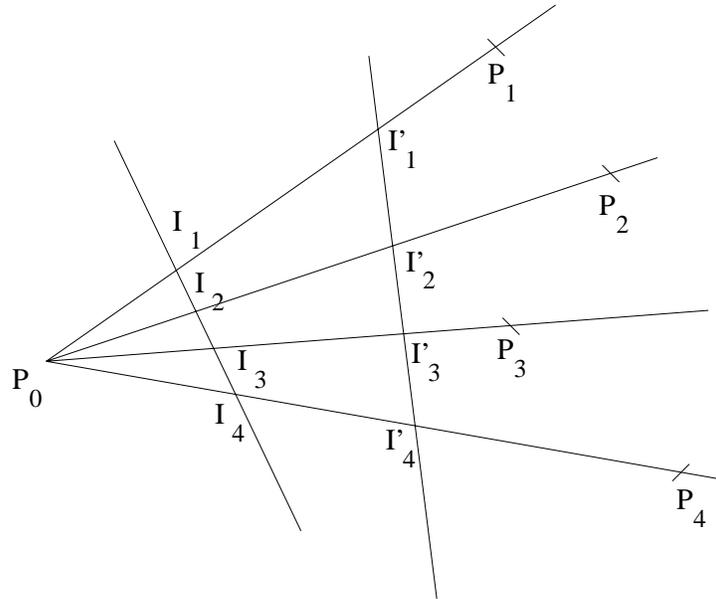


FIG. 1.4 - Définition du birapport de 4 droites.

- De la même manière, un faisceau de plans est un ensemble de plans ayant une droite en commun. Un faisceau de quatre plans dans  $\mathbb{P}^3$  définit un birapport, les quatre droites d'intersection de ces plans avec un plan n'appartenant pas au faisceau ayant un birapport ne dépendant pas de ce dernier plan (cf. FIG. 1.5).

La notion de birapport, ainsi que son calcul par la formule de Möbius, se généralisent pour tout faisceau de quatre hyperplans ayant en commun un sous espace de dimension  $n - 1$  [Bri 83].

- On peut, à partir de ces premières configurations ayant un birapport, en définir de nouvelles. Par exemple, 5 points du plan  $(P_i)_{i \in \{1, \dots, 5\}}$  dont trois quelconques d'entre eux ne sont pas alignés définissent deux faisceaux indépendants de quatre droites. Par exemple,  $(P_1P_2)$ ,  $(P_1P_3)$ ,  $(P_1P_4)$ ,  $(P_1P_5)$  et  $(P_2P_1)$ ,  $(P_2P_3)$ ,  $(P_2P_4)$ ,  $(P_2P_5)$  (cf. FIG. 1.6). Les birapports de tels faisceaux sont appelés **coordonnées projectives** du point  $P_5$  dans la base formée des quatre autres points. Si cette base est la base canonique de l'espace projectif, les coordonnées projectives sont reliées de manière très simple aux coordonnées homogènes du point [Mor 93].
- De manière duale, on peut définir le birapport de 5 droites coplanaires. Quatre de ces droites coupent la cinquième en quatre points qui définissent un birapport. On construit ainsi 5 birapports dont 2 sont indépendants (cf. FIG. 1.7). Ces birapports peuvent être calculés en utilisant les coordonnées homogènes des droites et la formule de Möbius (1.1).

La relation de conjugaison harmonique joue un rôle important en géométrie projective. Elle permet de définir de nombreux éléments. Rappelons que 4 éléments sont en relation harmonique si leur birapport vaut -1. Les deux premiers éléments sont dits conjugués harmoniques par rapport aux deux derniers, et vice versa.

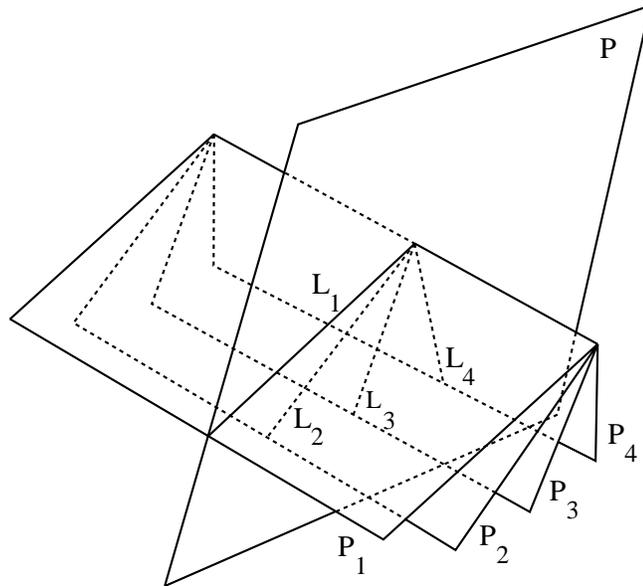


FIG. 1.5 - Définition du birapport de 4 plans.

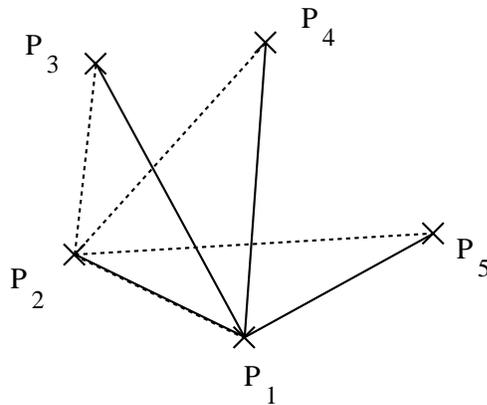


FIG. 1.6 - Birapports définis par 5 points coplanaires.

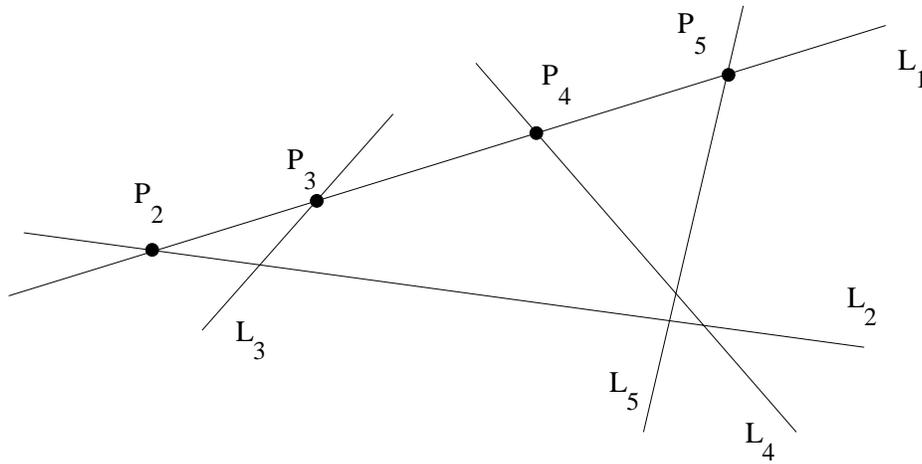


FIG. 1.7 - Birapport défini par 5 droites coplanaires.

### 1.1.5 Transformations projectives

La notion de transformation projective n'a pas la place centrale qu'on pourrait penser lui accorder. Cela vient du fait qu'une transformation projective d'un espace projectif dans lui-même, une homographie, est équivalente à un changement de représentation, notion qui a été introduite au début de ce paragraphe. Cette notion de transformation garde toutefois son intérêt dans le cas où on considère deux espaces différents, et particulièrement dans le cas de deux espaces de dimensions différentes. C'est le cas auquel on s'intéresse pour la vision, où l'opération de base est une projection de  $\mathbb{P}^3$  dans  $\mathbb{P}^2$ .

D'un point de vue géométrique, et comme cela sera détaillé plus loin, ce cas n'est pas très riche, car le seul invariant général existant est la colinéarité. Il faudra alors se ramener aux transformations de  $\mathbb{P}^2$  dans  $\mathbb{P}^2$  (observations de structures planes), ou passer aux transformations de  $\mathbb{P}^3$  dans  $\mathbb{P}^2 \times \mathbb{P}^2$  (observation d'une même structure dans deux images) qui fourniront un cadre beaucoup plus intéressant.

Ce paragraphe donne quelques définitions (collinéations, homographies et projections), et indique les liens entre ces différents objets.

**Définition 5** Soient deux entiers  $m$  et  $n$ . On appelle **transformation projective** de  $\mathbb{P}^n$  dans  $\mathbb{P}^m$ , ou **collinéation**, toute transformation linéaire pour les coordonnées homogènes. Cette définition ne dépend bien sûr pas des représentations choisies dans chacun des deux espaces.

Une telle transformation est donc définie par une matrice de taille  $(m+1) \times (n+1)$  déterminée de manière unique à un facteur de proportionnalité près. Rappelons le théorème fondamental démontré par Möbius dans le cas des espaces réels.

**Théorème 4 (Möbius)** Les transformations projectives sont celles qui transforment les points en points et les droites en droites. De telles transformations conservent le birapport.

Réciproquement, les transformations qui laissent le birapport invariant sont des transformations projectives. Dans les espaces complexes, certaines transformations faisant intervenir la conjugaison  $z \mapsto \bar{z}$ , transforment les points en points et les droites en droites, mais ne sont pas projectives.

D'une manière générale, on appellera **collinéation** une transformation projective générale, **homographie** une collinéation d'un espace dans lui-même, et **projection** les collinéations d'un espace dans un de ses sous-espaces. Les homographies d'un espace projectif forment un groupe non commutatif isomorphe au groupe projectif  $GP(n, \mathbb{C})$  et que nous confondrons donc avec celui-ci. Effectuer une homographie revient à faire un changement de représentation.

**Caractérisation d'une collinéation.** Une homographie étant équivalente à un changement de base, elle est déterminée de manière unique par la donnée des images des points d'une base. Dans le cas d'une collinéation quelconque, la donnée de l'image d'une base n'est pas nécessairement suffisante pour caractériser complètement cette collinéation.

Par exemple, dans le cas des projections de  $\mathbb{P}^3$  dans  $\mathbb{P}^2$ , l'ensemble des transformations forment un espace de dimension 11 (ce sont des matrices de taille  $3 \times 4$  définies à un facteur de proportionnalité près). Or, la donnée des images des points d'une base ne fournit que  $5 \times 2 = 10$  contraintes. Il faut donc au moins 6 points pour déterminer une telle projection.

**Collinéations de  $\mathbb{P}^3$  dans  $\mathbb{P}^2$  et projections.** Le fait d'appeler projection toute collinéation d'un espace projectif dans un autre de dimension inférieure n'est-il pas abusif? Pour répondre à cette légitime question, il faut étudier le lien entre ces collinéations, et les projections définies classiquement, par un centre de projection et un plan de projection. Rappelons donc les résultats démontrés dans [Fau 92b, Mor 93].

**Théorème 5** *Soit une collinéation de  $\mathbb{P}^3$  dans un de ses sous-espaces de dimension 2, représentée par sa matrice  $P$  de taille  $3 \times 4$  et soit  $A$  la matrice carrée formée des trois premières colonnes de  $P$ . On suppose l'espace muni d'un repère affine (la définition des notions affines sera faite au paragraphe 1.1.7).*

- Si le rang de  $A$  est 3,  $P$  représente une projection perspective, c'est à dire une projection centrale dont le centre de projection n'est pas à l'infini, et dont le plan de projection n'est pas l'infini et ne contient pas le centre de projection.
- Si  $A$  est de rang 2, alors  $P$  représente la composée d'une projection affine, c'est à dire une projection centrale dont le centre de projection est à l'infini, avec une homographie du plan image. Si, de plus, la dernière ligne de  $A$  est nulle, cette homographie est une transformation affine et  $P$  est une projection affine.
- Si  $A$  est de rang 1,  $P$  est une projection sur une droite.

Dans tous les cas, les collinéations sont donc des projections, suivies éventuellement d'un changement de représentation du plan, ce qui justifie l'emploi du mot projection.

### 1.1.6 Coniques

Avec les structures linéaires, droites, plans et hyperplans, les coniques sont des objets naturels de la géométrie projective. Tout d'abord, les coniques sont des objets projectifs : une collinéation ou un changement de représentation transforme une conique en une conique. D'autre part, et contrairement à ce qu'il se passe en géométrie affine, il y a une seule espèce de conique, et leur étude évite ainsi la multiplication des cas et des exceptions.

Le fait de travailler dans l'espace projectif complexe est aussi important.  $\mathbb{C}$  étant algébriquement clos, contrairement à  $\mathbb{R}$ , la géométrie projective complexe permet des énoncés très généraux d'une grande simplicité : deux coniques non dégénérées d'un plan se coupent en quatre points. De tels énoncés sont inimaginables en géométrie affine réelle où les cas se multiplient. Cela dit, les coniques de la géométrie affines restent des coniques en géométrie projective, c'est plutôt l'intuition qu'on s'en fait qu'il faut modifier (cf. FIG. 1.2). Les coniques sont dans les deux cas les courbes d'équations 2-homogènes, mais leurs propriétés diffèrent selon la géométrie dans laquelle on se place.

Après avoir donné leur définition, ce chapitre donne quelques notions sur les éléments conjugués par rapport à une conique. Ceux-ci ne sont pas fondamentaux pour la suite, et peuvent donc être ignorés pour une lecture rapide.

#### Définition, birapports

**Définition 6** On appelle *conique* du plan  $\mathbb{P}^2$  tout ensemble de points vérifiant une équation 2-homogène pour les coordonnées homogènes :

$$\sum_{i=1}^3 \sum_{j=1}^3 a_{ij} x_i x_j = 0 \quad (1.2)$$

Cette définition est indépendante de la représentation du plan choisie : la notion de conique est une notion projective. En symétrisant l'équation (1.2), on voit qu'une conique est caractérisée par une matrice  $3 \times 3$  symétrique définie à un facteur de proportionnalité près :

$$(x_1, x_2, x_3) = \begin{pmatrix} a_{11} & \frac{a_{12}+a_{21}}{2} & \frac{a_{13}+a_{31}}{2} \\ \frac{a_{21}+a_{12}}{2} & a_{22} & \frac{a_{23}+a_{32}}{2} \\ \frac{a_{31}+a_{13}}{2} & \frac{a_{32}+a_{23}}{2} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

Quand on associe une matrice à une conique, on choisit une telle matrice symétrique qui est unique à un facteur de proportionnalité près. Les coniques forment un espace projectif de dimension 5, les coefficients de cette matrice fournissant un système de coordonnées projectives pour cet espace. Dans le vocabulaire dual de l'espace des hyperplans (cf. Th. 2 page 30), on parle d'enveloppe de seconde classe.

Les coniques permettent de définir des birapports de plusieurs manières.

- Le birapport du faisceau des quatre droites qui joignent un point quelconque  $a$  d'une conique à quatre points fixes de cette conique  $\theta_1, \theta_2, \theta_3$  et  $\theta_4$  est indépendant du point  $a$ . Il est appelé birapport des quatre points de la conique (Théorème de Chasles). Réciproquement, étant donnés quatre points, l'ensemble des points sommets d'un faisceau de droites passant par ces quatre points et dont

le birapport est donné est une conique, appelée conique de Chasles (cf. FIG. 1.8).

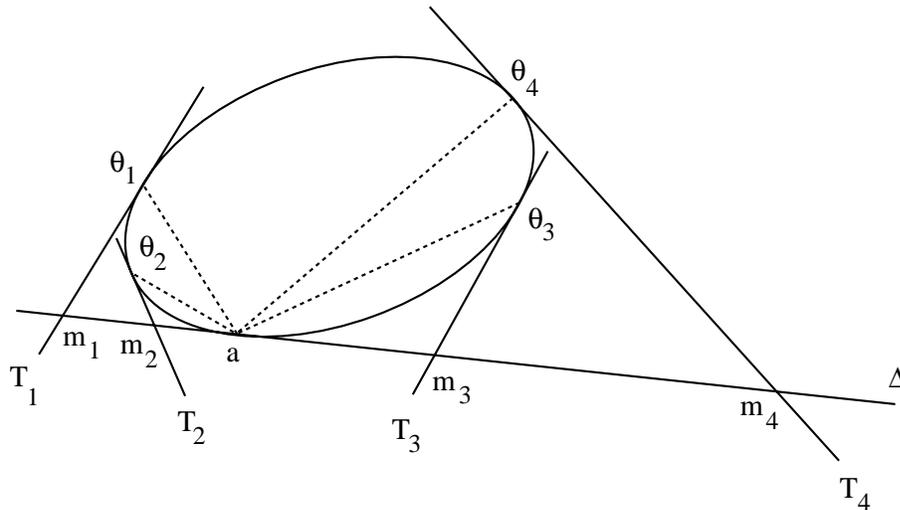


FIG. 1.8 - Définition de birapports à partir d'une conique.

- Soient  $T_1, T_2, T_3, T_4$  quatre tangentes quelconques à une conique. Le birapport des points d'intersection de ces droites avec une tangente quelconque  $\Delta$  à la conique ne dépend pas de cette dernière tangente et est appelé birapport des quatre tangentes. Cette construction est la duale de la précédente.

### Éléments conjugués par rapport à une conique

La notion de conjugaison harmonique peut être utilisée de manière féconde avec les coniques. Elle permet par exemple d'interpréter de manière géométrique certains invariants algébriques proposés dans la littérature [For 91b], ce qui sera fait au paragraphe 1.2.9.

**Définition 7** Deux points  $P$  et  $Q$  sont **conjugués par rapport à une conique  $\mathcal{C}$**  s'ils sont conjugués par rapport aux deux points  $M_1$  et  $M_2$  d'intersection de la droite  $(PQ)$  avec la conique  $\mathcal{C}$ . On a alors :  $\{P, Q, M_1, M_2\} = -1$ . Si on note  $\mathcal{M}_{\mathcal{C}}$  la matrice symétrique associée à la conique  $\mathcal{C}$ , et  $X$  et  $Y$  les vecteurs de coordonnées homogènes des points  $P$  et  $Q$ , on a :  ${}^t X \mathcal{M}_{\mathcal{C}} Y = 0$ . Cette dernière formule peut être prise comme définition de la conjugaison [Sem 52, page 108].

**Définition 8** Le lieu des points conjugués d'un point donné  $P$  par rapport à une conique est une droite appelée **polaire** de ce point.

Avec les notations de la définition précédente, si  $l$  est un vecteur de coordonnées homogènes de cette polaire, on a :  $\lambda l = \mathcal{M}_{\mathcal{C}} X$ , où  $\lambda$  est un facteur de proportionnalité complexe.

**Définition 9** Réciproquement, toutes les polaires des points d'une droite sont des droites qui concourent en un point fixe, appelé **pôle** de cette droite.

En géométrie affine, lorsque le point  $P$  n'est pas dans la concavité de la conique, on peut faire une construction simple de sa polaire (cf. FIG. 1.9). Il existe alors deux droites réelles passant par  $P$  et tangentes à la conique  $\mathcal{C}$  en deux points  $P_1$  et  $P_2$ . Ces deux points sont conjugués avec  $P$  par rapport à la conique ; ils appartiennent à la polaire de  $P$  qui est donc la droite  $(P_1P_2)$ . Lorsque le point est à l'intérieur de la concavité de la conique, la même construction est possible, mais avec des droites et des points imaginaires.

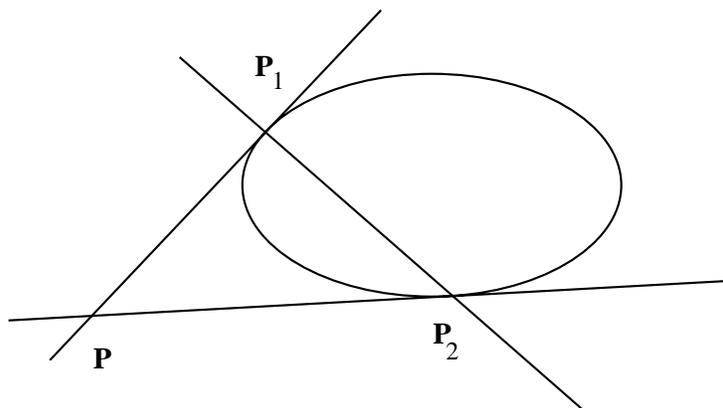


FIG. 1.9 - Construction de la polaire d'un point.

### Etude des triangles auto-polaires

Toujours dans le but d'interpréter géométriquement certains invariants algébriques, introduisons la notion de triangle auto-polaire.

**Définition 10** Trois points  $P, Q, R$  forment un **triangle auto-polaire** par rapport à une conique, si la polaire de chacun de ces points est la droite définie par les deux autres.

**Constructions des triangles auto-polaire d'une conique.** Soit une conique  $\mathcal{C}$  de matrice  $\mathcal{M}_C$  et soient quatre points  $A, B, C$  et  $D$  de cette conique. On construit alors les trois points suivants :  $P = (AB) \cap (CD)$ ,  $Q = (AC) \cap (BD)$  et  $R = (AD) \cap (BC)$  (cf. FIG. 1.10). Le triangle  $(P, Q, R)$  est auto-polaire.

**Démonstration** Pour démontrer l'auto-polarité de ce triangle, il faut remarquer au préalable que la polaire d'un point par rapport à une conique réduite à deux droites passe par l'intersection de ces deux droites. Soient  $E$  et  $F$  les points tels que  $P, E$  et  $P, F$  sont les conjugués harmoniques de  $C, D$  et  $A, B$  respectivement.  $E$  et  $F$  appartiennent d'une part à la polaire de  $P$  par rapport aux deux droites  $(AC)$  et  $(BD)$  qui se coupent en  $Q$ , et d'autre part à la polaire de  $P$  par rapport aux deux droites  $(AD)$  et  $(BC)$  qui se coupent en  $R$ . Les points  $E, F, Q$  et  $R$  sont donc colinéaires et  $Q$  et  $R$  appartiennent à  $(EF)$ , la polaire de  $P$  par rapport à la conique  $\mathcal{C}$ . Des constructions identiques avec les deux autres sommets mènent au résultat.

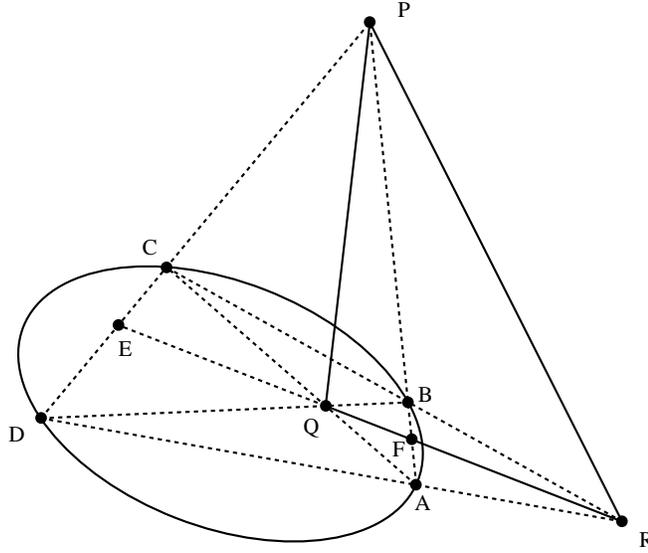


FIG. 1.10 - Construction d'un triangle auto-polaire.

**Equation d'une conique par rapport à un triangle auto-polaire.** Prenons comme points de référence de la représentation du plan, les trois sommets  $P$ ,  $Q$  et  $R$  d'un triangle auto-polaire par rapport à une conique  $\mathcal{C}$  de matrice  $\mathcal{M}_C$ .  $P$ ,  $Q$  et  $R$  ont respectivement comme coordonnées homogènes  $X_P = (1, 0, 0)$ ,  $X_Q = (0, 1, 0)$  et  $X_R = (0, 0, 1)$ , qui sont aussi les coordonnées homogènes  $l_{(QR)}$ ,  $l_{(PR)}$  et  $l_{(PQ)}$  des droites  $(QR)$ ,  $(PR)$  et  $(PQ)$ . On peut alors exprimer l'auto-polarité du triangle :

$$\exists(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{C}^3 \quad \begin{cases} \lambda_1 l_{(QR)} = \mathcal{M}_C X_P \\ \lambda_2 l_{(PR)} = \mathcal{M}_C X_Q \\ \lambda_3 l_{(PQ)} = \mathcal{M}_C X_R \end{cases} \quad (1.3)$$

La matrice  $\mathcal{M}_C$  étant symétrique, elle est diagonale, et l'équation de la conique est donc, dans cette représentation :

$$ax_1^2 + bx_2^2 + cx_3^2 = 0 \quad (1.4)$$

Réciproquement, si une conique  $\mathcal{C}$  a une équation de la forme (1.4) dans une base projective  $P$ ,  $Q$ ,  $R$ ,  $S$  où  $S$  est le point unité, alors le système (1.3) est vérifié et le triangle  $(P, Q, R)$  est auto-polaire par rapport à la conique  $\mathcal{C}$ .

**Vecteurs propres de la matrice  $\mathcal{M}_C$  d'une conique  $\mathcal{C}$  non dégénérée.** La matrice  $\mathcal{M}_C$  associée à une conique  $\mathcal{C}$  peut être choisie symétrique, sans restriction. Lorsque la conique n'est pas dégénérée, cette matrice est définie. Elle possède alors trois vecteurs propres orthogonaux  $X_P$ ,  $X_Q$  et  $X_R$  qui sont les vecteurs de coordonnées homogènes de trois points  $P$ ,  $Q$  et  $R$ .

La polaire de  $P$  par rapport à  $\mathcal{C}$  admet pour coordonnées homogènes :  $l = \mathcal{M}_C X_P$ . Puisque  $X_P$  est vecteur propre de  $\mathcal{M}_C$ , on a :  $l = \lambda X_P$ . Compte tenu de l'orthogonalité, on a  $X_P.X_Q = 0$  et  $X_P.X_R = 0$ , donc  $l.X_Q = 0$  et  $l.X_R = 0$  et  $Q$  et  $R$  sont sur cette

polaire. On raisonne de même avec les polaires de  $Q$  et de  $R$ . Les vecteurs propres de la matrice  $\mathcal{M}_C$  forment donc un triangle auto-polaire par rapport à la conique  $\mathcal{C}$ .

### 1.1.7 Géométrie affine

#### Introduction

La géométrie projective est un prolongement de la géométrie affine sous deux aspects. Tout d'abord, l'espace projectif contient un hyperplan de plus, celui des points à l'infini. Ensuite, l'ensemble des transformations est plus grand. De ces deux aspects, c'est le second qui est le plus important.

À l'inverse de cette démarche, on peut définir la géométrie affine comme une restriction de la géométrie projective. Au lieu de considérer toutes les transformations linéaires pour les coordonnées homogènes, on ne garde que des transformations en nombre plus restreint, celles qui laissent l'hyperplan de l'infini globalement invariant.

Les transformations affines étant moins nombreuses, les invariants affines sont plus riches. Par exemple, le parallélisme est un invariant. Quand deux droites du plan sont parallèles, elles se coupent sur la droite de l'infini (cf. FIG. 1.1) : on a ainsi trois droites concourantes. Elles sont transformées en trois nouvelles droites concourantes (par propriétés des homographies). La droite de l'infini étant invariante, une de ces nouvelles droites est celle de l'infini et les deux autres droites qui se coupent à l'infini sont donc parallèles.

Par ailleurs, toute droite coupe l'hyperplan de l'infini et possède un point canonique. Il ne suffit plus alors que de trois points pour définir un invariant, qui sera le birapport de ces points avec celui qui est à l'infini : on l'appelle le rapport des trois points. Ce rapport permet de définir le milieu de deux points. Toutes les propriétés affines peuvent être définies de même, à partir de la géométrie projective.

Pour une lecture rapide, on pourra omettre la fin de ce paragraphe, qui est une explicitation de ces mêmes idées sous une forme plus mathématique.

#### Approche mathématique

La géométrie projective étant présentée dans un cadre assez large, comme théorie des invariants pour le groupe projectif  $GP(n, \mathbb{K})$ , la géométrie affine apparaît simplement comme une restriction de la géométrie projective. Soit  $H_\infty$  un hyperplan quelconque de l'espace  $\mathbb{P}^n(\mathbb{K})$ , appelé hyperplan de l'infini. Les transformations de  $GP(n, \mathbb{K})$  qui laissent ce plan globalement invariant, c'est à dire qui ne changent pas l'équation de ce plan, forment un sous-groupe de  $GP(n, \mathbb{K})$  appelé **groupe affine**, et dénoté  $GA(n, \mathbb{K})$ . La géométrie affine est l'étude des invariants de ce groupe.

D'une manière générale, on peut distinguer dans  $\mathbb{P}^n(\mathbb{C})$ , des points dits réels car admettant un vecteur de coordonnées homogènes dont toutes les composantes sont réelles, et les points complexes qui n'admettent pas de telles coordonnées, même à un facteur de proportionnalité complexe près. Ces deux ensembles sont stables par les transformations de  $GP(n, \mathbb{K})$  qui ont des matrices réelles.

En géométrie affine, ces mêmes ensembles sont aussi stables, et on considère souvent uniquement la géométrie affine réelle, dans un espace réel, auquel peut être adjoint des points complexes. De même, les points de l'hyperplan de l'infini sont souvent considérés comme ne faisant pas partie proprement dit de l'espace affine, mais comme étant ajoutés

à cet espace. Dans la suite, nous considérerons la géométrie affine comme une restriction de la géométrie projective, mais en utilisant le même ensemble de points, complexes et réels, à distance finie ou infinie.

Une fois défini l'hyperplan de l'infini, les notions affines s'expriment facilement. Deux sous espaces **parallèles** sont deux droites qui se coupent à l'infini.

Trois points alignés  $A$ ,  $B$  et  $C$  définissent un unique point  $N$  aligné avec eux et appartenant à l'hyperplan de l'infini. Le rapport de ces trois points  $A$ ,  $B$  et  $C$  est le birapport de ces points auxquels on a adjoint  $N$ .

$$\{A, B, C, N\} = \{A, B, C, \infty\} = \frac{\overline{AC}}{\overline{BC}}$$

où  $\overline{AB}$  est une mesure algébrique. De même que le birapport de quatre points alignés est l'invariant fondamental de la géométrie projective, le rapport de 3 points alignés est celui de la géométrie affine.

Etant donnés deux points  $A$  et  $B$ , soit  $N$  le point d'intersection de la droite  $(AB)$  et de l'hyperplan de l'infini. Alors le conjugué de  $N$  par rapport à  $A$  et  $B$  est appelé le **milieu** de  $A$  et  $B$ . On a dans ce cas :  $\{A, B, C, N\} = -1$ , et  $\overline{AC} = \overline{CB}$ .

Dans le cas des transformations d'un espace  $E$  vers un espace  $F$  de dimension différente, les transformations affines sont celles qui envoient les points de l'infini de  $E$  dans celui de  $F$ , et ceux-là seulement.

La géométrie vectorielle peut être définie d'une manière analogue à la géométrie affine. On considère le groupe linéaire, sous groupe des transformations de  $GP(n, \mathbb{K})$  qui laissent l'hyperplan de l'infini et un point n'appartenant pas à cet hyperplan invariants globalement. On construit ainsi un espace affine (lorsqu'on définit l'hyperplan de l'infini), puis le vectorialisé de cet espace au point invariant [Ram 79, page 127].

### 1.1.8 Géométrie euclidienne

La géométrie euclidienne, peut se définir par restriction de la géométrie projective, d'une manière analogue à ce qui a été fait pour la géométrie affine. Nous appelons géométrie euclidienne la géométrie des similitudes et non celles des isométries. Il n'y a donc pas de différences entre les géométries affine et euclidienne en dimension 1.

De même que la géométrie affine est l'étude des transformations projectives qui laissent un hyperplan invariant, la géométrie euclidienne est l'étude des transformations qui laissent deux points invariants en 2D, ou une conique en 3D. En 2D ces deux points définissent une droite, qui est globalement invariante. De même, en 3D, la conique définit un plan globalement invariant. La géométrie euclidienne est donc aussi une restriction de la géométrie affine.

L'ensemble des transformations étant encore plus petit, les invariants sont plus nombreux. Ainsi, les angles et les rapports de distance sont invariants. Il n'y a pas de difficulté majeure dans cette nouvelle restriction, en dehors du fait que les éléments invariants sont le plus souvent choisis imaginaires, ce qui peut être gênant pour faire des représentations graphiques et aider son intuition !

### Géométrie euclidienne en dimension 2.

Fixons deux points  $I$  et  $J$ , appelés points cycliques. On appelle leur ensemble l'absolu. La géométrie euclidienne est la théorie des invariants du groupe des transfor-

mations qui laisse globalement invariant l'absolu. Ces transformations sont appelées similitudes.

Les transformations projectives, et par conséquent les similitudes, conservent la colinéarité. La droite  $(IJ)$  est donc globalement invariante, et en la choisissant comme droite de l'infini, on voit que la géométrie euclidienne est une restriction de la géométrie affine. En pratique, on prend souvent  $(1, i, 0)$  et  $(1, -i, 0)$  comme coordonnées homogènes des deux points cycliques.

On peut introduire facilement la notion d'orientation. Les similitudes qui laissent l'absolu invariant point par point, sont dites directes, les autres étant appelées indirectes.

On peut, à partir de la définition de l'absolu, donner une définition simple des concepts euclidiens. Soient deux droites  $l$  et  $l'$  et soit  $P$  leur point d'intersection. Soient alors  $j = (PI)$  et  $j' = (PJ)$  qui sont appelées droites isotropes (cf. FIG. 1.11). L'angle des deux droites  $l$  et  $l'$  peut être calculé en utilisant la formule de Laguerre :

$$\hat{l}l' = \frac{1}{2i} \log\{l, l', j, j'\}$$

L'angle peut aussi être défini en utilisant les deux points  $L$  et  $L'$  d'intersection entre la droite de l'infini et  $l$  et  $l'$  respectivement. On a :

$$\hat{l}l' = \frac{1}{2i} \log\{L, L', I, J\}$$

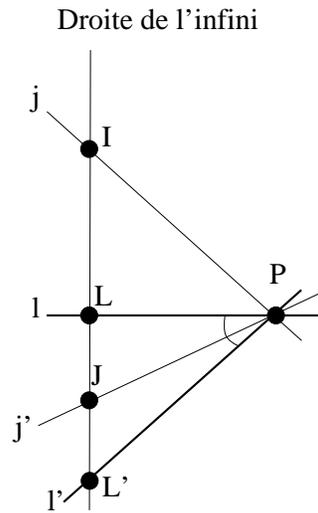


FIG. 1.11 - Définition de l'angle en géométrie euclidienne 2D.

Les deux droites  $l$  et  $l'$  sont dites orthogonales si elles sont conjuguées par rapport à  $j$  et  $j'$ . Dans ce cas, on a :  $\{l, l', j, j'\} = -1$  et :

$$\hat{l}l' = \frac{1}{2i} \log(-1) = \frac{1}{2i} i\pi = \frac{\pi}{2}$$

car  $e^{i\pi} = -1$ .

Dans la géométrie euclidienne telle que nous la définissons, la distance entre deux points n'est pas invariante. Par contre, le rapport des distances entre des points quelconques (et non pas seulement alignés) est un invariant.

### Géométrie euclidienne en dimension 3.

Dans l'espace à trois dimensions, on choisit comme absolu une conique. Le plan qui contient cette conique est pris comme plan de l'infini, et la géométrie euclidienne apparaît comme une restriction de la géométrie affine. De manière classique, on choisit comme absolu la conique imaginaire ( $\Omega$ ) d'équations suivantes :

$$\begin{cases} x^2 + y^2 + z^2 & = 0 \\ t & = 0 \end{cases}$$

Tout plan de l'espace coupe la conique absolue en deux points, que l'on choisit comme points cycliques de ce plan. Tout plan est ainsi muni d'une structure euclidienne canonique. Deux droites quelconques de l'espace  $l$  et  $l'$  définissent un plan  $P$ . Celui-ci coupe la conique absolue en deux points  $I$  et  $J$ . De plus, ces deux droites coupent le plan de l'infini en deux points  $L$  et  $L'$ . L'angle des deux droites est alors défini par ces quatre points (cf. FIG. 1.12) :

$$\hat{l}' = \frac{1}{2i} \log\{L, L', I, J\}$$

La présentation des géométries affine et euclidienne peut paraître rébarbative à quiconque est habitué à une présentation plus traditionnelle (partant de l'euclidien pour arriver vers le projectif). Notons toutefois que les concepts développés sont utilisés en vision. Ainsi, la connaissance de la projection de la conique absolue est équivalente à celle des paramètres intrinsèques ou intérieurs de la caméra.

#### 1.1.9 Autres géométries

En cas de lecture rapide, passer directement au paragraphe 1.2.

Les paragraphes précédents ont montré comment la géométrie projective pouvait être restreinte pour donner les géométries affine et euclidienne. Ce schéma ne se limite toutefois pas à ces deux restrictions. Ce paragraphe montre, sans entrer dans trop de détails, comment les géométries non euclidiennes planes peuvent aussi être obtenues par restriction de la géométrie projective.

Si la géométrie projective peut être restreinte, elle peut être aussi élargie. Déjà, Klein, dans son programme d'Erlangen, avait prévu de tels élargissements. Il suffit pour cela de considérer des groupes plus vastes que le groupe projectif. Ainsi peut-on considérer le groupe des transformations birationnelles (transformations rationnelles bijectives dont la réciproque est rationnelle) qui est à la base de la géométrie algébrique, ou même le groupe des transformations bicontinues dont les invariants sont étudiés en topologie.

Pour terminer, on peut signaler que, déjà du temps de Klein, les idées du programme d'Erlangen étaient dépassées par celles de Riemann, exprimées dans son mémoire d'admission à l'université de Göttingen en 1854, intitulées "*Sur les hypothèses qui servent*

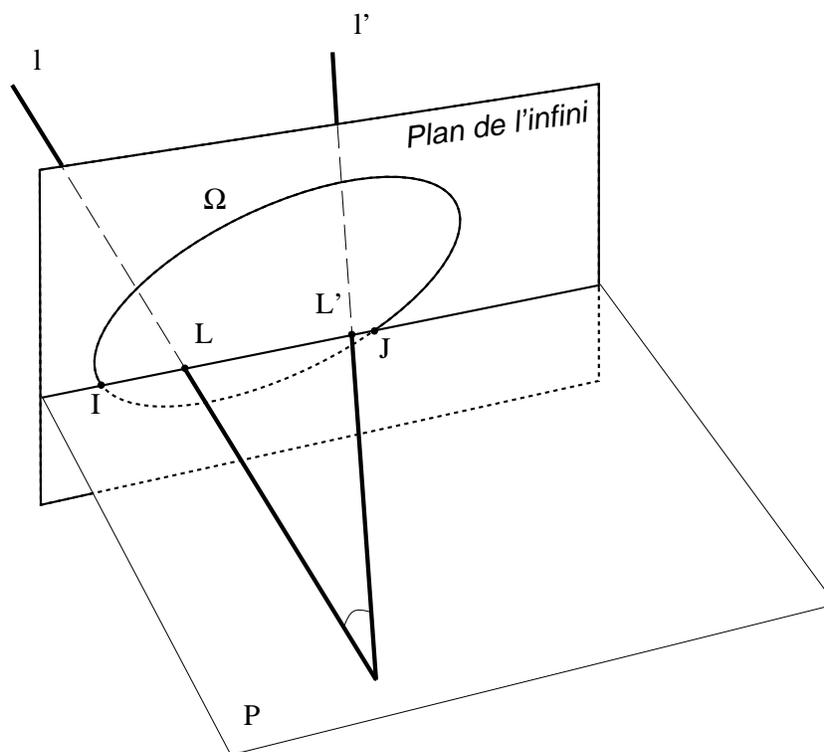


FIG. 1.12 - Définition de l'angle en géométrie euclidienne 3D.

*de fondement à la géométrie*” et publiées en 1867. En effet, Klein considère les espaces de manière globale, traitant de groupes ayant un invariant global : les droites, un hyperplan particulier, un absolu qui peut être une conique. Riemann considère, lui, des groupes qui conservent un invariant différentiel  $ds$ , ouvrant ainsi la voie à la géométrie différentielle qui ne rentre pas dans le cadre du programme d’Erlangen. Les invariants ne sont plus alors algébriques, mais infinitésimaux ou différentiels, et ne font plus seulement intervenir les formes différentielles  $dx_i$  mais aussi leurs dérivées. Notons que ces idées commencent à être reprises en vision pour trouver de nouveaux invariants.

### Géométries non euclidiennes planes.

En géométrie euclidienne plane, on considère les transformations qui laissent invariant un absolu qui est réduit à deux points. Lorsque cet absolu est une droite, on obtient la géométrie affine. Les géométries non euclidiennes découlent du choix d’une conique comme absolu. Dans l’espace, on peut choisir de manière analogue une quadrique.

Si l’absolu est une conique réelle, la géométrie obtenue est la géométrie hyperbolique de Gauss-Bolyai-Lobatchevski ; lorsqu’elle est imaginaire, on obtient la géométrie elliptique.

Ces deux géométries sont apparues très tôt, dès que certains mathématiciens ont voulu montrer par l’absurde que le postulat des parallèles d’Euclide n’était qu’un théorème. Ils prenaient une négation de ce postulat (par un point passent zéro ou plusieurs

parallèles à une droite donnée), et ils essayaient d'obtenir une contradiction. Même si beaucoup conclurent leurs essais par des succès, pensant réfuter les théorèmes qu'ils obtenaient à cause de leur monstruosité, il fallut attendre le dix-neuvième siècle pour voir ces deux géométries reconnues cohérentes. La difficulté de cette reconnaissance était due tant au fait d'une géométrie ne rendant pas compte des propriétés usuelles de l'espace ambiant, qu'aux présupposés philosophiques, dus entre autres à E. Kant, qui régnaient sans partage à l'époque.

Dans un tel espace non euclidien, la distance entre deux points peut être définie aisément. Soient deux points  $P_1$  et  $P_2$ . La droite  $(P_1P_2)$  coupe l'absolu en deux points (réels ou imaginaires)  $M_1$  et  $M_2$  (cf. FIG. 1.13). On définit alors la distance entre  $P_1$  et  $P_2$  comme étant proportionnelle au logarithme du birapport de ces quatre points :

$$d(P_1, P_2) = c \log\{P_1, P_2, M_1, M_2\} \quad \text{avec } c \in \mathbb{C}$$

On définit de même l'angle de deux droites  $l_1$  et  $l_2$  comme étant proportionnel au birapport de ces deux droites avec les deux droites  $\lambda_1$  et  $\lambda_2$  qui passent par le point d'intersection de  $l_1$  et  $l_2$  et qui sont tangentes à l'absolu (cf. FIG. 1.13) :

$$\widehat{l_1 l_2} = c' \log\{l_1, l_2, \lambda_1, \lambda_2\} \quad \text{avec } c' \in \mathbb{C}$$

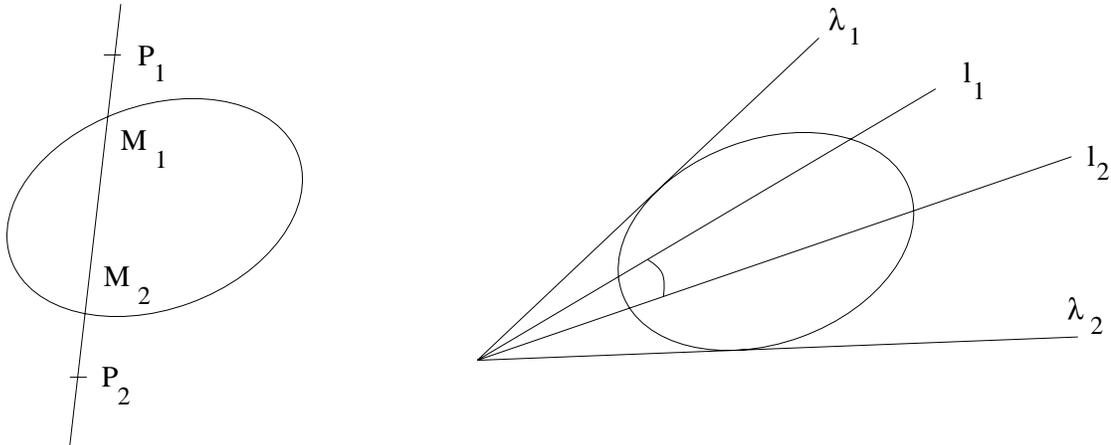


FIG. 1.13 - Définition de la distance et de l'angle en géométrie non euclidienne plane.

Deux droites sont dites parallèles lorsqu'elles se coupent sur l'absolu. Si celui-ci est une ellipse réelles, on peut voir simplement que par un point  $P$ , il peut passer deux parallèles  $D_1$  et  $D_2$  à une droite  $D$  donnée (cf. FIG. 1.14).

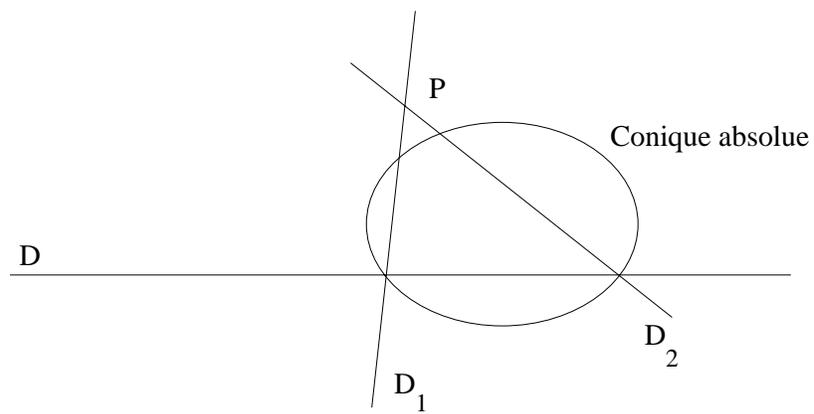


FIG. 1.14 - Par le point  $P$ , il passe deux parallèles à la droite  $D$ .



## 1.2 Théorie des invariants

La théorie des invariants prend place naturellement dans la géométrie. Elle a tout d'abord été développée par Cayley, puis par Sylvester et Hermite, mais n'a pris vraiment toute sa vraie dimension qu'avec Hilbert qui la présente sous sa forme algébrique générale. Elle est encore aujourd'hui un thème actif de recherche, tant en mathématiques [Mou 91] qu'en vision par ordinateur [Zis 91, Mun 92].

Le problème consiste à trouver des quantités numériques caractéristiques d'un objet, c'est à dire invariante par projection ou transformation géométrique. En vision, l'utilisation des invariants n'est pas récente. Les premiers invariants utilisés furent les invariants topologiques, tels la colinéarité et l'adjacence. Des invariants numériques ont aussi été utilisés pour le développement de systèmes industriels robustes de reconnaissance de formes [Bam 86, Hu 62]. Il s'agissait d'invariants globaux, des moments ou des rapports d'une aire sur le carré d'un périmètre, et les algorithmes nécessitaient de bonnes conditions d'éclairage et de contraste, et une absence d'occlusion partielle.

Récemment, de nombreux travaux ont été faits sur l'utilisation d'invariants locaux, utilisables par conséquent en cas d'occlusion partielle. Ces invariants peuvent être des angles et des rapports de distance associés à des paires de segments concourants [Gro 92a], des coordonnées affines associées à des quadruplets de points coplanaires [Lam 90, Wol 90], des invariants affines tels les coordonnées barycentriques [Koe 89, Lee 90], des birapports associés, par exemple, à cinq droites ou à deux coniques [Spa 90, Spa 91, Nie 91, Moh 91b, Spa 92, Rot 92], des quantités différentielles calculées en chaque point d'une courbe [Wei 88, Wei 91], ou une combinaison de quantités algébriques et différentielles [Goo 91, Goo 92b, Goo 92a].

Il est possible de trouver des quantités invariantes associées à la plupart des éléments extraits d'une image : points, segments de droites, courbes, facettes, coniques. Les recherches actuelles s'orientent dans deux directions principales : tout d'abord, la recherche de nouveaux invariants algébriques ou différentiels pour les courbes planes ou gauches ; d'autre part, l'utilisation de ces invariants pour résoudre des problèmes de vision. Ce dernier point amène, par exemple, à étudier la stabilité des invariants [Mor 93]. Les applications principales en vision concernent la modélisation, la reconnaissance et l'indexation.

Dans cette thèse, nous considérons des objets polyédriques, et nous étudierons donc uniquement les invariants algébriques. Tout d'abord, le problème des invariants est replacé dans un cadre théorique qui permet de résoudre le problème du dénombrement des invariants : combien existe-t-il d'invariants dans un problème donné ? Ensuite, nous nous intéressons au problème du calcul effectif des invariants. Notre approche est basée sur la recherche d'invariants canoniques. Cette méthode n'est pas la plus générale, mais elle s'avère facile, puissante et efficace pour les objets que nous manipulons. Signalons enfin qu'une méthode systématique de calcul existe : c'est la méthode symbolique [For 91b]. Cette méthode n'est cependant utilisable que pour des objets algébriques et pour autant que les calculs, souvent réalisés par des systèmes de calculs formels, restent de taille raisonnable. Pour les droites et les points, la méthode des invariants canoniques est plus simple d'emploi et a donc été préférée. Avec d'autres objets tels des courbes algébriques, le choix devrait sans doute être différent.

Ce paragraphe 1.2 contient les considérations générales. Après une présentation du problème, la méthode infinitésimale est exposée. Basée sur les notions de sous-variété et

de groupe de Lie, elle permet de dénombrer les invariants et de comprendre l'origine de la difficulté du cas des projections. Deux méthodes de calculs sont ensuite exposées : une méthode par généralisation puis contrainte, et la méthode des invariants canoniques qui est appliquée à deux exemples d'école. Des problèmes plus complexes et plus proches de la vision sont envisagés dans le chapitre 2 : coplanarité, invariants 3D, stabilité...

D'une manière générale, les notions de ce paragraphe 1.2 nécessitent des notions mathématiques assez fines pour pouvoir être utilisées de manière optimale. Cependant, ces notions mathématiques ne font qu'axiomatiser des concepts courants : une sous-variété est un ensemble qui admet un paramétrage local régulier. La plupart des objets de la géométrie classique forment ainsi des sous-variétés. Cette notion est toutefois nécessaire pour donner le cadre exact des théorèmes et résultats dont nous aurons besoin.

Pour une lecture rapide, on s'attachera donc aux présentations plus intuitives, en laissant de côté les définitions formelles (les définitions 13, 14 et 15 principalement).

### 1.2.1 Introduction aux invariants

Le problème des invariants est le suivant : on dispose d'un ensemble de configurations  $\alpha, \beta, \dots$  qui sont transformées par des éléments d'un ensemble de transformations. Ces configurations peuvent contenir des points, des droites et des courbes (des coniques en particulier). Un invariant est alors une quantité numérique  $f(\alpha)$  calculée à partir de chaque configuration, qui garde la même valeur lorsque la configuration est transformée par un des éléments  $\Phi$  de l'ensemble des transformations :  $f(\alpha) = f(\Phi(\alpha))$ .

Dans ce contexte, la configuration  $\alpha$  peut être indexée dans une base de données par la valeur de son invariant  $f(\alpha)$ . Une autre configuration  $\beta$  peut se déduire de  $\alpha$  par une des transformations seulement si  $f(\alpha) = f(\beta)$ . D'autre part, si la configuration  $\alpha$  est composée de sous-configurations  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ , la fonction peut être utilisée comme fonction de prédiction. Si  $\alpha_1, \dots, \alpha_{n-1}$  sont connus ainsi que la valeur  $f(\alpha)$ , la sous configuration  $\alpha_n$  peut être calculée si la fonction  $f$  est inversible.

Pour arriver à des résultats quantitatifs, on a toutefois besoin d'une définition et d'un cadre plus rigoureux. Après cette introduction aux invariants, deux méthodes sont expliquées : la méthode infinitésimale qui permet de répondre à la question du nombre des invariants indépendants, et une méthode par généralisation contrainte qui permet d'obtenir des résultats dans certains cas pratiques. Deux exemples utilisant la géométrie projective sont ensuite détaillés.

### Deux premiers exemples

1. Supposons que l'on dispose de plusieurs dessins plans contenant chacun 5 points. A l'aide d'une caméra, on prend des photos de ces dessins, de différents points de vue. Y a-t-il une méthode pour reconnaître quel dessin est représentée par chaque photo?

La théorie des invariants a pour but de trouver des indices numériques qui caractérisent chaque configuration et ses photos ou images. Si la caméra réalise une projection affine, les angles et les rapports de distance ne sont pas invariants, par contre les coordonnées affines de deux de ces points par rapport au repère affine formé des trois autres sont des invariants. Si la caméra effectue une projection

perspective, ces coordonnées affines ne sont plus invariantes. Il faut alors utiliser les coordonnées projectives d'un des points par rapport à la base projective formée des quatre autres points.

Ces invariants permettront de trier les photos, à moins qu'un des dessins ne puisse se déduire d'un autre par une transformation affine ou projective.

2. On considère le plan comme ensemble des configurations. Chaque configuration est donc formée d'un seul point. Les transformations sont les rotations de centre l'origine du plan. Les questions auxquelles nous aimerions répondre sont les suivantes.
  - Quelles sont les configurations que l'on peut atteindre à partir d'une configuration donnée  $C$ , si on lui applique toutes les transformations possibles ? L'ensemble de ces configurations s'appelle *l'orbite de  $C$* .
  - Etant données deux configurations, ont-elles la même orbite ? Leurs orbites se coupent-elles ?
  - Combien existe-t-il d'orbites ? Peuvent-elles être caractérisées par des fonctions constantes sur chaque orbite et dont la valeur change d'une orbite à l'autre ? Combien faut-il de ces fonctions pour caractériser complètement une orbite ?

Dans notre cas, les réponses sont simples : chaque point peut atteindre tous les points qui sont à même distance de l'origine et cette distance suffit à caractériser cet ensemble (cf. FIG. 1.15).

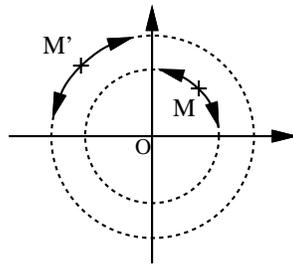


FIG. 1.15 - Cas des points du plan et des rotations de centre l'origine.

### Formulation générale

Soit  $E$  un espace appelé espace des configurations et soit  $G$  un espace de transformations sur cet espace  $E$  : les éléments de  $G$  sont des applications de  $E$  dans  $E$ .

On considère une configuration particulière  $x \in E$ , à laquelle on applique les différentes transformations de  $G$  (cf. FIG. 1.16). On s'intéresse alors aux questions suivantes.

- A partir de la configuration  $x$ , peut-on atteindre toutes les configurations de  $E$  ?
- Si tel n'est pas le cas, peut-on caractériser l'ensemble des configurations que l'on peut atteindre à partir d'une configuration initiale donnée ?

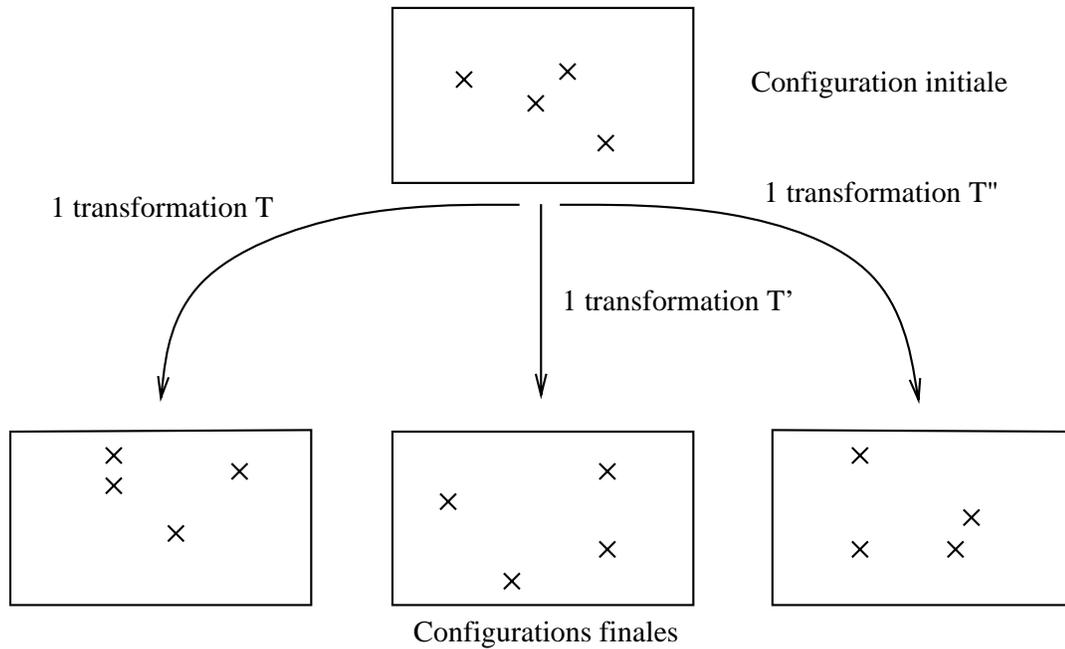


FIG. 1.16 - Le problème des invariants.

- Si une telle caractérisation est possible, combien faut-il de fonctions pour la réaliser ?

De telles fonctions sont appelées invariants pour l'ensemble des transformations de l'ensemble  $G$ . Ce sont des fonctions dont l'ensemble de départ est l'ensemble  $E$  des configurations, et dont l'ensemble d'arrivée est  $\mathbb{R}$  ou  $\mathbb{R}^n$ . Dans toute la suite, et conformément aux choix faits pour la présentation de la géométrie, nous n'envisagerons pas le cas des invariants différentiels.

### Deux exemples supplémentaires

1. On considère le plan affine comme ensemble des configurations et on prend comme transformations les translations de ce plan. Dans ce cas, n'importe quel point du plan peut atteindre tous les autres points (cf. FIG. 1.17). Il n'y a donc pas d'invariants dans ce cas.

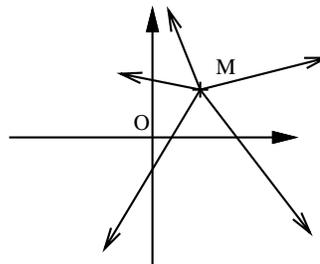


FIG. 1.17 - Cas des points du plan et des translations.

2. On considère maintenant l'ensemble des couples de points du plan (qui est isomorphe à  $\mathbb{R}^4$ ) et les isométries planes, c'est à dire les translations, rotations, symétries orthogonales et leurs composées. On peut représenter la situation soit dans le plan (cf. FIG. 1.18), soit dans l'espace à quatre dimensions. Tous les couples ne peuvent être atteints : il existe un invariant qui est la distance entre les deux points du couple.

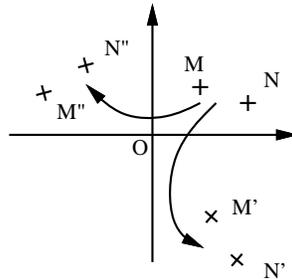


FIG. 1.18 - Cas des couples de points et des isométries.

### 1.2.2 Méthode infinitésimale de calcul des invariants.

Dans l'article de Forsyth et al. [For 91a], deux méthodes (théoriques) de calcul d'invariants sont décrites. La première est la méthode infinitésimale dont les grandes lignes sont indiquées, et qui est présentée ici en détail, car elle permet, entre autres, de dénombrer les invariants indépendants existant dans un problème donné. La deuxième est une méthode symbolique, basée sur des manipulations formelles, mais qui n'apporte pas de résultat utilisable simplement pour les cas auxquels nous nous intéressons. Elle n'est donc point reprise ici.

La méthode infinitésimale s'inscrit dans un cadre qui, bien que ne faisant appel qu'à des notions mathématiques très classiques (groupes de Lie, variétés différentielles), n'est pas toujours très familier dans la communauté de vision par ordinateur. De même que pour la géométrie, ce paragraphe ne vise donc pas tant la nouveauté, qu'une transmission de savoir et une formulation plus accessible envers un public donné.

#### Groupes opérant sur un ensemble

Les définitions qui suivent ont pour but de formaliser le cadre du programme d'Erlangen, dans lequel on s'intéresse aux invariants des groupes géométriques classiques. Commençons donc par définir les groupes. Dans le cas des groupes de fonctions, l'opération concernée est la composition des applications, qui est notée multiplicativement ou par le symbole  $\circ$ .

**Définition 11** On appelle **groupe** tout ensemble muni d'une loi de composition interne (pour nous la composition) qui vérifie les trois axiomes suivants :

– la composition est associative :

$$\forall (g, g', g'') \in G^3 \quad g \circ (g' \circ g'') = (g \circ g') \circ g'' \quad (1.5)$$

– il existe un élément neutre :

$$\exists e \in G \quad \forall g \in G \quad e \circ g = g \circ e = g \quad (1.6)$$

– chaque élément possède un inverse :

$$\forall g \in G \quad \exists g' \in G \quad g \circ g' = g' \circ g = e \quad (1.7)$$

Notons que s'il existe un élément de  $G$  qui est surjectif, ou, plus précisément, si :

$$\forall x \in E \quad \exists g \in G \quad \exists y \in E \quad x = g(y)$$

alors l'élément neutre de  $G$  est l'application identique de  $E$ , ce qui est le cas en général. On supposera qu'on est dans ce cas désormais.

Intuitivement, un groupe  $G$  opérant sur un ensemble  $E$  est donc un ensemble d'applications de  $E$  dans  $E$  qui possède les trois propriétés élémentaires des groupes (1.5), (1.6) et (1.7) : il contient la fonction identité de  $E$ , et toutes ses applications sont inversibles. Dans le deuxième exemple du paragraphe 1.2.1, le groupe considéré est celui des rotations de centre l'origine. Il opère sur le plan affine.

L'orbite d'un élément est l'ensemble de ses images par tous les éléments du groupe de transformations. Elle peut être définie mathématiquement comme suit.

**Définition 12** La relation :

$$\mathcal{R}(x, y) \text{ ssi } \exists g \in G \quad y = g(x)$$

est une relation d'équivalence sur  $E$ . La classe d'équivalence de l'élément  $x \in E$  est appelée **orbite** de  $x$  selon  $G$  et notée  $O(x)$  ou  $Gx$ .

$$O(x) = \{x' \in E \mid \exists g \in G \quad x' = g(x)\}$$

L'ensemble des orbites des éléments de  $E$  forme une partition de  $E$ , car les propriétés suivantes peuvent être déduites des axiomes (1.5), (1.6) et (1.7) :

$$\left\{ \begin{array}{ll} \forall (x, x') \in E^2 & x' \in O(x) \implies O(x) = O(x') \\ \forall (x, x') \in E^2 & x' \notin O(x) \implies O(x) \cap O(x') = \emptyset \\ \forall x \in E & x \in O(x) \end{array} \right.$$

Les invariants que l'on cherche sont, dans ce cadre, des fonctions dont la valeur est constante sur chaque orbite, mais dont la valeur change lorsqu'on change d'orbite. Dans le deuxième exemple du paragraphe 1.2.1, les orbites sont les cercles centrés sur l'origine. Dans le dernier exemple du paragraphe 1.2.1, les orbites sont les ensembles de paires de points équidistants.

## Sous-variétés et groupes de Lie

Voulant calculer les invariants analytiquement, on désire paramétrer les configurations et les transformations. Pour cela, on fait de nouvelles hypothèses : on suppose que l'espace des configurations est une *sous variété* de  $\mathbb{R}^n$  ou de  $\mathbb{C}^n$ , et que le groupe  $G$  est un *groupe de Lie*, qui opère analytiquement sur  $E$ .

Intuitivement, une sous-variété est un ensemble qui admet en tout point un paramétrage local régulier. Mais cette définition doit être précisée pour éviter des cas dégénérés. Dans l'espace affine 3D, la plupart des courbes et surfaces sont des sous-variétés, pour autant qu'elles ne contiennent pas de points singuliers. Un groupe de Lie est un groupe qui admet des paramétrages locaux compatibles avec sa structure de groupe. Pour plus de détails, on pourra se reporter à [Dem 87, Mne 86, Fav 57].

**Définition 13 ([Dem 87])** Soit  $F$  un espace vectoriel de dimension finie  $n$ ,  $V$  une partie de  $F$  et  $a$  un point de  $V$ . On dit que des fonctions  $\phi_1, \dots, \phi_d$  de classe  $C^\infty$ , définies au voisinage de  $a$ , forment un **système non-dégénéré d'équations locales** de  $V$  en  $a$  si les deux conditions suivantes sont satisfaites :

- il existe un ouvert  $U$  de  $F$  contenant  $a$  dans lequel les  $\phi_i$  sont définies et tel que  $V \cap U$  soit l'ensemble des points  $x \in U$  qui vérifient  $\phi_1(x) = 0, \dots, \phi_d(x) = 0$  ;
- les formes linéaires  $d\phi_i(a)$  sont linéairement indépendantes.

On démontre que l'entier  $d$  est uniquement déterminé. On l'appelle **codimension** de  $V$  en  $a$ . On appelle **dimension** de  $V$  en  $a$  et on note  $\dim_a V$  l'entier  $m = n - d$ . Par exemple, dans l'espace affine 3D, le cercle unité de centre l'origine et inclus dans le plan  $xOy$  admet l'ensemble  $\{\phi_1, \phi_2\}$  comme système non dégénéré d'équations locales, avec, les coordonnées d'un point  $M$  étant notées  $(x, y, z)$  :

$$\phi_1(M) = x^2 + y^2 - 1 \quad \text{et} \quad \phi_2(M) = z$$

La dimension du cercle est donc 1 en chacun de ses points. De la même manière, une surface régulière de l'espace a une dimension 2.

**Définition 14 ([Dem 87])** On dit que  $V$  est une sous-variété de  $F$  en  $a$  si elle possède un système non-dégénéré d'équations locales en  $a$  ; on dit que  $V$  est une **sous-variété** de  $F$  s'il en est ainsi en chacun de ses points, c'est à dire si elle possède en chacun de ses points un système non-dégénéré d'équations locales, ou encore s'il existe une famille d'ouverts de  $F$ , recouvrant  $V$ , telle que  $V$  possède dans chaque ouvert de la famille un système non-dégénéré d'équations.

Si  $V$  est connexe,  $V$  a même dimension en chacun de ses points. D'une manière générale, on pose :

$$\dim V = \sup_{a \in V} (\dim_a V)$$

De façon équivalente, dire que  $V$  est en  $a$  une sous-variété de  $F$  signifie qu'il existe un ouvert  $U$  de  $F$  contenant  $a$  et un difféomorphisme  $f$  de classe  $C^\infty$  de  $U$  sur un ouvert de  $\mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^d$  tel que  $f(U \cap V) = f(U) \cap (\{0\} \times \mathbb{R}^m)$ .  $f$  fournit ainsi un système local de  $m$  coordonnées curvilignes, qu'on appellera aussi paramètres du groupe.

**Définition 15** Un **groupe de Lie** est un ensemble qui est à la fois une sous-variété de  $\mathbb{R}^n$  ou de  $\mathbb{C}^n$  et un groupe tel que la multiplication et l'inversion sont continues.

Il faut bien remarquer que ces définitions sont locales. Ainsi, les rotations planes de centre l'origine admettent comme paramétrage local leur angle et forment une sous-variété, bien que ces angles ne forment pas un paramétrage continu global. Il en est de même pour les matrices inversibles, pour lesquelles le paramétrage canonique des matrices fournit un paramétrage local, ce qui est suffisant dans les définitions précédentes.

On appelle **ordre** d'un groupe  $G$  sa dimension en tant que sous-variété. Par mesure de simplicité, on le notera  $\dim G$ . Voici quelques exemples :

- le groupe des rotations planes de centre donné est d'ordre 1, le paramétrage se faisant par l'angle de rotation.
- le groupe des déplacements du plan est d'ordre 3 et contient les rotations et les translations ; les rotations ont pour coordonnées les deux coordonnées de leur centre et leur angle, les translations n'ont que deux coordonnées, celles du vecteur de translations.
- le groupe des applications affines planes bijectives, appelé encore groupe affine, est d'ordre 6. Un élément de ce groupe est en effet défini par une matrice  $2 \times 2$  inversible, la matrice de sa partie linéaire, et par deux coordonnées d'un vecteur de translation.
- le groupe des déplacements ou vissages de l'espace 3D est d'ordre 6 : 2 paramètres pour définir la direction de l'axe, 2 pour la position de l'axe, 1 pour l'angle de rotation et 1 pour la translation le long de l'axe ;
- le groupe des applications affines bijectives de l'espace 3D est d'ordre 12 (une telle application est définie par une matrice  $3 \times 3$  inversible et un vecteur de translation) ; celui des homographies du plan d'ordre 8 (une matrice  $3 \times 3$  définie à un facteur de proportionnalité près) et celui des homographies de l'espace d'ordre 15 (une matrice  $4 \times 4$  définie à un facteur de proportionnalité près).

Comme toute sous-variété, un groupe de Lie possède des systèmes locaux de *coordonnées*. On pose la convention suivante : l'élément neutre du groupe est défini par des coordonnées nulles, ce à quoi on peut toujours se ramener par translation sur les paramètres.

Le fait que  $G$  opère analytiquement sur  $E$  peut se traduire ainsi : les coordonnées  $(x'_1, \dots, x'_n)$  de l'image  $x'$  d'un élément  $x$  de  $E$  de coordonnées  $(x_1, \dots, x_n)$  par un élément  $g$  de  $G$  de coordonnées  $(a_1, \dots, a_m)$  peuvent s'écrire :

$$\forall h \in \{1, \dots, n\} \quad x'_h = \phi_h(x_1, \dots, x_n; a_1, \dots, a_m) \quad (1.8)$$

où les fonctions  $\phi_h$  sont analytiques, et vérifient d'après nos hypothèses :

$$\forall h \in \{1, \dots, n\} \quad x_h = \phi_h(x_1, \dots, x_n; 0, \dots, 0)$$

Les **orbites** sont les surfaces définies par les équations précédentes (1.8) lorsque les  $x_i$  sont fixes et que les  $a_j$  varient (voir l'exemple ci-dessous).

## Invariants et groupes de Lie

Dans le cadre des sous variétés et des groupes de Lie qui vient d'être présenté, les invariants sont des fonctions analytiques constantes sur les orbites, mais qui peuvent distinguer ces différentes orbites, c'est à dire des applications  $f_1, \dots, f_p$  qui vérifient :

$$\begin{cases} \forall x \in E & \forall y \in O(x) & \forall j \in \{1, \dots, p\} & f_j(x) = f_j(y) \\ \forall x \in E & \forall y \notin O(x) & \exists j \in \{1, \dots, p\} & f_j(x) \neq f_j(y) \end{cases} \quad (1.9)$$

Comme les applications  $f_i$  sont constantes sur chaque orbite, leur gradient est orthogonal aux espaces tangents aux orbites. Soit  $f_j$  une fonction invariante dont le gradient est noté  $\vec{\nabla} f_j$ , et soient  $(V_i(x))_i$  des vecteurs formant une base du sous espace tangent à l'orbite de  $x$  en  $x$  (il suffit en fait que ces vecteurs forment une famille génératrice de ce sous espace tangent). On a alors :

$$\forall j \quad \forall x \quad \forall i \quad \vec{\nabla} f_j \cdot V_i(x) = 0 \quad (1.10)$$

Il reste deux problèmes à résoudre : le calcul effectif des vecteurs  $(V_i(x))_i$ , et la résolution du système d'équations aux dérivées partielles obtenu.

**Un exemple simple.** Reprenons le deuxième exemple du paragraphe 1.2.1. Les configurations sont les points du plan et les transformations sont les rotations de centre l'origine. On paramètre les points par leurs coordonnées cartésiennes  $(x_1, x_2)$  et les rotations par leur angle  $\theta$ . Les équations (1.8) s'écrivent alors :

$$\begin{cases} x'_1 &= \phi_1(x_1, x_2; \theta) &= x_1 \cos \theta - x_2 \sin \theta \\ x'_2 &= \phi_2(x_1, x_2; \theta) &= x_1 \sin \theta + x_2 \cos \theta \end{cases}$$

Les orbites sont des cercles centrés sur l'origine. Leur équation est :

$$\begin{aligned} x_1 &= \phi_1(a, b; t) &= a \cos t - b \sin t \\ x_2 &= \phi_2(a, b; t) &= a \sin t + b \cos t \end{aligned}$$

où  $a$  et  $b$  sont des paramètres fixes et  $t$  varie.

Le sous-espace tangent à l'orbite du point de coordonnées  $(x_1, x_2)$  en ce point, est une droite de vecteur directeur  $(-x_2, x_1)$ . On obtient donc l'équation aux dérivées partielles suivante :

$$\forall (x_1, x_2) \in \mathbb{R}^2 \quad -x_2 \frac{\partial f}{\partial x_1} + x_1 \frac{\partial f}{\partial x_2} = 0 \quad (1.11)$$

La résolution de cette équation (1.11) mène à l'invariant suivant :  $f(x_1, x_2) = x_1^2 + x_2^2$ .

**Cas des groupes non connexes.** Certains groupes ne sont pas connexes. Par exemple, les groupes d'isométries du plan et de l'espace admettent comme composantes connexes l'ensemble des déplacements et celui des antidéplacements ; les groupes des applications bijectives du plan ou de l'espace ont pour composantes connexes les applications de déterminant strictement positif et celui des applications de déterminant strictement négatif. Pour prouver la non-connexité d'un groupe, il suffit de trouver une

fonction continue à valeurs réelles, le déterminant dans les cas précédents, dont l'image n'est pas un intervalle. Chacune des composantes connexes ne forme pas toujours un sous-groupe (c'est le cas des antidéplacements).

Dans ce cas, les orbites peuvent ne pas être connexes. La résolution des équations établies donne des fonctions invariantes sur les composantes connexes des orbites. Il est donc parfois nécessaire de pouvoir aussi retrouver les composantes qui forment une orbite. On négligera cette difficulté dans la suite.

### Calcul du sous-espace tangent à une orbite

Avant de pouvoir résoudre les équations (1.10), il faut calculer les vecteurs  $(V_i(x))_i$ . Comme le disent très justement Forsyth et al. dans leur article [For 90], la méthode la plus générale pour ce calcul utilise les notions d'algèbre de Lie et d'exponentielle. Mais, dans notre cas, l'introduction de ces concepts délicats n'est pas strictement nécessaire. Il existe une méthode par dérivation, dont le seul inconvénient (mineur) est de fournir des vecteurs qui ne sont pas indépendants du choix de la paramétrisation du groupe, mais dont on peut aisément se contenter.

**Méthode par dérivation.** On dispose d'un paramétrage du groupe, et on peut donc écrire l'action du groupe par une équation du type de l'équation (1.8) de la page 56, il suffit alors de dériver ces équations au point de coordonnées nulles.

$$\begin{aligned} V_i(x) &= \left( \frac{\partial x'_h}{\partial a_i} \Big|_{a_i=0} \right)_{h \in \{1, \dots, n\}} \\ &= \left( \frac{\partial \phi_h(x_1, \dots, x_n; 0, \dots, 0, a_i, 0, \dots, 0)}{\partial a_i} \Big|_{a_i=0} \right)_{h \in \{1, \dots, n\}} \end{aligned}$$

Il faut remarquer que les vecteurs ainsi obtenus ne sont pas forcément indépendants (voir le paragraphe 1.2.3), mais ils forment un système générateur du sous-espace tangent à l'orbite. Tout autre système de vecteurs générateurs obtenu à partir de celui ci fournit des équations équivalentes. Pour avoir un nombre minimal d'équations, il faut choisir un système de vecteurs qui soit une base du sous-espace tangent.

Autre remarque, tout paramétrage local du groupe autour de l'unité convient. Dans le cas de groupes n'ayant pas de paramétrage global simple comme celui des matrices inversibles, il suffit de remarquer que les matrices inversibles forment un ensemble ouvert : le paramétrage classique des matrices carrées fournit donc un paramétrage local du groupe dans un voisinage de l'unité, ce qui est suffisant.

### 1.2.3 Dénombrement des invariants

#### Intérêt d'un tel dénombrement

Après avoir développé une méthode de calcul des invariants, le problème à résoudre est celui du nombre des invariants.

Précisons, tout d'abord, la notion de nombre d'invariants. Il faut remarquer que lorsqu'on dispose de plusieurs fonctions invariantes  $f_1, \dots, f_p$ , on dispose aussitôt d'une infinité d'autres invariants qu'on peut exprimer sous la forme  $\Phi(f_1, \dots, f_p)$  où  $\Phi$  est

une fonction quelconque. Les deux questions importantes sont de savoir de combien de fonctions invariantes *indépendantes* on peut disposer au maximum, c'est ce qu'on appellera le nombre d'invariants, et s'il existe un ensemble de fonctions invariantes génériques, dont toutes les autres dériveraient [Dix 90]. Dans ce paragraphe, nous ne répondrons qu'à la première.

### Enumération

**Définition 16** Des fonctions  $f_1, \dots, f_p$  sont **fonctionnellement indépendantes** en un point  $a$ , si leur formes différentielles  $df_1(a), \dots, df_p(a)$  sont linéairement indépendantes.

Le nombre des invariants est relié à la dimension des orbites. En effet, si  $f_1, \dots, f_p$  est un système de  $p$  invariants indépendants, toute orbite de dimension  $n - p$ , où  $n$  est la dimension de  $E$ , admet un système non dégénéré d'équations de la forme  $f_1 - r_1, \dots, f_p - r_p$ , où  $r_1, \dots, r_p$  sont des réels.

Le nombre d'invariants existant dans un problème donné est égal au minimum des codimensions des orbites ; il indique le nombre de paramètres nécessaires pour caractériser une orbite. Pratiquement, si, dans un problème, on ne dispose que de  $k$  invariants alors que l'on sait que le problème en admet plus, on sait qu'on ne pourra pas distinguer toutes les orbites avec les invariants connus.

Il faut remarquer que, même dans les problèmes les plus simples, les orbites n'ont pas toutes la même dimension. Par exemple, dans le cas des rotations planes de centre fixé, toutes les orbites sont de dimension 1, sauf celle du centre de rotation qui est de dimension 0. C'est bien alors la dimension maximale, soit 1 dans l'exemple, qu'il faut considérer.

De façon plus mathématique, la première partie de la démonstration est facile. Soient  $f_1, \dots, f_p$  un système de  $p$  fonctions invariantes fonctionnellement indépendantes vérifiant les équations (1.9). Soit  $a$  une configuration constante. En notant  $r_i = f_i(a)$ , on obtient :

$$\forall x \in O(a) \quad f_i(x) - r_i = f_i(x) - f_i(a) = 0$$

Comme les fonctions  $f_i$  sont indépendantes :

$$\dim O(a) \leq n - p \quad \text{et} \quad p \leq n - \dim O(a)$$

La réciproque est plus délicate. Sa formulation finale est :  $p = n - \dim(O)$  où  $O$  est une orbite générique pour le problème. Voir [For 69, For 82] pour plus de détails.

### Sous-groupe d'isotropie d'un élément

En première approximation, le groupe des transformations et les orbites devraient avoir même dimension, à quelques exceptions près. Si tel était le cas, on aurait  $p = n - \dim(G)$ . Cependant, certaines transformations peuvent laisser des configurations invariantes ; elles forment le sous-groupe d'isotropie de ces configurations. Ces transformations ne sont pas "efficaces" pour ces configurations, dont l'orbite a donc comme dimension celle du groupe, moins celle de leur sous-groupe d'isotropie.

Intuitivement, le sous-groupe d'isotropie est le sous-ensemble des transformations n'ayant aucune action sur un élément donné. La dimension de l'ensemble des transformations "efficaces" dépend donc directement de la dimension de ce sous-groupe d'isotropie. Un exemple de sous-groupe d'isotropie est donné au paragraphe 1.2.8.

**Définition 17** Soit  $x$  un élément quelconque de  $E$ . On appelle **sous-groupe d'isotropie** ou **stabilisateur** de  $x$  suivant  $G$  et on note  $G_x$ , l'ensemble des éléments de  $G$  qui laissent  $x$  invariant.

$$G_x = \{g \in G \mid g(x) = x\}$$

On vérifie aisément que  $G_x$  est un sous-groupe de  $G$ . Tout d'abord, il n'est pas vide, puisque l'élément neutre de  $G$  qui est l'identité de  $E$  est dans  $G_x$ . Soient alors  $g$  et  $h$  deux éléments de  $G_x$ . Comme  $h(x) = x$ , on a  $x = h^{-1}(x)$  et :

$$x = g(x) = g \circ h^{-1}(x)$$

$g \circ h^{-1}$  appartient donc à  $G_x$ , qui est donc un sous-groupe [Ram 79].

La dimension  $n - p$  de l'orbite d'un point  $x$  est alors  $\dim G - \dim(G_x)$ , et le nombre d'invariants est :

$$\dim E - \left( \dim G - \min_{x \in E} \dim(G_x) \right) \quad (1.12)$$

**Remarque.** La dimension de l'orbite d'un point rappelle celle d'un groupe quotient ; pour trouver la dimension d'une orbite, il faut considérer les éléments de  $G$  modulo ceux de  $G_x$ . Ceci n'est toutefois qu'une analogie non mathématique, car pour parler de groupe quotient, il faudrait que le sous-groupe  $G_x$  soit distingué, ce qui n'est pas forcément le cas.

### 1.2.4 Quelques exemples

- On considère les rotations planes de centre l'origine. On peut paramétrer ce groupe par l'angle de la rotation. On a, dans ce cas, un seul invariant : le groupe des transformations est de dimension 1, l'espace des configurations est de dimension 2 et le sous-groupe stabilisateur d'un élément différent de l'origine, qui est réduit à l'identité de  $E$ , est donc de dimension 0. L'action du groupe s'écrit :

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{pmatrix}$$

On dérive alors cette formule par rapport à  $\theta$  :

$$\frac{\partial}{\partial \theta} \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} -x \sin \theta - y \cos \theta \\ x \cos \theta - y \sin \theta \end{pmatrix}$$

On prend alors cette équation pour  $\theta = 0$ . On obtient alors  $\frac{\partial x'}{\partial \theta} = -y$  et  $\frac{\partial y'}{\partial \theta} = x$ . L'équation que doit donc vérifier une fonction invariante est :

$$-y \frac{\partial f}{\partial x} + x \frac{\partial f}{\partial y} = 0$$

Cette équation aux dérivées partielles admet pour solutions toutes les fonctions composées de  $f(x, y) = x^2 + y^2$ .

- On prend comme configuration les points de l'espace affine réel 3D et comme transformations les translations de vecteur  $(a, b, 0)$  avec  $a, b \in \mathbb{R}$ . Les orbites étant de dimension 2, il y a un invariant. L'action du groupe s'écrit :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x + a \\ y + b \\ z \end{pmatrix}$$

En dérivant cette expression par rapport à  $a$  et à  $b$ , on trouve que le sous espace tangent à une orbite en un point a pour vecteurs de base  $(1, 0, 0)$  et  $(0, 1, 0)$ . Une fonction invariante doit donc vérifier :

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$

Les solutions sont les composées de la fonction  $f(x, y, z) = z$ .

- On considère toujours les points du même espace 3D, mais on ne prend plus que les translations de vecteur porté par le premier axe de coordonnées. On peut alors trouver 2 invariants indépendants. L'équation à laquelle on arrive est  $\partial f / \partial x = 0$ . On peut prendre comme solutions indépendantes  $f(x, y, z) = y$  et  $f(x, y, z) = z$ , ou bien  $f(x, y, z) = y + z$  et  $f(x, y, z) = yz$ .

### 1.2.5 Cas où les transformations ne forment pas un groupe

Dans tout ce qui précède, on est resté dans le cas classique où les transformations forment un groupe. Cependant, on s'intéresse parfois à des ensembles de transformations qui ne forment pas des groupes, les ensembles de départ et d'arrivée étant différents. Tel est le cas de la vision, où l'opération fondamentale est une projection, affine ou perspective. Les deux différences majeures existant avec ce qui précède, où on dispose d'un groupe de transformations, sont que *les orbites ne forment plus une partition* (elles peuvent se croiser par exemple), et que *tous les points d'une même orbite ne sont plus équivalents*, car on ne peut passer de l'un à l'autre par un élément de l'ensemble des transformations.

Burns et al. [Bur 92] démontrent même le résultat suivant.

**Théorème 6 (Burns)** *Dans la cas des configurations de  $n$  points et des projections perspectives, affines ou orthogonales, quel que soit l'entier  $n$ , les orbites se croisent de telle manière que toute fonction qui serait constante sur les orbites serait constante sur tout l'ensemble d'arrivée. Aussi, il n'y a pas d'invariants pour ce problème, en dehors des fonctions constantes qui n'apportent aucune information.*

Pour sortir de cette impasse, deux voies sont possibles : il faut soit disposer d'informations supplémentaires sur les configurations (leur planarité par exemple, comme dans [Mor 93]), soit travailler à partir de plusieurs images comme cela sera fait au paragraphe 2.2.

Dans le cas des groupes, la théorie est bien établie, venant du programme d'Erlangen et de Hilbert. Par contre, peu de travaux ont traité le cas où les transformations forment seulement une sous-variété sans former un groupe de Lie. Quelques travaux précurseurs existent toutefois, faits par Hilbert et Poincaré avec son étude qualitative des équations différentielles.

En restant dans ce cadre, on peut chercher des fonctions invariantes sur les orbites, bien qu'elles risquent d'avoir un pouvoir discriminant très réduit.

Malgré ce tableau peu encourageant, on peut essayer de mieux comprendre d'où viennent les limitations de ce cas.

**Méthode par dérivation.** On peut exprimer de manière différentielle le fait qu'une fonction  $f$  est constante sur une orbite. On suppose que l'action d'une transformation s'exprime à l'aide de formules analogues à l'équation (1.8) de la page 56. On note respectivement  $n$ ,  $n'$  et  $m$  les dimensions de l'espace de départ  $E$ , de l'espace d'arrivée  $F$  et de l'espace des transformations, qui appliquent les éléments de  $E$  dans  $F$ . On désire que la fonction  $f$  vérifie :

$$\begin{aligned} \forall a_1, \dots, a_m, a'_1, \dots, a'_m \\ f(\phi^1(x_1, \dots, x_n; a_1, \dots, a_m), \dots, \phi^{n'}(x_1, \dots, x_n; a_1, \dots, a_m)) \\ = f(\phi^1(x_1, \dots, x_n; a'_1, \dots, a'_m), \dots, \phi^{n'}(x_1, \dots, x_n; a'_1, \dots, a'_m)) \end{aligned}$$

Notons :

$$\Phi(x_1, \dots, x_n; a_1, \dots, a_m) = f(\phi^1(x_1, \dots, x_n; a_1, \dots, a_m), \dots, \phi^{n'}(x_1, \dots, x_n; a_1, \dots, a_m))$$

On obtient un système différentiel en dérivant cette dernière équation par rapport aux coordonnées des transformations. On trouve les  $m$  équations suivantes :

$$\begin{aligned} \forall j \in \{1, \dots, m\} \quad \frac{\partial \Phi}{\partial a_j}(x_1, \dots, x_n; a_1, \dots, a_m) = \\ \sum_{i=1}^{n'} \left( \frac{\partial f}{\partial x_i}(\phi^1(x_1, \dots, x_n; a_1, \dots, a_m), \dots, \phi^{n'}(x_1, \dots, x_n; a_1, \dots, a_m)) \right. \\ \left. \frac{\partial \phi^i}{\partial a_j}(x_1, \dots, x_n; a_1, \dots, a_m) \right) = 0 \end{aligned}$$

Malheureusement, ces équations qui ne font pas intervenir simplement les dérivées de  $f$ , mais la composée de ces dérivées avec les fonctions  $\phi$ , sont d'un degré de difficulté supplémentaire par rapport à celles qu'on obtenait dans le cas des groupes. Dans ce cas, on n'a pas pu se servir des espaces tangents et de la contraintes sur le gradient.

On peut illustrer cette difficulté par le schéma de la FIG. 1.19. Soit une configuration  $M$  et son image  $M'$  par une transformation donnée. Dans le cas d'un groupe de transformations, ce qui est représenté dans la partie gauche de la figure, l'orbite de  $M$  et celle de  $M'$  sont les mêmes et on peut exprimer l'équation de cette orbite en fonction d'un quelconque de ses points, soit  $M$ , soit  $M'$ . D'autre part, on peut passer d'un élément de l'orbite à un autre élément de l'orbite par une transformation du groupe. Dans le cas où les transformations ne forment pas un groupe, les points  $M$  et  $M'$  n'appartiennent pas au même espace.  $M'$  peut appartenir à plusieurs orbites. On ne peut donc espérer trouver une équation de l'orbite de  $M$  en ne connaissant que  $M'$ . Enfin, si  $M''$  est aussi un point de l'orbite de  $M$  ( $M''$  est image de  $M$  par une autre transformation de l'ensemble), on ne peut passer de  $M'$  à  $M''$  par une transformation de l'ensemble.

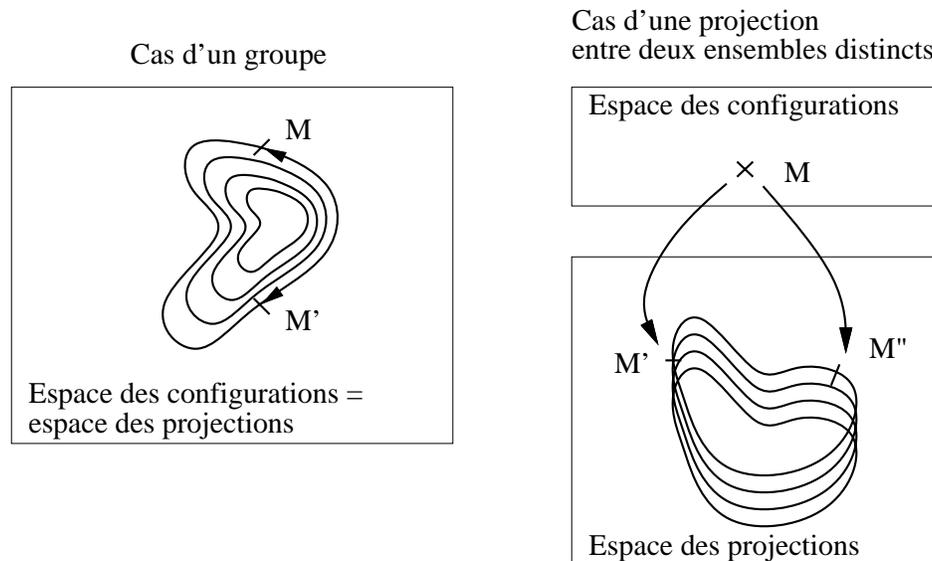


FIG. 1.19 - Une orbite dans le cas des groupes et dans le cas général

### 1.2.6 Conclusion sur la méthode infinitésimale

Le principal intérêt de cette méthode est son caractère rigoureux, qui permet, entre autres, un dénombrement précis des invariants. D'autre part, l'expression des invariants comme l'ensemble des solutions d'une équation aux dérivées partielles, permet d'appliquer aux invariants les résultats liés à ces équations.

Toutefois, l'utilisation pratique de cette méthode pour le calcul de nouveaux invariants dans des cas non triviaux est délicat. Tout d'abord, elle occulte l'aspect géométrique du problème traité. Ensuite, elle mène, dans le cas général, à des équations dont la résolution n'est pas aisée lorsqu'on n'a aucune idée de la solution. Ceci peut parfois être compensé par la recherche de solutions particulières, polynomiales par exemple.

### 1.2.7 Méthode par généralisation et contrainte

La méthode infinitésimale étant difficilement applicable pour calculer de nouveaux invariants dans des cas non triviaux, nous proposons deux méthodes pratiques de calcul de ces invariants ; la première est présentée ici, la deuxième est basée sur la recherche et l'utilisation des invariants géométriques canoniques (birapport, rapport, angle...), et elle sera utilisée dans les deux exemples des paragraphes 1.2.8 et 1.2.9.

#### Principe de la méthode

Cette méthode est fort classique parmi les mathématiciens travaillant sur les problèmes d'invariants [Dix 90], [Sem 52, pages 191 - 202]. On cherche à trouver les invariants associés à un problème donné. On va, pour cela, généraliser le problème que l'on doit résoudre, trouver les invariants pour cette généralisation, puis exprimer le fait que le problème de départ est un cas particulier. Le point crucial de la méthode est de trouver une généralisation du problème de départ pour laquelle on sache trouver les invariants. Nous allons illustrer cette méthode sur un exemple.

## Introduction et généralisation du problème

On se place dans le plan projectif et on s'intéresse aux invariants de deux coniques par changement de variables linéaire. Pour cela, on généralise le problème aux cas des formes quadratiques, et on reviendra à la fin au cas des coniques.

Soit une conique  $\mathcal{C}$  d'équation homogène :

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz = 0$$

On peut écrire cette dernière équation sous la forme  ${}^tXAX = 0$ , où  ${}^tX = (x, y, z)$  et  $A$  est la matrice symétrique :

$$\begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix}$$

Comme l'équation de la conique  $\mathcal{C}$  n'est définie qu'à un facteur multiplicatif près, la matrice  $A$  n'est, elle aussi, définie qu'à un facteur multiplicatif près. A la matrice  $A$ , on associe la forme quadratique dont  $A$  est la matrice, c'est à dire l'application suivante qui n'est plus définie à un facteur de proportionnalité près :

$$\begin{aligned} \mathbb{C}^3 &\longrightarrow \mathbb{C} \\ X &\longmapsto {}^tXAX \end{aligned}$$

## Les invariants des paires de formes quadratiques

Concernant ce paragraphe, on pourra se reporter à [Sem 52, page 194]. On prend comme ensemble de configurations l'ensemble des paires de formes quadratiques en trois variables, qu'on représente par leurs matrices  $3 \times 3$  symétriques associées  $A = (a_{ij})_{i,j \in \{1,2,3\}}$  et  $B = (b_{ij})_{i,j \in \{1,2,3\}}$ . On considère les transformations suivantes :

$$(A, B) \mapsto ({}^tPAP, {}^tPBP)$$

où on impose  $\det P = 1$ , qui correspondent aux changements de variables linéaires de déterminant égal à 1. En effet, par le changement de variable défini par la matrice  $P$ , la matrice  $A$  d'une forme quadratique est transformée en  ${}^tPAP$ . On pourra supprimer par la suite la contrainte sur le déterminant. On note  $A_{ij}$  et  $B_{ij}$  les cofacteurs de  $a_{ij}$  et  $b_{ij}$ .

On connaît deux invariants élémentaires pour ce problème,  $\det(A)$  et  $\det(B)$ , qui sont des fonctions homogènes de degré 3 :  $\det(tA) = t^3 \det(A)$ . Plus généralement, on dispose de l'invariant  $\det(\lambda A + \mu B)$ . On a :

$$\det(\lambda A + \mu B) = \lambda^3 I_1 + \lambda^2 \mu I_2 + \lambda \mu^2 I_3 + \mu^3 I_4 \quad (1.13)$$

avec

$$\begin{aligned} I_1 &= \frac{1}{3}(a_{11}A_{11} + a_{22}A_{22} + a_{33}A_{33} + 2a_{23}A_{23} + 2a_{13}A_{13} + 2a_{12}A_{12}) = \det A \\ I_2 &= b_{11}A_{11} + b_{22}A_{22} + b_{33}A_{33} + 2b_{23}A_{23} + 2b_{13}A_{13} + 2b_{12}A_{12} \\ I_3 &= a_{11}B_{11} + a_{22}B_{22} + a_{33}B_{33} + 2a_{23}B_{23} + 2a_{13}B_{13} + 2a_{12}B_{12} \\ I_4 &= \frac{1}{3}(b_{11}B_{11} + b_{22}B_{22} + b_{33}B_{33} + 2b_{23}B_{23} + 2b_{13}B_{13} + 2b_{12}B_{12}) = \det B \end{aligned}$$

Le polynôme (1.13) étant invariant, ses coefficients  $I_1, I_2, I_3$  et  $I_4$  le sont aussi. On dispose même du résultat plus fort suivant [Dix 90] : tout invariant polynomial de  $A$  et  $B$  est fonction polynomiale de  $I_1, I_2, I_3$  et  $I_4$ .

Si on relaxe la contrainte  $\det P = 1$ , les  $I_i$  ne sont plus des invariants, car au cours de la transformation,  $I_i$  devient  $(\det P)^2 I_i$ . On peut, par contre, prendre pour invariant les rapports  $I_i/I_j$ .

### Les invariants de paires de coniques

Revenons au cas des coniques. Une conique définit une forme quadratique à un coefficient multiplicatif près. On ne peut donc plus supposer que la matrice associée à cette conique a un déterminant égal à 1. Les invariants de coniques sont donc les invariants de formes quadratiques, qui sont aussi invariants lorsqu'on multiplie les matrices par un scalaire. Dans notre cas, on a :

$$\begin{aligned} I_1(\lambda A, \mu B) &= \lambda^3 I_1(A, B) \\ I_2(\lambda A, \mu B) &= \lambda^2 \mu I_2(A, B) \\ I_3(\lambda A, \mu B) &= \lambda \mu^2 I_3(A, B) \\ I_4(\lambda A, \mu B) &= \mu^3 I_4(A, B) \end{aligned}$$

On cherche alors les monômes  $I_1^{n_1} I_2^{n_2} I_3^{n_3} I_4^{n_4}$  vraiment invariants. Comme :

$$\forall (\lambda, \mu) \in \mathbb{R}^2 \quad \lambda^{3n_1} I_1^{n_1} \cdot \lambda^{2n_2} \mu^{n_2} I_2^{n_2} \cdot \lambda^{n_3} \mu^{2n_3} I_3^{n_3} \cdot \mu^{3n_4} I_4^{n_4} = I_1^{n_1} I_2^{n_2} I_3^{n_3} I_4^{n_4}$$

cela implique :

$$\begin{cases} 3n_1 + 2n_2 + n_3 &= 0 \\ n_2 + 2n_3 + 3n_4 &= 0 \end{cases}$$

On peut prendre pour solutions indépendantes  $n_1 = 1, n_2 = -2, n_3 = 1, n_4 = 0$ , ou  $n_1 = 0, n_2 = 1, n_3 = -2, n_4 = 1$ . Ces solutions donnent les invariants suivants :

$$\alpha = \frac{I_1 I_3}{I_2^2}, \quad \beta = \frac{I_2 I_4}{I_3^2}$$

### Conclusion sur la méthode

La méthode présentée ici a le double avantage d'être très classique mathématiquement et très puissante. Elle peut être une **très bonne approche pratique** dans de nombreux cas où la méthode infinitésimale est inapplicable, car menant à des équations qu'on ne sait pas résoudre. Cela dit, le fait qu'il n'y ait pas de méthode systématique de généralisation du problème limite grandement la possibilité d'utilisation de cette méthode.

#### 1.2.8 Invariants de deux points et deux droites

L'idée directrice de cet exemple et du suivant est qu'il existe un invariant canonique associé à certaines transformations. Tels sont la distance et l'angle pour les transformations euclidiennes, le rapport pour les transformations affines et le birapport pour les transformations projectives. On cherchera donc à ramener les problèmes d'invariants à

des recherches de distances, rapports ou birapports. Cela peut être considéré comme une troisième méthode de calcul des invariants. Bien qu'elle paraisse peu systématique, *elle s'avère la plus puissante des trois*, en particulier pour résoudre les problèmes de vision où n'apparaissent que des éléments d'ordre inférieurs à deux : points, droites, plans, coniques.

Il n'est certes pas assuré que cette recherche d'invariants canoniques puisse être faite systématiquement, en particulier lorsqu'on manipule des objets qui ne sont pas élémentaires, des courbes ou des surfaces cubiques par exemple. Cette méthode permet toutefois une recherche simple et une interprétation géométrique aisée des invariants. Dans les deux exemples qui suivent, on se placera dans le cas du birapport dans le plan projectif.

**Recherche d'un invariant.** On se place donc dans le plan projectif et on considère les configurations formées de deux points et de deux droites. On cherche un invariant pour ces configurations, avec l'ensemble de toutes les applications projectives du plan comme ensemble de transformations.

Il est très facile d'en trouver un. Pour cela, il suffit de construire, à l'aide de constructions stables par transformation projective, un ou plusieurs birapports qui seront les invariants cherchés. Dans le cas présent, on construit la droite passant par les deux points. Cette droite coupe les deux droites de la configuration en deux points, qui peuvent être à distance finie ou infinie. On dispose alors de quatre points alignés, les deux points de la configuration et les deux points d'intersection, et donc d'un birapport qui est l'invariant cherché.

Essayons d'appliquer la théorie faite sur les groupes de Lie. L'ensemble des applications projectives du plan, qui sont définies par des matrices  $3 \times 3$  à un facteur de proportionnalité près, est donc une sous-variété de dimension 8. Le fait de rajouter une contrainte (déterminant non nul) enlève une variété de transformation, mais ne fait pas varier la dimension. Une droite ou un point sont définis par 3 paramètres, mais à un facteur de proportionnalité près. Les configurations forment donc une sous-variété de dimension 8 aussi.

Bien que les dimensions des espaces de transformations et de configurations soient égales, l'existence d'un invariant est possible ; elle est due, d'après la formule (1.12) de la page 60, à la présence d'un groupe d'isotropie de dimension non-nulle, que nous allons rechercher.

**Recherche du groupe d'isotropie.** Pour mener à bien cette recherche, on considère une configuration générique (non dégénérée), c'est à dire telle que les droites ne soient pas confondues et que les points ne soient pas sur ces droites et ne soient pas confondus (cf. FIG. 1.20). On appelle  $P_1$  et  $P_2$  les deux points et  $D_1$  et  $D_2$  les deux droites de la configuration. On note  $D$  la droite  $(P_1P_2)$ ,  $M_1$  et  $M_2$  les intersections de la droite  $D$  avec  $D_1$  et  $D_2$  respectivement. On note  $M$  l'intersection de  $D_1$  et de  $D_2$ .

On choisit une représentation du plan telle que  $M$  soit l'origine et ait pour coordonnées  $(0, 0, 1)$  et telle que  $D$  soit la droite de l'infini, d'équation  $x_3 = 0$ .

Soit  $F$  une application projective qui laisse cette configuration invariante. Les points  $P_1$  et  $P_2$  sont invariants et la droite  $D$  est globalement invariante : en conséquence, le point  $M_1$  est invariant. De plus,  $(P_1, P_2, M_1)$  formant une base projective de la droite

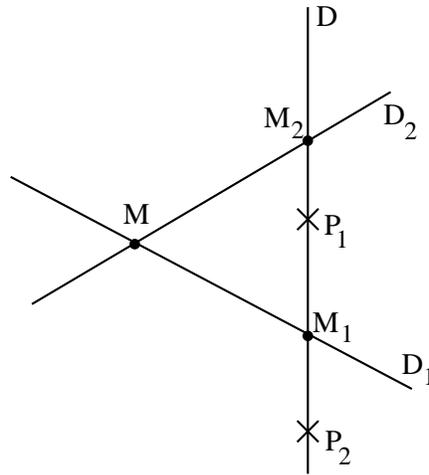


FIG. 1.20 - Une configuration de 2 points et de 2 droites

$D$ , cette dernière est stable (tous ses points sont invariants). Enfin, les droites  $D_1$  et  $D_2$  étant globalement stables, leur intersection le point  $M$  est invariant.

La droite de l'infini  $D$  étant globalement invariante par  $F$ ,  $F$  est une application affine. Comme l'origine  $M$  est invariante,  $F$  se réduit à sa partie linéaire:  $F$  est une application linéaire. Tous les points de la droite de l'infini, et en particulier les points cycliques sont invariants:  $F$  est une transformation euclidienne, donc une similitude (c'est à dire la composée d'une isométrie et d'une homothétie).

Toute droite qui passe par l'origine admet 2 points invariants: l'origine et son point à l'infini. Ces droites sont donc globalement invariantes, et toutes les directions vectorielles sont directions propres pour  $F$ . Les homothéties sont les seules similitudes vérifiant cette condition:  $F$  est donc une homothétie de centre  $M$ . L'ensemble de ces homothéties est un groupe d'ordre 1, qui est le groupe d'isotropie recherché.

**Conclusion.** Dans cet exemple qui concerne des configurations de deux points et deux droites, la sous-variété des transformations et celle des configurations ont même dimension. L'existence de l'invariant exhibé est due à la présence d'un sous-groupe d'isotropie pour chaque configuration, sous-groupe dont la nature a été déterminée.

### 1.2.9 Invariants de deux coniques

Soient deux coniques  $\mathcal{C}_1$  et  $\mathcal{C}_2$  auxquelles on associe les matrices  $\mathcal{M}_1$  et  $\mathcal{M}_2$ . On recherche les invariants de ces coniques, les transformations étant les homographies du plan. Les transformations forment une sous-variété de dimension 8, les coniques une de dimension 10. On devrait donc trouver 2 invariants au moins, et 2 seulement si le groupe stabilisateur d'une configuration est de dimension 0.

Le problème de la recherche des invariants de deux coniques a été abordé de diverses manières. Maybank [May 91] calcule les birapports des quatre points d'intersection. Forsyth et al. [For 91a] donnent comme invariants  $Trace(\mathcal{M}_2^{-1}\mathcal{M}_1)$  et  $Trace(\mathcal{M}_1^{-1}\mathcal{M}_2)$ . En fait, ces traces ne sont des invariants que des formes quadratiques associées à  $\mathcal{M}_1$  et  $\mathcal{M}_2$ , et non des coniques. Pour qu'elles le deviennent, il faut imposer aux matrices

$\mathcal{M}_1$  et  $\mathcal{M}_2$  d'être de déterminant 1, ce que font Forsyth et al. [For 91a]. D'un point de vue pratique, il serait souhaitable d'intégrer cette contrainte dans la formule même des invariants, ce que nous ferons dans la suite. Dans leur article [Kap 91], Kapur et al. interprètent ces invariants en fonction du triangle auto-polaire associé aux deux coniques, sans toutefois introduire explicitement ce triangle.

Pour notre part, nous montrons comment interpréter ces invariants en termes de birapports. Ce paragraphe est plus technique que le précédent. Pour une lecture rapide, on pourra donc le parcourir sans approfondir.

### Interprétation de la matrice $\mathcal{M}_2^{-1}\mathcal{M}_1$ en termes de birapports

Soient  $\mathcal{C}_1$  et  $\mathcal{C}_2$  deux coniques de matrices associées  $\mathcal{M}_1$  et  $\mathcal{M}_2$ . Ces deux coniques se coupent en quatre points  $A, B, C, D$ , distincts ou confondus, réels ou imaginaires. A partir de ces quatre points, on peut construire un triangle  $(P, Q, R)$  qui est auto-polaire par rapport à chacune de ces deux coniques. C'est le seul triangle auto-polaire commun (cf. le paragraphe 1.1.6, page. 39).

**Trace de la matrice  $\mathcal{M}_2^{-1}\mathcal{M}_1$ .** Soit  $\mathcal{N}$  une matrice  $3 \times 3$  non nulle. L'application projective qui à tout point de coordonnées  $X$  fait correspondre le point de coordonnées  $X'$  tel que  $X = \mathcal{N}X'$ , transforme  $\mathcal{M}_1$  et  $\mathcal{M}_2$  en  ${}^t\mathcal{N}\mathcal{M}_1\mathcal{N}$  et  ${}^t\mathcal{N}\mathcal{M}_2\mathcal{N}$ . On a :

$$\begin{aligned} \text{Trace}(({}^t\mathcal{N}\mathcal{M}_2\mathcal{N})^{-1} {}^t\mathcal{N}\mathcal{M}_1\mathcal{N}) &= \text{Trace}(\mathcal{N}^{-1}\mathcal{M}_2^{-1}\mathcal{M}_1\mathcal{N}) \\ &= \text{Trace}(\mathcal{M}_2^{-1}\mathcal{M}_1) \end{aligned}$$

La trace de la matrice  $\mathcal{M}_2^{-1}\mathcal{M}_1$  est donc un invariant de cette matrice, mais pas des coniques associées ; en effet, celles-ci ne définissent leurs matrices qu'à un facteur multiplicatif près, et la trace n'est pas invariante par homothétie sur une des matrices.

La trace n'étant pas un vrai invariant des coniques, on ne peut donc l'interpréter en termes de birapports.

**Vecteurs propres de la matrice  $\mathcal{M}_2^{-1}\mathcal{M}_1$ .** Soit  $l_P$  les coordonnées homogènes de la droite  $(QR)$ . Le triangle  $(P, Q, R)$  étant auto-polaire par rapport à chacune des deux coniques, on a :  $\lambda l_P = \mathcal{M}_1 P$  et  $\mu l_P = \mathcal{M}_2 P$ . Donc  $\mathcal{M}_2^{-1}\mathcal{M}_1 P = \lambda/\mu P$ . De même pour  $Q$  et  $R$ . Les sommets du triangle auto-polaire commun sont donc les vecteurs propres de la matrice  $\mathcal{M}_2^{-1}\mathcal{M}_1$ .

On peut ainsi lier les vecteurs propres de  $\mathcal{M}_2^{-1}\mathcal{M}_1$  au triangle auto-polaire commun, qui est lié aux coniques par des birapports.

**Valeurs propres de la matrice  $\mathcal{M}_2^{-1}\mathcal{M}_1$ .** Prenons les points  $P, Q$  et  $R$  comme références. Les deux coniques admettent alors chacune une équation réduite :

$$\begin{aligned} \mathcal{C}_1 &: a_1x^2 + b_1y^2 + c_1z^2 = 0 \\ \mathcal{C}_2 &: a_2x^2 + b_2y^2 + c_2z^2 = 0 \end{aligned}$$

Les valeurs propres d'une matrice ne dépendant pas de la base,  $a_1, b_1, c_1$  sont les valeurs propres de  $\mathcal{M}_1$  et  $a_2, b_2, c_2$  sont celles de  $\mathcal{M}_2$ . Il faut remarquer que ce ne sont pas les valeurs propres qui sont caractéristiques des coniques, mais leurs rapports (à cause du coefficient multiplicatif).

Les valeurs propres de  $\mathcal{M}_2^{-1}\mathcal{M}_1$  sont donc :

$$\lambda_1 = \frac{a_1}{a_2}, \quad \lambda_2 = \frac{b_1}{b_2}, \quad \lambda_3 = \frac{c_1}{c_2}$$

Là aussi, seuls les rapports de valeurs propres sont caractéristiques des coniques. Ces rapports peuvent d'ailleurs s'interpréter géométriquement.

Soient  $M_1, N_1$  et  $M_2, N_2$  les intersections de la droite  $(PQ)$  avec  $\mathcal{C}_1$  et  $\mathcal{C}_2$  respectivement (cf. FIG. 1.21). Soient respectivement  $\alpha_1, \alpha_2, \beta_1, \beta_2$  des racines carrées, réelles

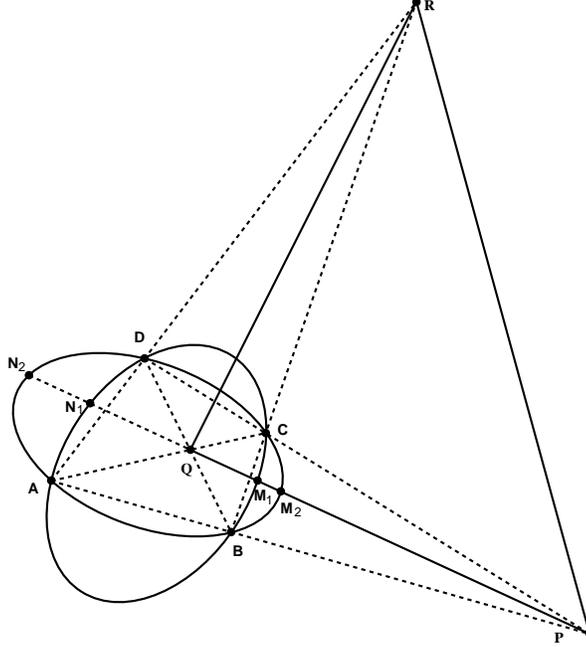


FIG. 1.21 - Interprétation des matrices de coniques en termes de birapports

ou imaginaires, de  $a_1, a_2, b_1, b_2$  et soit  $i^2 = -1$ . L'équation de la droite  $(PQ)$  est :  $z = 0$ .  $M_1, N_1, M_2$  et  $N_2$  sont définis par les systèmes suivants :

$$\begin{cases} a_1x^2 + b_1y^2 + c_1z^2 = 0 \\ z = 0 \end{cases} \quad \text{et} \quad \begin{cases} a_2x^2 + b_2y^2 + c_2z^2 = 0 \\ z = 0 \end{cases}$$

On a alors :

$$\begin{aligned} M_1 &= (\beta_1, i\alpha_1, 0) \\ N_1 &= (\beta_1, -i\alpha_1, 0) \\ M_2 &= (\beta_2, i\alpha_2, 0) \\ N_2 &= (\beta_2, -i\alpha_2, 0) \end{aligned}$$

Donc :

$$\{M_1, M_2; P, Q\} = \left\{ -i\frac{\beta_1}{\alpha_1}, -i\frac{\beta_2}{\alpha_2}; \infty, 0 \right\} = \frac{\alpha_1\beta_2}{\beta_1\alpha_2}$$

et :

$$\frac{\lambda_1}{\lambda_2} = \{M_1, M_2; P, Q\}^2$$

On peut de même interpréter les autres rapports de valeurs propres.

Dans le cas où les deux coniques ont un point double d'intersection la même construction est possible. Si, par exemple,  $A = B$ , il suffit de remplacer la droite  $(AB)$  par la tangente à la conique au point  $A$ .

### Nouvelle interprétation des résultats du paragraphe 1.6.7

Dans ce paragraphe, on va interpréter les invariants trouvés en fonction de birapports. Pour cela, on prend les sommets du triangle auto-polaire commun comme points de référence et on calcule  $I_1, I_2, I_3$  et  $I_4$  avec les équations réduites des deux coniques :

$$\begin{aligned} I_1 &= a_1 b_1 c_1 \\ I_2 &= a_2 b_1 c_1 + b_2 a_1 c_1 + c_2 a_1 b_1 \\ I_3 &= a_1 b_2 c_2 + b_1 a_2 c_2 + c_1 a_2 b_2 \\ I_4 &= a_2 b_2 c_2 \end{aligned}$$

On peut exprimer les rapports de ces nombres en fonction des valeurs propres de la matrice  $\mathcal{M}_2^{-1}\mathcal{M}_1$  :

$$\begin{aligned} \frac{I_1}{I_4} &= \lambda_1 \lambda_2 \lambda_3 \\ \frac{I_2}{I_4} &= \lambda_2 \lambda_3 + \lambda_1 \lambda_3 + \lambda_1 \lambda_2 \\ \frac{I_3}{I_4} &= \lambda_1 + \lambda_2 + \lambda_3 \end{aligned}$$

On a alors de manière évidente :

$$\begin{aligned} \text{Trace}(\mathcal{M}_2^{-1}\mathcal{M}_1) &= \frac{I_3}{I_4} \\ \text{Trace}(\mathcal{M}_1^{-1}\mathcal{M}_2) &= \frac{I_2}{I_1} \\ \frac{\det \mathcal{M}_1}{\det \mathcal{M}_2} &= \frac{I_1}{I_4} \end{aligned}$$

Pour obtenir les invariants de coniques, il faut faire apparaître des birapports, c'est à dire des rapports de valeurs propres comme on l'a vu. Par exemple, on peut prendre :

$$\begin{aligned} \alpha &= \frac{I_1 I_3}{I_2^2} = \frac{(\lambda_1 + \lambda_2 + \lambda_3)(\lambda_1 \lambda_2 \lambda_3)}{(\lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_1 \lambda_3)^2} = \lambda_3 / \lambda_2 \frac{1 + \lambda_2 / \lambda_1 + \lambda_3 / \lambda_1}{(1 + \lambda_3 / \lambda_1 + \lambda_3 / \lambda_2)^2}, \\ \beta &= \frac{I_4 I_2}{I_3^2} = \frac{\lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_1 \lambda_3}{(\lambda_1 + \lambda_2 + \lambda_3)^2} = \lambda_2 / \lambda_1 \frac{1 + \lambda_3 / \lambda_1 + \lambda_3 / \lambda_2}{(1 + \lambda_2 / \lambda_1 + \lambda_3 / \lambda_1)^2}. \end{aligned}$$

Les invariants donnés au paragraphe 1.2.7 sont ainsi reliés à des birapports.

Il faut remarquer qu'on a aussi :

$$\begin{aligned} \alpha &= \frac{I_1 I_3}{I_2^2} = \frac{\text{Trace}(\mathcal{M}_2^{-1}\mathcal{M}_1) \det \mathcal{M}_2}{(\text{Trace}(\mathcal{M}_1^{-1}\mathcal{M}_2))^2 \det \mathcal{M}_1} \\ \beta &= \frac{I_4 I_2}{I_3^2} = \frac{\text{Trace}(\mathcal{M}_1^{-1}\mathcal{M}_2) \det \mathcal{M}_1}{(\text{Trace}(\mathcal{M}_2^{-1}\mathcal{M}_1))^2 \det \mathcal{M}_2} \end{aligned}$$

Ainsi, les invariants donnés par Forsyth et al. [For 91a] sont liés à des birapports, dès qu'on les a combinés pour en faire de vrais invariants de coniques et non pas de formes quadratiques seulement.

### Birapports et intersection des coniques

Deux coniques se coupent en quatre points. D'après le paragraphe 1.1.6, quatre points sur une conique définissent un birapport. Les quatre points d'intersection qui appartiennent à chacune des deux coniques définissent donc deux birapports.

Nous allons faire le lien entre ces birapports et ceux rencontrés au paragraphe précédent. Soient deux coniques  $\mathcal{C}_1$  et  $\mathcal{C}_2$  de matrices  $\mathcal{M}_1$  et  $\mathcal{M}_2$ ,  $A, B, C, D$  leurs points d'intersection et  $k_1$  et  $k_2$  les deux birapports formés. Dans son article [May 91], Maybank montre que ces birapports s'expriment en fonction des racines de l'équation :

$$\det(\mathcal{M}_1 + \xi\mathcal{M}_2) = 0 \quad (1.14)$$

Pour cela, il choisit les coordonnées de telle sorte que les quatre points d'intersection soient  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$  et  $(1, 1, 1)$ . Dans ces coordonnées, les coniques ont pour équations :

$$\begin{aligned} \mathcal{C}_1 & : \phi_1 xy + \psi_1 yz - (\phi_1 + \psi_1)xz = 0 \\ \mathcal{C}_2 & : \phi_2 xy + \psi_2 yz - (\phi_2 + \psi_2)xz = 0 \end{aligned}$$

où  $\phi_1, \phi_2, \psi_1$  et  $\psi_2$  sont des constantes. Les trois racines de l'équation (1.14) sont donc :

$$\begin{aligned} \xi_1 & = -\frac{\psi_1}{\psi_2} \\ \xi_2 & = -\frac{\phi_1}{\phi_2} \\ \xi_3 & = -\frac{\phi_1 + \psi_1}{\phi_2 + \psi_2} \end{aligned}$$

D'autre part, dans la paramétrisation suivante de  $\mathcal{C}_1$  :

$$s \mapsto (s[\psi_1(1-s) - s\phi_1], \psi_1(1-s) - s\phi_1, -s\phi_1)$$

les quatre points d'intersection ont pour paramètres :  $\infty, 0, \psi_1/(\psi_1 + \phi_1), 1$ . On a donc :

$$\begin{aligned} k_1 & = \{A, B; C, D\}_{\mathcal{C}_1} \\ & = \{\infty, 0; \psi_1/(\psi_1 + \phi_1), 1\} \\ & = 1 + \frac{\phi_1}{\psi_1} \\ & = \{0, \xi_2; \xi_3, \xi_1\} \end{aligned}$$

On a de même :

$$k_2 = \{\infty, \xi_2; \xi_3, \xi_1\}$$

En reprenant les notations du paragraphe précédent, on a :

$$\det(\mathcal{M}_1 + \xi\mathcal{M}_2) = (a_1 + \xi a_2)(b_1 + \xi b_2)(c_1 + \xi c_2)$$

et par conséquent :

$$\xi_1 = \lambda_1, \quad \xi_2 = \lambda_2, \quad \xi_3 = \lambda_3$$

On en conclut :

$$k_1 = \frac{\lambda_3(\lambda_2 - \lambda_1)}{\lambda_1(\lambda_2 - \lambda_3)} = \frac{\frac{\lambda_2}{\lambda_1} - 1}{\frac{\lambda_2}{\lambda_3} - 1}$$

$$k_2 = \frac{\lambda_2 - \lambda_1}{\lambda_2 - \lambda_3} = \frac{1 - \frac{\lambda_1}{\lambda_2}}{1 - \frac{\lambda_3}{\lambda_2}}$$

Le lien avec les paragraphes précédents est donc fait.

### Conclusion sur les deux exemples précédents

L'idée que nous avons voulu développer dans ces deux exemples, est que les invariants de certaines classes de problèmes peuvent être trouvés très simplement. Ces problèmes sont ceux où interviennent un ensemble classique de transformations (transformations affines, euclidiennes ou projectives) et des éléments simples (points, droites, coniques). On peut alors chercher les invariants sous formes de rapports, angles, distances ou birapports suivant le type des transformations utilisées. Le lien entre ces invariants simples et d'autres formes invariantes utilisées lorsque l'on manipule des éléments moins élémentaires, peut, toutefois, être plus complexe.

Dans le cas des transformations projectives planes, les résultats suivants sont ainsi immédiats :

- la configuration constituée de deux droites et une conique a un invariant qui est le birapport des quatre points d'intersection entre les droites et la conique ;
- deux coniques concentriques et une droite ont pour invariant le birapport des quatre points intersection de la droite et des coniques ;
- une conique et deux points ont un invariant, le birapport de ces deux points avec les deux intersections de la droite qu'ils définissent et de la conique ;
- quatre droites et un point se trouvant sur l'une d'elle ont un invariant, le birapport de ce point avec les intersections de la droite sur laquelle il se trouve avec les trois autres droites ;
- ...

D'après la formule (1.12), on sait qu'on a obtenu le maximum d'invariants indépendants dans tous ces problèmes.

## 1.3 Éléments de vision

Les paragraphes précédents, où étaient développés la géométrie projective et la théorie des invariants, supposaient implicitement que la géométrie projective est utilisable pour modéliser la formation d'une image par une caméra. Ce paragraphe a pour but de justifier cette supposition, en étudiant différents modèles optiques et géométriques de caméras et en exhibant les projections perspectives ou parallèles qui réalisent des approximations de ces différents modèles.

Après cela, les traitements des images dits de bas niveau sont brièvement présentés.

Il importe, dans tout ce paragraphe, de bien distinguer les modèles optiques et géométriques. Le modèle optique d'une caméra, et plus particulièrement de son objectif, est une description d'un modèle physique. On s'intéresse alors à la focale réelle de l'objectif, à l'axe optique réel. Le modèle géométrique modélise quant à lui la formation de l'image. Peu importe alors la réalité de la focale ou de l'axe principal par rapport aux caractéristiques réelles. On peut supposer que l'image s'est formée en avant de la caméra plutôt qu'au fond de celle-ci. Le modèle géométrique ne rend compte que du résultat, l'image, et non du processus de sa formation. On essaiera donc d'utiliser un vocabulaire différent pour ces deux modèles, et on évitera de vouloir relier les grandeurs mesurables ou données par le constructeur (focale, interstice) avec les grandeurs intervenant dans le modèle géométrique.

### 1.3.1 L'approximation des systèmes minces

Le premier modèle que nous étudions, est celui des systèmes minces. Ce modèle sert généralement à étudier les lentilles convergentes ou divergentes, ou des associations simples de telles lentilles, dont on suppose l'épaisseur négligeable. Pour une présentation générale de l'optique géométrique, on pourra se reporter à [Sua 74].

Soit un système mince placé dans le plan  $\Omega xy$  et centré au point  $\Omega$  (cf. FIG. 1.22). On définit un tel système par son centre optique  $\Omega$ , son foyer objet  $F$ , son foyer image  $F'$  et sa distance focale  $f = \overline{\Omega F} = -\overline{\Omega F'}$ , où  $\overline{\Omega F}$  est une mesure algébrique. L'axe  $\Omega z$  est appelé axe optique du système, c'est un axe de symétrie de révolution du système.

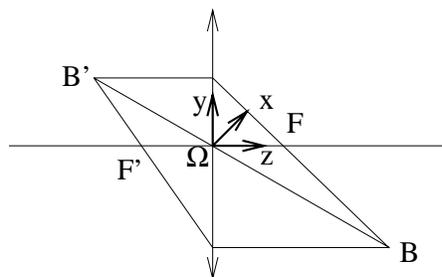


FIG. 1.22 - Le modèle des systèmes minces.

On place un plan de capteurs  $P$  : ce plan est orthogonal à l'axe optique, son intersection avec cet axe est le point  $C$ , la cote de ce point est  $\gamma$ . Soit un point  $B$ , dit point objet. On cherche à étudier l'image de ce point sur le plan des capteurs  $P$  (cf. FIG. 1.23).

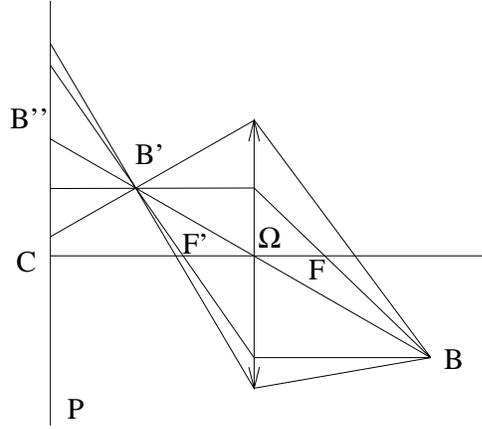


FIG. 1.23 - Le modèle des systèmes minces avec capteur.

Le point  $B'$ , appelé point image de  $B$ , est défini comme étant l'intersection de trois rayons lumineux :

1. le rayon  $(B\Omega)$  qui passe aussi par  $B'$ ,
2. le rayon parallèle à  $\Omega z$  qui coupe le rayon  $(BF)$  sur la lentille,
3. le rayon passant par  $F'$  qui coupe sur la lentille le rayon parallèle à  $\Omega z$  qui passe par  $B$ .

Cela donne, dans un système de coordonnées centrées en  $\Omega$  :

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto B' \begin{pmatrix} \frac{xf}{f-z} \\ \frac{yf}{f-z} \\ \frac{zf}{f-z} \end{pmatrix} \quad \text{et} \quad B \begin{pmatrix} \frac{x'f}{f+z'} \\ \frac{y'f}{f+z'} \\ \frac{z'f}{f+z'} \end{pmatrix} \mapsto B' \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \quad (1.15)$$

Le point  $B$  donne une image nette dans le plan des capteurs si  $B' \in P$ . L'ensemble des points qui ont une image nette dans  $P$ , s'appelle le plan de netteté, c'est un plan parallèle à  $P$ , de cote  $z = \gamma f / (f + \gamma)$ . Quand  $\gamma$  tend vers  $-f$ , ce plan s'éloigne vers l'infini.

Dans le cas où le système de coordonnées est centré en  $C$  avec les mêmes axes que précédemment,  $\gamma$  désigne alors la cote du point  $\Omega$ . Les formules (1.15) deviennent alors :

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto B' \begin{pmatrix} \frac{xf}{\gamma+f-z} \\ \frac{yf}{\gamma+f-z} \\ \gamma + \frac{(z-\gamma)f}{\gamma+f-z} \end{pmatrix} \quad \text{et} \quad B \begin{pmatrix} \frac{x'f}{f+z'-\gamma} \\ \frac{y'f}{f+z'-\gamma} \\ \frac{\gamma z' - \gamma^2 + z'f}{f+z'-\gamma} \end{pmatrix} \mapsto B' \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \quad (1.16)$$

Dans ces coordonnées, le plan de netteté a pour cote  $\gamma^2 / (\gamma - f)$ . En inversant cette formule, on peut calculer la cote que doit avoir la lentille pour qu'un point de cote  $z$  soit net :

$$\gamma = \frac{z - \sqrt{z^2 - 4zf}}{2}$$

Le phénomène de mise au point correspond au fait que le plan de netteté a une cote variable lorsque  $\gamma$  varie. Pour  $\gamma = f$ , il est à l'infini, et il se rapproche du point  $F$  lorsque  $\gamma$  augmente.

### Image d'un point sur l'écran

Dès que les objets observés ne sont pas parfaitement plans et orthogonaux à l'axe optique, tous les points observés ne peuvent appartenir au plan de netteté. Quelle est alors leur image sur le plan des capteurs ?

Supposons que la lentille est circulaire de rayon  $R$ , et centrée sur l'axe optique. Considérons le rayon lumineux qui part d'un point  $B$  observé de coordonnées  $(x, y, z)$  et qui passe par le point  $L$  de la lentille de coordonnées  $(l \cos \theta, l \sin \theta, \gamma)$ . Ce rayon coupe le plan des capteurs en un point de coordonnées :

$$\begin{cases} x'' &= l \cos \theta \left(1 + \frac{\gamma z - \gamma - f}{f \gamma - z}\right) + \frac{\gamma x}{\gamma - z} \\ y'' &= l \sin \theta \left(1 + \frac{\gamma z - \gamma - f}{f \gamma - z}\right) + \frac{\gamma y}{\gamma - z} \\ z'' &= 0 \end{cases}$$

En considérant tous les rayons qui partent de  $B$  et qui passent par un point de la lentille, on obtient l'image de  $B$ . Pour cela, on fait varier  $l$  et  $\theta$  dans les formules précédentes : l'image de  $B$  est un cercle de centre  $B''$  et de rayon  $R_B$  avec :

$$B'' \left( \begin{array}{c} \frac{\gamma x}{\gamma - z} \\ \frac{\gamma y}{\gamma - z} \\ 0 \end{array} \right) \quad \text{et} \quad R_B = R \left( 1 + \frac{\gamma z - \gamma - f}{f \gamma - z} \right)$$

L'existence d'un rayon  $R_B$  explique le phénomène de flou. Lorsque le point projeté n'est pas dans le plan de netteté, il se projette non en un point, mais en un cercle de rayon non nul. Les points projetés formant un continuum, leurs cercles images se chevauchent et on ne distingue plus bien la contribution de chacun : l'image est floue.

Le centre de l'image de  $B$  se déduit de ce point par une transformation projective. En coordonnées homogènes :

$$\left( \begin{array}{c} \frac{\gamma x}{\gamma - z} \\ \frac{\gamma y}{\gamma - z} \\ 1 \end{array} \right) \propto \left( \begin{array}{cccc} \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & -1 & \gamma \end{array} \right) \left( \begin{array}{c} x \\ y \\ z \\ 1 \end{array} \right)$$

Si on néglige les flous, on peut donc représenter la transformation réalisée par la caméra par une transformation projective.

#### 1.3.2 L'approximation des systèmes épais

Lorsque le système optique de la caméra est plus épais, le modèle précédent n'est plus applicable. Il faut utiliser un modèle à un paramètre supplémentaire, l'interstice. Celui-ci est relié à l'épaisseur du système optique, mais aussi aux propriétés des dioptries qui le forment [Sua 74, Lav 91].

Dans l'approximation des systèmes épais, un système est caractérisé par son axe optique, qui est un axe de révolution cylindrique, par ses focales image et objet, et par deux plans orthogonaux à l'axe optique, appelés plans principaux image et objet. Ces plans sont distants d'une longueur  $i$  qui est l'interstice, chaque point focal étant distant du plan qui lui correspond d'une distance  $f$  appelée focale du système (cf. FIG. 1.24).

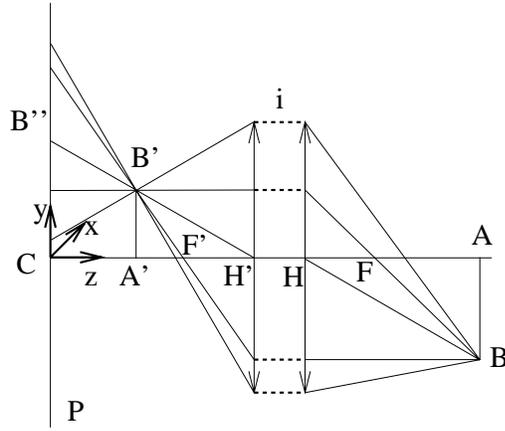


FIG. 1.24 - Système optique épais avec plan de capteurs.

On note  $F$  et  $F'$  les points focaux,  $H$  et  $H'$  les intersections des plans principaux avec l'axe optique,  $B$  un point quelconque,  $B'$  son image par le système optique, et  $A$  et  $A'$  les projections orthogonales de  $B$  et  $B'$  sur l'axe optique. On a :  $f = \overline{HF} = -\overline{H'F'}$ . On se place dans un repère centré en  $C$ , point d'intersection entre le plan des capteurs et l'axe optique. En notant  $(x, y, z)$  les coordonnées du point  $B$  et  $(x', y', z')$  celles de  $B'$ , le point  $B'$  est déterminé par les deux lois suivantes :

$$\frac{1}{\overline{HA}} - \frac{1}{\overline{H'A'}} = \frac{1}{f} = \frac{1}{z - \gamma - i} - \frac{1}{z' - \gamma}$$

$$\frac{\overline{A'B'}}{\overline{AB}} = \frac{\overline{H'A'}}{\overline{HA}} = \frac{y'}{y} = \frac{z' - \gamma}{z - \gamma - i}$$

Par rapport au modèle précédent, cela revient à couper la lentille en deux et à séparer les deux sous-espaces image et objet d'une distance  $i$ , sans modifier les angles d'incidence des rayons lumineux sur chacun des "moitié de plan" qui sont les plans principaux. A partir des formules :

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto B' \begin{pmatrix} \frac{x f}{\gamma + f + i - z} \\ \frac{y f}{\gamma + f + i - z} \\ \gamma + \frac{(z - \gamma - i) f}{\gamma + f + i - z} \end{pmatrix} \quad \text{et} \quad B \begin{pmatrix} \frac{x' f}{f + z' - \gamma} \\ \frac{y' f}{f + z' - \gamma} \\ \gamma + i + \frac{f(z' - \gamma)}{z' - \gamma + f} \end{pmatrix} \mapsto B' \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

on déduit l'existence d'un plan de netteté de cote  $z$  définit par :

$$z = \frac{\gamma^2 + \gamma i - i f}{\gamma - f} \quad \text{ou} \quad \gamma = \frac{z - i - \sqrt{(z - i)^2 - 4f(z - i)}}{2}$$

Si le système optique est circulaire de rayon  $R$  autour de l'axe optique, l'image d'un point  $B$  quelconque est un cercle de centre  $B''$  et de rayon  $R_B$ , où les coordonnées homogènes de  $B''$  dans le plan des capteurs sont :

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

et :

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

D'un point de vue géométrique, c'est à dire lorsqu'on ignore les caractéristiques réelles du système optique, ce deuxième modèle n'apporte rien de nouveau par rapport au modèle précédent : il y a des flous, et lorsqu'on les néglige, la transformation est projective. Certes, cette transformation a des paramètres différents, mais soit il faut les déterminer par étalonnage, ce qui était aussi nécessaire avec le modèle précédent, soit il suffit de savoir que la transformation est projective.

### 1.3.3 Autres modèles

On peut utiliser d'autres modèles, qui sont de nature géométrique plutôt qu'optique. Ils sont utilisés soit comme approximation géométrique des modèles optiques qui précèdent, soit comme approximation des systèmes optiques eux-mêmes.

**Modèle de projection perspective.** Une caméra est définie par un centre de projection  $\Omega$  et un plan des capteurs  $P$  (cf. FIG. 1.25).

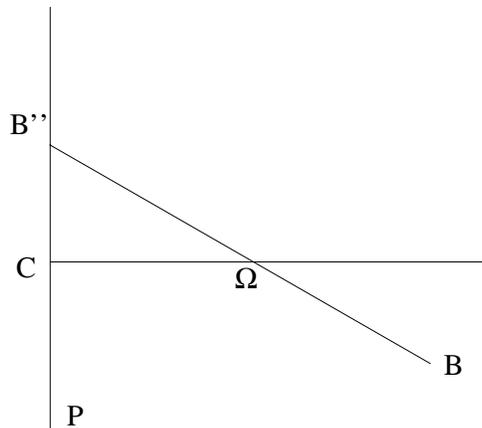


FIG. 1.25 - Modèle de la projection perspective.

On appelle  $C$  la projection orthogonale de  $\Omega$  sur  $P$ . La distance  $C\Omega = \gamma$  est appelée distance focale de la projection. Il convient de bien différencier la focale d'une projection perspective, qui est la distance entre la centre de projection et le plan de projection, de la focale d'un système optique mince ou épais, qui ne dépend pas, elle, du plan des capteurs. La droite  $(C\Omega)$  est appelée axe optique ou plus justement axe principal de la projection.

La transformation effectuée peut s'exprimer comme suit dans un repère centré en  $C$  :

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto B'' \begin{pmatrix} \frac{x f}{\gamma - z} \\ \frac{y f}{\gamma - z} \\ 0 \end{pmatrix}$$

**Modèle de projection affine.** La projection affine, dite encore projection parallèle ou orthographique, est un cas particulier de projection perspective où le centre de projection est à l'infini. Une caméra est alors définie par un plan de capteurs ou plan de projection et par une direction de projection. La transformation réalisée est affine. Un changement de coordonnées dans le plan permet de faire des changements d'échelle dans la projection obtenue.

Lorsque la direction de projection est orthogonale au plan des capteurs et que la transformation effectuée dans le plan est euclidienne, la projection obtenue est orthogonale.

### 1.3.4 De la caméra aux données

Les modèles précédents, qu'ils soient de nature plutôt optique ou géométrique, ne sont malheureusement que des modèles ne rendant pas complètement compte de la réalité. Tout d'abord, la chaîne d'acquisition des données ne comporte pas la seule partie optique. Ensuite, du bruit vient brouiller le signal à chaque étape de traitement. Il convient donc d'étudier cette chaîne d'acquisition et le bruit qu'elle apporte, en faisant un petit panorama des solutions proposées dans la littérature pour remédier à ces effets.

#### Chaîne d'acquisition des données

Elle se compose schématiquement des étapes suivantes :

1. La lumière émise par les différentes parties de la scène observée est projetée sur un plan de capteurs par l'intermédiaire d'un système optique.
2. Ce plan de capteurs est une matrice de diodes photomultiplicatrices ou de CCD, qui transforment le signal lumineux reçu en un courant électrique, proportionnel au nombre de photons reçu pendant un temps donné.
3. Le signal obtenu est analogique sur une durée très brève et discret en espace. Il est converti en signal analogique bidimensionnel, puis converti en digital par une carte spécialisée.
4. L'information se trouve alors sous la forme d'une matrice entière, chaque élément indiquant un niveau de gris. Les lignes de contraste, c'est à dire de plus fort gradient sont extraites par lissage, calcul du gradient et hystérésis [Can 86, Der 87a].
5. Les points de contour ainsi obtenus sont chaînés. Les chaînes sont approchées par des lignes droites ou des courbes, telles les B\_splines [Arb 91a, Arb 91b].

#### Quelques problèmes et solutions

**Correction des distorsions.** Chacune des étapes précédentes est cause de défauts. Tout d'abord, le système optique utilisé diffère du modèle de la projection perspective en deux points : certains points sont projetés de manière floue, et même en négligeant ces flous, la projection n'est pas perspective. On peut toutefois se ramener à une telle projection, en transformant l'image point à point, selon une méthode développée par Peuchot [Peu 92] et Brand [?].

On procède pour cela en deux temps. On prend une image d'une grille parfaitement carrée. L'image de cette grille n'est pas parfaite : les projections des droites sont courbes et ne sont plus équidistantes, les angles ne sont plus droits. On calcule donc la transformation qu'il faut faire subir à chaque point de la grille pour le ramener sur un point d'une grille parfaite arbitraire (parfaite signifie que les lignes sont droites et équidistantes, et les angles droits) (cf. FIG. 1.26). La transformation associée de chaque autre point de l'image est calculé par interpolation des transformations des points de la grille entourant ce point.

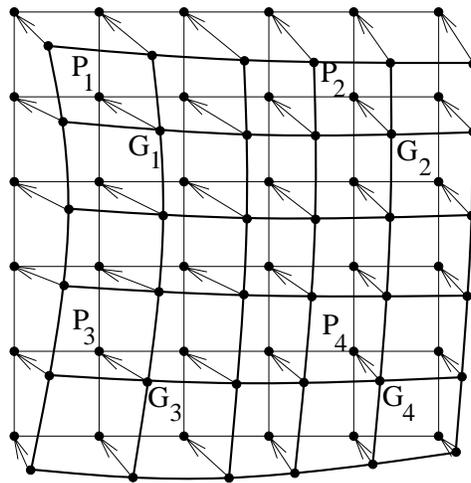


FIG. 1.26 - Calcul d'un déplacement point par point.

On choisit alors quatre points  $G_1, G_2, G_3$  et  $G_4$  de la projection de la grille initiale comme point de référence. Lors de la transformation calculée précédemment, ces points sont devenus  $P_1, P_2, P_3$  et  $P_4$ . On calcule alors la transformation projective qui transforme ces quatre derniers points en  $G_1, G_2, G_3$  et  $G_4$  respectivement (cf. FIG. 1.27).

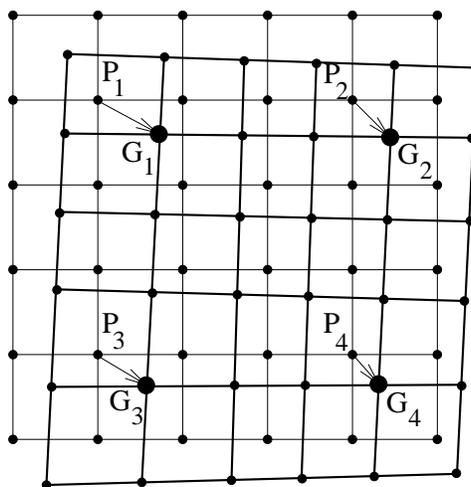


FIG. 1.27 - Calcul d'une transformation projective.

La composée de la première transformation et de cette deuxième transformation projective donne la correction à appliquer à chaque image. Lorsqu'on dispose d'une nouvelle image, on lui applique cette correction, et on obtient ainsi l'image qu'aurait donné une "caméra projective". Cette méthode peut être mise en œuvre avec une très grande précision (0,02 pixels).

Une autre possibilité consiste à ne pas utiliser le modèle projectif, mais un modèle plus riche, tenant compte de la distorsion radiale. Il est alors nécessaire d'étalonner les caméras pour déterminer ces paramètres [Bro 71, Tsa 87].

**Divers autres bruits.** En dehors des distorsions, de nombreux autres bruits affectent le signal [Bey 92]. L'objectif de la caméra est cause des flous, des distorsions, des aberrations chromatiques.

La partie de traitement du signal par des dispositifs électroniques est riche en sources de bruit. Les CCD ne sont pas jointifs, et de petits cristaux déforment le signal pour que ces CCD captent plus de photons. Les éléments électroniques sont sensibles à la température : un changement de celle-ci peut entraîner une variation de 5 pixels dans le résultat final.

Tous ces traitements agissent à la fois sur la géométrie et la radiométrie du signal. La carte de digitalisation procède par rééchantillonnage, ce qui entraîne une perte à la fois sur l'intensité de chaque pixel et sur leur localisation exacte.

**Autres traitements.** Le flou ne peut être supprimé pour toute l'image dès que la scène observée a une profondeur non nulle. Ce flou peut dans certains cas être utilisé. Tout d'abord pour calculer la profondeur relative des objets de la scène [Pen 87], ou pour faire une mise au point automatique (autofocus optique passif) [Kro 87]. Cette mise au point permet de limiter les effets de flou, à défaut de les supprimer. L'effet de ces flous est renforcé par le fait que les capteurs ne sont pas ponctuels.

La détection des contours est généralement faite en utilisant des filtres qui calculent le gradient ou le laplacien de la fonction d'intensité lumineuse, puis qui cherche les régions homogènes ou les contours de discontinuité. Ces filtres sont optimaux dans certaines hypothèses qui ne sont malheureusement que très mal respectées dans les images. Par exemple, le filtre de Deriche [Der 87a] est optimal pour la détection en 1D du signal en marche d'escalier (fonction de Heavyside) bruitée par un bruit blanc gaussien. Dans les images, le signal est bidimensionnel et le filtre n'est pas isotrope. De plus, les fonctions de Heavyside ne sont pas les seules qu'il faudrait détecter. Parmi les approches novatrices, on peut citer [Sal 92] qui part d'une analyse plus fine de la notion de contour et qui développe un système donnant des résultats nettement meilleurs. La contre partie est un temps de calcul cinq à dix fois plus long, et un algorithme non vectorisable, même si parallélisable sur une machine MIMD.

Une fois le modèle de caméra établi, deux attitudes sont possibles. Tout d'abord, l'étalonnage, qui consiste à calculer les paramètres de ce modèle [Len 87, Tsa 87, Fau 92c]. On peut alors reconstruire la scène observée en 3D en utilisant plusieurs images. Malheureusement, cet étalonnage est souvent instable, peu précis ou impossible, dans le cas d'une caméra mobile par exemple. Certaines techniques essaient alors de n'utiliser que certains des paramètres, les paramètres intrinsèques ou intérieurs par exemple [Hor 89, Yua 89, Liu 90]. D'autres auteurs ont essayé de ne se servir que des

propriétés de la géométrie projective : si on suppose que la transformation est projective, cela fournit suffisamment de contraintes pour faire du positionnement relatif [Moh 91a, Moh 92a], de la reconstruction projective [Fau 92a, Moh 92b, Har 92a] ou du calcul d'indices tridimensionnels [Gro 92b].

## 1.4 Conclusion du chapitre

Ce premier chapitre décrit deux théories classiques : la géométrie projective et la théorie des invariants. Celles-ci sont détaillées en fonction de leurs applications en vision. La géométrie est ainsi présentée en mettant l'accent sur les différentes manières d'associer et de construire des birapports et des invariants à partir de figures simples, la théorie des invariants permet, elle, de répondre aux questions suivantes : combien existe-t-il d'invariants dans un problème donné, et comment trouver ces invariants ?

La dernière partie du chapitre justifie l'emploi de ces deux théories en vision, et décrit très succinctement les traitements de bas niveau des images. Les applications de ces théories à la vision font l'objet du chapitre suivant.



## Chapitre 2

# Géométrie, invariants et vision : applications

Le présent chapitre présente trois applications des théories du chapitre précédent à des problèmes de vision : ces applications, bien qu'illustrant des théories mathématiques, sont au cœur de véritables problèmes de vision, et elles sont donc abordées tant du point de vue théorique qu'expérimental. En particulier, les algorithmes proposés sont testés avec des données réelles.

La première application est la recherche des points coplanaires d'une scène 3D. La deuxième est le calcul d'invariants 3D caractéristiques d'un objet, à partir de deux images de cet objet. La troisième est l'appariement de deux images. Cet appariement, supposé réalisé dans ce qui précède, est un problème difficile. Comment reconnaître dans deux images, les éléments qui se correspondent ? Pour résoudre ce problème, nous proposons une méthode basée sur l'approximation du mouvement 2D des indices dans l'image par des transformations simples, puis sur l'utilisation des invariants associés à ces transformations.

Chacune de ces applications fait l'objet d'un paragraphe du chapitre. Chacun de ces paragraphes contient une partie théorique et des expérimentations. Le deuxième paragraphe utilise la construction des projections de points coplanaires exposée dans le premier, et est donc une suite logique de celui-ci. Le troisième paragraphe est par contre indépendant des deux premiers. Il se situe dans un cadre moins mathématique et plus expérimental, plus proche des problématiques classiques de la vision.

## 2.1 Géométrie épipolaire et coplanarité

Dans ce paragraphe, on s'intéresse au problème suivant : lorsque des points de l'espace 3D sont observés dans deux images, peut-on retrouver, à partir de la seule observation des deux images, en supposant ces images appariées, quels sont les points qui sont coplanaires dans l'espace 3D ? Des images sont appariées lorsque l'on sait retrouver dans chaque images les traces d'un même point de la scène 3D observée. Ces traces, qui sont des points, sont dites en correspondance ; appairer ou mettre en correspondance des images consiste à retrouver ces correspondances.

Pour résoudre ce problème, trois méthodes sont envisageables. Si la caméra utilisée est étalonnée, c'est à dire que l'on connaît tous les paramètres de la transformation qu'elle réalise, on peut reconstruire en trois dimensions la scène observée à partir des deux images, en utilisant une technique classique de triangulation. On peut alors travailler sur la reconstruction obtenue.

Lorsque la caméra n'est pas étalonnée, il est encore possible de faire une reconstruction de la scène à une homographie de l'espace près [Fau 92a, Moh 92b]. La coplanarité étant invariante par transformation projective, il est possible de rechercher les points coplanaires à partir de cette reconstruction partielle. La dernière méthode, que nous présentons ici, consiste à rechercher la coplanarité sans reconstruction ni étalonnage, mais en utilisant les contraintes fournies par la géométrie épipolaire.

Les deux premières méthodes présentent quelques inconvénients. La première nécessite un étalonnage : ceci n'est théoriquement pas nécessaire, comme le prouve la seconde méthode, et est contraignant et peu robuste en pratique. En particulier, l'incertitude sur la profondeur des points peut être grande si le mouvement de la caméra entre les deux prises de vues est faible, ce qui peut entraîner une mauvaise détection des plans fronto-parallèles.

Dans la deuxième méthode, la difficulté est l'établissement d'une mesure de coplanarité. Les points reconstruits sont dans un espace projectif ni affine ni euclidien, et on ne peut donc utiliser ni angle, ni volume, ni distance. Deux points proches dans la scène initiale peuvent être reconstruits à une distance euclidienne infinie. La coplanarité exacte peut toujours être vérifiée aisément, mais on ne dispose d'aucune mesure de coplanarité approchée qui puisse être utilisée lorsque les données sont bruitées. En effet, on ne peut calculer aucune quantité numérique projective à partir de quatre points non strictement coplanaires.

La méthode que nous présentons pallie ces inconvénients. On suppose qu'on ne connaît que les projections des points dans les deux images et la correspondance entre ceux-ci. Aucune hypothèse n'est faite sur le mouvement de la caméra ou sur son étalonnage. Il suffit que le mouvement de la caméra ne soit pas nul, et que la géométrie épipolaire des deux images puisse être calculée.

La recherche est alors basée sur l'approximation par une projection perspective de transformation réalisée par la caméra. Le fait que la projection est perspective fournit des contraintes, réunies sous le vocable de géométrie épipolaire. Les deux projections d'un même point 3D sont liées par une relation quadratique, mais qui n'est que linéaire en fonction des coordonnées de chacun des points. Une utilisation astucieuse de ces contraintes et des mesures de birapports permet alors de retrouver les points coplanaires.

La suite du paragraphe est organisée comme suit. Après une brève présentation de

la géométrie épipolaire et de quelques méthodes de calcul de la matrice fondamentale, le test de coplanarité est exposé. Une utilisation inversée de ce test permet de “construire les projections de nouveaux points coplanaires”. La stabilité du test est précisément étudiée sur des données simulées, et sa validité est montrée sur des données réelles.

### 2.1.1 Introduction à la géométrie épipolaire

Lorsque l'on prend deux images d'un même ensemble de points 3D, il existe des relations géométriques entre les projections de ces points. La géométrie épipolaire exprime ces relations.

**Définition 18** Soient deux caméras définies par leur centre de projection  $O$  et  $O'$  respectivement, et leur plan de projection respectif  $P$  et  $P'$ . Soit  $M$  un point de l'espace et  $m$  et  $m'$  ses projections par chacune des caméras (cf. FIG. 2.1).

Si, parmi les trois points  $M$ ,  $m$  et  $m'$ , on ne connaît que la position de  $m$ , on sait alors que le point  $M$  est sur sa ligne de vue, c'est à dire la droite  $(Om)$ , et par conséquent, le point  $m'$  appartient à la projection de cette droite sur la deuxième image. Cette droite est appelée **droite épipolaire**. Elle passe par la projection  $e'$  de  $O$  sur la deuxième image, qui est appelé **épipole**.

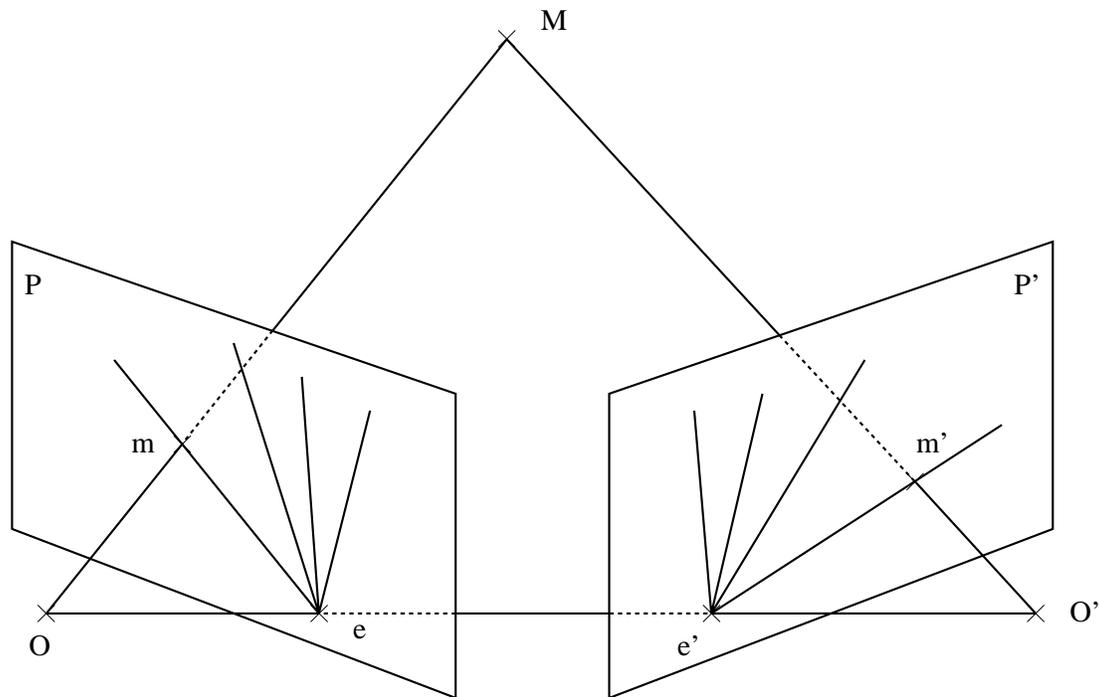


FIG. 2.1 - Définition de la géométrie épipolaire.

Réciproquement, on peut aussi associer une droite épipolaire à tout point de la deuxième image, toutes ces droites passant par la projection  $e$  de  $O'$  sur la première image.

En considérant le plan  $(OO'M)$ , on voit que tous les points de la droite  $(em)$  ont pour droite épipolaire associée la même droite  $(e'm')$ . On dispose ainsi d'une relation

entre droites épipolaires. Cette relation exprime le fait que les droites  $(em)$  et  $(e'm')$  sont les projetées d'un même plan  $(OO'M)$ . Cette relation est donc homographique.

La FIG. 2.2 montre deux images d'une même scène. Des paires de points appariés ont été sélectionnées, et les droites épipolaires correspondantes ont été tracées.

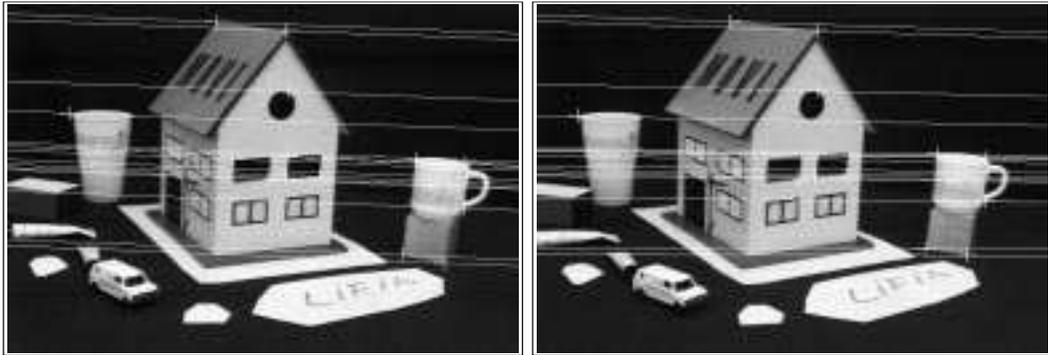


FIG. 2.2 - Droites épipolaires associées à quelques points de deux images.

**Définition 19** On introduit une matrice  $F$  appelée **matrice fondamentale** qui caractérise cette relation [Fau 92c, Luo 92]. Si  $x$  et  $x'$  sont des vecteurs de coordonnées homogènes de  $m$  et  $m'$  respectivement, on a la relation :  ${}^t x' F x = 0$ . Si  $l'$  est un vecteur de coordonnées homogènes de la droite épipolaire correspondant à  $m$ , on a :  $l' = F x$ . Toutes les droites engendrées par cette équation lorsque  $x$  varie passent par un point fixe,  $e'$ . L'image de  $F$  est donc un espace de dimension 1, exprimé en coordonnées homogènes : le rang de la matrice  $F$  est 2. Cette matrice étant définie à un facteur de proportionnalité près, elle possède donc sept degrés de liberté.

### Calcul de la matrice fondamentale

Chaque fois que l'on sait que deux points, chacun appartenant à une image, sont appariés, c'est à dire sont les projections dans les deux images d'un même point 3D, on peut écrire une relation  ${}^t x' F x = 0$ . Ceci nous fournit une équation. La matrice  $F$  possédant sept degrés de liberté, sept telles équations devraient permettre de déterminer entièrement cette matrice. Tel est bien le cas, à condition que les points 3D ne soient pas dans des positions dégénérées, ou sur des surfaces critiques. La résolution mène alors à un nombre fini de possibilités [Fau 92c].

Lorsque l'on connaît huit appariements de points entre les deux images, on peut négliger la contrainte  $\text{rg} F = 2$  et résoudre le système de manière linéaire [Fau 92a, Har 92b, Har 92a]. Une autre méthode linéaire est possible à partir de six appariements seulement, si on sait que quatre des points 3D sont coplanaires. Cette méthode est présentée ici à titre de simple illustration de la notion de géométrie épipolaire. Pour une lecture rapide, on peut passer directement à la conclusion et au paragraphe 2.1.2.

### Calcul de la matrice fondamentale à partir de six appariements de points.

Soient  $A, B, C, D, E$  et  $F$  six points 3D, tels que quatre d'entre eux, par exemple  $A, B, C$  et  $D$ , soient dans un même plan  $P$  (cf. FIG. 2.3). A partir de deux points  $O_1$  et  $O_2$ , on prend deux images des six points. Notons  $E' = P \cap (O_1 E)$ ,  $E'' = P \cap (O_2 E)$ ,

$F' = P \cap (O_1F)$  et  $F'' = P \cap (O_2F)$ . Les projetés de tous ces points sont notés avec des minuscules et un indice 1 ou 2 suivant l'image à laquelle ils appartiennent.

Comme  $O_1$ ,  $E$  et  $E'$  sont colinéaires,  $e_1$  et  $e'_1$  sont en fait le même point. Les coordonnées projectives de  $e'_1$  dans la base  $(a_1, b_1, c_1, d_1)$  peuvent être calculées. Elles sont aussi les coordonnées de  $E'$  dans la base  $(A, B, C, D)$  et celles de  $e'_2$  dans la base  $(a_2, b_2, c_2, d_2)$ . Cela est dû à la conservation des coordonnées projectives par les collinéations. Les points  $e''_1$ ,  $f'_2$  et  $f''_1$  peuvent être construits de la même manière.

Par construction, les points  $O_1, O_2, E, E'$  et  $E''$  sont coplanaires. Ainsi sont aussi  $O_1, O_2, F, F'$  et  $F''$ . L'intersection de ces deux plans est la droite  $(O_1O_2)$  et l'intersection de cette droite avec le plan  $P$  est l'intersection des deux droites  $(E', E'')$  et  $(F', F'')$ . Notons  $G$  ce point.

Il est possible de construire les projections de  $G$  sur chaque image. :

$$\begin{cases} g_1 &= (e'_1e''_1) \cap (f'_1f''_1) \\ g_2 &= (e'_2e''_2) \cap (f'_2f''_2) \end{cases}$$

Comme les trois points  $O_1, O_2$  et  $G$  sont colinéaires,  $g_1$  et  $g_2$  sont les épipoles des deux images. Un point quelconque  $h_1$  de la première image peut être interprété comme la projection d'un point 3D  $H$  appartenant au plan  $P$ . La droite épipolaire associée à  $h_1$  dans la deuxième image est la droite  $(g_2h_2)$ , où  $h_2$  est le point ayant les mêmes coordonnées projectives dans la base  $(a_2, b_2, c_2, d_2)$  que  $h_1$  dans  $(a_1, b_1, c_1, d_1)$  et  $H$  dans  $(A, B, C, D)$ .

Par cette méthode, il est donc possible de calculer l'épipole de chaque image et de déterminer la droite épipolaire associée à chaque point d'une des images. La géométrie épipolaire est donc complètement déterminée.

**Conclusion.** Le grave inconvénient pratique des méthodes évoquées ci-dessus est de ne se servir que d'un nombre minimal de données et de calculer une solution de manière exacte. Tout bruit dans les données est reporté dans le résultat. La solution à ce problème, et qu'il faut donc utiliser dans la pratique, consiste à partir d'appariements plus nombreux et à utiliser un algorithme d'optimisation non linéaire avec contrainte [Fau 92c, Moh 92b]. La redondance des données permet alors de minimiser les effets du bruit et d'obtenir des résultats précis. On se reportera aux articles cités pour le détail des calculs mathématiques.

### 2.1.2 Un test de coplanarité

Supposons que nous disposons de deux images et que nous connaissons la géométrie épipolaire associée à ces deux images. Etant données les projections de quatre points dans ces deux images, il est possible de dire si ces quatre points, dont on ne connaît pas la position tridimensionnelle, sont coplanaires ou non. Les applications d'un tel test de coplanarité sont nombreuses en vision. Pour notre part, nous l'utiliserons comme prédiction, pour calculer les projections dans les deux images de l'intersection d'une droite et d'un plan. Cela nous permettra de calculer de nombreux invariants 3D directement à partir des images sans aucune reconstruction en 3D. Cela sera fait au paragraphe 2.2.

**Notations.** Soient  $A, B, C$  et  $D$  quatre points de l'espace tridimensionnel, et soient  $a_1, b_1, c_1, d_1$ , et  $a_2, b_2, c_2, d_2$  leurs projections respectives dans deux images

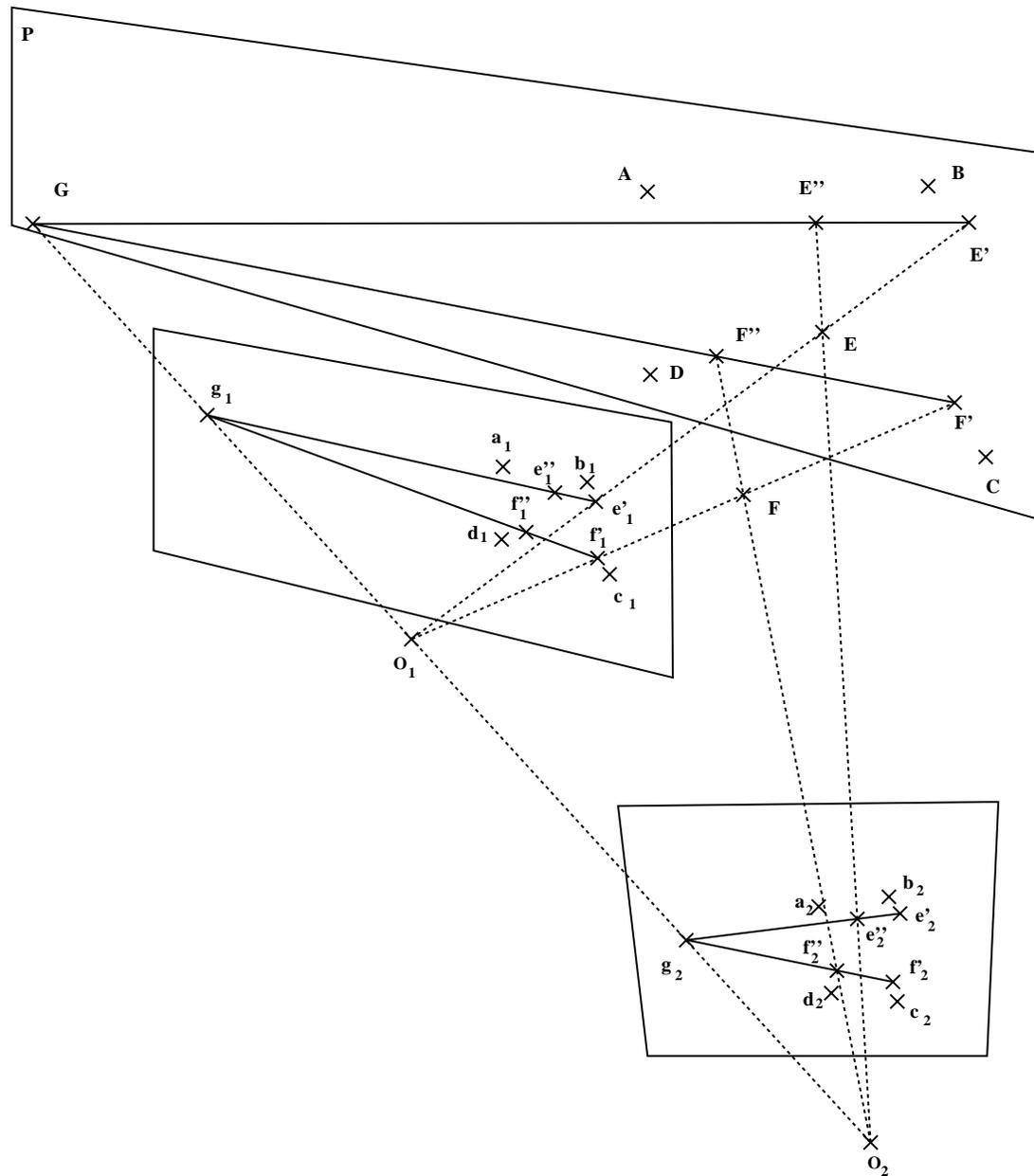


FIG. 2.3 - Détermination de la géométrie épipolaire à partir de six appariements de points.

(cf. FIG. 2.4). Soient  $e_1$  et  $e_2$  les épipoles des deux images. On suppose que les projections des points ne sont pas confondues avec les épipoles et que ces points sont étiquetés de telle sorte que  $(a_1, c_1, e_1)$ ,  $(b_1, d_1, e_1)$ ,  $(a_2, c_2, e_2)$  et  $(b_2, d_2, e_2)$  ne soient pas des triplets de points colinéaires. Notons  $f_1 = (a_1c_1) \cap (b_1d_1)$  et  $f_2 = (a_2c_2) \cap (b_2d_2)$ .

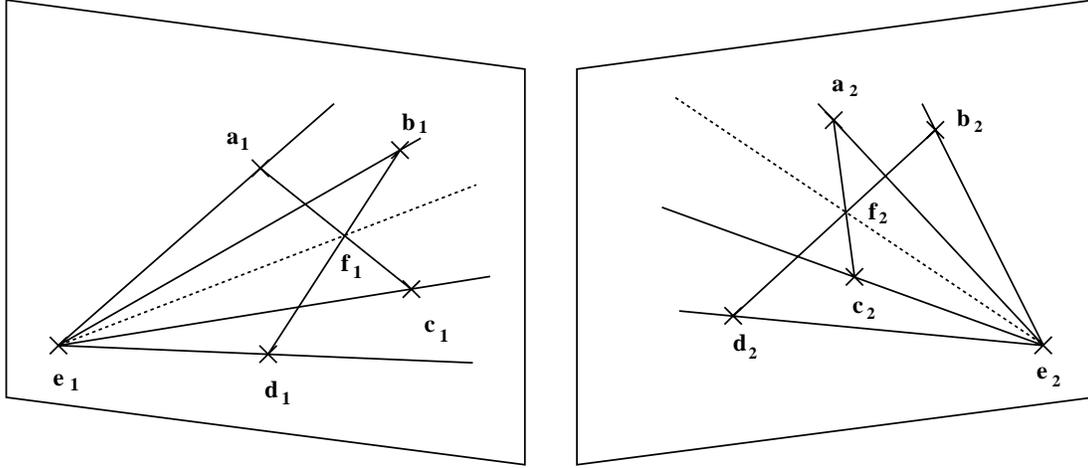


FIG. 2.4 - Test de coplanarité.

**Théorème 7 (Test de coplanarité)** *Les quatre points  $A$ ,  $B$ ,  $C$ , et  $D$  sont coplanaires si et seulement si les deux faisceaux de droites  $(e_1a_1, e_1b_1, e_1c_1, e_1d_1, e_1f_1)$  et  $(e_2a_2, e_2b_2, e_2c_2, e_2d_2, e_2f_2)$  définissent les mêmes birapports (les birapports d'un faisceau de droites sont définis au paragraphe 1.1.4 page 31).*

**Démonstration.** En effet, si ces quatre points sont effectivement coplanaires, les deux droites  $(AC)$  et  $(BD)$  se coupent en un point  $F$ . Soient  $f_1$  et  $f_2$  les projections de  $F$  sur les deux images. Les droites épipolaires formant deux faisceaux de droites homographiques, les faisceaux  $((e_1a_1), (e_1b_1), (e_1c_1), (e_1d_1), (e_1f_1))$  et  $((e_2a_2), (e_2b_2), (e_2c_2), (e_2d_2), (e_2f_2))$  sont eux-mêmes homographiques et définissent les mêmes birapports.

Deuxièmement, supposons les quatre points  $A, B, C$  et  $D$  non coplanaires, et plus précisément, supposons que  $A, B$  et  $C$  sont fixes, et que le point  $D$  se déplace le long de sa ligne de vue par rapport à la première image. Dans ces conditions,  $a_1, b_1, c_1$  et  $d_1$  sont des points fixes de la première image, ainsi que  $a_2, b_2$  et  $c_2$  dans la deuxième. Par contre, le point  $d_2$  se déplace le long de la droite épipolaire sur laquelle il se trouve. Comme nous avons supposé que  $a_2, c_2$  et  $e_2$  ne sont pas colinéaires,  $(a_2, c_2)$  n'est pas une droite épipolaire et le point  $f_2$  ne reste pas sur la même droite épipolaire. Les deux faisceaux de droites considérés ne sont plus homographiques, et ils ne définissent donc plus les mêmes birapports.

**Calcul.** Notons  $(x_{p_i}, y_{p_i})$  les coordonnées non homogènes du point  $p_i$  de l'image  $i$ , et  $(x_{p_i}, y_{p_i}, t_{p_i})$  ses coordonnées homogènes. Les coordonnées homogènes du point

$f_1$  sont définies comme suit à partir des coordonnées non-homogènes des points  $a_1, b_1, c_1$  et  $d_1$  ;

$$\begin{aligned} x_{f_1} &= (y_{a_1}x_{c_1} - x_{a_1}y_{c_1})(x_{d_1} - x_{b_1}) - (y_{b_1}x_{d_1} - x_{b_1}y_{d_1})(x_{c_1} - x_{a_1}) \\ y_{f_1} &= (y_{a_1}x_{c_1} - x_{a_1}y_{c_1})(y_{d_1} - y_{b_1}) - (y_{b_1}x_{d_1} - x_{b_1}y_{d_1})(y_{c_1} - y_{a_1}) \\ t_{f_1} &= (y_{d_1} - y_{b_1})(x_{c_1} - x_{a_1}) - (y_{c_1} - y_{a_1})(x_{d_1} - x_{b_1}) \end{aligned}$$

Les coordonnées du point  $f_2$  sont obtenues en remplaçant les indices 1 par 2 dans les formules précédentes.

Le test consiste à vérifier que deux faisceaux de cinq droites définissent les mêmes birapports. Prenons le cas d'un de ces birapports, les autres se déduisant par substitution d'un des points. On veut vérifier, par exemple :

$$\{(e_1a_1), (e_1b_1); (e_1c_1), (e_1f_1)\} = \{(e_2a_2), (e_2b_2); (e_2c_2), (e_2f_2)\}$$

ou, en utilisant les notations du paragraphe 1.1.4 :

$$\{e_1; a_1, b_1, c_1, f_1\} = \{e_2; a_2, b_2, c_2, f_2\}$$

Pour cela, on se sert de la formule de Möbius (1.1), ce qui donne :

$$\frac{|e_1a_1c_1||e_1b_1f_1|}{|e_1a_1f_1||e_1b_1c_1|} = \frac{|e_2a_2c_2||e_2b_2f_2|}{|e_2a_2f_2||e_2b_2c_2|}$$

avec, par exemple :

$$|e_1a_1c_1| = \begin{vmatrix} x_{e_1} & x_{a_1} & x_{c_1} \\ y_{e_1} & y_{a_1} & y_{c_1} \\ t_{e_1} & t_{a_1} & t_{c_1} \end{vmatrix}$$

**Utilisation.** Ce test de coplanarité peut s'utiliser de deux manières. Tout d'abord, pour rechercher des configurations de points coplanaires, ces points pouvant ensuite servir de balises pour un positionnement relatif, comme dans [Mor 93]. Dans ce cas, il suffit de construire les points  $f_1$  et  $f_2$  et de calculer les birapports définis par les faisceaux de droites ainsi constitués. On peut aussi se servir de ce test de manière constructive pour fabriquer des points coplanaires, ce qui va être fait dans la suite.

### Construction de points coplanaires

On reprend les notations du paragraphe précédent.

**Théorème 8 (Réciproque du test de coplanarité)** *Supposons que nous connaissons les projections dans deux images de trois points  $A, B$  et  $C$  et d'une droite  $L$ . Soient  $a_1, b_1, c_1, l_1, a_2, b_2, c_2$  et  $l_2$  ces projections. Si la géométrie épipolaire est connue, si  $(a_1, c_1, e_1), (b_1, d_1, e_1), (a_2, c_2, e_2)$  et  $(b_2, d_2, e_2)$  ne sont pas des triplets de points colinéaires, et si  $l_1$  et  $l_2$  ne passent pas par  $e_1$  et  $e_2$  respectivement,  $e_1$  et  $e_2$  étant les épipoles des deux images, alors il est possible de construire les projections du point  $D$  d'intersection entre  $L$  et le plan  $(ABC)$ .*

Deux méthodes sont possibles pour une telle construction. La première est simple et intuitive, mais elle n'est pas linéaire. La deuxième est linéaire et constructive, mais elle est géométriquement plus sophistiquée. Elle peut donc être omise en cas de lecture rapide.

**Première méthode de construction.** Cette méthode est une simple inversion du test de coplanarité.

1.  $e_1$  et  $e_2$  étant les épipoles des deux images, soit  $d_1$  un point quelconque de  $l_1$  (cf. FIG. 2.5). On calcule le birapport  $k$  du faisceau de droites  $(e_1a_1, e_1b_1, e_1c_1, e_1d_1)$ .

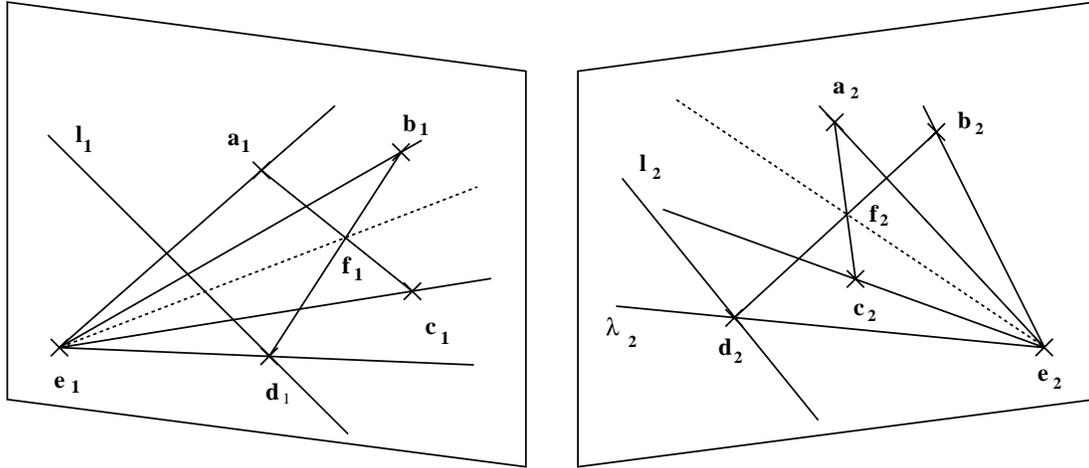


FIG. 2.5 - Construction de points coplanaires : première méthode.

2. Ce birapport définit une unique droite  $\lambda_2$  de l'image 2, qui passe par  $e_2$  et telle que le faisceau  $(e_2a_2, e_2b_2, e_2c_2, \lambda_2)$  ait  $k$  comme birapport.  $\lambda_2$  est donc la droite épipolaire associée dans la deuxième image au point  $d_1$ .
3. Soit  $d_2$  le point d'intersection entre  $\lambda_2$  et  $l_2$ , et soient  $f_1$  et  $f_2$  ceux de  $(a_1c_1)$  et  $(b_1d_1)$ , et  $(a_2c_2)$  et  $(b_2d_2)$  respectivement.
4. D'après le test de coplanarité,  $d_1$  et  $d_2$  sont les projections d'un point  $D$  de  $L$  coplanaire avec  $A, B$  et  $C$ , si et seulement si les deux faisceaux  $(e_1a_1, e_1b_1, e_1c_1, e_1f_1)$  et  $(e_2a_2, e_2b_2, e_2c_2, e_2f_2)$  ont les mêmes birapports.

Dans cette méthode, il faut résoudre une équation à une seule inconnue, la coordonnée de  $d_1$  sur  $l_1$ . Malheureusement, cette équation est quadratique, Il y a en effet une deuxième solution, qui correspond au point  $d_1$  tel que la droite  $(b_1d_1)$  passe par l'épipole  $e_1$ . Les calculs sont détaillés dans l'annexe A. On peut éviter cette non linéarité par la deuxième méthode.

**Deuxième méthode de construction.** Cette deuxième méthode est une méthode de construction géométrique des points, qui est linéaire, mais un peu plus compliquée que la précédente dans sa formulation. De même que précédemment, les calculs sont détaillés dans l'annexe A.

Etant données les projections de cinq points  $A, B, C, D$  et  $E$  dans deux images, on cherche dans ces images les projections du point  $J$  qui est l'intersection du plan  $(ABC)$  avec la droite  $(DE)$  (cf. FIG. 2.6). On suppose que la géométrie épipolaire entre les deux images est connue.

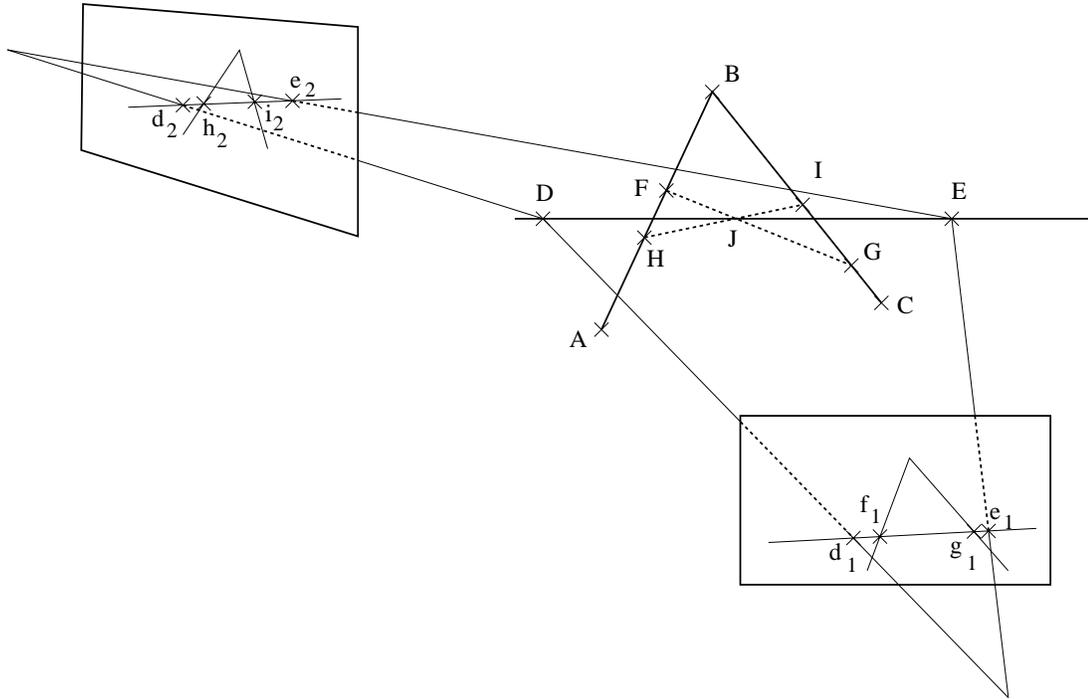


FIG. 2.6 - Deuxième méthode de construction de points coplanaires : vue 3D.

Le plan d'interprétation de la droite  $(DE)$  par rapport à la première image, c'est à dire le plan qui contient cette droite et le centre de projection de la première image, coupe respectivement  $(AB)$  et  $(BC)$  en  $F$  et  $G$ . Les droites  $(DE)$  et  $(FG)$  se coupent en un point qui est coplanaire avec  $A, B$  et  $C$  : c'est le point  $J$ . Dans la première image, la construction des projections de  $F$  et  $G$ ,  $f_1$  et  $g_1$  respectivement, est immédiate (cf. FIG. 2.7).

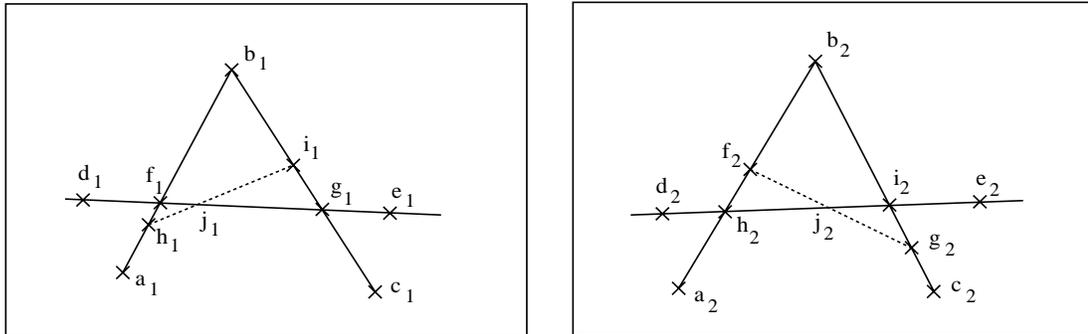


FIG. 2.7 - Deuxième méthode de construction de points coplanaires : les deux images.

On peut construire de manière similaire  $h_2$  et  $i_2$ , les projections dans la deuxième image des points  $H$  et  $I$ , intersections de  $(AB)$  et  $(BC)$  avec le plan d'interprétation de  $(DE)$  par rapport à cette deuxième image. A l'aide de la géométrie épipolaire, on peut alors construire  $h_1, i_1, f_2$  et  $g_2$ . Par exemple,  $h_1$  est l'intersection de la droite épipolaire associée à  $h_2$  dans la première image avec la droite  $(a_1b_1)$ . Comme  $J = (FG) \cap (HI)$ ,

on a  $j_1 = (f_1 g_1) \cap (h_1 i_1)$  et  $j_2 = (f_2 g_2) \cap (h_2 i_2)$ . Ce sont les points que nous cherchons.

Les calculs correspondants peuvent être effectués très simplement dans un formalisme vectoriel si l'on confond les points et les droites avec les vecteurs tridimensionnels de leurs coordonnées projectives. La droite définie par deux points ou le point intersection de deux droites se calculent alors par un simple produit vectoriel. Par exemple, la droite  $(d_1 e_1)$  a pour coordonnées  $d_1 \wedge e_1$ .

On obtient alors successivement, en notant  $F$  la matrice fondamentale :

$$\begin{aligned} f_1 &= (d_1 \wedge e_1) \wedge (b_1 \wedge a_1) & \text{et} & & f_2 &= (b_2 \wedge a_2) \wedge F.f_1 \\ g_1 &= (d_1 \wedge e_1) \wedge (b_1 \wedge c_1) & \text{et} & & g_2 &= (b_2 \wedge c_2) \wedge F.g_1 \\ h_2 &= (d_2 \wedge e_2) \wedge (b_2 \wedge a_2) & \text{et} & & h_1 &= (b_1 \wedge a_1) \wedge {}^t F.h_2 \\ i_2 &= (d_2 \wedge e_2) \wedge (b_2 \wedge c_2) & \text{et} & & i_1 &= (b_1 \wedge c_1) \wedge {}^t F.i_2 \end{aligned}$$

et finalement :

$$\begin{aligned} j_1 &= (h_1 \wedge i_1) \wedge (f_1 \wedge g_1) = (h_1 \wedge i_1) \wedge (d_1 \wedge e_1) \\ j_2 &= (f_2 \wedge g_2) \wedge (h_2 \wedge i_2) = (f_2 \wedge g_2) \wedge (d_2 \wedge e_2) \end{aligned}$$

### 2.1.3 Expérimentation

L'existence mathématique du test de coplanarité et de la construction de l'image de points coplanaires n'indique malheureusement guère si ces calculs sont utilisables dans la pratique. Les images dont on dispose sont en effet bruitées, et il est nécessaire de vérifier la stabilité des calculs vis à vis de ce bruit.

Pour cela, nous utilisons des expérimentations sur des données artificielles. Ces données permettent de faire varier précisément le niveau de bruit, et de voir ainsi quel est le niveau de bruit admissible en fonction des performances attendues. Cela permet aussi d'étudier plus particulièrement certaines configurations, ou, au contraire, d'éviter des configurations dégénérées.

D'un autre côté, la réalisation de tests sur de nombreuses données pseudo-aléatoires peut rendre l'interprétation délicate. Des configurations dégénérées peu probables dans la pratique peuvent en effet apparaître, et détériorer les résultats des calculs s'ils sont mesurés en terme d'erreur maximale par exemple.

Le bruit existant dans les images ne pouvant pas être parfaitement modélisé, des images réelles sont aussi utilisées pour étudier la pertinence du test.

#### Type de bruit utilisé

Pour notre part, nous avons choisi de travailler avec un bruit uniforme. Ce choix a été fait pour permettre un contrôle précis de l'amplitude du bruit, ce qui n'est pas possible avec un bruit gaussien, dont la variance peut changer, mais pour lequel toute amplitude est toujours possible même si peu probable. De plus, une partie du bruit affectant l'image, le bruit de discrétisation par exemple, est par nature plus proche du bruit uniforme que gaussien.

#### Comparaison des birapports

Beaucoup des tests faits conduisent à comparer des birapports. Ces birapports n'étant pas des grandeurs métriques, une comparaison par différence ou rapport n'a

pas de signification. A cause des propriétés de symétrie, on voudrait pouvoir poser que le birapport de valeur 1 est aussi éloigné de celui de valeur 0 que de celui de valeur  $\infty$ .

Pour faire ces comparaisons, nous avons utilisé la méthode des birapports aléatoires [Åst 92, Mor 93]. On choisit quatre points de manière uniforme dans un intervalle, et on calcule leur birapport. Cela permet de calculer la fonction de densité des valeurs du birapport sur la droite réelle, montrée sur la FIG. 2.8, et la fonction de répartition de ces valeurs, montrée sur la FIG. 2.9. Ces deux FIG. montrent le résultat du calcul théorique des deux fonctions et celui d'une simulation numérique à partir de tirages aléatoires de points.

Pour comparer deux birapports, on calcule alors la différence  $|F(x) - F(y)|$ , c'est à dire le "nombre" de birapports se situant entre les deux valeurs. Comme le point de l'infini joue le même rôle que les autres dans la droite projective, il faut envisager deux chemins pour aller de  $x$  à  $y$  : celui qui reste à distance finie, et celui qui passe par le point à l'infini. On prend donc comme distance entre birapports :

$$d(x, y) = \min(|F(x) - F(y)|, 1 - |F(x) - F(y)|) \quad (2.17)$$

avec :

$$F(x) = \begin{cases} 0 & \text{si } x = \infty \\ F_1(x) + F_3(x) & \text{si } \infty < x < 0 \\ 1/3 & \text{si } x = 0 \\ 1/2 + F_2(x) + F_3(x) & \text{si } 0 < x < 1 \\ 2/3 & \text{si } x = 1 \\ 1 + F_1(x) + F_2(x) & \text{si } 1 < x < \infty \end{cases}$$

$$F_1(x) = \frac{1}{3} \left( x(1-x) \ln \left( \frac{x-1}{x} \right) - x + \frac{1}{2} \right)$$

$$F_2(x) = \frac{1}{3} \frac{x - x \ln(x) - 1}{(x-1)^2}$$

$$F_3(x) = \frac{1}{3} \frac{(1-x) \ln(1-x) + x}{x^2}$$

Maybank, qui a travaillé indépendamment sur le problème, part d'une distribution de probabilité différente, mais arrive à une fonction de répartition très proche dans son aspect qualitatif et quantitatif, même si son expression analytique diffère nettement.

On pourrait penser utiliser d'autres fonctions, telles celle donnée par Berger [Ber 90, pages 159 et 160, tome 1] :

$$f : x \mapsto \frac{4}{27} \frac{(x^2 - x + 1)^3}{x^2(1-x)^2}$$

Une telle fonction est un invariant du birapport pour les permutations de l'ordre des points. Elle ne permet pas d'interpréter une expression telle  $|f(a) - f(b)|$  et donc de comparer des birapports, surtout lorsque l'ordre des points doit être pris en compte.

### Calcul de la géométrie épipolaire

Pour le calcul de la géométrie épipolaire et de la matrice fondamentale, nous avons utilisé une méthode d'optimisation non linéaire sous contrainte. La solution trouvée

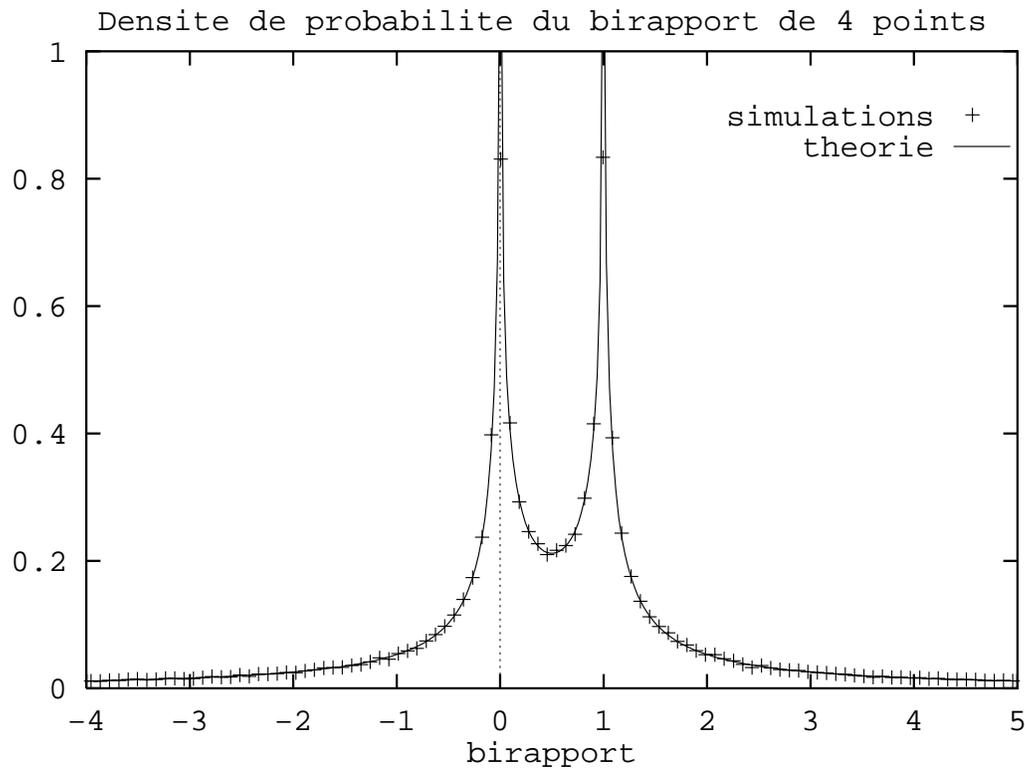


FIG. 2.8 - Fonction de densité des valeurs du birapport.

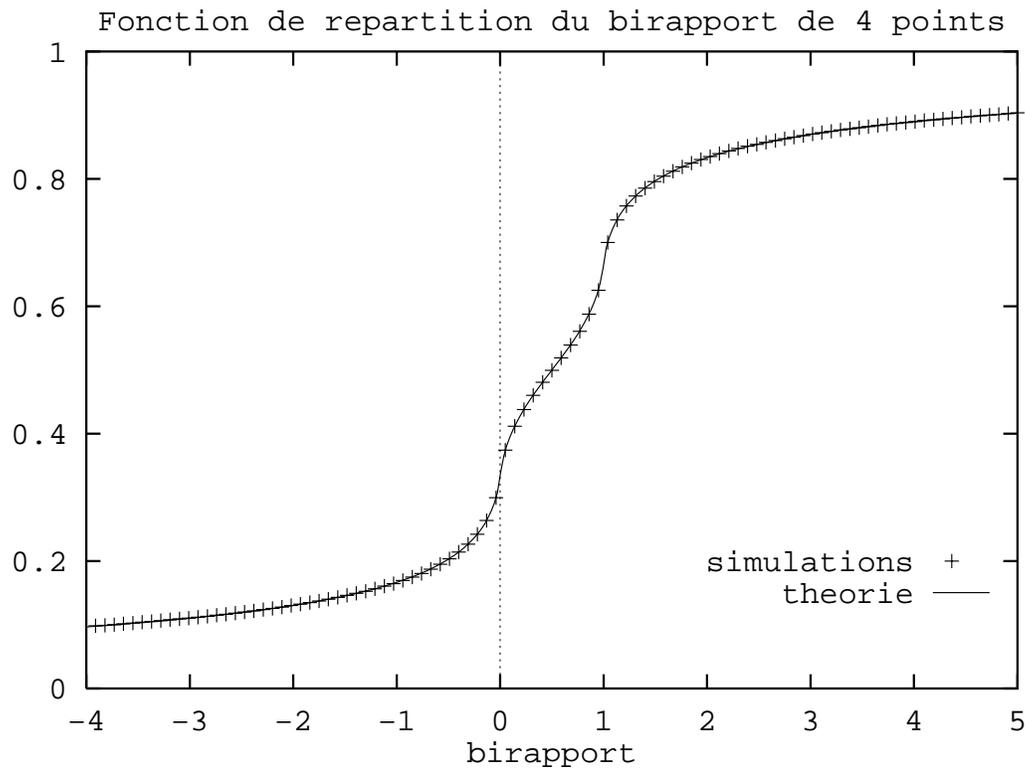


FIG. 2.9 - Fonction de répartition des valeurs du birapport.

par une première résolution par une méthode linéaire (SVD) est ensuite améliorée par une méthode itérative basée sur l'algorithme de Levenberg-Marquardt. La fonction d'erreur utilisée est la somme des carrés des distances des points à la droite épipolaire sur laquelle ils devraient se trouver. La contrainte portant sur le rang de la matrice est imposée au niveau de la représentation de la matrice. Se reporter à [Pre 88, Fau 92c] pour plus de détails.

### Test utilisé

On considère quatre points  $A, B, C$  et  $D$  de l'espace 3D, qui sont projetés dans deux images. On dispose donc d'une configuration de quatre points dans les deux images,  $a_1, b_1, c_1, d_1$  et  $a_2, b_2, c_2, d_2$  respectivement. On calcule alors  $f_1 = (a_1c_1) \cap (b_1d_1)$  et  $f_2 = (a_2c_2) \cap (b_2d_2)$ .  $e_1$  et  $e_2$  étant les épipoles des deux images, on calcule la quantité suivante, ou  $d$  est la distance entre birapports présentée ci-dessus :

$$\begin{aligned} \Delta &= d(\{e_1; a_1, b_1, c_1, d_1\}, \{e_2; a_2, b_2, c_2, d_2\}) \\ &+ d(\{e_1; a_1, b_1, c_1, f_1\}, \{e_2; a_2, b_2, c_2, f_2\}) \\ &+ d(\{e_1; a_1, b_1, f_1, d_1\}, \{e_2; a_2, b_2, f_2, d_2\}) \\ &+ d(\{e_1; a_1, f_1, c_1, d_1\}, \{e_2; a_2, f_2, c_2, d_2\}) \\ &+ d(\{e_1; f_1, b_1, c_1, d_1\}, \{e_2; f_2, b_2, c_2, d_2\}) \end{aligned}$$

Cette quantité  $\Delta$  est appelée, dans la suite, “distance entre les configurations de 4 points de chaque images”, ou “distance entre les birapports de ces configurations”, ou “distance” tout simplement.

### Méthode

Dans ce paragraphe, nous n'avons expérimenté que la partie “test de coplanarité”. La partie “construction de points coplanaires”, dont l'utilité principale est le calcul des invariants 3D, sera étudiée expérimentalement avec ces invariants, au paragraphe 2.2.

La première étape de l'expérimentation consiste à déterminer un seuil de coplanarité. En effet, le test de coplanarité est basé sur la comparaison de birapports. Ceux-ci sont égaux en cas de coplanarité, et différents dans le cas inverse. Cela dit, si on calcule ces birapports à partir d'images, c'est à dire de données bruitées, en n'utilisant pas une précision infinie lors des calculs, les birapports obtenus en cas de coplanarité ne seront plus strictement identiques. De combien diffèrent-ils ?

C'est ce que mesure le seuil que l'on recherche. Des birapports dont la distance est inférieure au seuil seront considérés “égaux”, et le quadruplet de points 3D correspondant sera considéré coplanaire. Dans le cas contraire, on conclura à la non coplanarité.

L'existence d'un seuil entraîne des erreurs de détection : des quadruplets coplanaires ne sont pas reconnus, des quadruplets non coplanaires sont au contraire détectés comme coplanaires. L'étude qui suit a pour but d'étudier ces erreurs pour certains seuils, sachant qu'il revient ensuite à l'expérimentateur de choisir le seuil à utiliser en fonction des performances qu'il souhaite et des erreurs qu'il peut accepter.

L'étude est réalisée en plusieurs étapes.

1. Tout d'abord, on étudie des quadruplets coplanaires et on mesure la distance entre les configurations qu'ils engendrent dans chaque image, pour différents niveaux

de bruit dans les images (de 0 à 5 pixels). Cela permet d'étudier la robustesse du test et de proposer différentes valeurs de seuil pour chaque niveau de bruit. Dans la suite, on étudie les erreurs auxquels ces seuils donnent lieu.

2. Pour chacun des seuils proposés à chaque niveau de bruit, on étudie la réponse au test de quadruplets coplanaires. Cela permet de déterminer, pour chaque seuil, le pourcentage des quadruplets coplanaires qui sont reconnus comme tels.
3. Pour ces mêmes seuils, on étudie aussi la réponse de quadruplets quelconques. Combien sont-ils reconnus ? Les quadruplets étudiés sont quelconques et n'excluent donc pas les quadruplets presque coplanaires. Pour compléter l'étude, on essaye de corrélérer la distance entre configurations et le volume défini par les quatre points dans l'espace 3D.
4. Enfin, on valide l'approche en utilisant des images réelles. Cela élimine les effets statistiques et permet de tester l'effet du bruit réel qui n'est pas complètement modélisable.

#### 2.1.4 Résultats obtenus avec des données simulées

##### Proposition de seuils

Cette première étape a pour but de proposer des seuils de coplanarité pour chaque niveau de bruit, seuils qui seront étudiés par la suite. Pour proposer ces seuils, on commence par étudier la distance entre des configurations qui sont les projections de quadruplets de points coplanaires.

On tire au hasard 2500 quadruplets de points coplanaires de la manière suivante : les trois premiers points sont tirés uniformément dans un cube, et le quatrième est tiré uniformément dans la partie du plan défini par les trois premiers points qui est dans le cube. Tous ces points sont ensuite projetés sur deux images, par deux caméras qui diffèrent l'une de l'autre par une rotation de 30 degrés. Les projetés sont alors bruités et la géométrie épipolaire est calculée.

Chaque quadruplet donne ainsi une configuration de quatre points dans chaque image. Ces configurations sont filtrées, certaines d'entre elles, considérées comme instables, étant supprimées. On supprime alors aussi la configuration correspondante dans l'autre image. Les cas de suppression sont les suivants :

1. les configurations qui contiennent des points distants de moins de dix pixels ;
2. celles qui sont trop colinéaires, ce que l'on détecte en mesurant la distance entre la droite définie par les deux points les plus éloignés et les deux autres : si cette distance est inférieure à cinq pixels, la configuration est éliminée : ce cas correspond à des plans vus "sur la tranche" et donc de manière très instable ;
3. les quatre points de chaque image définissent un faisceau de droites dont le sommet est l'épipole. On élimine les configurations dont les faisceaux dans les deux images n'ont pas des birapports très proches, ce qui est le cas lorsque l'ordre des points n'est pas conservé.

Il faut bien remarquer que ces trois tests sont réalisés sur les points projetés, et non sur les points 3D. Ils consistent juste à éliminer des cas pour lesquels on sait que le calcul du test de coplanarité sera instable et donnera donc de mauvais résultats.

Sur les configurations restantes, on réalise le test de coplanarité en mesurant la distance entre configurations correspondantes. La FIG. 2.10 montre les valeurs minimales, moyennes et maximales de cette distance, obtenues pour différents niveaux de bruit compris entre 0 et 5 pixels.

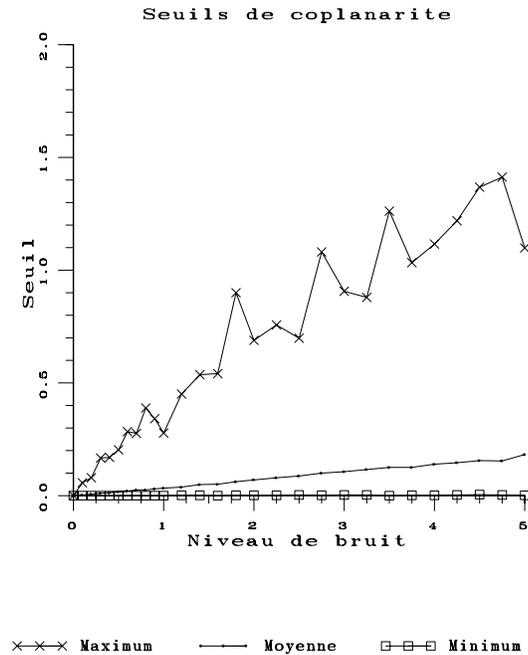


FIG. 2.10 - Seuils pour le test de coplanarité.

- La courbe de valeurs maximales dépend de l'existence de configurations dégénérées non filtrées à chaque niveau de bruit. Plus il y a du bruit, plus de telles configurations existent. Notons tout de même que ces configurations sont beaucoup plus nombreuses si on n'utilise pas les trois filtres présentés ci-dessus.

Cette courbe des valeurs maximales n'est, en fait, pas très significative. Pour s'en convaincre, il suffit de regarder un histogramme des valeurs obtenues à divers niveaux de bruit (cf. FIG. 2.11). Les valeurs maximales sont isolées et en petit nombre, ce qui explique la régularité de la courbe des valeurs moyennes.

- La faible pente de la courbe des valeurs moyennes montre la robustesse au bruit du calcul effectué pour le test. Pour chaque niveau de bruit, c'est la valeur moyenne représentée par cette courbe qui servira d'étalon : on considérera dans la suite que des points sont coplanaires si leur leur réponse au test est plus petite que la valeur moyenne, 1,5 ou 2 fois cette valeur, suivant les performances de détection que l'on désire.

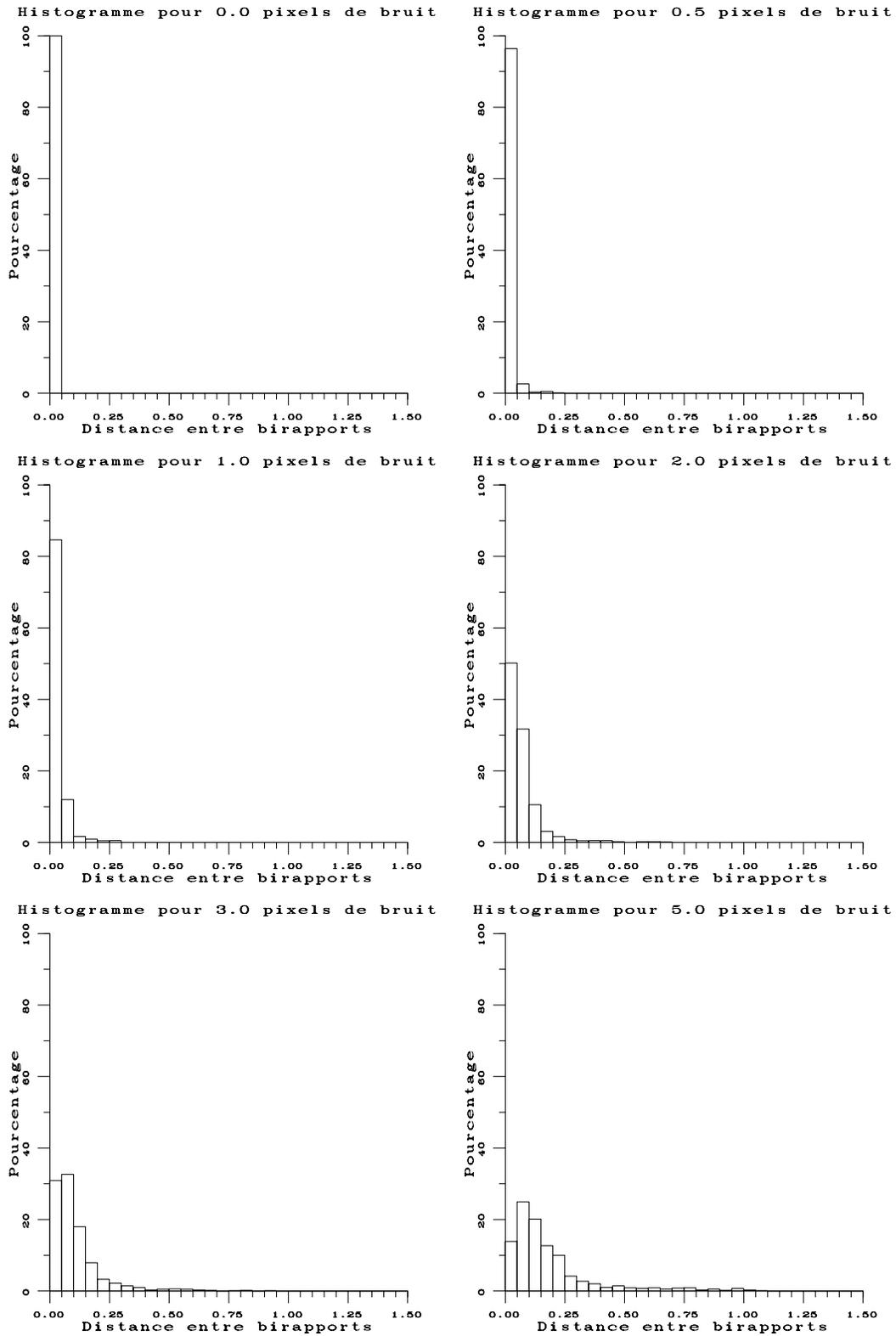


FIG. 2.11 - Histogrammes des valeurs au test de coplanarité pour des quadruplets coplanaires.

- La courbe de valeur minimale montre qu'à tous les niveaux de bruit, certaines configurations ne sont pas affectée par ce bruit.

On propose donc d'étudier trois seuils pour chaque niveau de bruit :  $v_{moy}$  la valeur moyenne obtenue au niveau de bruit considéré sur la FIG. 2.10,  $1,5.v_{moy}$  et  $2.v_{moy}$ .

### Réponse des quadruplets coplanaires

On étudie maintenant les seuils proposés pour chaque niveau de bruit et, plus précisément, on veut déterminer les erreurs de détection que ces seuils entraînent. Le présent paragraphe concerne la détection des quadruplets coplanaires, le suivant celle des quadruplets quelconques.

On reprend les 2500 quadruplets du paragraphe précédent et les trois seuils proposés pour chaque niveau de bruit. On mesure alors le pourcentage de quadruplets qui donnent, dans les images, des configurations dont la distance est inférieure à chacun des seuils. Les résultats sont montrés sur la FIG. 2.12.

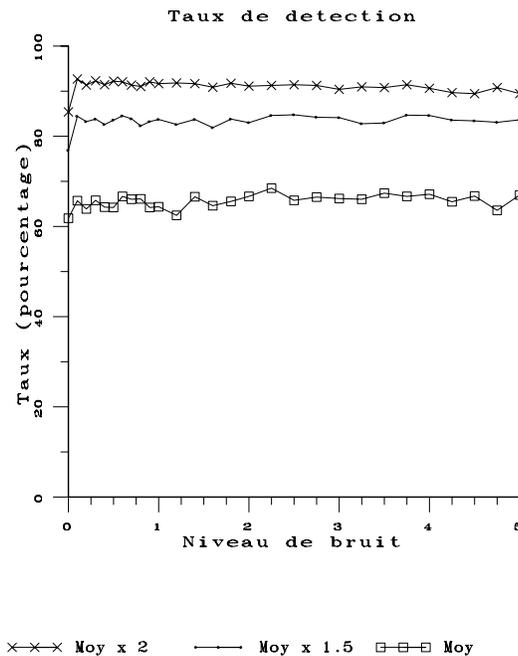


FIG. 2.12 - Taux de résultats positifs au test pour des quadruplets coplanaires.

L'horizontalité de ces courbes montrent que le bruit ne fait que dilater les histogrammes de la FIG. 2.11 selon l'axe des valeurs, mais sans changer le type de distribution.

En conclusion, l'accroissement du niveau de bruit présent dans les images implique l'augmentation de la valeur des seuils si l'on veut garder un taux constant de détection des configurations coplanaires.

### Réponse de quadruplets quelconques

Dans ce paragraphe, on étudie la réponse au test de quadruplets quelconques. Le terme “quelconque” est plus adéquat que “non coplanaire”. Le tirage effectué n'exclut pas, en effet, les quadruplets coplanaires. En supposant que la proportion de ces quadruplets presque coplanaires reste constante, on peut étudier l'évolution du taux de réponse positive à différents niveaux de bruit. Un autre aspect intéressant est l'étude de la corrélation entre le résultat au test et le volume du tétraèdre des points 3D initiaux, ceci à un niveau de bruit donné.

Dans ce but, et pour des niveaux de bruit variant entre 0 et 5 pixels, on tire uniformément dans un cube 2500 quadruplets de points 3D que l'on projette dans deux images, les projections sont bruitées, la géométrie épipolaire est calculée et le test de coplanarité est effectué. Avant le calcul du test, on a préalablement filtré certaines configurations : celles dont les points de projections étaient trop proches les uns des autres, et celles qui sont trop colinéaires. Les filtres utilisés sont les mêmes qu'avec les quadruplets coplanaires. On n'a pas, par contre, utilisé le filtre basé sur la ressemblance des birapports des projections de quatre points initiaux. Ce birapport étant intégré dans le test lui-même, les configurations correspondantes sont éliminées lors du seuillage.

La FIG. 2.13 montre les résultats positifs obtenus pour les mêmes seuils qu'au paragraphe précédent, soit pour le seuil moyen  $v_{moy}$ ,  $1,5.v_{moy}$ , et  $2.v_{moy}$ .

Il faut bien remarquer que le test et les conditions expérimentales utilisés avec les quadruplets quelconques sont les mêmes que ceux utilisés avec les quadruplets coplanaires. La FIG. 2.13 montre une augmentation du taux de détection. Cette augmentation n'est pas due à une plus grande proportion de quadruplets coplanaires, mais à la variation des seuils utilisés en fonction du bruit (cf. FIG. 2.10).

On peut se convaincre de cela en observant la FIG. 2.14 qui montre les histogrammes de valeurs obtenues par le test pour quelques niveaux de bruit. Ces histogrammes ont des formes tout à fait similaires, ne laissant pas apparaître une augmentation du nombre de configurations donnant de faibles résultats au test.

En conclusion, le taux de détection des quadruplets quelconques ne dépend pas du niveau de bruit, mais uniquement de la valeur de seuil utilisée.

### Corrélation avec les volumes des tétraèdres initiaux

Pour affiner la conclusion obtenue, on étudie maintenant la corrélation entre le volume du tétraèdre formé par quatre points dans l'espace 3D et le résultat au test de coplanarité obtenu par les projections de ces mêmes quatre points.

Dans cette expérimentation, on est confronté à la difficulté signalée au début de ce paragraphe 2.1. Le volume du tétraèdre formé par quatre points est-il une bonne mesure de la coplanarité de ces quatre points ? Si l'échelle d'observation n'a pas de signification, la réponse est clairement négative. Deux configurations homothétiques n'ont pas le même volume, mais elles sont aussi peu coplanaires l'une que l'autre à leur échelle respective.

Si l'échelle d'observation a une signification, la réponse est positive. Un tétraèdre de volume très faible est soit presque ponctuel, soit presque linéaire, soit presque plan. Dans ces trois cas, on peut considérer que les points sont presque coplanaires.

Dans le cas de la vision, il y a une échelle d'observation. Ne serait-ce qu'à cause des

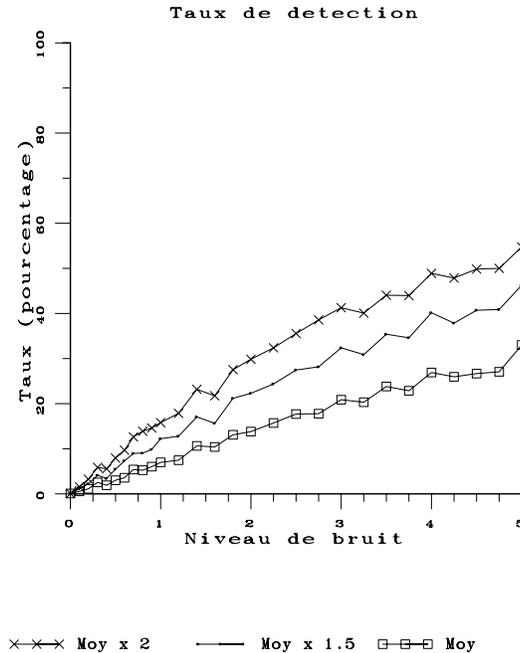


FIG. 2.13 - Taux de résultats positifs au test pour des quadruplets quelconques.

problèmes de discrétisation et de bruit dans l'image, deux points séparés de un ou deux pixels et deux points séparés de 100 pixels ne peuvent être traités de la même manière. L'incertitude relative sur leur distance est proche de 1 dans le premier cas et n'est que de quelques pour-cent dans l'autre. Aussi, nous utilisons le volume du tétraèdre comme mesure de coplanarité.

Pour quelques niveaux de bruit, nous avons tracé trois graphiques, montrés en colonnes dans les FIG. 2.15, 2.16 et 2.17. Dans chacun des graphiques, chaque configuration de quatre points de l'espace donne lieu à une croix, dont l'abscisse est son résultat au test, et dont l'ordonnée est son volume.

L'unité de volume a été choisie ainsi : un pixel d'une caméra a une taille de l'ordre de  $10\mu m$ , ce qui correspond à  $0,5mm$  vu à  $80cm$  par une caméra dont l'objectif a une distance focale de  $16mm$  ou à  $1mm$  vu à  $1,2m$  avec un objectif de  $12mm$ . Reprenant cette dernière correspondance, on peut considérer que l'unité de volume utilisée dans les graphiques est celle d'un cube dont le côté mesure  $1mm$ , c'est à dire  $1mm^3$ , ce cube étant observé à  $1,2m$  avec un objectif de focale  $12mm$ .

Les graphiques de haut de colonnes montrent tous les points de corrélation obtenus à partir des configurations quelconques utilisées précédemment. Outre l'accumulation de points autour de l'origine, le seul résultat clair de ces graphiques est, qu'en dehors éventuellement de l'origine, il n'y a pas de corrélation entre le volume et le test de coplanarité. Autrement dit, le test de coplanarité ne peut servir à mesurer un degré de non-coplanarité de quatre points dont on connaît les projections.

Pour obtenir les graphiques de milieu de colonnes, on a fait subir aux graphiques

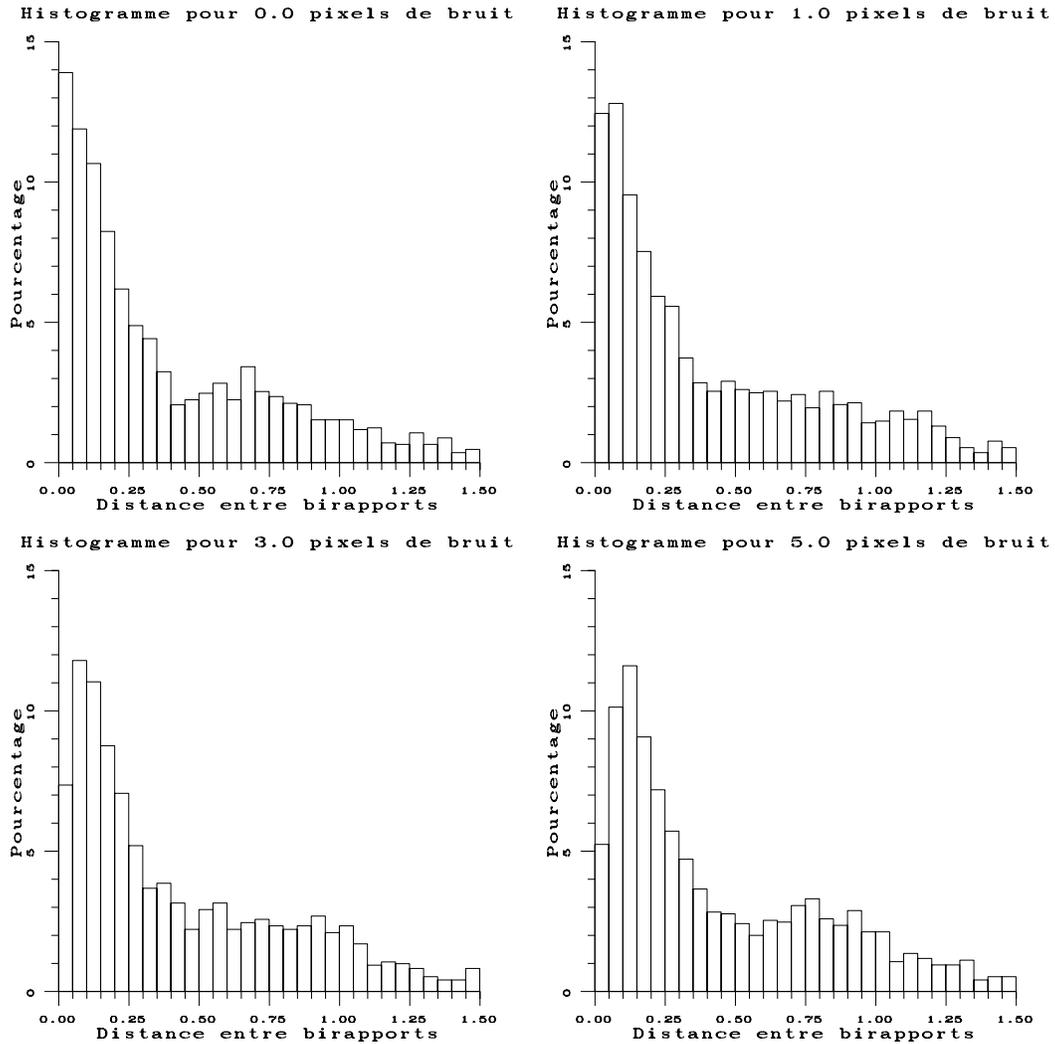


FIG. 2.14 - Histogramme des valeurs au test de coplanarité pour des quadruplets quelconques.

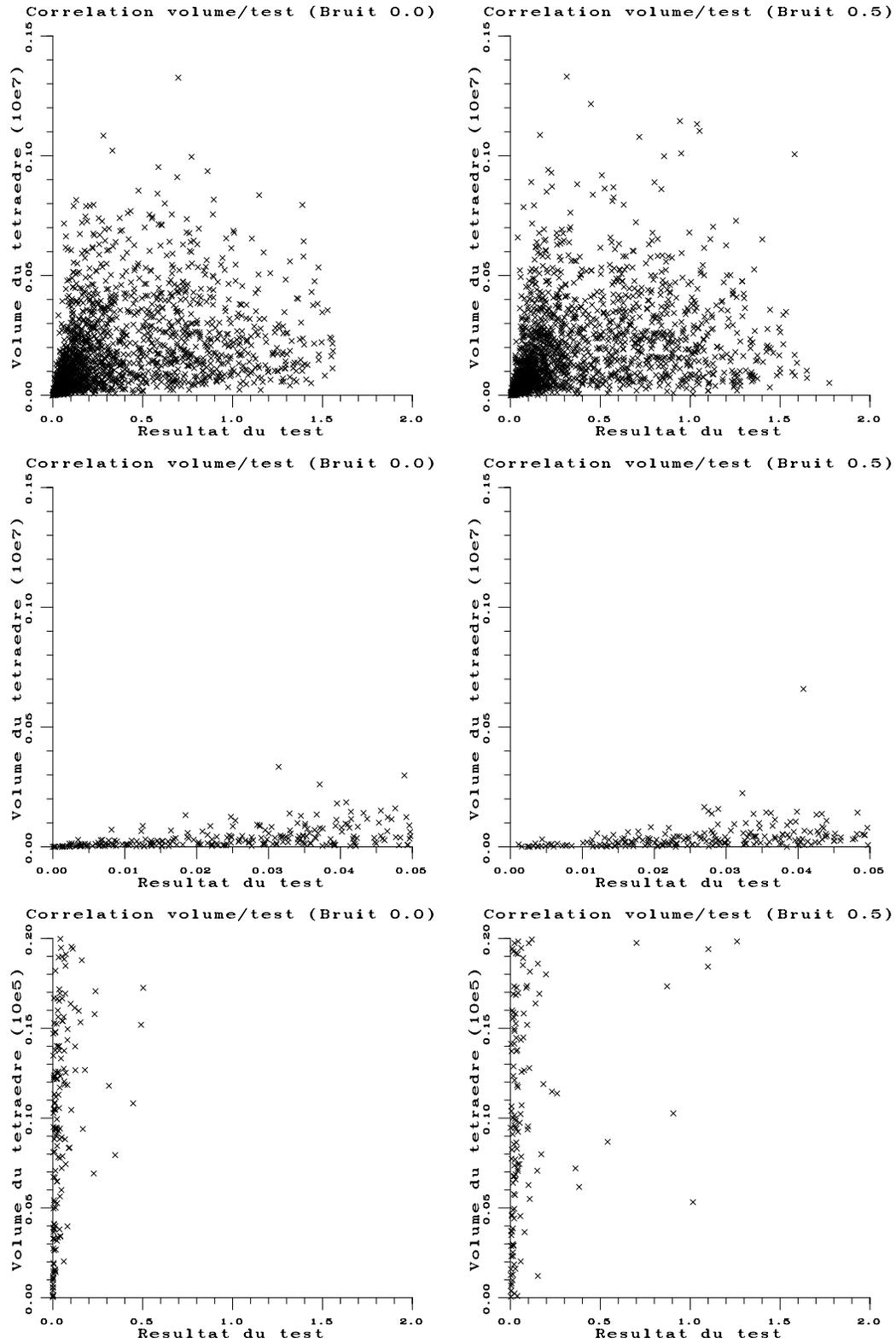


FIG. 2.15 - Corrélation entre les volumes de tétraèdres et les résultats au test pour des bruits de 0 et 0,5 pixels dans les images.

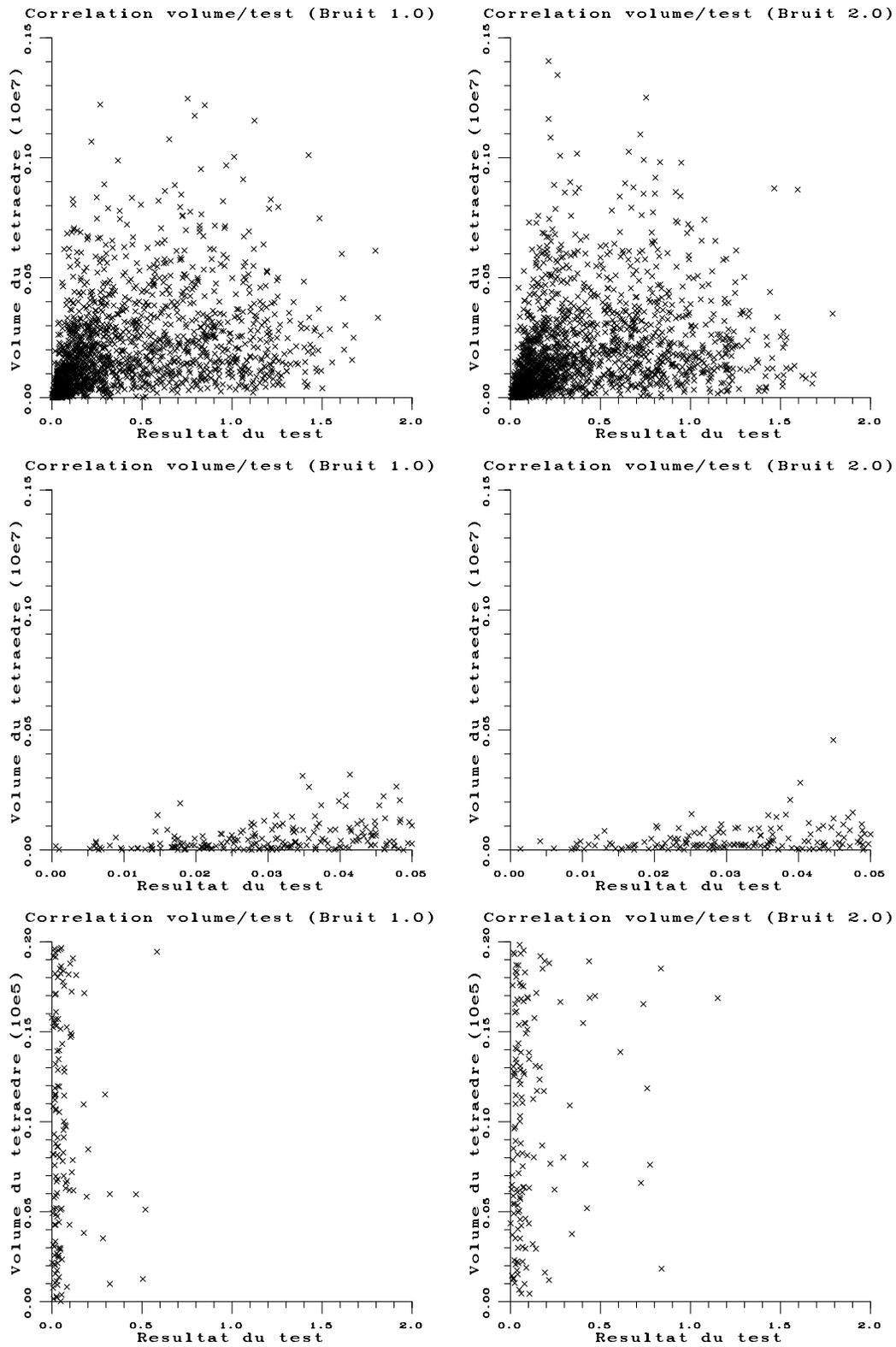


FIG. 2.16 - Corrélation entre les volumes de tétraèdres et les résultats au test pour des bruits de 1 et 2 pixels dans les images.

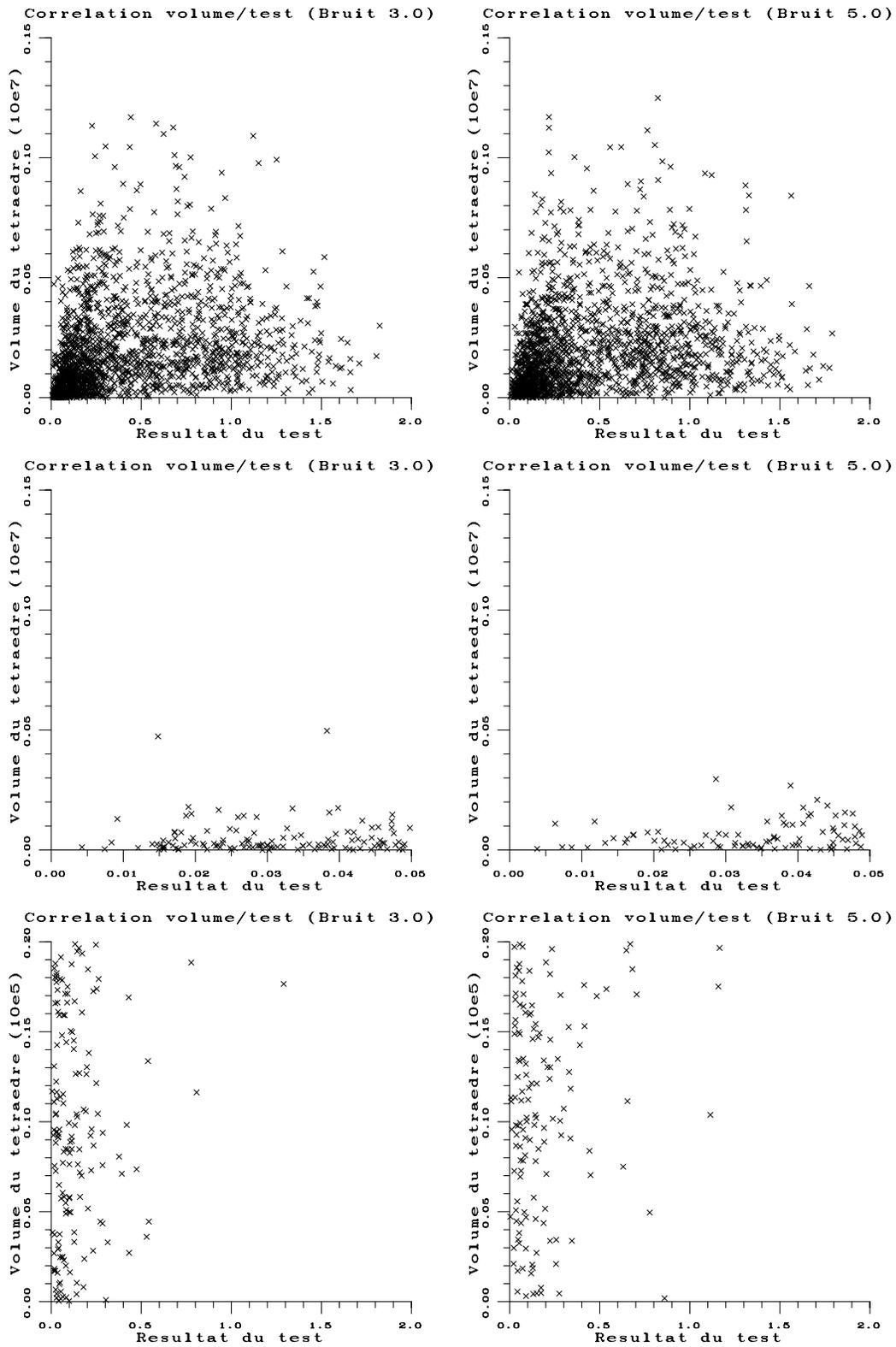


FIG. 2.17 - Corrélation entre les volumes de tétraèdres et les résultats au test pour des bruits de 3 et 5 pixels dans les images.

précédents une dilatation sur l'axe des abscisses, l'axe des ordonnées restant le même. Ces graphiques permettent donc de répondre à la question suivante : à quels volumes correspondent les configurations qui ont une petite différence entre birapports au test de coplanarité? Réponse : à de petits volumes. Lorsque le bruit augmente, la courbe est moins plate, vers l'origine en particulier : certaines configurations ayant un volume non négligeable donnent quand même une faible différence entre birapports au test. On peut aussi remarquer que le nombre de configurations donnant une différence faible au test baisse avec l'augmentation du bruit.

Les graphiques de bas de colonnes sont, eux, obtenus par dilatation de l'autre axe, celui des ordonnées. Ils permettent donc de répondre à la question : que donne le test avec des configurations de faible volume? Réponse : de faibles distances lorsque le bruit est faible, mais certaines d'entre elles, de nombre croissant avec le niveau de bruit, donnent des résultats quelconques.

En conclusion, la corrélation entre volume du tétraèdre des points 3D initiaux et résultats au test des projections de ces points est donc forte pour les configurations de petit volume, ou donnant de faibles différences au test, mais n'existe pas ailleurs. Le bruit tend à supprimer la corrélation entre ces deux grandeurs, même lorsqu'elles sont faibles.

### Récapitulatif

Pour conclure cette partie de l'étude utilisant des données simulées, on peut faire un bilan des résultats du test pour différents seuils et divers niveaux de bruit selon les quatre catégories classiques : configurations coplanaires acceptées et rejetées, configurations quelconques acceptées et rejetées. La FIG. 2.18 montre les pourcentages de configurations coplanaires acceptées (colonnes marquées Copl) et de configurations quelconques acceptées (colonnes marquées Qcq).

Niveau de bruit	Seuils								
	moy			moy $\times 1.5$			moy $\times 2$		
	Valeur	Copl	Qcq	Valeur	Copl	Qcq	Valeur	Copl	Qcq
0.0	1.5e-05	61.8	0.06	2.25e-5	76.9	0.06	3e-5	85.4	0.06
0.5	0.017	64.2	3.0	0.026	83.6	5.47	0.034	92.3	7.95
1.0	0.033	64.4	7.0	0.049	83.8	12.2	0.065	91.6	15.8
2.0	0.069	66.7	13.8	0.10	83.0	22.3	0.14	91.1	29.8
3.0	0.11	66.0	20.9	0.16	84.1	32.4	0.21	90.4	41.2
4.0	0.14	67.1	26.9	0.21	84.6	40.2	0.28	90.6	48.9
5.0	0.18	67.0	33.0	0.27	83.7	46.3	0.37	89.5	54.7

FIG. 2.18 - Récapitulatif des résultats du test de coplanarité sur des données simulées : taux de réponses positives des quadruplets coplanaires et quelconques.

Dans ce tableau, l'effet du bruit apparaît clairement : pour garder un taux constant de détection des configurations coplanaires, il faut augmenter les seuils lorsque le niveau de bruit augmente, et cela entraîne un taux de mauvaises détections de plus en plus élevé. Garder ce dernier taux constant entraîne une baisse du taux de détection des configurations coplanaires. Il y a donc un compromis à faire, compromis de plus en plus

défavorable lorsque le bruit augmente.

Ce tableau montre aussi la qualité des données nécessaire pour garder de bonnes performances de détection. Pour avoir un taux de mauvaises détections inférieur à 10% et un taux de bonnes détections supérieur à 90%, il faut choisir un seuil de 0,04 et ne pas dépasser un niveau de bruit de 0.7 pixels, ce qui est relativement faible.

### 2.1.5 Résultats obtenus avec des images réelles

Pour terminer l'étude du test de coplanarité, voici quelques expérimentations faites avec des images réelles. Elles n'ont pas pour but une étude quantitative du test. En particulier, on ne peut mesurer simplement le taux de bonnes ou de mauvaises détections, car les données 3D ne sont pas connues, et certaines des coplanarités qui apparaissent ne sont pas évidentes et ne peuvent pas être vérifiées précisément.

Il s'agit plutôt de valider le processus en testant si les résultats sont "raisonnables" lorsque l'on utilise des données réelles. En cela, cette expérimentation est complémentaire de celle faite sur les données simulées.

On considère deux images d'une maison en bois entourée de divers objets (cf. FIG. 2.19). Après extraction des contours, les sommets sont extraits par un détecteur de coins paramétrique et appariés. 29 d'entre eux ont été sélectionnés à la main pour cette étude. Ces sommets sont montrés sur la FIG. 2.21.

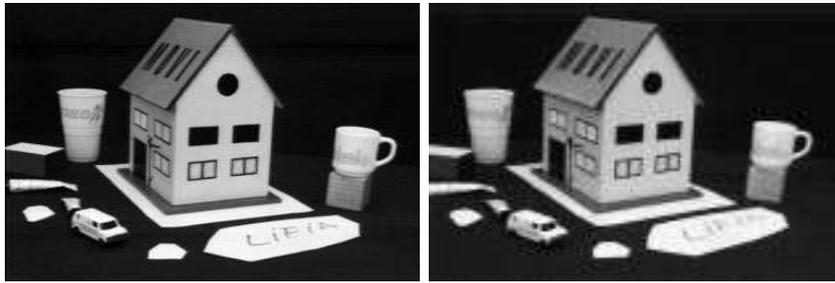


FIG. 2.19 - Deux images d'une maison en bois.

A partir de 29 points, on peut former 23 751 configurations de 4 points. Ces configurations ont été filtrées de même que précédemment : celles ayant des points trop proches ou qui sont presque linéaires ont été supprimées. Après ce filtrage, restent 17 745 configurations. On applique alors le test à toutes ces configurations. La FIG. 2.20 montre l'histogramme des valeurs obtenues. Celles-ci forment un continuum et il n'y a pas de seuil naturel de coplanarité.

La FIG. 2.22 donne quelques résultats plus précis. Ces résultats ont été choisis parmi les configurations donnant les plus petites valeurs au test. La première colonne contient les numéros des quatre sommets de la configuration et la deuxième le résultat obtenu au test. On peut séparer ces configurations en deux classes :

- la première contient des configurations dont la coplanarité est difficile à évaluer et dont on ne peut donc dire aisément si ce sont de bonnes ou de mauvaises détections ; de telles configurations apparaissent dans la première partie du tableau ;
- la deuxième contient les configurations dont on connaît la coplanarité (seconde partie du tableau).

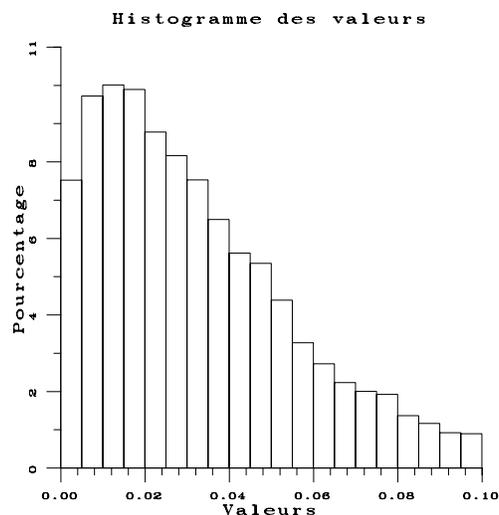


FIG. 2.20 - Histogramme des résultats du test de coplanarité pour les images de la maison.

Une autre manière d'évaluer le test consiste à regarder les résultats de toutes les configurations contenant trois points particuliers. La FIG. 2.23 montre ainsi les résultats obtenus par toutes les configurations contenant les points 5, 8 et 14 et pour toutes celles contenant les points 21, 23 et 28. Les première et troisième colonnes contiennent le numéro du quatrième point de la configuration et la deuxième et la quatrième le résultat du test. Les configurations contenant les trois points 5, 8 et 14 ou 21, 23 et 28 et qui n'apparaissent pas dans le tableau ont été supprimées au filtrage.

Les configurations ont été triées dans l'ordre de leur résultat. Pour les configurations de la partie gauche du tableau, le seuil égal à 0.005 est matérialisé par un trait horizontal. Il en est de même dans la partie droite pour le seuil égal à 0.03. Ces deux seuils partagent les configurations coplanaires de celles qui ne le sont point. La seule configuration que l'on peut éventuellement considérer comme mal classée est celle contenant les points 5, 8, 14 et 15. En dehors de ce cas limite, les résultats sont conformes à ce qui peut être estimé "à l'œil" sachant que les deux images ne contiennent aucun effet spécial (effet d'optique ou trompe-l'œil).

D'une manière générale, les résultats obtenus sont bons : peu de mauvaises détections, les configurations de points coplanaires sont détectées. La difficulté se situe plutôt au niveau de la scène 3D elle-même (jusqu'à quelle limite considère-t-on des points coplanaires ?) et au niveau du choix préalable d'un seuil de coplanarité.

### 2.1.6 Conclusion

De toute cette étude expérimentale du test de coplanarité, on peut retenir trois points principaux.

- Le test est validé par les expérimentations sur données réelles, même lorsque le mouvement de la caméra entre les deux prises d'image est faible (le calcul de la géométrie épipolaire peut alors être instable).

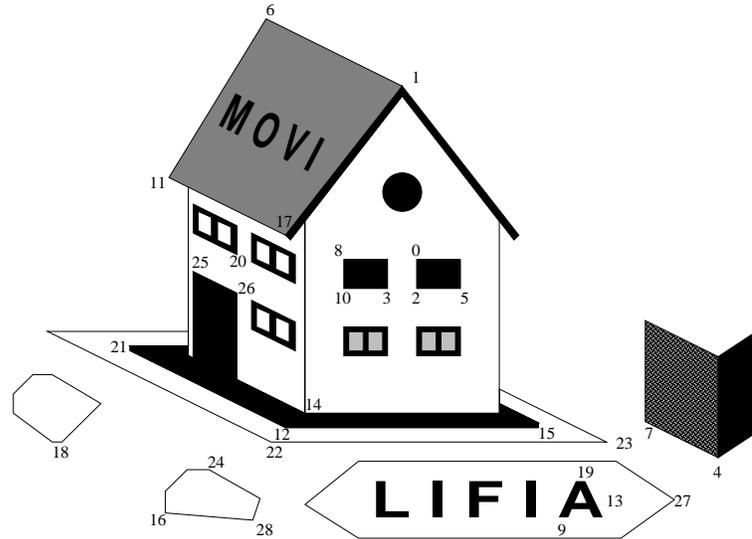


FIG. 2.21 - Numérotation des sommets sélectionnés.

Sommets				Résultat
2	24	25	28	0.000250446
0	11	15	27	0.000313107
7	11	12	20	0.000345824
0	5	12	17	0.00026739
2	10	14	17	0.000313721
1	8	10	15	0.000314295
3	10	12	15	0.000347512
1	2	10	17	0.000366849
1	2	14	17	0.000370531
7	16	19	24	0.00039199
0	1	10	14	0.000396603

FIG. 2.22 - Résultats obtenus pour quelques configurations.

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

- Le test est robuste au bruit ; un léger bruit dans les images n'est pas catastrophique, mais la légère dégradation des résultats engendrée par ce bruit fera tout de même préférer des techniques d'extraction des points très précises : extracteur de coins paramétrique, techniques sub-pixel. . .
- Le problème le plus délicat posé par ce test est la détermination du seuil de coplanarité. L'exemple des images de la maison montre en effet qu'il n'y a pas de seuil naturel, et que le seuil optimal, celui qui sépare au mieux les configurations coplanaires des autres, n'est pas le même suivant les plans considérés.

Ce dernier point doit-il disqualifier ce test pour une utilisation pratique ? Le test n'est certainement pas suffisant comme base d'un système automatique de détection "sans erreur" de toutes les configurations de quatre points coplanaires d'une scène. Il peut par contre se révéler utile pour choisir une configuration coplanaire, il suffit alors de prendre celle donnant le plus petit résultat, ou pour, au contraire, choisir une configuration qui ne l'est pas. Il peut être aussi utilisé conjointement avec d'autres outils, tel un détecteur de polygones, afin d'introduire dans la recherche des coplanarités d'autres informations que la seule géométrie.

Par ailleurs, ce test peut être un auxiliaire utile et pratique pour un système interactif : le test permet de trier automatiquement les configurations en fonction de leur résultat, puis l'utilisateur choisit alors le seuil ou sélectionne les configurations qu'il désire, opérations largement facilitées par le tri préalable. L'ordinateur n'est plus alors le système automatique qui remplace l'opérateur humain, mais une aide qui facilite le travail en réalisant rapidement les opérations ou calculs fastidieux. L'opérateur reste, quant à lui, maître des choix principaux et des jugements qualitatifs.

## 2.2 Calcul d'invariants 3D à partir de plusieurs images

Ce paragraphe concerne le problème de l'extraction de caractéristiques numériques d'objets 3D à partir d'images de ces objets. Autrement dit, il s'agit de répondre à la question : peut-on faire des mesures sur un objet lorsqu'on ne connaît que des images de cet objet ? C'est le problème que résolvent les photogrammètres à partir de photographies.

Dans le cas général, si l'on ne dispose que d'une seule image de l'objet et que la caméra qui a servi à prendre l'image n'est pas étalonnée, aucune mesure n'est possible, ou, en exprimant cela plus mathématiquement, il n'y a pas d'invariant dans cette image lorsque le point de vue change [Bur 92]. Plusieurs voies sont alors possibles.

Tout d'abord, on peut étalonner la caméra. L'utilisation de plusieurs images permet alors de reconstruire complètement l'objet en 3D. C'est le problème classique de la stéréovision, avec ses non moins classiques ennuis. Si le problème de l'appariement des images stéréographiques est maintenant résolu de manière satisfaisante [Sko 88, Fua 90], celui de la stabilité numérique de l'étalonnage et des quantités qui en sont déduites ne l'est encore que partiellement.

Une deuxième approche consiste à utiliser des contraintes connues a priori, telles la coplanarité [Mor 93, Rot 93]. Il est alors possible de calculer des invariants à partir d'une seule image. La difficulté réside alors dans la formulation et l'utilisation de ces connaissances a priori.

Une troisième méthode consiste à utiliser plusieurs images, sans étalonnage préalable de la caméra. Comme déjà signalé au début du paragraphe 2.1, on peut soit reconstruire l'objet en 3D à une homographie de l'espace près, soit calculer les invariants directement à partir des images. Dans ce dernier cas, on peut soit poser les équations et utiliser des méthodes algébriques, avec l'aide précieuse d'un système de calcul formel, soit utiliser des constructions géométriques pour faire apparaître des invariants simples, des birapports dans notre cas. C'est cette dernière possibilité qui est explorée ici.

Ces différentes voies possibles ne sont guère comparables entre elles. Elles correspondent en effet à des hypothèses différentes sur l'étalonnage et les connaissances préalables, et donc à différents cas d'utilisation.

Pour notre part, nous supposons donc que l'on dispose de deux images représentant la même configuration tridimensionnelle de points, droites, plans et coniques, et que l'appariement des deux images et le calcul de la géométrie épipolaire sont réalisés au préalable. Il est alors possible de calculer directement à partir des deux images, c'est à dire sans reconstruction, des invariants, qui seront bien entendu des birapports. Cette technique est l'équivalent 3D des méthodes 2D utilisées aux paragraphes 1.2.8 et 1.2.9.

Les applications d'un tel calcul d'invariants sont nombreuses : ces invariants 3D peuvent servir à modéliser non seulement topologiquement mais géométriquement des objets, à les indexer dans une base de données, à les reconnaître, à calculer des indices utilisables facilement en vision à partir de données CAO, etc. Il faut remarquer que tout cela peut être fait avec des images prises d'une caméra non calibrée, ou même de plusieurs caméras différentes. La seule hypothèse est le caractère projectif de la transformation réalisée par cette ou ces caméras.

Dans tout ce paragraphe, nous considérons des invariants de points, droites et coniques, qui peuvent être observés aisément dans les images, à l'inverse des plans. Il est

rappelé que le groupe des homographies de l'espace 3D est un groupe de Lie d'ordre 15, chaque homographie étant caractérisée par une matrice  $4 \times 4$  définie à un facteur de proportionnalité près. Les invariants de diverses configurations sont tout d'abord calculés de manière théorique, puis l'un d'entre eux, l'invariant de deux paires de droites coplanaires testé avec des données simulées et réelles.

Ce paragraphe utilise de manière intensive la “construction des projections de nouveaux points coplanaires” présentée au paragraphe 2.1.2. Une fois cette construction comprise, ce paragraphe ne présente plus guère d'autres difficultés mathématiques, en dehors de l'étude du cas de quatre droites quelconques, qui pourra être laissé de côté pour une lecture rapide.

## 2.2.1 Etude de différentes configurations

### Cas de six points

Soient six points d'un espace 3D. Chaque point est caractérisé par quatre coordonnées homogènes définies à un facteur de proportionnalité près, et a donc trois degrés de liberté. L'ensemble des configurations forme donc une sous-variété de dimension 18. Une homographie étant entièrement définie par les images d'une base projective de cinq points, une configuration de six points a un stabilisateur réduit à l'identité. L'espace des homographies a pour dimension 15, et il y a trois invariants dans ce cas.

Notons  $A, B, C, D, E$  et  $F$  les six points (cf. FIG. 2.24). On construit trois nouveaux points de manière projective (c'est à dire en n'utilisant que des constructions projectives) :  $D' = (EF) \cap (ABC)$ ,  $E' = (DF) \cap (ABC)$  et  $F' = (DE) \cap (ABC)$ , où  $(ABC)$  représente le plan contenant les trois points  $A, B$  et  $C$ . Les six points  $A, B, C, D', E'$  et  $F'$  sont coplanaires ; quatre d'entre eux forment une base projective, et les deux autres définissent chacun deux coordonnées projectives dans cette base, c'est à dire quatre birapports au total. Trois d'entre eux sont indépendants et peuvent être pris comme invariants pour la configuration initiale.

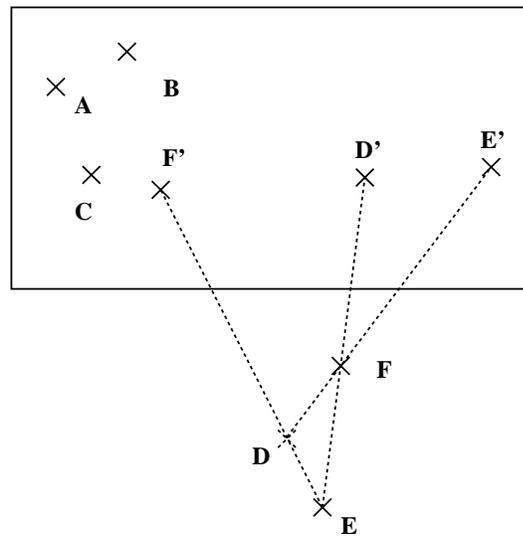


FIG. 2.24 - Cas de six points dans l'espace 3D.

Le calcul direct de ces invariants à partir de deux images est possible. Il suffit pour cela d'utiliser une des deux méthodes exposées au paragraphe 2.1.2 pour construire les projections dans ces deux images des points  $D', E'$  et  $F'$ . Les birapports définis par des points coplanaires étant invariants par projection perspective, les invariants de la configuration peuvent donc être calculés à partir des deux seules images, et ce de façon simple.

### Cas de quatre points et une droite

Une droite dans l'espace possède quatre degrés de liberté. Les configurations de quatre points et une droite forment donc un ensemble de dimension 16. Comme le stabilisateur d'une telle configuration est réduit à l'identité, il existe un unique invariant dans ce cas.

Trois des quatre points définissent un plan qui coupe la droite en un point (cf. FIG. 2.25). Il y a quatre façons de choisir ces trois points, et on peut ainsi construire quatre points sur la droite. Ces quatre points définissent un birapport.

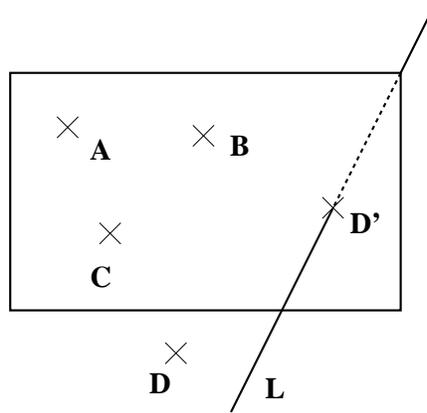


FIG. 2.25 - Cas de quatre points et une droite.

Comme dans le cas précédent, les méthodes de construction des projections de l'intersection d'une droite et d'un plan permettent de calculer ce birapport directement à partir des deux images.

On peut calculer l'invariant de cette configuration de manière différente. On considère l'ensemble des plans contenant la droite de la configuration. Les quatre points caractérisent quatre de ces plans, qui définissent un birapport. Pour calculer ce birapport, il faut construire l'intersection de ces quatre plans avec une droite les coupant en quatre points distincts (cf. §1.1.4).

### Cas de trois points et de deux droites.

Une telle configuration a  $3 \times 3 + 2 \times 4 = 17$  degrés de liberté et il y a deux invariants. Les trois points définissent un plan qui coupe les deux droites en deux points (cf. FIG. 2.26). On a ainsi cinq points coplanaires, qui définissent donc deux birapports indépendants, qui peuvent être calculés à partir de deux images seulement.

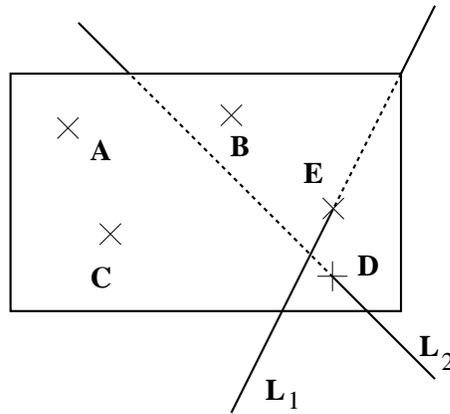


FIG. 2.26 - Cas de trois points et deux droites.

### Cas de deux points et trois droites.

Ce type de configuration a  $2 \times 3 + 3 \times 4 = 18$  degrés de liberté, soit trois invariants. Notons  $P_1, P_2, D_1, D_2$  et  $D_3$  les points et droites de la configuration. On peut construire quatre plans  $\pi_1, \pi_2, \pi_3$  et  $\pi_4$  de la manière suivante : ils contiennent respectivement  $D_1$  et  $P_1$ ,  $D_1$  et  $P_2$ ,  $D_2$  et  $P_1$ , et  $D_2$  et  $P_2$ . Ces quatre plans coupent  $D_3$  en quatre points qui définissent un birapport.

Par permutation circulaire sur les indices des droites, on définit ainsi trois birapports, qui peuvent, comme dans les cas précédents, être calculés directement à partir de deux images. Cela implique toutefois qu'aucune des projections de ces droites ne passe par un des épipoles.

### Cas de quatre droites dont deux coplanaires

Dans ce cas, une configuration est composée de deux droites coplanaires et de deux autres droites. Deux droites coplanaires ont sept degrés de liberté et la configuration a donc 15 degrés de liberté.

Un invariant peut être trouvé facilement pour cette configuration : le plan défini par les droites coplanaires coupe les deux autres droites en deux points. On obtient deux droites et deux points dans le même plan, ce qui est la configuration étudiée au paragraphe 1.2.8. La présence d'un invariant est due à l'existence d'un groupe d'isotropie pour chaque élément.

Comme dans les cas précédents, le birapport invariant peut être calculé directement à partir de deux images, si la géométrie épipolaire est connue.

Un cas particulier est constitué par les configurations comprenant deux paires de droites coplanaires (cf. FIG. 2.27). Chacune des paires de droites définit un plan. L'intersection de ces deux plans est une droite qui coupe les quatre droites de départ en quatre points, ce qui définit un birapport. Dans ce cas, le groupe d'isotropie de chaque élément est d'ordre 2. C'est cet invariant particulier qui est l'objet des expérimentations des paragraphes 2.2.2 et 2.2.3.

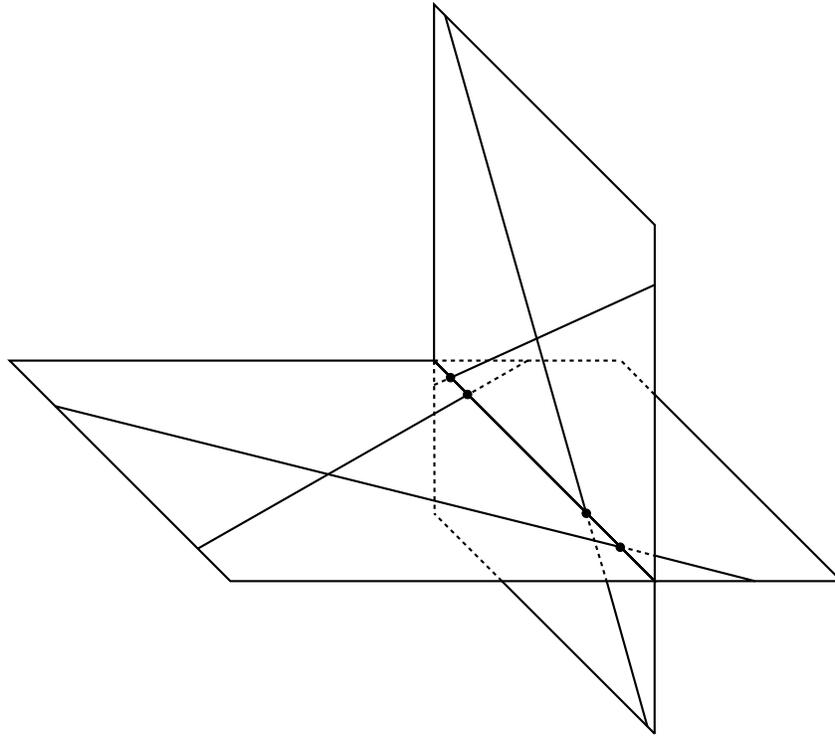


FIG. 2.27 - Cas de deux paires de coplanaires.

### Cas de quatre droites quelconques

Ce cas est plus compliqué que le précédent et il pourra être omis en cas de lecture rapide. Comme les configurations de quatre droites forment un espace de dimension 16, il doit exister au moins un invariant. Une construction géométrique de birapports invariants est possible. Voilà les différentes étapes de cette construction lorsqu'on se place dans un espace réel. La même construction est possible analytiquement dans un espace complexe, mais la visualisation est alors délicate.

1. Le fait qu'une droite soit incluse dans une quadrique impose trois contraintes à cette quadrique. Soient  $A$  et  $B$  deux points et  $\mathcal{Q}$  une quadrique d'équation :

$$ax^2 + by^2 + cz^2 + dyz + exz + fxy + gx + hy + jz + k = 0$$

Supposons que la droite  $(AB)$  est incluse dans la quadrique  $\mathcal{Q}$  : un point quelconque de cette droite peut s'écrire sous la forme  $P = \lambda A + B$  (avec  $\lambda$  infini pour  $P = A$ ). On a :

$$\begin{aligned} (\forall \lambda \quad P \in \mathcal{Q}) \implies \\ a(\lambda x_A + x_B)^2 + b(\lambda y_A + y_B)^2 + c(\lambda z_A + z_B)^2 \\ + d(\lambda y_A + y_B)(\lambda z_A + z_B) + e(\lambda x_A + x_B)(\lambda z_A + z_B) + f(\lambda x_A + x_B)(\lambda y_A + y_B) \\ + g(\lambda x_A + x_B) + h(\lambda y_A + y_B) + i(\lambda z_A + z_B) + j = 0 \end{aligned}$$

On obtient les trois contraintes en annulant les termes en  $\lambda^2$ ,  $\lambda^1$  et  $\lambda^0$ .

2. Une quadrique est caractérisée par dix paramètres définis à un facteur de proportionnalité près. Le fait que trois droites appartiennent à cette quadrique fournit neuf contraintes, ce qui définit complètement la quadrique.
3. Une quadrique contenant des droites est appelée quadrique réglée. C'est soit un hyperboloïde à une nappe, soit un parabololoïde hyperbolique. Les droites incluses sont appelées génératrices et se répartissent en deux familles. En tout point de la quadrique passe une unique droite de chaque famille. Deux droites de la même famille ne se coupent pas, mais une droite d'une famille coupe toutes les droites de l'autre famille.
4. Quatre droites de la même famille définissent un birapport. En effet, toute droite  $D$  de l'autre famille coupe ces quatre droites en quatre points qui définissent un birapport. Ce birapport ne dépend pas du choix de la droite  $D$ , mais seulement des quatre droites initiales.
5. On considère quatre droites de l'espace 3D  $D_1, D_2, D_3$  et  $D_4$ , telles que deux quelconques d'entre elles ne soient pas coplanaires. Trois d'entre elles,  $D_1, D_2$  et  $D_3$  par exemple, définissent une quadrique réglée (cf. FIG. 2.28). Comme ces trois droites ne sont pas coplanaires, elles ne se coupent pas et appartiennent donc à la même famille  $F_1$  de génératrices. La quatrième droite coupe cette quadrique en deux points  $P_1$  et  $P_2$ . Par chacun de ces points, il passe une droite de la famille  $F_1$  :  $\Delta_1$  et  $\Delta_2$ . On obtient ainsi cinq droites de la même famille soit deux birapports.

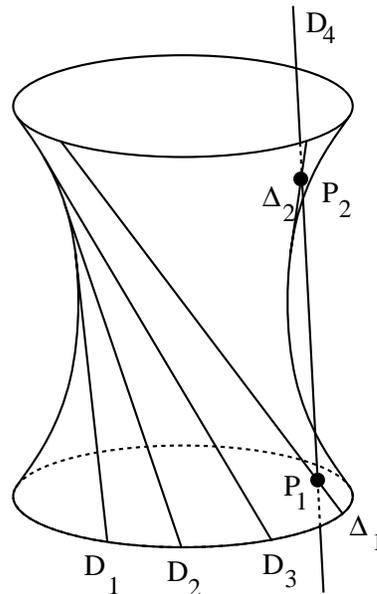


FIG. 2.28 - Cas de quatre droites quelconques.

Si on permute le rôle des droites, on trouve ainsi huit birapports. On peut se poser alors les questions suivantes : combien y a-t-il de birapports indépendants ? Ces birapports sont-ils calculables à partir de deux images lorsqu'on connaît la géométrie épipolaire ?

On a démontré qu'il existe deux droites  $L_1$  et  $L_2$  qui coupent quatre droites quelconques. Ces deux droites comportent donc quatre points d'intersection et définissent chacune un birapport.

Réciproquement, il n'existe pas plus de deux telles droites. Si une troisième droite  $L_3$  coupait aussi les quatre droites, elle couperait la quadrique  $\mathcal{Q}$  en trois points et serait donc incluse dans  $\mathcal{Q}$ . Elle couperait donc la quatrième droite en  $P_1$  ou  $P_2$  et serait égale à  $L_1$  ou  $L_2$  d'après les propriétés des familles de génératrices.

Comme les configurations de quatre droites quelconques possèdent 16 degrés de liberté alors que le groupe des homographies de l'espace n'en possède que 15, il doit exister un sous-groupe d'isotropie pour chaque configuration. Hartley<sup>1</sup>, qui a travaillé indépendamment sur ces invariants, a trouvé ce groupe. Si on choisit la représentation de l'espace de telle manière que les équations de  $L_1$  et  $L_2$  soient respectivement :

$$x = y = 0 \quad \text{et} \quad z = t = 0$$

le sous groupe d'isotropie est formé des transformations dont la matrice est diagonale de la forme  $\text{diag}(\alpha, \alpha, \beta, \beta)$ , où  $\alpha$  et  $\beta$  sont deux constantes définies à un facteur multiplicatif près. Ce groupe est d'ordre 1.

### Configurations de points, droites et coniques

**Cas d'une conique et trois points.** Une conique de l'espace tridimensionnel est définie par huit paramètres: trois pour définir le plan dans lequel elle se trouve, et cinq pour définir la conique dans ce plan. Chaque point ayant trois paramètres, une configuration a 17 degrés de liberté. L'ordre du groupe étant 15, il devrait y avoir deux invariants au moins.

Les trois points définissent trois droites, qui coupent le plan en trois points alignés. La droite ainsi formée coupe la conique en deux points: on dispose de cinq points alignés et donc de deux birapports (cf. FIG. 2.29). Ce sont les seuls invariants.

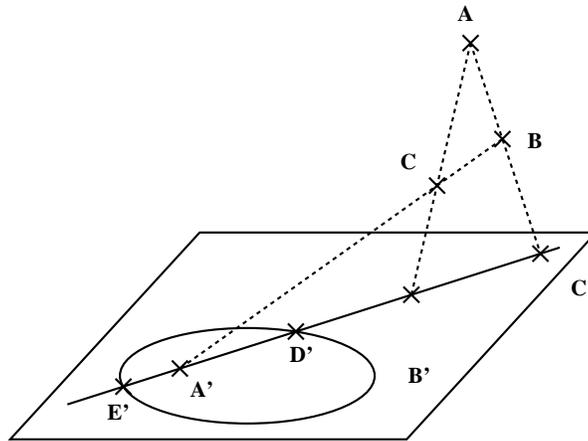


FIG. 2.29 - Cas d'une conique et de trois points.

<sup>1</sup>Il a présenté ses résultats à la journée workshop qui a suivi le congrès ECCV92, mais n'a pas encore publié d'article sur le sujet.

**Autres cas.** Dans le cas d'une conique, d'une droite et de deux points, les deux points définissent une droite. On a donc deux droites qui coupent le plan de la conique en deux points, qui définissent à leur tour une droite qui coupe la conique deux fois. Cela fait quatre points alignés, d'où un birapport. Pour trouver les deux autres invariants, considérons les plans comprenant la droite de la configuration. Deux de ces plans sont tangents à la conique, deux passent par un des points de la configuration et un passe par le point d'intersection du plan de la conique et de la droite définie par les deux points. On a ainsi cinq plans d'un même faisceau, d'où deux faisceaux de quatre plans et deux birapports.

D'une manière similaire, on peut trouver les invariants d'une conique et deux droites, de deux coniques...

### 2.2.2 Expérimentations à partir de données simulées

#### Méthode

De même qu'avec le test de coplanarité, la théorie a besoin d'être validée expérimentalement. On a donc procédé de manière tout à fait similaire. Ce paragraphe présente quelques résultats obtenus à partir d'images de synthèse, les images réelles étant étudiées dans le suivant.

On a choisi d'étudier l'invariant de deux paires de droites coplanaires. Cette configuration présente, en effet, plusieurs facilités pour l'expérimentation : on peut énumérer toutes les configurations de ce type de manière simple (en énumérant les jonctions communes de chacune des deux paires), il n'y a qu'un birapport défini par cette configuration, les segments extraits d'une image sont souvent plus fiables que leurs extrémités.

On a utilisé les mêmes méthodes que pour les tests sur la coplanarité : le bruit est uniforme d'amplitude variable, les birapports sont comparés en employant la méthode des birapports aléatoires, la géométrie épipolaire est calculée avec le même algorithme.

Le plan d'expérience suivi a pour but principal d'étudier la stabilité et la robustesse des invariants. Pour cela, on fait plusieurs expériences qui permettent des évaluations différentes : comparaison du calcul mené dans deux images différentes, dans deux paires d'images d'un même objet, comparaison entre le calcul dans les images et le calcul dans la scène 3D, entre notre méthode et une autre méthode. On a donc procédé de la manière suivante.

1. On étudie tout d'abord la stabilité du calcul de l'invariant. Pour cela, on se sert du fait que le calcul peut être mené dans les deux images : cela donne deux résultats qui devraient être égaux, et qui diffèrent en fait selon le bruit qui affecte les images. Cette étape permet donc de déterminer des seuils de stabilité.
2. Pour jauger ces seuils, on compte le nombre d'invariants dont la stabilité est inférieure à chacun d'eux. On appelle stabilité la différence des valeurs obtenues lors du calcul dans chaque image.
3. Une deuxième manière d'estimer la stabilité du résultat est utilisée : on calcule l'invariant 3D directement à partir des points 3D non bruités, et on compare cet invariant à celui calculé à partir des images bruitées. La première méthode teste la stabilité du calcul, celle-ci la stabilité du résultat.

4. On peut alors comparer notre méthode de calcul purement 2D avec une méthode par reconstruction. A partir des données bruitées, on reconstruit la scène en 3D par une méthode de reconstruction projective, et on compare les invariants calculés à partir de cette reconstruction avec ceux calculés à partir de la scène 3D initiale.
5. On teste enfin l'invariance par changement de point de vue. On prend quatre images d'une scène et on calcule les invariants pour les deux paires d'images. En comparant ces invariants, on peut alors apprécier "l'invariance" des invariants.

### Détermination de seuils de stabilité

Dans cette première expérimentation et dans toutes les suivantes, les images sont obtenues de manière identique. On tire 450 points aléatoires dans un cube puis on ajoute des segments entre les points 0 et 1, 0 et 2, puis 3 et 4, 3 et 5... de manière à obtenir les paires de droites coplanaires à partir desquelles sont calculés les invariants. Cette scène est projetée dans plusieurs images (deux ou quatre), la caméra effectuant une rotation de 30 degrés entre deux prises de vue. Les projections sont bruitées, puis la géométrie épipolaire est calculée.

On énumère alors les 22 350 configurations de deux paires de droites coplanaires. Ces configurations sont filtrées pour éliminer certaines configurations :

1. celles qui comportent des segments trop courts ou trop proches les uns des autres ;
2. celles qui comportent des colinéarités gênantes : angle trop faible entre deux droites, épipole sur une des droites...
3. celles qui comportent des coplanarités accidentelles : trois des segments appartiennent alors à un même plan et les quatre points d'intersection ne sont pas tous définis.

Les invariants sont alors calculés pour les configurations restantes, comme indiqué au paragraphe précédent. On utilise la deuxième méthode de construction des projections de nouveaux points coplanaires du paragraphe 2.1.2 pour calculer les projections des points d'intersection entre plans et droites, puis on calcule dans chaque image le birapport des quatre points ainsi construits. Les expérimentations menées permettent donc de tester aussi l'algorithme de construction de points coplanaires.

Dans un premier temps, on mesure juste la distance entre les birapports calculés dans chaque image, pour différents niveaux de bruit. On utilise pour cela la distance de la formule (2.17). Il faut noter que cette formule n'a pas été calculée dans le cas des birapports de quatre droites concourantes, mais dans celui de quatre points alignés. Son utilisation ici est justifiée par le fait que c'est la mesure "la plus raisonnable" dont on dispose pour comparer des birapports (qui tiennent par exemple compte des symétries du birapport), et par l'étude expérimentale faite dans [Åst 92] qui montre que les fonctions de répartition du birapport dans le cas de 4 points alignés et dans celui de 4 droites concourantes ont des allures similaires.

La FIG. 2.30 montre les valeurs minimales, moyennes et maximales obtenues aux différents niveaux de bruit. Les valeurs moyennes forment une courbe de très faible pente, ce qui indique donc une très bonne stabilité du calcul effectué. Les valeurs

maximales correspondent à des cas dégénérés non filtrés, et ce même en l'absence de bruit rajouté (il reste le bruit numérique). La qualité du calcul vient, entre autres, du nombre de points présents dans les images. Même avec un fort niveau de bruit, le calcul de la géométrie épipolaire est stable : la méthode utilisée n'est pas linéaire et elle utilise la redondance des données, ce qui permet d'éliminer les effets du bruit.

Pour compléter ces données, la FIG. 2.31 montre des histogrammes des valeurs obtenues pour quelques niveaux de bruit. A partir de ces histogrammes, on a alors compté les configurations qui donnent des distances inférieures à trois valeurs : la première valeur est le seuil moyen déterminé pour chaque niveau de bruit et représenté sur la FIG. 2.30, la deuxième valeur est cette première valeur multipliée par deux et la troisième est cette même valeur multipliée par trois. Le résultat est montré sur la FIG. 2.32.

Ces nouvelles données confirment le caractère d'isolement des valeurs maximales : 85% des valeurs sont inférieures à deux fois la valeur moyenne et 90% sont inférieures à trois fois cette valeur moyenne.

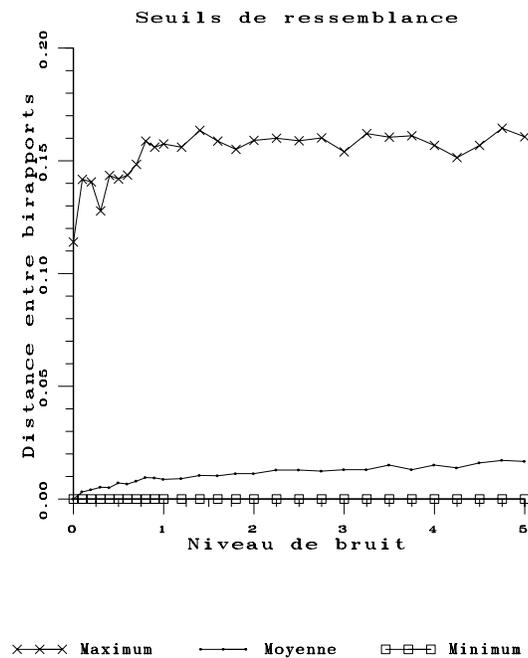


FIG. 2.30 - Seuils de stabilité des invariants de deux paires de droites coplanaires.

### Comparaison avec les invariants exacts

A cette étape, on compare les invariants calculés à partir des images avec ceux que l'on calcule directement sur la scène 3D. Le calcul est simple dans ce dernier cas : on calcule les deux plans définis par la configuration, les quatre points d'intersection et leur birapport (cf. FIG. 2.27). On calcule alors la distance entre ce birapport et celui obtenu dans une des deux images.

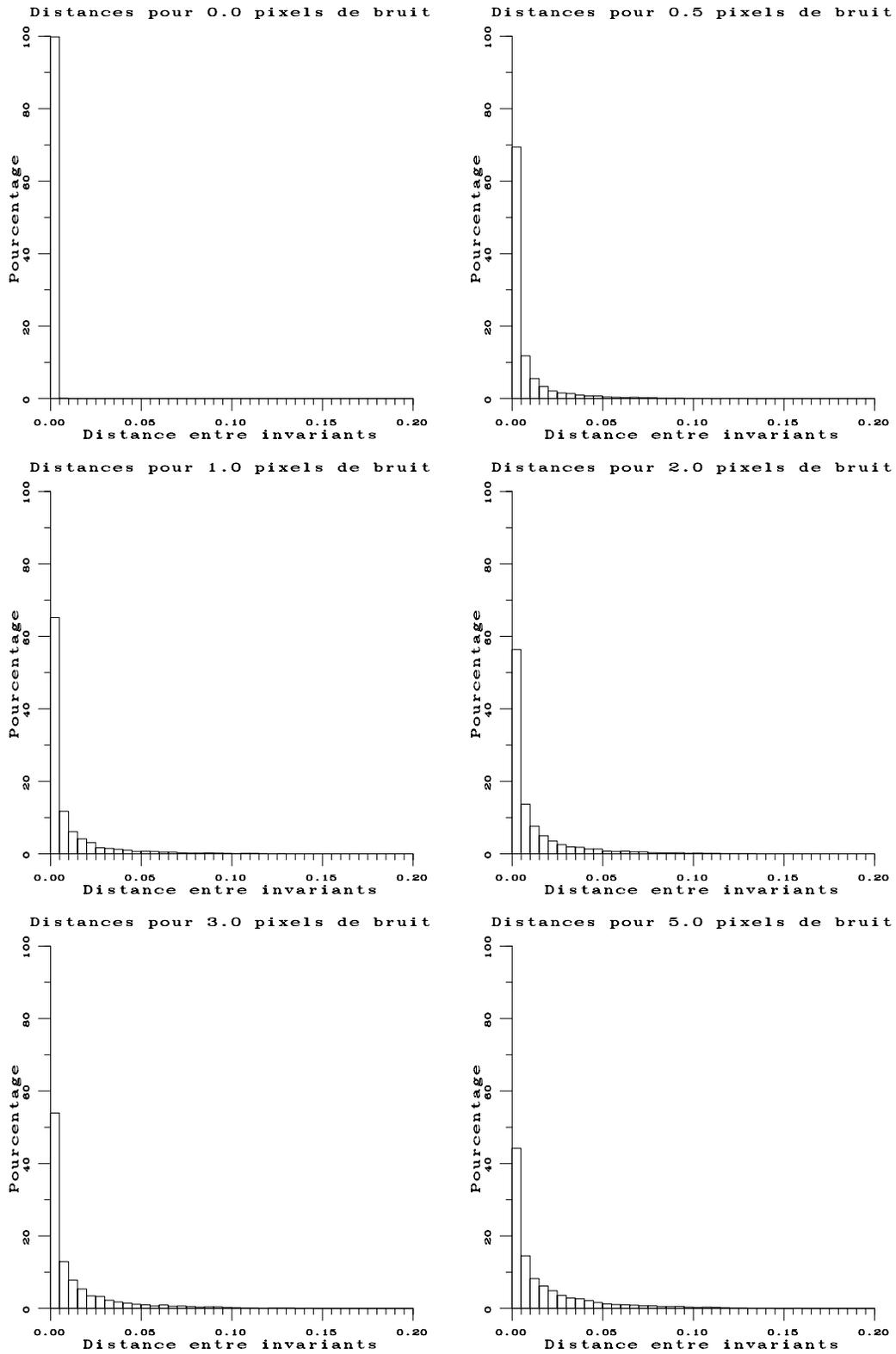


FIG. 2.31 - Histogrammes des distances obtenues lors du calcul de l'invariant de deux paires de droites coplanaires.

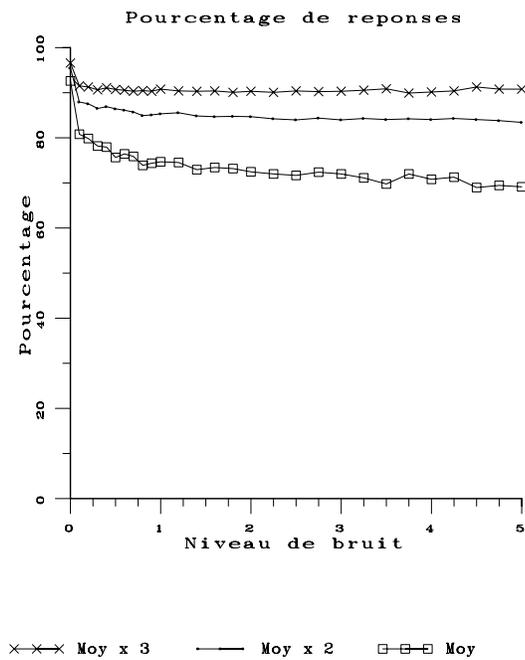


FIG. 2.32 - Pourcentage de configurations de deux paires de droites coplanaires donnant des distances inférieures au seuil moyen et à deux et trois fois cette valeur.

La FIG. 2.33 montre les valeurs minimale, moyenne et maximale qui sont obtenues aux différents niveaux de bruit. Il faut bien noter que le calcul à partir des images est fait avec des données bruitées, contrairement à celui fait à partir de la scène.

En moyenne, le calcul des invariants est donc précis. De même que précédemment, certaines configurations donnent de mauvais résultats dès qu'il existe un peu de bruit dans les images. Le bruit amène une dégradation régulière des résultats : il n'y a pas de valeur critique. Cela est confirmé par la FIG. 2.34 qui montre les histogrammes des valeurs obtenues pour quelques niveaux de bruit.

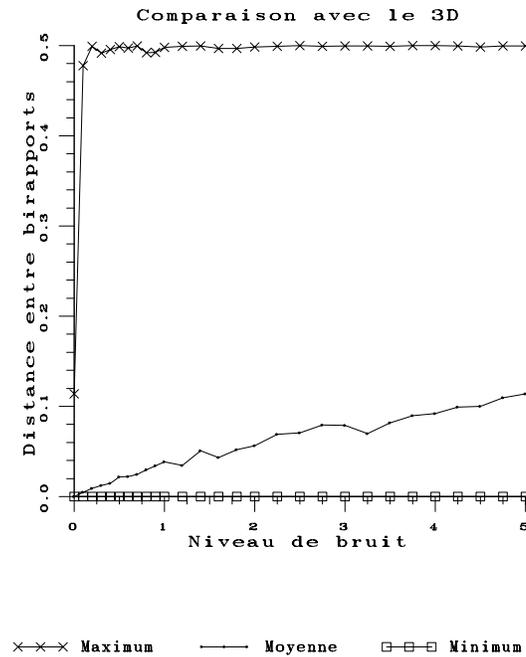


FIG. 2.33 - Comparaison entre le calcul à partir des images et celui à partir de la scène 3D.

### Comparaison avec une méthode par reconstruction

On compare maintenant les résultats obtenus par notre méthode avec ceux obtenus par une méthode utilisant une reconstruction de la scène. Dans notre méthode, bien qu'on calcule des informations 3D, le calcul est purement 2D : on fait des constructions dans chaque image, les points coplanaires construits ne sont connus que par leurs projections.

Dans l'autre méthode, on reconstruit la scène à partir des données contenues dans les images, puis les constructions et calculs sont faits à partir de cette reconstruction. La méthode de reconstruction utilisée ici est celle proposée dans [Fau 92a].

Cette méthode est basée sur le fait que, lorsqu'on dispose de suffisamment de points dans plusieurs images, il est possible de résoudre les équations de projections. Lors de cette résolution, on détermine à la fois les paramètres des caméras et les coordonnées des

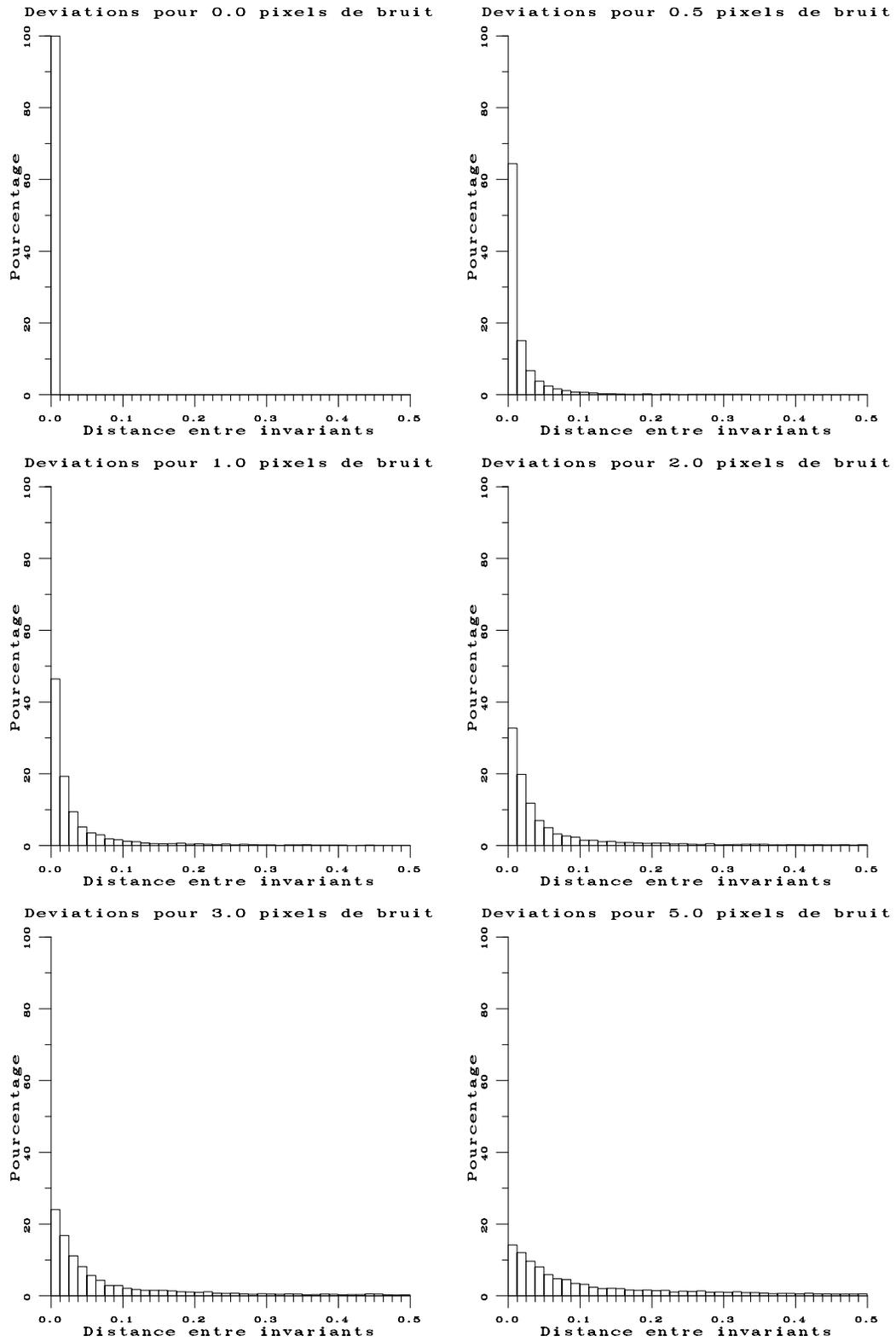


FIG. 2.34 - Histogrammes des distances entre les invariants calculés à partir de la scène et ceux calculés à partir des images.

points 3D de la scène. Cette reconstruction, basée sur l'utilisation d'un algorithme de résolution d'équations non linéaires, ne suppose donc aucune hypothèse supplémentaire par rapport à notre méthode : l'étalonnage des caméras, par exemple, n'a pas à être connu.

A partir des mêmes images que celles ayant servi aux tests précédents, on reconstruit donc la scène puis on calcule les invariants des configurations. La méthode de reconstruction fonctionnant à partir de points, on reconstruit les points à partir de leurs projections bruitées dans les deux images, puis les droites sont déduites de ces points. Les points d'intersection des droites faisant partie des points reconstruits, la coplanarité des droites reconstruites est donc assurée. Le calcul des invariants ne présentent alors pas de difficulté.

On compare alors ces invariants à ceux calculés à partir de la scène 3D initiale. Les valeurs minimale, moyenne et maximale de la distance obtenue sont montrées sur la FIG. 2.35. Il faut bien noter que la reconstruction est faite à partir des données bruitées des images.

On peut alors comparer les résultats des deux méthodes. Les courbes des minima et des maxima sont très semblables. Par contre, la courbe des valeurs moyennes présente deux différences majeures :

1. même avec le grand nombre de points présents dans chaque image, la courbe n'est pas régulière avec la méthode par reconstruction. Cette méthode utilise aussi un algorithme utilisant la redondance des informations, mais celui-ci doit être plus instable que le seul algorithme de calcul de la géométrie épipolaire ;
2. les valeurs moyennes sont supérieures pour la méthode par reconstruction : la phase de reconstruction, au lieu de supprimer encore les effets du bruit, tend à détériorer les résultats. Une autre explication de ces résultats est le changement de dimension : les calculs géométriques en 3D ne sont peut être pas aussi stables que ceux fait en 2D, les calculs d'intersection en particulier.

Cette comparaison montre l'intérêt de notre approche 2D. Premièrement les résultats sont plus stables et de meilleure qualité. Deuxièmement, il est possible de ne calculer que quelques invariants, alors que la méthode par reconstruction nécessite d'abord une reconstruction complète et coûteuse, avant tout calcul d'invariant.

### **Invariance par changement de point de vue**

Pour ce dernier test, on étudie "l'invariance" des invariants par changement de point de vue et leur caractère discriminant. Pour cela, on prend quatre images de la même scène 3D, chaque image différant de la précédent par une rotation de la caméra de 30 degrés. On considère alors les deux paires d'images formées des deux premières et des deux dernières images.

Pour chacune de ces paires d'images, on énumère les configurations de deux paires de droites coplanaires, on les filtre de la même manière que précédemment, puis on calcule les invariants. Avant de mesurer la distance entre les invariants obtenus dans chaque paire d'images, on a symétrisé ces birapports.

Quatre points sur une droite définissent par permutation sur l'ordre de ces points six valeurs de birapport, qui peuvent se déduire les unes des autres (cf. 1.1.4). Pour

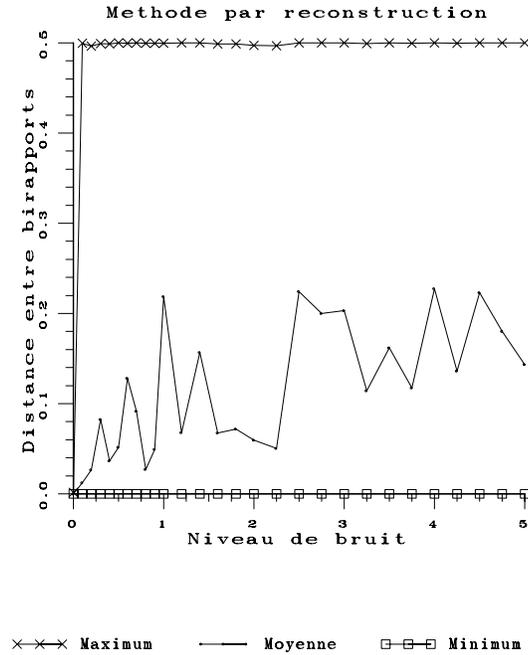


FIG. 2.35 - Résultats obtenus avec une méthode de reconstruction.

symétriser une de ces valeurs, on la remplace par celle des six qui est, par exemple, comprise entre 0 et 0,5. Cette valeur ne dépend plus alors de l'ordre des points. On peut symétriser de même tous les birapports.

On a mesuré les invariants dans chacune des deux paires d'images, puis symétrisé les valeurs obtenues. On calcule les distances entre tous les invariants de la première paire avec tous ceux de la deuxième, et on compare ces invariants à trois seuils : deux et trois fois la valeur du seuil de stabilité moyen et la valeur du seuil de stabilité maximal, déterminés au début de cette expérimentation.

Pour présenter ces résultats, on a séparé deux cas. Le premier cas est celui des invariants qui se correspondent : pour chaque configuration 3D, projetée dans les quatre images, on a calculé les invariants dans chaque paire d'images et leur distance, puis comparé cette distance aux trois seuils. Le pourcentage de ces configurations qui donnent une distance inférieure à chacun des seuils est indiqué sur la FIG. 2.36. Cette mesure permet donc d'estimer la stabilité des invariants lors d'un changement de point de vue.

Le deuxième cas est celui de toutes les autres distances calculées. Elles ne correspondent donc pas à la stabilité d'un invariant 3D, mais permettent d'estimer le caractère discriminant des invariants : si on se sert des invariants pour appairer les configurations entre les deux paires d'images, la FIG. 2.36 indique donc le pourcentage d'appariements corrects effectivement détectés, et la FIG. 2.37 celui des appariements incorrects détectés.

Les pourcentages des appariements corrects détectés forment une courbe avec une

pente fortement négative pour les faibles niveaux de bruit. La courbe a ensuite une pente négative, mais moins élevée en valeur absolue. Par contre, les pourcentages d'appariements incorrects détectés forment une courbe de faible pente positive. L'effet principal du bruit est donc principalement de dégrader la stabilité des invariants. Dès les plus faibles niveaux de bruit, le pourcentage d'appariements corrects détectés baisse ainsi de moitié pour le premier seuil et de plus de 30% pour le deuxième.

Les courbes obtenues avec le troisième seuil montrent que l'obtention de forts pourcentages dans la première figure (supérieur à 80% pour moins de 3 pixels de bruit) a pour conséquence un fort pourcentage d'appariements incorrects détectés: environ 30%.

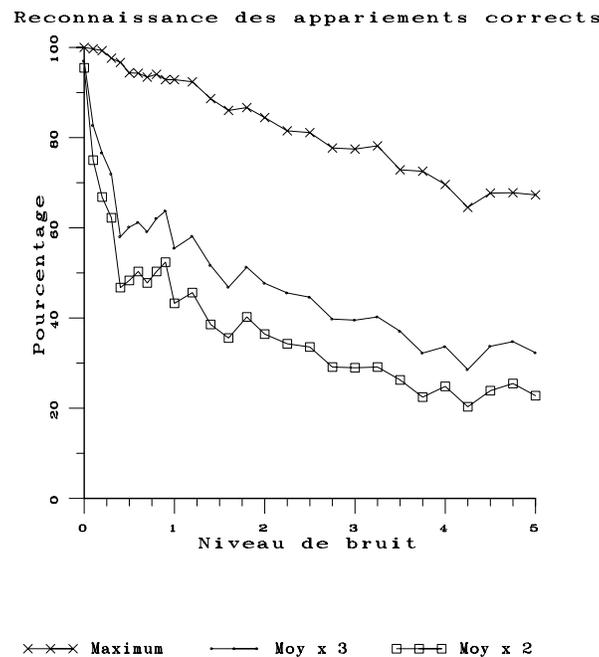


FIG. 2.36 - Pourcentage d'appariements corrects détectés.

## Conclusion

Cette première étude des invariants 3D fait ressortir quelques points clés :

- l'influence nette du bruit qui dégrade la précision et le caractère discriminant des invariants ;
- le fait qu'il vaut mieux calculer en 2D que faire une reconstruction préalable de la scène avant de calculer les invariants.

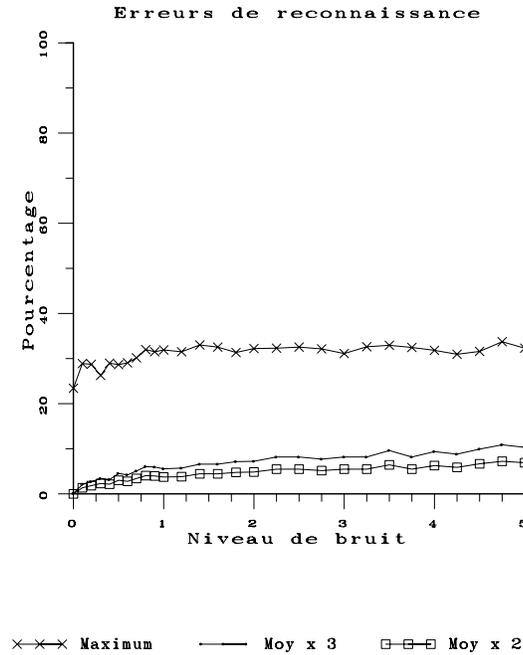


FIG. 2.37 - Pourcentage d'appariements incorrects détectés.

### 2.2.3 Expérimentations à partir de données réelles

Pour terminer cette étude expérimentale, on a testé les mêmes invariants 3D de deux paires de droites coplanaires sur une paire d'images réelles. La difficulté d'une telle étude provient du fait que cet invariant n'est pas adapté à la plupart des objets polyédriques habituels. En effet, les arêtes de ceux-ci n'ont souvent que trois directions différentes. Lors de la construction, des points sont donc à l'infini, ce qui pose des problèmes de précision et d'arrondi dans les calculs menés.

Pour remédier à cela, on étudie deux images d'un dièdre sur lequel on a tracé des triangles. Cela donne de nombreuses directions et des calculs plus stables. On a pris trois images de cet objet (cf. FIG. 2.38). Les contours ont été extraits en utilisant le détecteur de Deriche, puis on a réalisé une approximation polygonale. Ces segments sont numérotés comme indiqué sur la FIG. 2.39.



FIG. 2.38 - Trois images d'un dièdre.

On a alors calculé, en utilisant les deux premières images ou les deux dernières, les

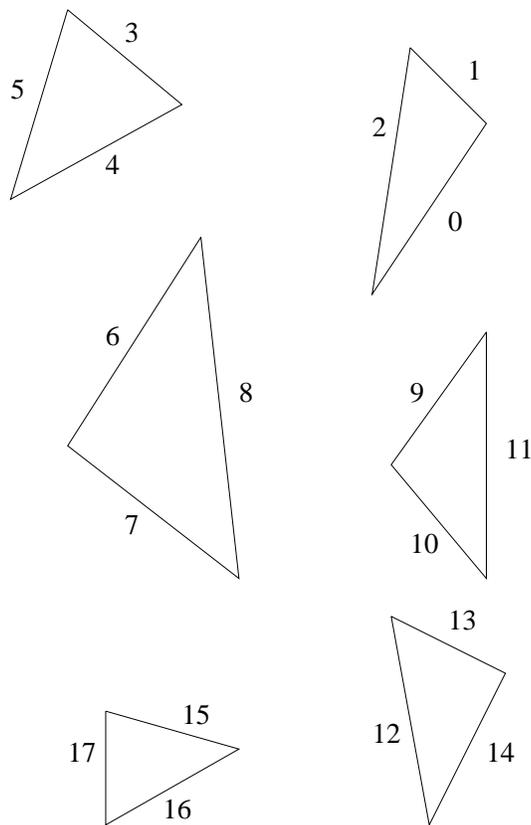


FIG. 2.39 - Numérotation des segments du dièdre.

invariants des configurations suivantes de segments :

$$\begin{aligned}
 S_1 &= \{0, 2, 3, 5\} \\
 S_2 &= \{1, 2, 15, 16\} \\
 S_3 &= \{3, 5, 9, 11\} \\
 S_4 &= \{3, 5, 12, 14\} \\
 S_5 &= \{7, 8, 12, 14\} \\
 S_6 &= \{9, 11, 15, 17\} \\
 S_7 &= \{12, 13, 15, 17\}
 \end{aligned}$$

La FIG. 2.40 montre alors les distances entre ces invariants. L'élément  $(i, j)$  du tableau contient la distance entre l'invariant de la configuration  $S_i$  calculée à partir des deux premières images et l'invariant de la configuration  $S_j$  calculé à partir des deux dernières images.

L'invariant apparaît très discriminant, en prenant, par exemple, un seuil égal à 0,01. La seule ambiguïté est formée par l'appariement entre les configurations entre  $S_1$  et  $S_7$ . Nous employons le terme ambiguïté et non erreur, car il est possible que ces deux configurations diffèrent l'une de l'autre par une homographie de l'espace. Cela est difficile à exclure ou à vérifier effectivement.

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

FIG. 2.40 - Distance entre les invariants calculés dans les deux paires d'images.

## 2.2.4 Conclusion

Ce paragraphe concernant les invariants 3D s'oriente selon deux axes :

1. il montre comment la théorie des invariants et la méthode de calcul par recherche des invariants canoniques peuvent être appliquées dans des cas non triviaux, avec des données bruitées ;
2. il expose comment extraire des informations 3D à partir d'images, sans étalonnage des caméras, ni reconstruction 3D préalable de la scène.

Les tests, réalisés tant sur les données artificielles que réelles, valident la méthode, avec toutefois deux réserves. Tout d'abord, le calcul est sensible au bruit. Il faut donc veiller à la qualité des données : cela exige de surveiller l'éclairage de la scène, l'extraction des contours, l'approximation polygonale... Les techniques employées ici ne sont pas les plus précises, et on pourrait avec profit les remplacer par d'autres : détection des segments à l'aide de splines, détection de coins par un algorithme spécialisé...

Ensuite, les invariants testés ne sont pas très bien adaptés aux objets dont les arêtes ont seulement trois directions. Il faudrait donc compléter cette étude par le test d'autres invariants.

Ces réserves étant données, les résultats sont tout à fait satisfaisants. Ils montrent, en effet, qu'il est possible d'extraire des données numériques d'images, sans utiliser un protocole très lourd nécessitant l'étalonnage des caméras, l'utilisation de capteurs 3D, des techniques de reconstruction, la connaissances d'informations préalables...

## 2.3 Appariement d'images

### 2.3.1 Introduction

La troisième et dernière application des invariants, qui est traitée dans ce chapitre, est géométriquement bien moins spectaculaire que les deux autres. On y utilise, en effet, des invariants fort simples, les invariants pour les similitudes 2D. Par contre, le problème à résoudre, l'appariement des images, est un problème de base difficile de la vision. C'est donc à l'aune de la vision que ce paragraphe doit être apprécié, plutôt qu'à celle de la géométrie.

Le problème de l'appariement, ou mise en correspondance, des images est le suivant. On dispose de deux images représentant partiellement le même objet ou la même scène. Chacune de ces images contient des primitives, points et segments dans notre cas, courbes et régions dans d'autres applications, qui sont les projections de primitives de la scène observée. Mettre en correspondance deux images consiste à trouver les primitives de chaque image qui correspondent à une même primitive de la scène 3D. Ainsi, deux points sont appariés s'ils sont les projections dans chacune des images d'un même point 3D.

L'appariement des images est un problème central de la vision. On peut, en effet, classer grossièrement les algorithmes de vision en trois catégories.

1. Tout d'abord les algorithmes de bas niveau traitant une seule image, pour en extraire des points d'intérêt [Der 90b], des contours [Can 86, Der 87a], des régions [Naz 84, Mon 87, Fra 90] ou un mélange des trois [Bar 86, Der 87b, Mon 88]. Notons que ces primitives peuvent être regroupées ou traitées pour obtenir des primitives de plus haut niveau (approximation polygonale des contours...) [Hor 90].
2. Une deuxième catégorie regroupe les algorithmes utilisant plusieurs images. A partir de ces images, on peut obtenir des informations 3D (reconstruction) de mouvement [Zha 90a] ou de position : c'est le problème classique de la stéréovision [Aya 89, Alo 88, Vai 89, Coc 90]. Ce problème peut être traité en utilisant des caméras étalonnées [Tsa 87, Fau 92c, May 92], ou être résolu de manière relative, c'est à dire à une transformation projective de l'espace près [Fau 92a, Moh 92b].
3. La dernière catégorie fait le lien entre les précédentes en réalisant l'appariement des images. Cet appariement peut être réalisé tant pour les contours [Sha 85a, Sko 88] que pour les régions [Coc 87, Wro 87].

Plusieurs approches sont classiquement utilisées pour résoudre le problème de la mise en correspondance.

Une première approche, par corrélation, est utilisée lorsque le mouvement de la caméra entre les deux prises de vue est faible ou approximativement connu [Gru 85, Ana 89, Fua 90]. On connaît alors la position approchée de la primitive recherchée dans la deuxième image, et on peut affiner cette approximation en calculant le maximum d'une convolution entre une région de la première image et une de la seconde.

Cette méthode peut être étendue au cas des séquences d'images, lorsqu'on suppose que le mouvement de la caméra est régulier entre les différentes prises de vue. Lorsque deux ou trois images ont déjà été appariées, le mouvement de la caméra, ou celui des primitives, peut être prédit dans chaque image en fonction de ce qui est observé dans

les images précédentes de la séquence. Cette prédiction est réalisée au moyen d'un filtre, comme celui de Kalman par exemple. Une telle méthode est appelée "poursuite d'indices" ou "tracking" [Mat 89, Cro 90, Der 90a, Vi 90, Zha 90a, Zha 90b, Zha 92].

Une deuxième approche, très utilisée en stéréovision, suppose connue la géométrie épipolaire entre les 2 ou 3 images à apparier. On sait alors que le correspondant d'un point d'une image doit se trouver sur une certaine droite de l'autre image, sa droite épipolaire. La recherche s'effectue alors le long de cette droite [Zha 89, Aya 89]. En cas de stéréovision trinoculaire, le système est presque totalement contraint et les ambiguïtés d'appariement disparaissent presque complètement.

Une troisième approche est basée sur une description topologique, voire symbolique, des images. Les images sont, par exemple, décrites comme des graphes de segments, et l'appariement est basé sur les contraintes topologiques [Aya 85, Sko 88, H91].

La méthode proposée ici est basée sur une approximation géométrique du mouvement apparent des primitives entre les deux images, par une similitude dans le cas présent. On peut alors se servir des invariants associés à ce type de transformations. A chaque paire de segments connectés, on associe un angle et un rapport de distances. Les valeurs numériques obtenues dans chaque image sont comparées, et cela fournit de premières propositions d'appariements.

On peut filtrer ces propositions en exprimant la contrainte d'unicité du mouvement apparent. A chaque appariement, on peut, en effet, associer une transformation, représentée par un point dans l'espace des paramètres des transformations. Chaque proposition donne ainsi un point dans cet espace des paramètres. L'unicité du mouvement apparent se traduit alors par l'existence d'un point d'accumulation dans ce même espace. En ne gardant que les points se trouvant autour du point d'accumulation, on peut filtrer les bons appariements des autres.

Cette technique, à la fois locale par ses invariants, et globale par l'expression de la contrainte d'unicité, présente plusieurs avantages.

- L'objet observé peut ne pas être seul à apparaître sur les images : le "fond" peut aussi être représenté dans l'image. Ce même objet peut aussi être partiellement occulté dans l'une des images.
- Toutes les primitives ne peuvent être appariées, dès que le bruit perturbe la connexité entre les segments. Par contre, les mauvais appariements sont très peu nombreux. Cette technique d'appariement peut donc être considéré comme un premier pas d'une méthode plus globale, qui la compléterait en utilisant d'autres sources d'information, obtenant ainsi un appariement de meilleure qualité (voir le chapitre 3).

La suite de ce paragraphe contient une description précise de l'algorithme proposé et une justification mathématique de son principe, puis des résultats expérimentaux.

### 2.3.2 Algorithme proposé

On considère deux images contenant des points et des segments reliant ces points. On cherche les segments qui sont les projections d'une même arête de l'objet observé. Le résultat attendu est donc une relation entre les segments et les points de chacune des deux images. L'algorithme proposé comprend plusieurs étapes qui sont détaillées successivement.

### Approximation du mouvement apparent

Soit un point 3D  $P$  dont les projections  $P_1$  et  $P_2$  dans chacune des deux images ont respectivement pour coordonnées dans les images  $(x_1, y_1)$  et  $(x_2, y_2)$ . Le mouvement apparent de  $P$  entre les deux images est la transformation qui amène  $P_1$  sur  $P_2$ , c'est à dire la translation de vecteur  $(x_2 - x_1, y_2 - y_1)$ .

Si on considère plusieurs points ou arêtes, le mouvement apparent est la transformation entre les images qui amènent les projections de la première image en superposition sur celles de la deuxième.

Dans la suite, on suppose que ce mouvement apparent est, entre les deux images que l'on considère, une similitude directe, c'est à dire la composée d'une rotation, d'une translation et d'une homothétie. Ce choix est commenté plus loin. Il existe donc des constantes  $k$ ,  $\theta$ ,  $t_x$  et  $t_y$ , telles que, pour tous les couples de points appariés :

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = k \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

L'algorithme consiste à calculer cette similitude pour retrouver les primitives des images qui se correspondent.

### Calcul des invariants locaux

Dans chaque image, on considère les paires de segments qui ont une extrémités commune. Une telle paire est définie par trois points  $A, B$  et  $C$ ,  $A$  étant l'extrémité commune au deux segments. On associe à cette paire deux invariants pour les similitudes : un angle  $\alpha$ , et un rapport de longueur  $\rho$ .

L'angle est celui des deux angles définis par les vecteurs  $\vec{AB}$  et  $\vec{AC}$  qui est compris entre 0 et  $\pi$ . Pour le calculer, on utilise le produit scalaire :

$$(\widehat{\vec{AB}, \vec{AC}}) = \arccos \frac{\vec{AB} \cdot \vec{AC}}{\|\vec{AB}\| \|\vec{AC}\|}$$

Le rapport de longueurs calculé est :

$$\rho = \min\left(\frac{\|\vec{AB}\|}{\|\vec{AC}\|}, \frac{\|\vec{AC}\|}{\|\vec{AB}\|}\right)$$

### Appariement des invariants

Les couples d'invariants des deux images sont comparés. Pour cela, on utilise deux seuils :  $\alpha_{max} = 15^\circ$  et  $\rho_{max} = 1,5$ . Ces seuils sont fixés en fonction du bruit existant dans les images. Une paire de segments de la première image d'invariants  $(\alpha_1, \rho_1)$  est appariée avec une paire de la deuxième image d'invariants  $(\alpha_2, \rho_2)$  si :

$$\begin{aligned} |\alpha_2 - \alpha_1| &< \alpha_{max} \\ \frac{\rho_2}{\rho_1} &< \rho_{max} \\ \frac{\rho_1}{\rho_2} &< \rho_{max} \end{aligned}$$

A cette étape, une paire de segments d'une image est classiquement appariée avec une dizaine de paires de segments de l'autre image.

### Filtrage des appariements

Parmi tous les appariement réalisés, seule une petite partie est correcte. Pour la séparer des mauvais appariements, on se sert du fait que tous les appariements corrects doivent définir le même mouvement apparent, et on utilise une technique de Hough.

Lorsque deux paires de segments sont appariées, on peut calculer la similitude qui transforme la première paire en la deuxième. Cette similitude possède quatre paramètres, deux de translation, un angle de rotation et un facteur d'homothétie, et peut donc être représentée par un point dans l'espace 4D des paramètres de similitude. On associe donc un tel point à tous les appariements réalisés à l'étape précédente.

Les mauvais appariements ne sont pas corrélés, et les points correspondants dans l'espace des paramètres sont donc répartis de manière très uniforme dans une grande région de cet espace. A l'opposé, les bons appariements correspondent eux à un même mouvement apparent. Les points correspondants sont donc regroupés dans une petite région de l'espace. Ce "point d'accumulation" peut être trouvé facilement : on projette tous les points obtenus sur chacun des axes de coordonnées. Un simple calcul de convolution permet de trouver l'intervalle de chaque axe où se trouvent le plus grand nombre de ces projections. Un appariement qui donne un point dont les coordonnées sont dans ces intervalles sera considéré comme correct et conservé. Les autres appariements sont éliminés. Détaillons cela.

**Calculs de paramètres de similitude.** Soient deux paires de segments, définies respectivement par les points  $(A_1, B_1, C_1)$  et  $(A_2, B_2, C_2)$ . On suppose que ces points sont dénommés de telle sorte que les angles utilisés comme invariants soient :  $A_1\vec{B}_1, A_1\vec{C}_1$  et  $A_2\vec{B}_2, A_2\vec{C}_2$ . Soient  $k, \theta, t_x$  et  $t_y$  les paramètres de la similitude qui constitue le mouvement apparent entre les deux paires de segments. Ces paramètres peuvent être calculés de la manière suivante :

$$\begin{aligned} \begin{pmatrix} t_x \\ t_y \end{pmatrix} &= A_1\vec{A}_2 \\ \theta &= \frac{1}{2} \left( (A_1\vec{B}_1, A_2\vec{B}_2) + (A_1\vec{C}_1, A_2\vec{C}_2) \right) \\ k &= \frac{1}{2} \left( \frac{A_2\vec{B}_2}{A_1\vec{B}_1} + \frac{A_2\vec{C}_2}{A_1\vec{C}_1} \right) \end{aligned}$$

L'utilisation des seuls points  $A_1$  et  $A_2$  pour calculer la translation est motivée par le fait que les points extrémités de segments ne sont précis que lorsqu'ils sont l'intersection de deux segments.

**Recherche du point d'accumulation.** A chaque appariement de paires de segments, on associe donc un quadruplet  $(k, \theta, t_x, t_y)$ . Ces quadruplets sont des points d'un espace à quatre dimensions, dit espace des paramètres de similitude. On recherche alors dans cet espace, une accumulation de points.

Pour cela, on projette tous les quadruplets sur chacun des axes de l'espace des paramètres, selon la direction définie par les trois autres axes. Sur chaque axe, on obtient un ensemble de valeurs  $v_i$ . On définit alors la fonction :

$$\phi(v, s) = \text{Card}\{v_i \mid v - s < v_i < v + s\}$$

et on calcule :

$$v_m = \sup_{v_i} \phi(v_i, s)$$

où  $s$  est une constante qui dépend de l'axe considéré. Pour les trois premiers axes, on prend  $s = 20$ , et pour le dernier  $s = 1$ . Ce paramètre  $v_m$  est le paramètre recherché, qui est celui de la similitude qui approche le mieux le mouvement apparent global entre les deux images.

**Filtrage des appariements.** De tous les appariements de paires de segments réalisés, on ne garde alors que ceux qui définissent des paramètres de similitude qui, sur chaque axe, ne diffère pas de la valeur  $v_m$  de plus de  $s$ .

### Dernière étape

Les appariements de segments sont déduits aisément des appariements entre paires. Quelques ambiguïtés demeurent, mais elles peuvent être facilement levées : un segment apparié dix fois à un segment  $A$  et une fois à un segment  $B$  doit correspondre au segment  $A$ .

### Epilogue

Cette méthode permet un appariement sans connaissance préalable des images, en autorisant un mouvement apparent des primitives de grande amplitude. En cela, elle est plus puissante que les méthodes de tracking ou de corrélation. De plus, elle utilise uniquement des invariants associés à des structures de très petite taille et est donc robuste au bruit et aux occlusions partielles. Sa première limite importante est liée au choix du type de transformation utilisée comme approximation du mouvement apparent. Avec des similitudes, on ne peut appairer que des images assez semblables. La deuxième limite est due aux primitives utilisées : des segments reliés par des sommets. La méthode n'est donc pas utilisable avec des images texturées. Le paragraphe suivant justifie l'utilisation des similitudes et envisage d'autres choix.

#### 2.3.3 A propos du choix des similitudes

Même en supposant que les caméras utilisées suivent le modèle de la transformation projective, le mouvement apparent exact des primitives entre deux images n'est pas une transformation classique : ni euclidienne, ni affine, ni projective. Ce mouvement peut, par contre, être souvent approché de manière précise par une transformation d'un des types cités ci-dessus. Thompson et al. [Tho 87] montrent, par exemple, que les distorsions projectives sont négligeables si la profondeur de l'objet est inférieure au dixième de sa distance à la caméra.

Le mouvement apparent est ainsi une similitude à deux conditions. Premièrement, l'objet observé est plat et orthogonal à l'axe principal de la caméra. Deuxièmement, l'axe principal de la caméra est globalement le même lors des deux prises de vue ; il peut par contre subir une translation selon sa direction.

Dans la pratique, le respect de ces hypothèses, en particulier de la première, n'a nul besoin d'être très strict (cf. 2.3.4). Cela est dû, entre autres, au fait que les invariants ne sont calculés qu'à partir de segments connectés. Les hypothèses peuvent en effet

être respectées localement, même si elles ne le sont pas globalement. Le calcul des invariants à partir de toutes les paires de segments aurait augmenté drastiquement la combinatoire de l'algorithme. Les expériences montrent que la deuxième hypothèse peut être comprise ainsi : l'axe principal ne doit pas subir une rotation de plus de quinze degrés.

### Considérations mathématiques.

On peut justifier plus mathématiquement l'emploi des similitudes. Cette justification n'apporte cependant aucune conclusion nouvelle par rapport à ce qui est énoncé ci-dessus de manière plus "intuitive", et peut donc être sautée pour une lecture rapide.

**Définition des paramètres.** On considère un modèle projectif de caméra, défini par la donnée d'un point principal, centre de la projection perspective, d'un axe principal et d'un plan de projection. On définit les trois repères orthonormés suivants (cf. FIG. 2.41) :

1. Le repère normé de la scène  $(O, X, Y, Z)$  lié à l'objet qui est observé ;
2. le repère de l'image  $(o, u, v)$  dont l'orientation est fournie par l'axe principal ; on suppose ce repère orthogonal, ce qui n'est pas toujours le cas et qui ramène le nombre de paramètres de la caméra de 11 à 10 ;
3. le repère normé de la caméra  $(C, x, y, z)$  où  $C$  est le point principal et  $Cz$  l'axe principal de la caméra ; on suppose ici que l'axe  $Cx$  est parallèle à l'axe  $ou$ .

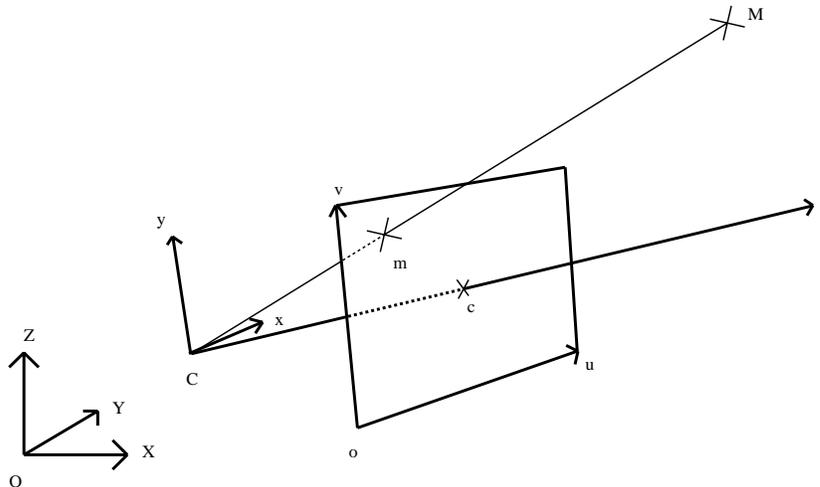


FIG. 2.41 - Définition des repères liés à une projection perspective.

Il faut bien remarquer que  $o$  est différent de  $c$ , projection orthogonale de  $C$  sur le plan de formation des images. On a donc besoin de dix paramètres pour définir la projection d'un objet sur le plan des capteurs :

- six pour définir le repère de la scène dans le repère de la caméra (trois composantes  $(a, b, c)$  d'une translation et trois angles de rotation par rapport aux trois axes

- $(\alpha, \beta, \gamma)$ ); ce dernier paramétrage d'une rotation n'est pas canonique, mais il est suffisant pour notre démonstration ;
- deux paramètres donnent les rapports de distance entre les axes  $Cx$  et  $ou$ , et  $Cy$  et  $ov$ , ces paramètres pouvant être pris égaux en se ramenant au cas où les pixels du plan des capteurs sont carrés. On note alors  $k$  cette valeur. Ces deux paramètres rendent compte à la fois de la distance focale et du nombre de pixels par millimètre dans les deux directions  $ou$  et  $ov$  ;
  - les deux derniers paramètres  $(d, e)$  définissent la position de  $o$  dans le plan par rapport au point  $c$ .

Un onzième paramètre apparaît lorsque le repère du plan de l'image n'est pas orthogonal. Les six premiers paramètres, ou plus exactement leurs homologues qui définissent le repère de la caméra par rapport au repère de la scène, sont appelés paramètres extrinsèques ou extérieurs de la caméra. Les quatre derniers paramètres sont appelés paramètres intrinsèques ou intérieurs de la caméra.

**Equations de la transformation** On se place dans des systèmes de coordonnées homogènes, et on suppose les pixels carrés (hypothèse à laquelle on peut toujours se ramener). Soit un point  $M$  de coordonnées homogènes  $(x_M, y_M, z_M, t_M)$  dans le repère  $(O, X, Y, Z)$ . Son image a pour coordonnées homogènes  $(u_m, v_m, w_m)$  dans  $(o, u, v)$  avec :

$$\begin{pmatrix} u_m \\ v_m \\ w_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix} \begin{pmatrix} x_M \\ y_M \\ z_M \\ t_M \end{pmatrix}$$

et :

$$\begin{aligned} a_{11} &= k \cos \gamma \cos \beta + d \sin \beta \\ a_{21} &= k \sin \gamma \cos \beta + e \sin \beta \\ a_{31} &= \sin \beta \\ a_{12} &= -k \cos \gamma \sin \beta \sin \alpha - k \sin \gamma \cos \alpha + d \cos \beta \sin \alpha \\ a_{22} &= -k \sin \gamma \sin \beta \sin \alpha - k \cos \gamma \cos \alpha + e \cos \beta \sin \alpha \\ a_{32} &= \sin \alpha \cos \beta \\ a_{13} &= -k \cos \gamma \sin \beta \cos \alpha - k \sin \gamma \sin \alpha + d \cos \beta \cos \alpha \\ a_{23} &= -k \sin \gamma \sin \beta \cos \alpha - k \cos \gamma \sin \alpha + e \cos \beta \cos \alpha \\ a_{33} &= \cos \alpha \cos \beta \\ a_{14} &= ka + dc \\ a_{24} &= kb + ec \\ a_{34} &= c \end{aligned}$$

**Hypothèses restrictives** On simplifie ces équations en se donnant des hypothèses restrictives. Celles-ci ne sont que la traduction mathématique de celles qui ont été

exposées auparavant.

1. On suppose tout d'abord que les distorsions projectives sont faibles, ce qui peut se traduire par :

$$t \ll z$$

2. On suppose ensuite que le déplacement correspondant aux paramètres extrinsèques de la caméra est faible dans les directions transversales :

$$\alpha \approx 0$$

$$\beta \approx 0$$

$$a \ll c$$

$$b \ll c$$

A l'ordre 0, l'équation de projection devient donc :

$$\begin{pmatrix} u_m \\ v_m \\ w_m \end{pmatrix} = \begin{pmatrix} k \cos \gamma & -k \sin \gamma & d & 0 \\ k \sin \gamma & k \cos \gamma & e & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_M \\ y_M \\ z_M \\ t_M \end{pmatrix}$$

Cette expression peut être décomposée comme suit :

$$\begin{pmatrix} u_m \\ v_m \\ w_m \end{pmatrix} = \begin{pmatrix} k \cos \gamma & -k \sin \gamma & d \\ k \sin \gamma & k \cos \gamma & e \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_M \\ y_M \\ z_M \\ t_M \end{pmatrix}$$

On reconnaît la **composée de la projection perspective avec une similitude affine directe du plan**, qui est elle-même la composée d'une homothétie, d'une rotation et d'une translation. On peut donc résumer cette formule par :

$$m = S \circ P(M)$$

où  $S$  est la matrice de similitude et  $P$  la matrice de projection. Lorsqu'un point est projeté deux fois, on obtient ;

$$m_1 = S_1 \circ P(M)$$

$$m_2 = S_2 \circ P(M)$$

On a, par conséquent :

$$m_2 = S_2 \circ S_1^{-1}(m_1)$$

L'inversion et la composition étant des lois de composition interne à l'ensemble des similitudes, les projections d'un même point dans deux images différentes se déduisent l'une de l'autre par une similitude indépendante de ce point. C'est cette similitude que l'on estime dans l'algorithme d'appariement. Dans le cadre de nos hypothèses, elle est une approximation à l'ordre 0 du mouvement apparent.

### Autres transformations utilisables.

La distorsion entre les projections d'un même objet 3D sur plusieurs images prises sous des angles de vue différents ne peut être modélisée exactement par une application entre ces deux images. L'utilisation des similitudes est une approximation. Les transformations affines ou projectives possèdent plus de paramètres, de degrés de liberté (6 dans le cas affine, 8 dans le cas projectif), et devraient permettre une meilleure approximation du mouvement apparent.

Cependant, ces transformations plus complexes ont des invariants plus compliqués : rapports et coordonnées barycentriques dans le premier cas, birapports et coordonnées projectives dans le second. De tels invariants peuvent être calculés à partir de configurations de 4 ou 5 points. Les trois problèmes posés par ces configurations sont le choix de ces points, la stabilité des invariants calculés [Mor 93], et celle de la transformation. Même en restreignant les configurations de points utilisées aux sommets de certains sous-graphes particuliers, la combinatoire est bien plus grande que pour les similitudes.

#### 2.3.4 Expérimentations

Ce paragraphe contient deux exemples d'utilisation de l'algorithme. Ils en valident le principe et en montrent parallèlement les limites.

##### Premier exemple

Les FIG. 2.42, 2.43 et 2.44 montrent trois exemples. Les images du haut montrent les images initiales, les primitives extraites de ces images sont montrées au milieu, et les primitives appariées sont montrées en bas.

Il s'agit de quatre images d'une même séquence, et on a successivement apparié la première avec les trois autres. Les deux images ayant servi au premier appariement sont très proches. La seule différence notable est due au bruit. La plupart des segments qui ne sont pas dus au bruit ont été appariés, en dehors de ceux qui sont divisés en plusieurs petits segments dans une image et non dans l'autre. Le bruit est, par contre, éliminé presque totalement.

Les images ayant servi au deuxième appariement diffèrent plus nettement, mais le résultat reste de qualité comparable. Là encore, le fait que divers segments ne sont pas appariés s'explique par le bruit qui "casse" ces segments en plusieurs parties.

Lors du troisième appariement, la différence entre les angles de vue des deux images devient plus importante. En plus du bruit, la cause majeure de non-appariement est le fait que les invariants utilisés ne sont plus invariants dans ce cas. Les similitudes ne sont plus une approximation valide du mouvement. Dans ce cas, on est sorti du cadre des hypothèses de l'algorithme.

L'utilisation d'un autre type de transformation doit permettre d'améliorer ces résultats. Par ailleurs, on peut remarquer que les quelques segments appariés le sont convenablement. Cela montre bien que les mauvais appariements ne forment aucun point d'accumulation dans l'espace des paramètres des transformations, et que ce sont bien les appariements cohérents qui sont trouvés même s'il sont peu nombreux.

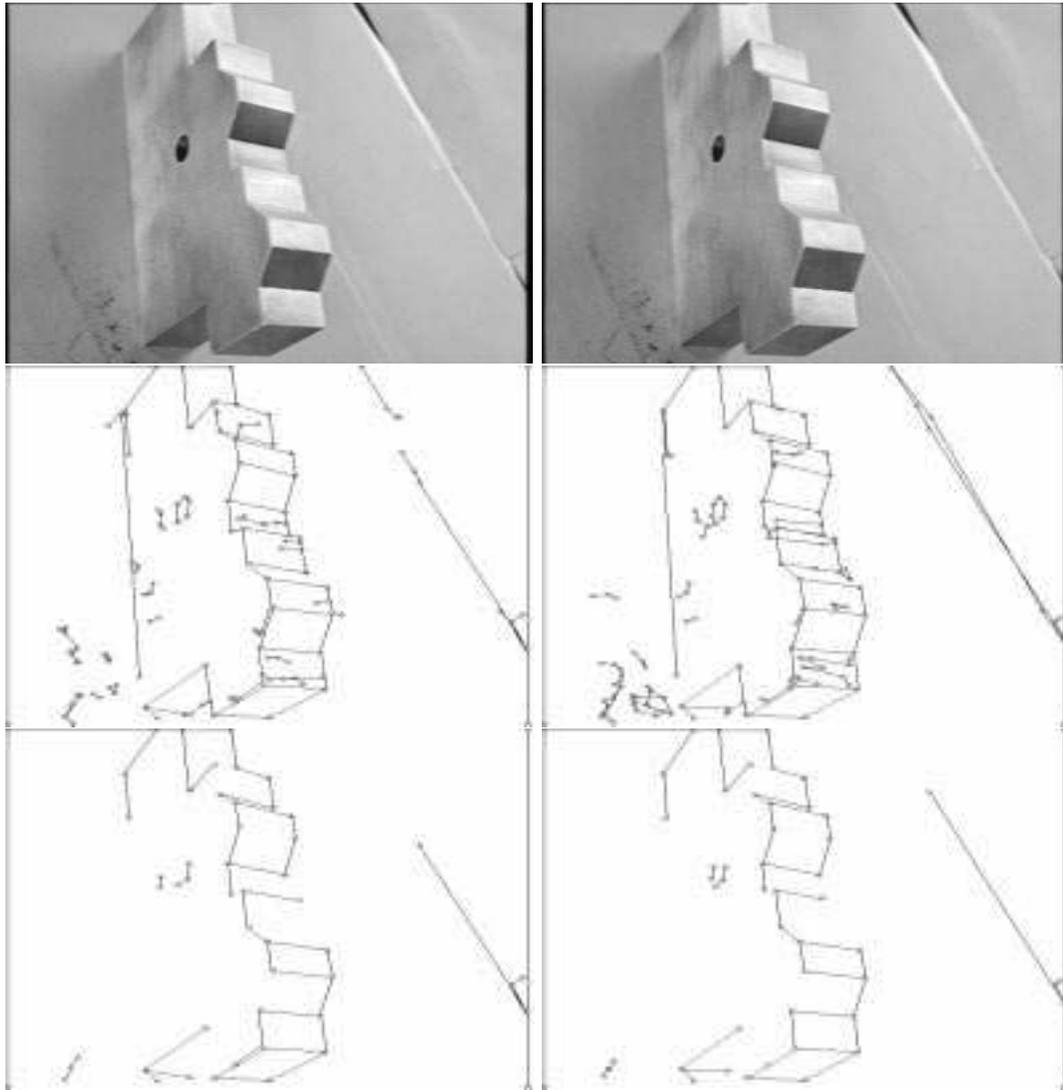


FIG. 2.42 - Premier exemple d'appariement.

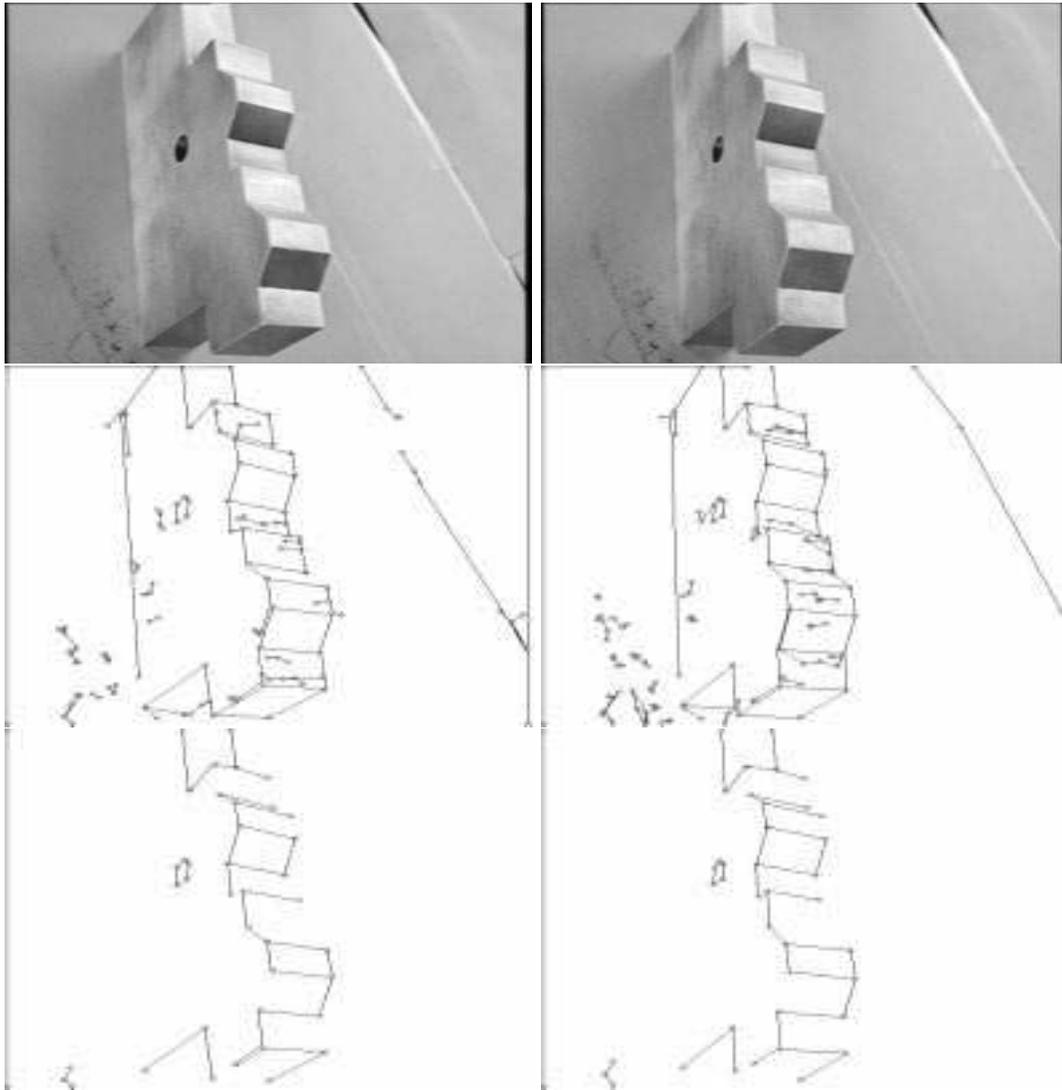


FIG. 2.43 - Deuxième exemple d'appariement.

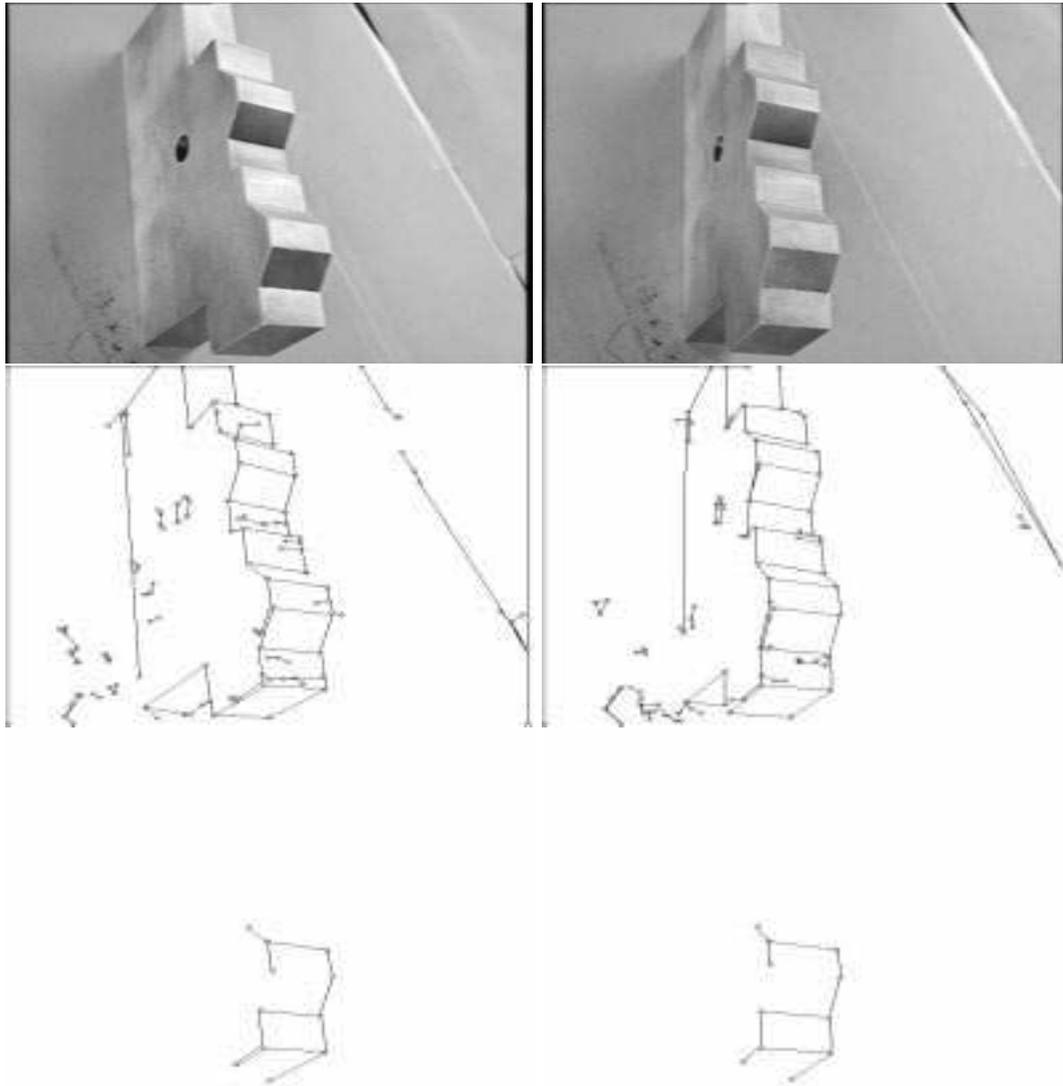


FIG. 2.44 - Troisième exemple d'appariement.

## Deuxième exemple

La FIG. 2.45 montre un deuxième exemple d'utilisation de l'algorithme sur deux images plus complexes que précédemment. Ces images contiennent, outre l'objet étudié qui est la maison, divers petits objets, certains n'étant pas polyédriques.

L'appariement des segments des projections de ces petits objets n'est pas très satisfaisant. Cela est dû à l'algorithme d'approximation polygonale, dont les résultats ne sont stables que si les objets sont polyédriques avec des arêtes de longueur suffisante. En ce qui concerne la maison, le non appariement des segments est dû au bruit, à l'instabilité de l'approximation polygonale, et au fait que l'algorithme d'appariement n'est lui-même pas très stable avec les petits segments.

Il faut aussi remarquer, au niveau des fenêtres de la maison, qu'un segment d'une image peut être apparié avec deux segments de l'autre image. Ces cas correspondent au fait que ces deux segments ont les mêmes sommets extrémités.

### 2.3.5 Conclusion

La méthode d'appariement d'images proposée dans ce paragraphe est validée par les résultats obtenus sur des images réelles. Ces résultats mettent parfaitement en valeur les atouts et défauts de l'algorithme.

- Le mouvement de la caméra entre les deux prises de vue peut être quelconque dans certaines directions (rotation autour de l'axe principal par exemple), mais doit être très limité transversalement.
- La précision des primitives extraites des images est fondamentale pour la qualité des résultats. Les textures, les objets non polyédriques et les arêtes de petite dimension sont cause de nombreux non-appariements de segments.
- Une fois les paramètres de recherche du point d'accumulation ajustés, la méthode ne nécessite aucun seuil dépendant des images pour fonctionner.
- L'algorithme ne suppose aucune connaissance préalable sur le contenu des images : pas de premier appariement, pas d'étalonnage, même partiel de la caméra, pas de connaissance de la géométrie épipolaire, pas d'estimation du mouvement de la caméra.

Malgré quelques imperfections dans les résultats obtenus par cet algorithme, ceux-ci sont tout à fait satisfaisants si l'on considère cet algorithme comme un premier pas d'une méthode plus globale. Avec les premiers appariements trouvés, on peut calculer la géométrie épipolaire ou se servir de contraintes de connexité pour poursuivre l'appariement. Il ouvre par ailleurs la voie à des traitements de plus haut niveau, tels la modélisation qui fait l'objet du chapitre suivant.

## 2.4 Conclusion du chapitre

Le premier chapitre de cette thèse donne les indications théoriques relatives aux invariants géométriques et à la justification de leur emploi en vision. Ce deuxième



FIG. 2.45 - Exemple d'appariement d'une scène complexe.

chapitre montre des exemples de calcul effectif et d'emploi de ces invariants, pour résoudre des problèmes réels de vision. Les trois applications traitées sont

1. la recherche des quadruplets de points coplanaires d'une scène,
2. le calcul d'invariant 3D caractéristiques d'un objet,
3. l'appariement d'images.

Des conclusions ont été tirées pour chacun de ces points. Rappelons toutefois un point commun. Les données des images sont bruitées, et les primitives à partir desquelles sont calculées les invariants ne sont localisées qu'approximativement. De plus certaines de ces primitives ne correspondent pas à des éléments de la scène observée, d'autres éléments ne donnent lieu à aucune détection de primitives. Il est donc nécessaire d'utiliser des algorithmes précis et robustes à toutes les étapes.

Malgré ce constat, de bons résultats ont pu être obtenus sur des données réelles, et ce, pour les trois applications. Il est donc possible de calculer une information numérique à partir d'images.



## Chapitre 3

# Modélisation automatique

### 3.1 Introduction

Ce chapitre concerne le problème de la modélisation automatique des objets polyédriques. Cette modélisation est la première étape d'un système de reconnaissance qui comporte les parties suivantes : un système de modélisation qui construit un modèle d'un objet à partir d'images de cet objet, un système d'indexation de ces modèles dans une base de données, et une procédure de reconnaissance des objets utilisant cette base de données [Sos 92].

C'est la partie modélisation qui est l'objet de ce chapitre. Le problème est le suivant. On veut modéliser les différents **aspects visuels** d'un objet, pour pouvoir le reconnaître ultérieurement, quelque soit l'angle de vue sous lequel il sera perçu. On entend le mot aspect dans son acception courante : un objet change d'aspect lorsqu'on le regarde de différents points de vue.

On commence par prendre des images de cet objet sous de nombreux angles de vue. On essaye alors de retrouver celles de ces images qui appartiennent à la même vue. Dans la suite, on appellera **vue** l'ensemble des images représentant un même aspect visuel de l'objet. Chacune de ces vues donne lieu à l'établissement d'un modèle de l'aspect en question. Le modèle total de l'objet est l'ensemble des modèles de tous ces aspects.

Le but d'un tel système est de réduire la taille de la représentation d'un objet. La première représentation dont on dispose est l'ensemble des images, mais ces dernières contiennent une quantité d'information très importante, redondante, peu maniable et difficile à stocker. L'extraction des modèles des aspects de l'objet permet des gains de place, de vitesse et de maniabilité très importants. D'une centaine d'images, on

construit classiquement une dizaine de modèles d'aspects.

Les applications d'un tel système de modélisation et reconnaissance sont nombreuses dans le cadre de la robotique : reconnaissance pour l'identification ou la saisie d'un objet, positionnement d'un robot mobile par rapport à son environnement, reconnaissance des obstacles à éviter lors d'un mouvement ou d'un déplacement. . . L'apport spécifique de la modélisation est la possibilité d'un lien entre un vocabulaire de bas niveau et un vocabulaire plus symbolique, dit de haut niveau. Au lieu de décrire l'environnement d'un robot en termes de surfaces ou de volumes, on peut le faire en nommant les objets : il suffit pour cela d'associer un nom à chaque modèle ou groupe de modèles.

### 3.1.1 Etat de l'art

Les approches actuelles pour résoudre le problème de la modélisation et de la reconnaissance peuvent être classées selon deux critères : quel est le type des données utilisées pour la construction du modèle ? De quel type est le modèle construit ? Les données peuvent, en effet, être 2D ou 3D, être obtenues par un ou plusieurs capteurs, auquel cas on parlera de "données réelles", ou être construites par un opérateur humain, auquel cas on parlera de "données artificielles". Quant à lui, le modèle peut être 2D ou 3D. Une telle présentation des systèmes de modélisation peut être trouvée dans l'article de Flynn et al. [Fly 91]. Elle est utilisée ici pour permettre de situer notre propre système au sein de l'existant.

#### Données 3D artificielles.

De telles données proviennent, la plupart du temps, de systèmes CAO. Dans ces systèmes, l'objet est décrit en fonction de ses caractéristiques géométriques (volumiques et métriques), mécaniques et parfois électriques. Le principal problème consiste alors à déduire l'aspect visuel de l'objet de ces propriétés. Cette approche utilisant la CAO a été très étudiée en vision, donnant même naissance à un domaine particulier de la vision, appelé en anglais "CAD-based vision" [Bha 87].

Plusieurs stratégies sont possibles pour la reconnaissance. Tout d'abord, la programmation automatique. Goad [Goa 86] écrit directement un programme de reconnaissance d'un objet à partir d'une description 3D de celui-ci. La reconnaissance est basée sur l'appariement des arêtes et des segments, et sur le calcul des aspects 2D de l'objet à partir de son modèle 3D. Le programme est écrit dans un premier temps, seule son exécution est effectuée pendant la phase de reconnaissance proprement dite.

Une deuxième stratégie consiste à construire un modèle contenant toutes les informations visuelles nécessaires à la reconnaissance. Shneier et al. [Shn 87] ont développé un tel système qui utilise des informations à la fois volumiques et surfaciques pour construire des modèles. Bolles et al. [Bol 86] calculent les différents aspects 3D (les aspects ont une représentation 3D) de l'objet, perçu sous différents points de vue (contrairement à Goad qui calculait des aspects 2D). D'autres systèmes sont plus spécifiques [Kak 88, Fly 91] : les modèles construits sont adaptés à une reconnaissance des objets à partir d'images 3D obtenues par des capteurs de profondeur.

**Limites de ces méthodes.** La principal problème avec ces méthodes est l'inférence des aspects visuels d'un objet à partir d'une description purement volumétrique de

celui-ci. Même en utilisant une description physique des matériaux constituant la surface de l'objet et des modèles de réflexion de la lumière, il n'est pas possible d'obtenir une description convenable des aspects visuels [Che 91]. La difficulté vient des modèles de réflexion de la lumière (modèle lambertien, ou ceux, plus sophistiqués, utilisés par Chen), qui ne sont pas valables pour la plupart des matériaux utilisés, les métaux en particulier.

### **Données 2D artificielles.**

Ce cas n'est qu'une variante du précédent. On commence d'abord par calculer tous les aspects 2D de l'objet, et ce sont ces aspects qui sont utilisés pour la modélisation. La limite entre cette technique et la précédente n'est pas très nette (cf. le système de Goad). Les aspects sont définis topologiquement comme ils peuvent l'être en morphogénèse [Dem 87]. Deux images représentent le même aspect si l'on peut passer de l'une à l'autre sans catastrophe (discontinuité topologique). Ces aspects sont alors ordonnés en un graphe, appelé graphe d'aspects, selon leurs points de vue associés. Des objets même simples peuvent avoir plusieurs centaines d'aspects. De manière identique au cas des données 3D, deux stratégies sont possibles : programmation automatique [Ike 88], ou construction de modèles [Bur 88]. Ces méthodes de calcul des graphes d'aspects sont maintenant aussi utilisées pour les objets non polyédriques [Pon 90].

**Limites de ces méthodes.** Les méthodes utilisant les graphes d'aspect partent d'une description de l'objet de type CAO. Elles souffrent donc des mêmes limites que les méthodes précédentes : les aspects topologiques calculés ne rendent pas compte des aspects visuels.

Un deuxième problème s'ajoute à cela. Le nombre des aspects topologiques est, en général, nettement supérieur au nombre des aspects visuels. En effet, les aspects topologiques peuvent différer à cause de détails de taille très petite, non visibles par une caméra ordinaire observant la totalité de l'objet. Outre ce problème d'échelle, la complexité en temps du calcul des aspects, et celle en taille du graphe obtenu, sont des obstacles vite rédhibitoires dès que l'objet n'est pas trop simple. Bowyer [Bow 91] fait une critique complète de ces méthodes.

### **Données 3D réelles.**

Elles proviennent principalement de deux types de capteurs : les appareils d'investigation biomédicaux (scanners, IRM) et les capteurs de profondeur, tels les télémètres laser. Dans le premier cas, les données sont des images 3D denses, alors que ce ne sont que des cartes de profondeur dans le second. Les techniques utilisées sont souvent spécifiques à ces capteurs. On en trouvera une synthèse dans les articles de Besl [Bes 88] et Nitzan [Nit 88].

**Limites de ces techniques.** Ces techniques sont bien adaptées aux applications où l'aspect 3D de l'information est primordial, par exemple pour les applications médicales. Leur principale limite vient de la spécificité des capteurs.

### Données 2D réelles.

Ces données sont, dans la majeure partie des cas, des images obtenues sous différents angles de vue par des caméras. Malgré quelques travaux notables utilisant des images couleur [Lee 92], la plupart des travaux utilisent des images en noir et blanc. Les systèmes de modélisation à partir d'images sont fort nombreux. Ils diffèrent par le type d'information qu'ils extraient des images et par la dimension du modèle construit. Connell et al. [Con 87] utilisent l'intensité lumineuse en chaque point ; Arbogast et al. [Arb 91c] et Cipolla et al. [Cip 92] utilisent la projection des contours d'occultation de surfaces courbes ; Faugeras [Fau 92a] et Mohr et al. [Moh 92c] utilisent des points ; Chabbi et al. [Cha 91] et Paris [Par 91] utilisent des régions pour calculer les facettes 3D de la scène observée ; Forsyth et al. [For 91a] utilisent des invariants associés à certaines configurations de points, droites et courbes ; Weiss [Wei 88, Wei 91] utilise les invariants différentiels associés aux courbes algébriques.

**Limites de ces méthodes.** Les limites sont aussi diverses que les méthodes. On peut les classer en trois catégories :

1. les primitives employées ne sont pas stables. Cela est particulièrement vrai pour les intensités lumineuses, ou les dérivées partielles de courbes d'ordre supérieur à deux qui servent à calculer les invariants différentiels ;
2. certaines, parmi ces méthodes, nécessitent des connaissances préalables, l'étalonnage des caméras par exemple : c'est le cas des méthodes utilisant les surfaces d'occultation de surfaces courbes ;
3. Quelques méthodes considèrent que l'appariement des primitives a été réalisé au préalable. C'est le cas des méthodes de Faugeras et de Mohr qui utilisent des points appariés. Cette hypothèse est très forte et suppose, en fait, qu'une grande partie du travail de modélisation a été fait.

### 3.1.2 Méthode proposée

Notre approche entre dans cette dernière catégorie : modélisation à partir de données 2D réelles. Ses sources d'inspiration sont multiples.

- Modéliser pour reconnaître de nouvelles images revient à faire un apprentissage, c'est à dire une induction à partir d'exemples. On dispose d'exemples à partir desquels on tire un modèle dont on postule la pertinence pour d'autres images qui ne sont pas dans les exemples : en cela, il s'agit d'une induction.
- Cette induction ne porte pas sur des concepts ou des grandeurs bien établis, comme c'est le cas, par exemple, pour un système d'apprentissage utilisant uniquement des invariants. Il s'agit, dans la méthode proposée, de modéliser des aspects visuels. En cela, cette méthode se rapproche des techniques neuronales : on fournit au système un ensemble de primitives, et c'est au système de trouver ce qui est caractéristique et permet la reconnaissance.
- Enfin, on réalise l'apprentissage directement à partir de ce qui est perçu, c'est à dire à partir de données de même type que celles qui seront à reconnaître. Il

s'agit d'un apprentissage à partir d'exemples. Ceci est un principe classique en classification automatique qui est nécessaire pour la robustesse du système au cas où le bruit qui affecte les images n'est pas modélisable.

Les différentes étapes de la méthode sont les suivantes (elles sont résumées sur la Fig. 3.1).

1. On prend un ensemble d'images d'un objet depuis de nombreux points de vue. On extrait alors les contours de chacune de ces images, puis on réalise une approximation polygonale de ces contours. Les segments obtenus et les sommets qui les joignent sont les primitives de base.
2. Les images sont appariées deux par deux, en utilisant la méthode exposée au paragraphe 2.3.
3. Deux images représentent le même aspect de l'objet si elles sont presque similaires, c'est à dire si elles contiennent beaucoup de primitives appariées. On définit une mesure de ressemblance entre images basées sur le pourcentage de primitives appariées entre les deux images considérées. On calcule alors la matrice de ressemblance de l'ensemble des images.
4. Une méthode classique de regroupement ("clustering" en anglais), utilisant la matrice de ressemblance, permet de déterminer les différentes vues, c'est à dire les ensembles d'images représentant un même aspect.
5. Chaque vue donne lieu à l'établissement d'un modèle qui contient l'ensemble des primitives communes aux images de la vue. Ce modèle peut être complété avec des primitives n'apparaissant pas dans toutes les images, des informations probabilistes (probabilités d'apparition de primitives, probabilités conditionnelles d'appariement) ou géométriques (invariants 2D ou 3D associés à certaines configurations de primitives).

Ajoutons à cela deux étapes qui ne sont pas détaillées dans ce chapitre, mais qui font partie du système complet de modélisation et reconnaissance.

6. L'indexation des modèles construits dans une base de données. Cette indexation est réalisée actuellement sur la base de propriétés topologiques par hachage de graphes [Sos 92], mais elle devrait pouvoir être complétée en utilisant les propriétés géométriques et probabilistes présentes dans le modèle.
7. L'algorithme de reconnaissance proprement dit découle en fait directement de la fonction d'indexation utilisée à l'étape précédente.

Cette méthode se distingue donc de celles proposées dans la littérature sur plusieurs points.

- Tout d'abord, elle ne nécessite ni connaissances préalables ni étalonnage, même partiel, des caméras.
- Elle ne nécessite aucune reconstruction de la forme de l'objet en 3D.

Etape	Traitement	Résultat du traitement
1	acquisition	objet ↓
2	traitement des images	images ↓
3	appariement des images 2 à 2	primitives ↓
4	mesure de ressemblance	relations entre primitives ↓
5	regroupement	matrice de ressemblance ↓
6	modélisation	vues ↓
7	indexation	modèles des aspects ↓
8	reconnaissance	base de modèles ↓ identification d'objets

FIG. 3.1 - Le processus de modélisation et reconnaissance.

- Les images de départ peuvent ne pas former une séquence ordonnée d'images où la caméra ne ferait qu'un petit mouvement entre deux prises de vue.
- Enfin, elle utilise tant l'information topologique et visuelle (notion d'aspect visuel) que l'information géométrique (pour le matching). Cela permet de modéliser des objets complexes ce qui n'est, pour l'instant, pas le cas des techniques n'utilisant que les invariants.

Au total, notre méthode est donc très peu contrainte. Sa seule limite est liée au type de primitives utilisées : des points et des segments. La localisation de ces primitives est sensible au bruit. D'autre part, il est très difficile de modéliser des objets courbes : l'approximation des contours par des segments est souvent instable dans ce cas.

### 3.1.3 Plan du chapitre

Ce chapitre est organisé selon les étapes de la méthode proposée. Un premier paragraphe concerne l'appariement des images. L'algorithme d'appariement proprement dit ayant fait l'objet du paragraphe 2.3, il n'est point repris ici. Ce paragraphe donne plutôt quelques indications sur les possibilités d'amélioration et de poursuite du premier appariement effectué. Le deuxième paragraphe expose ensuite la mesure de ressemblance entre images et son utilisation pour regrouper les images représentant le même aspect. Le dernier paragraphe aborde succinctement les étapes suivantes : modélisation, indexation, reconnaissance.

## 3.2 Appariement et correction d'images

La première phase de l'algorithme de modélisation que nous proposons est l'appariement des images deux à deux. On part d'un ensemble d'images représentant un objet, ensemble qui n'est pas ordonné en séquence, et on souhaite en appairer toutes les paires d'images. Pour cela, on procède en deux temps. On utilise d'abord l'algorithme d'appariement exposé au paragraphe 2.3, puis on complète ce premier appariement.

Les images qu'on utilise sont, en effet, dégradées par tous les traitements effectués. Aussi, le graphe des segments et sommets que nous utilisons n'est pas la simple projection des arêtes et coins des objets représentés dans l'image. Pour s'en convaincre, la FIG. 3.2 montre, en haut, 3 images d'un même objet : entre les prises de vue, ni l'objet ni la caméra n'ont bougé, seul l'éclairage a varié. En bas, se trouvent le résultat du traitement de ces trois images : le graphe formé par les segments et les sommets est très différent d'une image à l'autre.

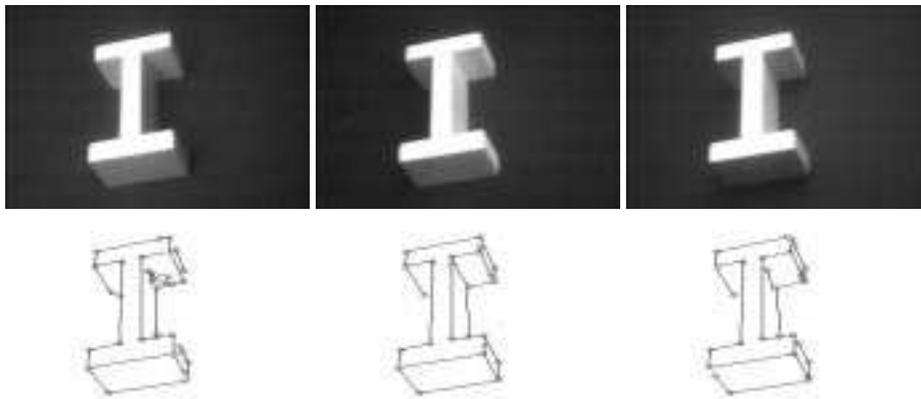


FIG. 3.2 - Trois images d'un même objet observé du même point sous trois éclairages différents et le résultat de leur traitement.

La conséquence de ces bruits qui affectent l'image est l'incapacité de l'algorithme du paragraphe 2.3 à appairer tous les segments de chacune des deux images qu'il traite. Divers techniques de correction de données sont proposées dans la littérature. Par exemple, Porrill [Por 88] développe un formalisme statistique pour prendre en compte les erreurs de précision dans les données qu'il obtient. Un tel travail diffère cependant de nos propres objectifs sur plusieurs points :

- il ne considère que des données 3D obtenues à partir de plusieurs paires d'images stéréographiques, et il n'y a donc pas de mouvement apparent entre les différentes données ;
- seules les erreurs de position sont prises en compte à l'exclusion de celles liées à de mauvais appariements ou à une mauvaise segmentation des contours dans les images.

Dans ce paragraphe, après une brève analyse de quelques sources de bruit, des techniques d'amélioration de l'appariement déjà réalisés sont proposées.

### 3.2.1 Analyse de quelques sources de bruit

Dans la suite, nous appellerons bruit tout ce qui fait que le graphe des segments et des sommets obtenu à partir d'une image diffère de la projection perspective des arêtes et sommets de l'objet représenté dans cette image (cf. FIG. 3.2). Les sources de bruit sont fort nombreuses : elle ne seront donc pas toutes citées ici. On trouvera, par exemple, dans [Bey 92] une étude complète du bruit qui affecte l'image lors de son acquisition par une caméra.

Les étapes d'acquisition et de traitement d'une image sont nombreuses et variées, et le bruit qui en résulte est aussi varié dans ses formes et ses conséquences. Le but de cette étude est de déterminer quelques conséquences typiques du bruit dans les images, pour ensuite corriger ces effets de manière systématique. Pour cela, on parcourt les phases successives d'acquisition et de traitement.

- La première source de bruit est l'objet lui-même. Lors de l'approximation polygonale des lignes de contraste ou lignes de contour, la texture des faces et le fait que les arêtes de l'objet ne sont pas parfaitement rectilignes sont la source de l'instabilité des segments extraits. Ces segments peuvent, en effet, varier fortement entre deux images représentant des aspects même très proches. En définitive, la difficulté est de savoir ce qui, dans l'objet, peut être considéré comme caractéristique stable, discriminante et utilisable et doit donc être soigneusement conservé, et ce qui est du bruit ou qui n'est pas utilisable et qui peut donc être éliminé.
- La scène est la deuxième source de bruit. On entend par scène tout ce qui est visible et qui n'est pas l'objet à étudier proprement dit. Autre élément de la scène, l'éclairage est la source des ombres et des reflets qui créent nombre de segments généralement considérés comme du bruit. Certaines études ont toutefois essayé de tirer de ces éléments des informations sur la structure tridimensionnelle des objets observés [Nay 90]. Dans l'algorithme exposé au chapitre précédent, c'est le fait de partir de plusieurs images et d'utiliser la redondance de l'information qui permet de différencier ce qui est dû aux ombres ou reflets de ce qui est caractéristique de l'objet.
- La caméra qui sert à prendre les images est, elle-même, source de nombreux bruits. Son optique n'est pas parfaite et crée des flous qui entraînent une imprécision sur la localisation des points et un effet d'arrondi des angles. Dans l'image finale, cet arrondi se traduit souvent par le fait qu'un angle est remplacé par deux sommets liés par un petit segment, comme montré par sur la FIG. 3.3. Les composants électroniques de la caméra sont aussi sources de nombreux bruits.
- Une fois l'image en niveau de gris obtenue, les contours sont extraits en utilisant un filtre de Deriche. Comme indiqué au paragraphe 1.3.4, les conditions d'optimalité de ce filtre ne sont pas respectées en pratique. Cela est particulièrement le cas pour les sommets où 3 segments devraient se joindre et où l'un d'eux peut ne pas être détecté jusqu'au sommet. L'approximation polygonale faite après cette détection des contours peut entraîner une perte importante de précision dans leur localisation. De nouvelles techniques, passant par une approximation préalable des contours par des B-splines sont maintenant utilisées [Arb 91a].

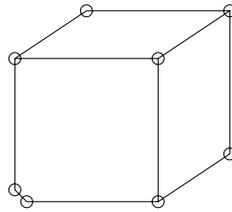


FIG. 3.3 - Exemple d'arrondi dû à un flou

### 3.2.2 Correction des images

Une fois réalisé un premier appariement entre deux images, on peut le compléter en corrigeant certains effets du bruit, et en utilisant une information différente de celle fournie par les invariants. En particulier, la topologie du graphe des segments de chacune des images et la géométrie épipolaire apportent des indications complémentaires à l'information géométrique déjà utilisée.

Les corrections proposées ici n'ont pas un caractère mathématique ou rigoureux. Ce sont plutôt des méthodes d'usage général, dont la justification se trouve dans l'étude du bruit qui affecte les images et dans les résultats pratiques obtenus lors de leur utilisation. Ce paragraphe n'est pas central dans le processus de modélisation, contrairement au suivant. Il peut être omis pour une lecture rapide.

#### Traitement des extrémités libres

On appelle extrémité libre, toute extrémité d'un segment non reliée à d'autres segments. Lors de la projection d'objets polyédriques, de telles extrémités libres ne devraient pas apparaître dans les images. Pourtant elles ne sont point rares. Deriche et al. [Der 87b] ont traité ce problème dans le cas où l'on désire obtenir des régions à partir des lignes de contours.

L'existence de ces extrémités libres est due principalement au filtre de Deriche qui n'est pas optimal aux endroits où trois lignes de contours se rejoignent. Ce mauvais fonctionnement du filtre en ces points peut être, en partie seulement, contrebalancé par le réglage des deux seuils du filtre.

Le cas le plus fréquent d'extrémités libres est celui des jonctions à 3 segments (cf. FIG. 3.4). La correction appliquée est simple : on cherche si, en prolongeant un peu le segment qui possède une extrémité libre, on ne rejoint pas un sommet entre deux autres segments. Si tel est le cas, le segment est effectivement prolongé jusqu'à ce sommet.

Un autre traitement possible consisterait à réutiliser l'information des niveaux de gris.

#### Traitement des dédoublements

Les dédoublements ont déjà été signalés comme le résultat de la finitude de la profondeur de champ de l'objectif de la caméra utilisée. On observe alors le remplacement d'un sommet par deux autres très proches reliés par un petit segment.

Une correction est alors possible lorsqu'un tel dédoublement apparaît dans une des images et non dans les autres. Il faut, de plus, pour ne pas trop détériorer la position et l'orientation impliqués dans ce dédoublement, qu'un des deux sommets ne relie que

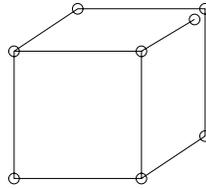


FIG. 3.4 - Un exemple d'extrémité libre.

deux segments. Deux exemples de corrections possibles et un de correction impossible sont indiqués sur la FIG. 3.5.

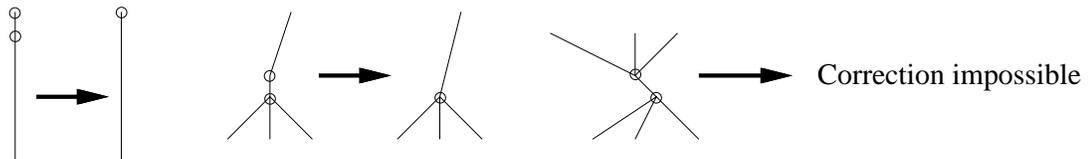


FIG. 3.5 - Deux exemples de corrections possibles des dédoublements (à gauche), et un cas où la correction est impossible (à droite).

### Traitement des jonctions en T

On appelle jonction en T toute jonction reliant trois segments dont deux sont dans le prolongement exact l'un de l'autre : l'angle entre ces segments doit être de l'ordre de l'erreur de discrétisation sur les réels ( $10^{-6}$ ). Elles ne correspondent pas, en général, à la projection de points physiques, mais au fait qu'une arête en occulte partiellement une autre. Le problème lié à ces jonctions est leur instabilité. Comme les deux arêtes qui réalisent cette jonction peuvent être très éloignées l'une de l'autre, leurs conditions d'éclairément peuvent varier fortement d'une image à l'autre, même pour des images proches prises dans une même séquence.

L'inconvénient majeur de ces jonctions est de diviser en deux parties un segment, rendant son appariement impossible. On peut remédier à ce problème en remplaçant les deux segments qui se prolongent par un seul segment, laissant ainsi le troisième segment avec une extrémité libre.

### Utilisation de la topologie et de la géométrie épipolaire

Utiliser la topologie consiste à comparer les graphes de segments qui constituent les deux images. Ce faisant, on peut chercher des correspondants à des segments non appariés en se servant du fait que l'une au moins de leurs extrémités est appariée.

La géométrie épipolaire fournit, elle, une contrainte sur les positions respectives de deux éléments appariés. Suite au premier appariement réalisé, si cette géométrie peut être calculée, on peut l'utiliser pour prédire de nouveaux appariements de points.

### Exemple expérimental

La FIG. 3.6 montre un exemple d'utilisation de toutes les techniques précédemment citées. Les deux images du haut sont deux images d'un même objet. L'image en bas à gauche représente l'ensemble des primitives appariées avant utilisation des corrections. L'image en bas à droite, représente ces primitives après emploi des correction.

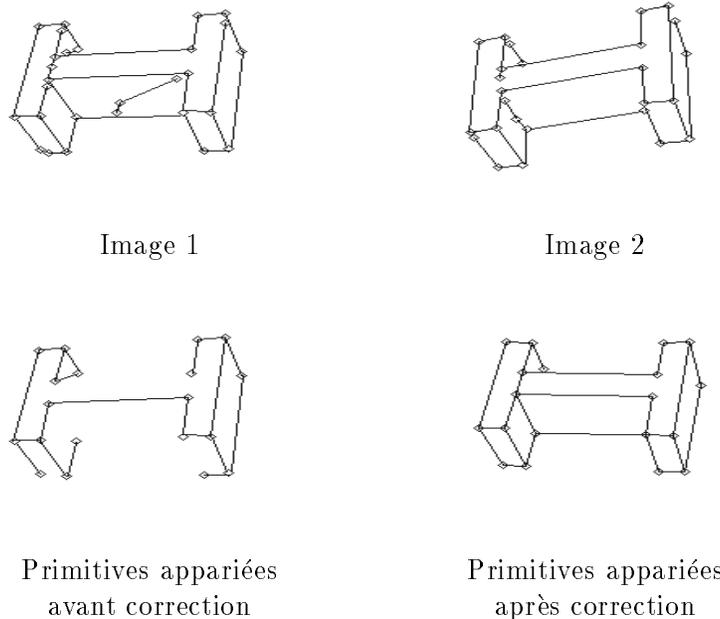


FIG. 3.6 - Un exemple d'emploi des techniques de correction.

## 3.3 Distance entre images et regroupement

Ce paragraphe détaille les étapes 4 (mesure de ressemblance) et 5 (regroupement) du processus de modélisation montrée sur la FIG. 3.1. On suppose donc que les étapes précédentes ont été effectuées.

### 3.3.1 Evaluation de la ressemblance des images

On dispose d'un ensemble d'images d'un objet, images qui sont appariées deux par deux. On souhaite alors les comparer, afin de déterminer automatiquement celles qui représentent un même aspect de l'objet. Pour cela, on cherche à déterminer une mesure de ressemblance entre images. Le terme mesure de ressemblance sera utilisé de préférence à celui de distance, car il n'est point assuré que cette mesure vérifie les axiomes mathématiques d'une distance.

Ce problème d'évaluation quantitative de la ressemblance entre images est généralement abordé de deux façons. Tout d'abord, on peut comparer directement les images en niveau de gris, et rechercher les correspondances entre pixels. C'est la base des techniques de corrélation. La deuxième voie est basée sur la comparaison entre graphes.

Lorsque les images sont décrites en termes de relations entre primitives, on peut mesurer la ressemblance des images en comparant les graphes formés [Sha 85b]. De plus, cette approche reste encore valable si l'on apparie un modèle et une image, par exemple lors de la phase de reconnaissance qui n'est pas traitée dans ce travail.

Pour notre part, nous proposons de déduire la mesure de ressemblance de l'appariement des images. Si deux images se ressemblent, une grande partie de leurs primitives peut être appariée. Au contraire, leur dissemblance entraîne un faible taux de primitives appariées. Faisant cela, on utilise la même information que lors de l'appariement, c'est à dire une information géométrique et topologique locale avec une contrainte géométrique globale.

### Diverses mesures

Le principe général des diverses mesures testées est d'établir un rapport entre la quantité de primitives appariées et celle des primitives présentes dans les images. Cela peut être résumé dans la formule suivante :

$$\text{distance}(\text{image1}, \text{image2}) = \frac{\text{contenu}(\text{image1}) \times \text{contenu}(\text{image2})}{\text{contenu\_appariement}(\text{image1}, \text{image2})^2}$$

Les mesures utilisées diffèrent par les éléments qu'elles prennent en compte dans les images ou les appariements.

Dans une première mesure, on n'a utilisé qu'un comptage des primitives. Etant données deux images  $I_1$  et  $I_2$ , leur mesure est définie par :

$$d_1(I_1, I_2) = \frac{a \text{nbseg}_1 \cdot \text{nbseg}_2}{d \text{nbseg}_{\text{appariée}}^2} + \frac{b \text{nbsmt}_1 \cdot \text{nbsmt}_2}{d \text{nbsmt}_{\text{appariée}}^2} + \frac{c \text{sumdeg}_1 \cdot \text{sumdeg}_2}{d \text{sumdeg}_{\text{appariée}}^2} \quad (3.18)$$

où  $\text{nbseg}_1$  et  $\text{nbseg}_2$  sont les nombres de segments présents dans chacune des deux images,  $\text{nbseg}_{\text{appariée}}$  est le nombre de segments appariés entre les images ;  $\text{nbsmt}_1$ ,  $\text{nbsmt}_2$  et  $\text{nbsmt}_{\text{appariée}}$  sont les grandeurs similaires relatives aux sommets entre segments ;  $\text{sumdeg}_1$  et  $\text{sumdeg}_2$  sont les sommes des degrés de tous les sommets de chacune des deux images, et  $\text{sumdeg}_{\text{appariée}}$  est la somme des degrés des sommets de l'appariement (lorsque deux sommets sont appariés, on compte le nombre de paires de segments appariés partant de ces sommets).

$a$ ,  $b$  et  $c$  sont trois coefficients à déterminer expérimentalement. Suite à nos expériences, nous avons choisi  $a = b = 2$  et  $c = 1$ .  $d$  est égal à  $a + b + c$ . Cette mesure vaut 1 lorsque les images sont complètement appariées, et elle tend vers l'infini dans le cas inverse (c'est donc une mesure de dissemblance pour être exact).

On a aussi essayé une deuxième mesure qui tient compte d'informations supplémentaires : les invariants et la fiabilité des primitives considérées. Pour cela, on définit les éléments suivants.

- On appelle, dans ce paragraphe, jonction d'une image toute paire de segments de cette image ayant un sommet en commun. Lors de l'appariement, ce sont de telles jonctions qui sont utilisées pour calculer les invariants locaux.
- On note  $\text{Jonc}(I)$  l'ensemble des jonctions de l'image  $I$ , et  $\text{Jonc\_appariée}(I_1, I_2)$  l'ensemble des jonctions appariées entre les deux images  $I_1$  et  $I_2$ . Cet ensemble est déterminé lors de l'étape d'appariement entre les deux images.

- Chaque jonction  $i$  définit deux invariants : un angle  $A_i$  et le rapport  $R_i$  des longueurs de deux segments qui forment la jonction. On associe aussi une fiabilité  $f_i$  à la jonction, qui est définie en fonction des longueurs  $l_1$  et  $l_2$  des segments de la jonction par la formule suivante :

$$f_i = \frac{\text{arctg}(l_1 + l_2)}{\pi/2}$$

- $\alpha$  est un facteur de pondération entre angles et rapports de longueurs.

La mesure est alors définie de la manière suivante :

$$d_2(I_1, I_2) = \frac{\sum_{i_1 \in \text{Jonc}(I_1)} \left(2f_{i_1} + \frac{1}{2\alpha}\right) \times \sum_{i_2 \in \text{Jonc}(I_2)} \left(2f_{i_2} + \frac{1}{2\alpha}\right)}{\left(\sum_{i_1, i_2 \in \text{Jonc\_apparie}(I_1, I_2)} \left(f_{i_1} + f_{i_2} + \frac{1}{|A_{i_1} - A_{i_2}| + \alpha * \left(\frac{R_{i_1}}{R_{i_2}} + \frac{R_{i_2}}{R_{i_1}}\right)}\right)\right)^2} \quad (3.19)$$

Le numérateur caractérise chacune des deux images, et le dénominateur rend compte des appariements réalisés et de la qualité de ceux-ci, en comparant les invariants des jonctions appariées.

### Résultats expérimentaux

On considère la séquence de 10 images montrée sur la FIG. 3.7. Ces images sont appariées deux à deux, puis on calcule la matrice de ressemblance de la séquence avec chacune des deux mesures sus-citées.



FIG. 3.7 - Séquence d'images du premier objet

Avec la mesure  $d_1$  de la formule (3.18), on obtient la matrice suivante :

$$M_1 = \begin{pmatrix} 1 & 2.2 & 5.8 & 75.6 & 80.2 & 25.5 & 103.6 & 34.3 & \infty & 52.3 \\ & 1 & 2.1 & 63.4 & 27.9 & 98.9 & 30.4 & 115 & 103.3 & 90.1 \\ & & 1 & 3.9 & 17.8 & 49.7 & 59.7 & 73.1 & \infty & 18.8 \\ & & & 1 & 1.9 & 2.2 & 11.3 & 5.9 & 17.6 & 16.0 \\ & & & & 1 & 2 & \infty & 21.2 & 15.5 & 139 \\ & & & & & 1 & 1.8 & 6.9 & 9.3 & 89.9 \\ & & & & & & 1 & 7 & 32.2 & 99.8 \\ & & & & & & & 1 & 1.6 & \infty \\ & & & & & & & & 1 & 7.3 \\ & & & & & & & & & 1 \end{pmatrix}$$

Avec la mesure  $d_2$  (3.19), la matrice obtenue est la suivante :

$$M_2 = \begin{pmatrix} 1 & 3.2 & 10 & 640 & 2600 & 62 & 1800 & 91 & \infty & 180 \\ & 1 & 1.5 & 1300 & 380 & 3000 & 370 & 1800 & 1000 & 930 \\ & & 1 & 7.6 & 270 & 1300 & 1300 & 1200 & \infty & 52 \\ & & & 1 & 1.5 & 5.6 & 240 & 18 & 37 & 120 \\ & & & & 1 & 2.8 & \infty & 190 & 140 & \infty \\ & & & & & 1 & 1.5 & 14 & 43 & 230 \\ & & & & & & 1 & 12 & 190 & 590 \\ & & & & & & & 1 & 1.4 & \infty \\ & & & & & & & & 1 & 36 \\ & & & & & & & & & 1 \end{pmatrix}$$

Pour comparer ces résultats, on a tracé sur la FIG. 3.8, la courbe des mesures des neuf dernières images avec la première d'entre elles. La principale différence entre ces deux courbes est un changement d'échelle verticale, mais les résultats sont qualitativement les mêmes. La prise en compte des invariants n'apporte donc pas d'éléments prépondérants dans l'établissement de la ressemblance des images. Cela s'explique par le fait que seules les jonctions ayant des invariants proches sont appariées. Les différences entre invariants ont donc déjà été prises en compte lors de l'appariement. Les réutiliser pour la mesure de la ressemblance n'apporte pas d'informations nouvelles.

Dans la suite, on utilisera uniquement la première mesure qui est la plus simple.

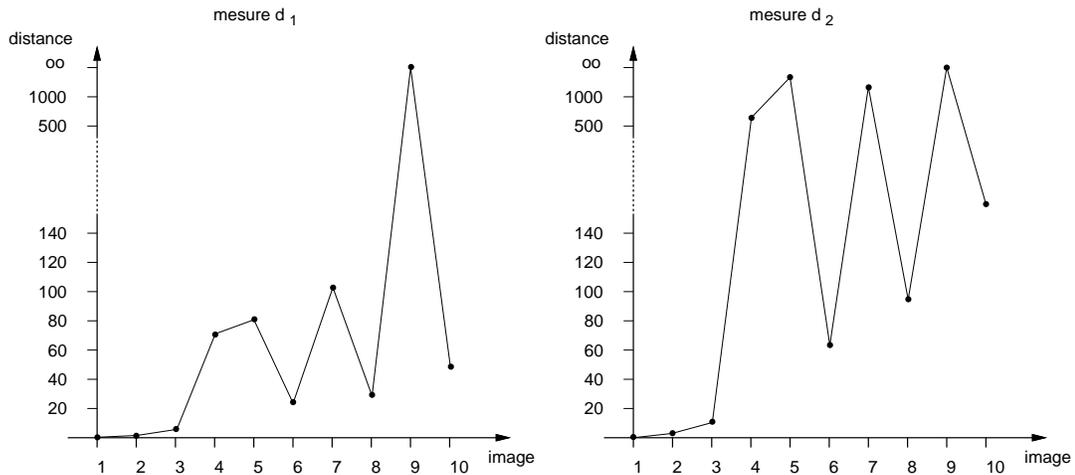


FIG. 3.8 - Comparaison des deux mesures.

### 3.3.2 Regroupement des images représentant un même aspect

Le stade de regroupement consiste à retrouver, à l'aide de la matrice de ressemblance qui vient d'être calculée, les images qui représentent un même aspect, c'est à dire celles qui forment des blocs de ressemblance dans la matrice. Il s'agit d'un problème très classique en reconnaissance des formes ou en classification automatique par exemple, et connu sous son nom anglais "clustering".

Diverses techniques de regroupement ont été proposées en vision. Par exemple, Scott et al. [Sco 90] utilisent les éléments propres d'une matrice déduite de la matrice des distances. L'inconvénient de sa méthode est qu'il faut pouvoir fixer au préalable soit le nombre de groupes que l'on souhaite, soit un seuil qui détermine la dissemblance maximale entre deux points qui peuvent être regroupés.

Diday et al. [Did 76] propose lui une méthode itérative, nommée nuées dynamiques, mais qui exige aussi de connaître au préalable le nombre de groupes que l'on souhaite former.

La difficulté du regroupement dans le processus de modélisation est que le nombre d'aspects que l'on veut isoler n'est pas connu au préalable, et qu'aucun seuil n'est connu non plus. Nous avons donc cherché une méthode qui permette de déterminer automatiquement ce seuil.

### Méthode de regroupement agglomérative

On utilise une méthode agglomérative formant des partitions. Cette méthode peut se résumer ainsi : il s'agit de regrouper les points les plus proches, jusqu'à un certain seuil qu'il faudra déterminer. De manière plus détaillée, les étapes de la méthode sont les suivantes.

1. On dispose d'un ensemble d'images dont on connaît une matrice de ressemblance. Chacune de ces images est mise dans un groupe. On a donc autant de groupes que d'images à cette première étape.
2. Les deux groupes les plus ressemblants sont rassemblés en un groupe unique si leur mesure de ressemblance est inférieure à un seuil donné.
3. La matrice des ressemblances est mise à jour. La mesure de ressemblance entre deux groupes est, par définition, la moyenne des mesures de ressemblance entre les images de ces deux groupes.
4. On boucle jusqu'à ce qu'aucun nouveau regroupement ne soit possible.

L'étape de mise à jour de la matrice des ressemblances peut être effectuée très simplement. En effet, soient  $I, J$  et  $K$  trois groupes. Supposons qu'on réunit  $I$  et  $J$ . On désire alors calculer la ressemblance entre le nouveau groupe formé et le groupe  $K$ . Notons  $|L|$  le cardinal du groupe  $L$ , c'est à dire le nombre d'images qu'il contient.

On a :

$$\begin{aligned}
 d(I \cup J, K) &= \frac{1}{|I \cup J||K|} \sum_{i \in I \cup J, k \in K} d(i, k) \\
 &= \frac{1}{|I \cup J||K|} \sum_{i \in I, k \in K} d(i, k) + \frac{1}{|I \cup J||K|} \sum_{j \in J, k \in K} d(j, k) \\
 &= \frac{|I|}{|I \cup J|} \left( \frac{1}{|I||K|} \sum_{i \in I, k \in K} d(i, k) \right) + \frac{|J|}{|I \cup J|} \left( \frac{1}{|J||K|} \sum_{j \in J, k \in K} d(j, k) \right)
 \end{aligned}$$

On obtient donc :

$$d(I \cup J, K) = \frac{|I|}{|I \cup J|} d(I, K) + \frac{|J|}{|I \cup J|} d(J, K)$$

Cette méthode, telle qu'exposée ici, forme une partition des images et nécessite la détermination d'un seuil. Le fait d'obtenir une partition permet de déterminer le seuil de manière automatique : des modifications apportées à l'algorithme pour ne pas former de partition, en permettant à une image d'appartenir à plusieurs groupes à la fois, ont entraîné la disparition de ce seuil.

Lorsqu'on considère la mesure de ressemblance entre les deux groupes qui sont réunis à chaque étape, on obtient une suite croissante de réels. De manière purement expérimentale, on a remarqué que cette suite possède toujours un saut parmi les valeurs successives. Voici quelques exemples de telles suites où le saut a été marqué par des nombres en caractères gras :

- 1,115 1,242 1,257 1,367 1,380 1,425 1,425 1,541 1,757 1,849  
2,120 2,130 2,485 **3,939** **36,097** 471,307 905,698
- 3,537 3,690 4,309 4,435 4,704 5,026 **10,488** **47,163** 75,783
- 1,18 1,26... 4,7 11,5 **13,14** **296,12**
- 1,04 1,14 1,20 1,47... 3,31 3,34 3,56... 6,87 7,65 12,57 13,73  
14,88 16,57 18,33 22,22 **22,75** **64,30**

Cette découverte expérimentale d'un saut correspond aussi à la détermination d'un seuil qui semble convenir pour les ensembles d'images utilisés, qu'on peut prendre égal à 25 par exemple. Ce seuil mériterait d'être affiné en fonction des séquences, car il est fort probable qu'un seuil légèrement différent puisse parfois donner des résultats qui semblent plus satisfaisants. Mais la détermination de la qualité du résultat est visuelle... et très qualitative !

### Résultats expérimentaux

La méthode décrite ci-dessus est appliquée à la recherche des aspects dans deux ensembles d'images. Le premier ensemble est formé des huit premières images montrées sur la FIG. 3.7. Trois groupes sont formés et montrés sur la FIG. 3.9 : chacun des groupes contient les images montrées sur une des lignes.

Le deuxième ensemble d'images est constitué des quatre-vingts images d'un objet, prises comme indiqué sur la FIG. 3.10. Les vingt premières images sont prises tous les 2,5 degrés, les soixante autres tous les 5 degrés. Sept aspects sont différenciés lors du regroupement :

**Aspect 1:** images 1 2 3 4 33 34 35 36 37 38 39 68 69 70 72 73

**Aspect 2:** images 5 22 23 24 25 26 28 58 59 60 61 62

**Aspect 3:** images 6 7 8 9 10 40 41 42 43 44 74 75 76 77 78

**Aspect 4:** images 11 12 13 14 15 27 45 46 47 48 49

**Aspect 5:** images 16 17 50 71

**Aspect 6:** images 18 19 20 21 51 52 53 54 55 56 57

**Aspect 7:** images 29 30 31 32 63 64 65 66 67

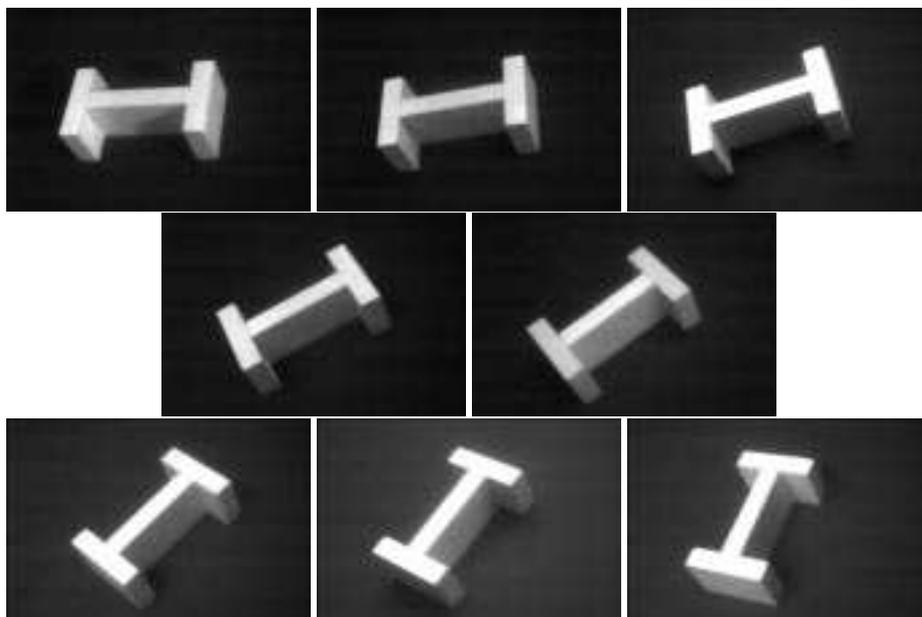


FIG. 3.9 - Les trois aspects d'un premier ensemble d'images.

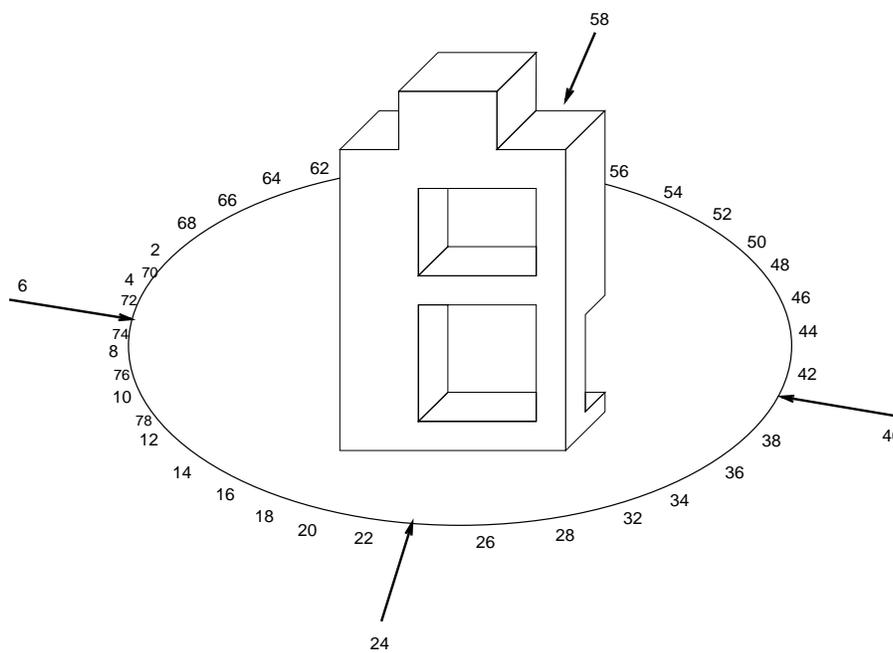


FIG. 3.10 - Quatre-vingts vues d'un objet.

La FIG. 3.11 montre deux images pour chacun de ces groupes et les segments et sommets extraits de ces images. On peut remarquer que l'objet observé est presque symétrique par rapport à un plan vertical. On trouve donc dans chacun des groupes des images prises des deux cotés de l'objet. Cela confirme le caractère pragmatique de la notion d'aspect utilisée. Pour séparer ces aspects, il suffit de baisser le seuil limite utilisé lors du regroupement. On peut même concevoir un algorithme interactif qui procéderait aux regroupements, proposant à chaque étape à un opérateur humain le soin de décider la terminaison du processus.

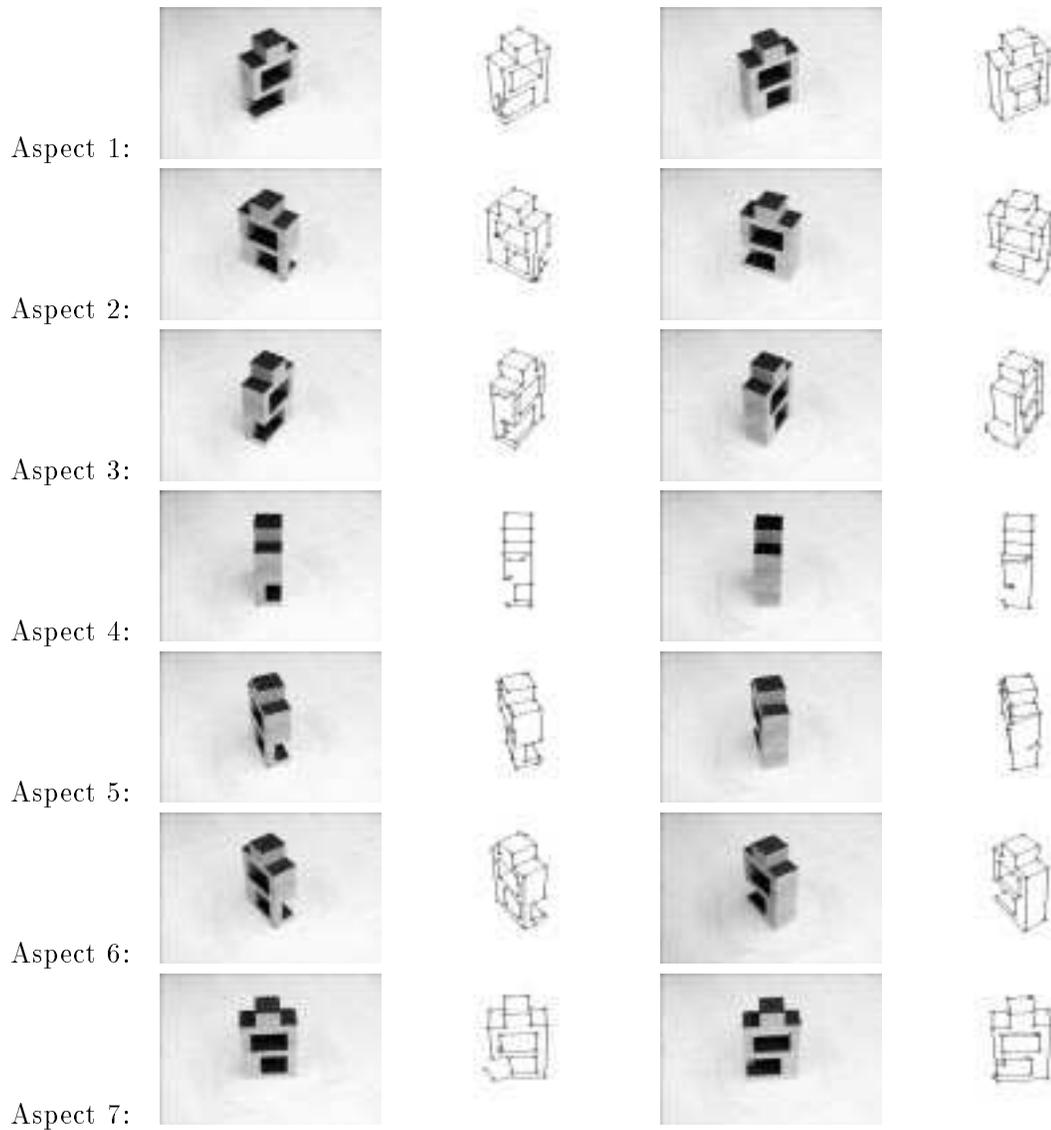


FIG. 3.11 - Deux représentants pour chacun des sept aspects détectés du deuxième ensemble d'images.

La deuxième remarque que l'on peut faire est que les aspects trouvés sont définis

en fonction de leur dissemblance pour les similitudes. Ainsi, les aspects 2 et 6 qui sont topologiquement semblables sont ici séparés car ils ne peuvent se déduire l'un de l'autre par une similitude de l'image. L'emploi de transformations plus générales lors de l'appariement doit entraîner une réduction du nombre des groupes formés : il doit être ainsi possible d'obtenir moins de modèles, mais ceux-ci, construits à partir d'images plus nombreuses, doivent donc être plus robustes.

Dernière remarque : l'algorithme n'est point parfait ; prenant en entrée des images bruitées, donc une matrice de ressemblance qui l'est aussi, certaines images n'ont pas été classées dans le groupe correct. Ainsi en est-il des images 5 et 71. Cela fait 2,5% d'erreur, ce qui est très bon vu la qualité des images initiales et les résultats de l'algorithme de segmentation visibles sur la FIG. 3.11.

## 3.4 Les étapes ultérieures

Le processus de modélisation et reconnaissance se décompose en plusieurs phases. Jusqu'ici, les phases suivantes ont été détaillées, ou supposées réalisées : acquisition et traitement des images, appariements des primitives, mesure des ressemblances entre images, regroupement de ces images selon l'aspect de l'objet qu'elles représentent. Le présent paragraphe concerne les trois dernières étapes : la modélisation des aspects, l'indexation des modèles et la reconnaissance. Ces trois étapes dépendent, de manière primordiale, de l'algorithme d'indexation. Celui-ci n'étant pas dans le champ de ce présent travail, les trois étapes sus-citées ne seront donc qu'esquissées.

Une fois le regroupement des images d'une même vue effectué, on dispose de l'information suivante sur l'objet : ce dernier est connu par les images qu'on a prises de lui, ces images sont appariées deux à deux et celles qui représentent un même aspect sont réunies dans un même groupe d'images. Par rapport à la situation de départ, on a donc su mettre les images en correspondance, c'est à dire retrouver les parties communes de l'objet qu'elles représentent, et on dispose maintenant de la notion d'aspect de l'objet.

### 3.4.1 La modélisation des aspects

La phase suivante dans le processus de modélisation et reconnaissance est la modélisation proprement dite. Cette phase consiste à regrouper l'information dont on dispose sur chaque aspect de l'objet. En effet, les différentes images qui représentent un même aspect sont similaires et la redondance de l'information sur cet aspect est donc forte.

Il faut remarquer que les étapes précédentes ont considérablement déblayé le travail. Les images sont déjà regroupées par aspect et appariées deux à deux. Il ne reste, en fait, qu'à collecter l'information dont on a besoin. La seule difficulté de cette tâche vient du fait que la cohérence des informations présentes n'est pas assurée. Cette difficulté doit pouvoir être surmontée grâce à la très forte redondance des informations entre les différentes images.

Le modèle créé peut, selon les besoins de son utilisation future, comprendre différents éléments. Tout d'abord, on peut y placer les primitives, segments et sommets, communs à toutes les images de la vue (on rappelle qu'on dénomme vue l'ensemble des images qui représente un même aspect de l'objet). L'inconvénient d'une telle méthode est d'éliminer toutes les primitives qui n'apparaissent pas dans au moins une image,

même si elles sont présentes dans la majorité des autres. Par ailleurs, on peut ajouter aux modèles des indices de nature géométrique ou statistique.

Les modèles, qui regroupent l'information disponible sur un aspect, ont une structure très proche de celle des images. Il s'agit de segments reliés par leurs extrémités. Les informations supplémentaires peuvent être ajoutées sous forme d'attribut supplémentaire de chaque segment (fiabilité, variance sur la position...), ou sous forme de relation entre ces segments (invariants...).

Tout d'abord, on peut associer aux primitives une fréquence d'apparition et mettre dans le modèle celles dont la fréquence est supérieure à un taux donné, 30% ou 50% par exemple. Cette fréquence d'apparition peut être stockée comme caractéristique supplémentaire associée aux primitives.

On peut ajouter au modèle d'autres informations disponibles. Tout d'abord, les aspects ont été calculés sur la base de la similarité d'invariants locaux 2D. Toutes les images d'une même vue doivent donc avoir des invariants locaux 2D, égaux au bruit près, qui peuvent être mémorisés. D'autre part, disposant de plusieurs images, nous pouvons calculer la géométrie épipolaire pour chaque paire d'images et calculer des invariants 3D ou rechercher des coplanarités. Ces informations peuvent aussi être stockées dans le modèle.

Les différentes informations numériques de position, d'orientation, de fiabilité, les invariants, peuvent être complétées avec leur variance. D'autres mesures, non stables en général, mais qui le seraient, vus le type des images traitées et la méthode suivie, sont sans doute imaginables.

L'intégration de toutes ces données dans le modèle est sujette à deux restrictions. Tout d'abord, il faut veiller à ne pas trop augmenter la taille des modèles. La détermination des informations réellement nécessaires dépend des algorithmes d'indexation et de reconnaissance utilisés. De toutes les possibilités évoquées ci-dessus, toutes ne devront sans doute pas être retenues. D'autre part, et cela est surtout valable pour les calculs de coplanarité et d'invariants, il faut veiller avec grand soin à la précision des calculs effectués, le fait de prendre simplement des moyennes des grandeurs observées dans les images ou paires d'images d'une vue pouvant, en effet, leur faire perdre tout caractère discriminant [Mor 93].

### 3.4.2 L'indexation des modèles

Cette étape n'est, ici, citée qu'à titre de rappel. Elle ne fait pas partie du champ de cette thèse. Cela dit, elle constitue une étape importante du processus de modélisation. Une fois les modèles construits, il est nécessaire de les stocker pour permettre une reconnaissance la plus rapide et la moins combinatoire possible.

L'établissement de bases de données d'images est un thème de recherche très actif à l'heure actuelle. La différence entre notre système et les nombreux systèmes de classement d'images existant est que nous ne souhaitons pas avoir accès aux modèles stockés en fonctions de mots clés ou de caractéristiques stockées avec ces images, mais en fonction de leur aspect visuel.

Le problème de l'indexation d'images pour la reconnaissance en vision a déjà fait l'objet de divers travaux. Le lecteur trouvera dans [Sos 92] une bibliographie et une proposition d'indexation topologique. Les systèmes d'indexation d'images pour la vision sont généralement basés soit sur une description géométrique des images donnant

lieu aux techniques de hachage géométrique [Kal 86, Lam 88, Wol 90], soit sur une description structurelle [Nev 77, Ett 88, Ste 92, Sos 92].

La limite de ces méthodes, en vue d'une utilisation dans notre système, est qu'elles n'utilisent pas simultanément l'information structurelle et l'information géométrique. La conséquence en est une combinatoire élevée pour les algorithmes géométriques et une faible robustesse pour les algorithmes structurels. Le problème de l'indexation des modèles est donc encore ouvert.

Une solution adaptée à nos données semble possible à construire à partir du système de Sossa [Sos 92]. Dans ce travail, les images sont considérées comme des graphes, les segments étant les arêtes et leurs extrémités étant les sommets. Ces graphes sont décomposés en petits sous-graphes. A chacun de ces sous-graphes, on associe sa matrice d'adjacence, puis un polynôme caractéristique de cette matrice (qui n'est pas "le" polynôme caractéristique, mais son deuxième immanent). Deux des coefficients de ce polynôme sont alors choisis comme clé d'indexation du sous-graphe.

Lors de la phase d'indexation, on associe donc à un graphe de nombreuses paires de clés. Lors de la reconnaissance, l'image à reconnaître est aussi décomposée en sous-graphes codés par des clés. On compare ces clés avec celles de la base de modèles. Cela indique de nombreux modèles qui comportent un sous-graphe commun avec l'image. Un système de vote permet alors de ne garder que les modèles les plus probables.

On peut reprendre ce schéma, en y intégrant simplement les invariants. On prend des sous-graphes qui peuvent être plus simples, donc plus robustes que ceux utilisés dans le système de Sossa, et on les caractérise par leurs invariants. L'indexation et la reconnaissance sont alors faites sur le même principe, en utilisant un système de vote.

### 3.4.3 Reconnaissance des images

La dernière étape est la reconnaissance. Celle-ci se décompose en deux temps. Tout d'abord, l'image à reconnaître est traitée, puis le système d'indexation fournit des prédictions : il recherche dans la base de données les modèles les plus susceptibles de convenir. Cette recherche doit être effectuée sur des critères locaux pour être utilisable en cas d'occlusion partielle. Elle peut donc être faite sur la base des sous-graphes ayant servi à l'indexation.

La qualité du système d'indexation peut être jugée à la réduction des prédictions qu'il réalise : quel est le pourcentage des modèles proposés comme candidats par rapport à l'ensemble des modèles de la base ?

Le deuxième temps est la vérification. C'est une étape exhaustive qui procède par appariement. On apparie l'image à reconnaître avec tous les modèles proposés, le modèle retenu finalement est celui qui donne le meilleur appariement.

Il faut remarquer que les modèles ayant une structure similaire à celle des images, l'algorithme d'appariement du paragraphe 2.3 peut être utilisé sans beaucoup de modifications à cette étape. Pour évaluer l'appariement, on peut utiliser la mesure de ressemblance du paragraphe 3.3.1. Cette mesure peut être légèrement modifiée pour tenir compte des fréquences d'apparition des primitives du modèle : les primitives de forte fréquence ont alors un poids plus fort.

L'information de coplanarité et les invariants 3D peuvent aussi servir comme facteurs de discrimination. L'image à reconnaître étant appariée avec un modèle, on peut

calculer la géométrie épipolaire puis calculer des invariants 3D ou rechercher les éléments coplanaires et comparer ces résultats avec ceux stockés dans le modèle.

Pour terminer, on peut envisager un système incrémentiel : l'image ayant été reconnue, les données qu'elle contient peuvent être incluses dans le modèle par une mise à jour des fréquences d'apparition par exemple.

### 3.5 Conclusion du chapitre

Les contributions de ce chapitre sont de deux ordres. Tout d'abord, nous avons proposé une méthode de modélisation d'objets pour la reconnaissance. Cette méthode se distingue de celles proposées dans la littérature par des hypothèses peu restrictives (il suffit que les images contiennent des segments qui représentent des arêtes 3D), une faible combinatoire et une grande robustesse.

Ensuite, nous avons développé les premiers algorithmes de cette méthode. L'appariement des images a été exposé dans le chapitre précédent, mais il fait partie intégrante de la modélisation. Suite à ce premier appariement utilisant les invariants locaux, on peut utiliser d'autres techniques pour compléter et corriger cet appariement. Les images sont alors comparées et regroupées selon les aspects qu'elles représentent.

Tous les algorithmes présentés sont validés par leur utilisation correcte avec des images réelles.

# Conclusion

Le développement d'un système de modélisation et reconnaissance automatique d'objets polyédriques est, encore aujourd'hui, un objectif ambitieux. Les différentes méthodes essayées jusqu'à présent se sont souvent révélées peu robustes, ou trop spécialisées, ou dépendantes d'hypothèses trop fortes pour une utilisation pratique.

Cette thèse apporte deux contributions principales pour résoudre ce problème.

1. La première est d'ordre méthodologique. Nous proposons un cadre et une méthode pour modéliser des objets, et les premiers algorithmes de cette méthode ont été développés : il s'agit de l'appariement des images, de la mesure de leur ressemblance et de leur regroupement selon les aspects de l'objet observé qu'elles représentent.
2. La deuxième concerne l'application de la théorie des invariants en vision. Nous montrons comment cette théorie permet d'extraire des quantités numériques stables des images, et donc d'apparier ces images, ou de calculer des caractéristiques 3D des objets observés.

Nous avons proposé le schéma d'un système nouveau dont la robustesse vient des principes adoptés. Ces principes sont classiques dans certaines disciplines comme la classification automatique, mais ils n'avaient pas encore été mis à profit dans un système de vision. Ces principes sont les suivants :

- on doit apprendre les modèles à partir de données de même type que celles qui seront à reconnaître, c'est à dire à partir d'images et non de données artificielles, de type CAO par exemple ;
- il faut rester proche des données : les étapes supplémentaires, telle une reconstruction 3D, ont tendance à dégrader fortement l'information contenue dans les données ;

- ne pas avoir besoin de connaissances préalables pour la modélisation permet de faire une reconnaissance à partir de données non contraintes : la généralité de la reconnaissance dépend de celle de la modélisation.

On est donc parti des hypothèses suivantes : on connaît un objet par un ensemble non ordonné de vues prises de lui sous différents angles de vues. On essaye alors d'extraire de ces images les informations qui caractérisent cet objet et qui permettent donc de le reconnaître.

L'algorithme central de notre système est celui d'appariement. Cet appariement, basé sur l'utilisation d'invariants locaux et d'une contrainte géométrique globale, permet de retrouver les parties communes à deux images lorsque le mouvement relatif des primitives dans les images peut être approché par une similitude.

Une fois cet appariement réalisé, les images peuvent être comparées et regroupées selon les aspects de l'objet à modéliser qu'elles représentent. Chaque aspect donne alors lieu à un modèle qui condense et organise l'information dont on dispose à son propos. Cette information est géométrique (position, orientation, invariants...), mais aussi statistique (fiabilité, variance...).

A partir de l'appariement des images, on peut, de manière parallèle à ce qui précède, calculer la géométrie épipolaire de la paire d'images, puis calculer, à partir de ces images et sans reconstruction, des invariants 3D et des propriétés de coplanarité caractéristiques de l'objet 3D.

Ce calcul d'invariants 3D a été l'occasion d'une étude plus générale de la théorie des invariants dans le cadre de la vision. Celle-ci permet de répondre aux questions relatives au nombre et à l'indépendance des invariants existant dans un problème de vision. Du point de vue calcul de ces invariants, nous avons préconisé l'emploi d'une méthode de recherche des invariants canoniques qui, si elle n'est pas systématique, s'est révélée souple et extrêmement puissante dans tous les cas que nous avons traités.

Les différentes étapes de ce système ont été développées avec un grand souci de robustesse au bruit. Celui-ci est, en effet, une caractéristique incontournable de toute image. On a ainsi étudié les invariants 3D avec des données simulées afin de définir la précision que l'on peut atteindre à différents niveaux de bruit. Les principaux algorithmes proposés ont aussi fait l'objet d'une validation avec des images et données réelles.

Les principales avancées faites dans cette thèse se situent à plusieurs niveaux. Tout d'abord, nous avons montré que l'utilisation des invariants canoniques permettait la recherche et le calcul aisés des invariants dans les problèmes où n'interviennent que des éléments d'ordre inférieur à deux : points, droites, plans, coniques. Cette méthode a même été appliquée avec succès au calcul d'invariants 3D à partir de deux images dont on ne connaît que la géométrie épipolaire, sans reconstruction aucune.

Du côté de la modélisation, notre système reprend des éléments de divers systèmes existant d'une manière originale, qui permet de passer outre les limites de ces systèmes. Ainsi, on reconnaît un objet en modélisant ses aspects (cf. les graphes d'aspects), en partant de ce qui est perçu (cf. l'apprentissage), la topologie des images étant une information très riche (cf. les méthodes structurelles) mais peu robuste, ce à quoi on remédie en utilisant la géométrie (cf. les méthodes utilisant les invariants).

Toutefois, le défi n'est pas complètement relevé, et quelques questions sont encore

ouvertes.

1. Tout d'abord, nous avons utilisé des objets polyédriques. La méthode d'appariement dépend des invariants locaux dans les images, invariants facilement calculables et énumérables dans le cas des objets polyédriques. Qu'en est-il des objets courbes ? La question se pose différemment pour les objets ayant des arêtes courbes planes, ceux ayant des arêtes courbes gauches et ceux n'ayant point d'arêtes du tout.

Pour les premiers, une solution semble être l'emploi d'invariants semi-différentiels : on calcule alors des invariants classiques à partir d'éléments différentiels telles les droites de bitangence et d'inflexion. Pour les seconds, la question est de trouver des invariants par projection perspective. Pour les derniers, toutes les méthodes les concernant à l'heure actuelle font des hypothèses bien plus fortes que les nôtres : l'étalonnage des caméras semble encore indispensable.

2. Le problème de l'indexation des modèles n'a pas été traité. En dépit de quelques travaux précurseurs dans le domaine, aucun système n'arrive à indexer de manière robuste des modèles ou images non simplistes, réduits à quelques points ou droites, ou caractérisés par seulement 2 ou 3 invariants. Ce problème est incontestablement un des problèmes durs de la vision, à l'image du problème d'appariement des images : l'homme voit et se rappelle de très nombreuses images mais il ne connaît pas son algorithme de reconnaissance. Tout au plus, on sait que ce dernier est massivement parallèle, ce en quoi il diffère des algorithmes informatiques classiques d'indexation et de reconnaissance.
3. En de nombreux passages de cette thèse, il a été signalé que la précision et la qualité des données étaient primordiales pour les algorithmes étudiés. La vision, dont le passé se trouve tant du côté du traitement du signal que de la reconnaissance des formes et de l'intelligence artificielle, traite de plus en plus des problèmes réels, dans des environnements variés. Il reste toutefois encore du travail pour rendre plus robustes les algorithmes de vision afin de les rendre utilisables dans un contexte industriel. La vision par ordinateur, science pour l'ingénieur, recevrait, sans conteste, dans ce passage à l'industrie son plus beau titre de gloire.



## Annexe A

# Construction de points coplanaires

Cette annexe détaille les calculs des deux méthodes de construction des projections de nouveaux points coplanaires du paragraphe 2.1.2. On reprend les notations de ce paragraphe.

### A.1 Première méthode

#### A.1.1 Principe de la méthode

Cette première méthode procède par simple inversion du test. Rappelons-en les étapes.

1.  $e_1$  et  $e_2$  étant les épipoles des deux images, soit  $d_1$  un point quelconque de  $l_1$  (cf. FIG. A.1). On calcule le birapport  $k$  du faisceau de droites  $(e_1a_1, e_1b_1, e_1c_1, e_1d_1)$ .
2. Ce birapport définit une unique droite  $\lambda_2$  de l'image 2, qui passe par  $e_2$  et telle que le faisceau  $(e_2a_2, e_2b_2, e_2c_2, \lambda_2)$  ait  $k$  comme birapport.  $\lambda_2$  est donc la droite épipolaire associée à  $d_1$  dans la deuxième image.
3. Soit  $d_2$  le point d'intersection entre  $\lambda_2$  et  $l_2$ , et soient  $f_1$  et  $f_2$  ceux de  $(a_1c_1)$  et  $(b_1d_1)$ , et  $(a_2c_2)$  et  $(b_2d_2)$  respectivement.

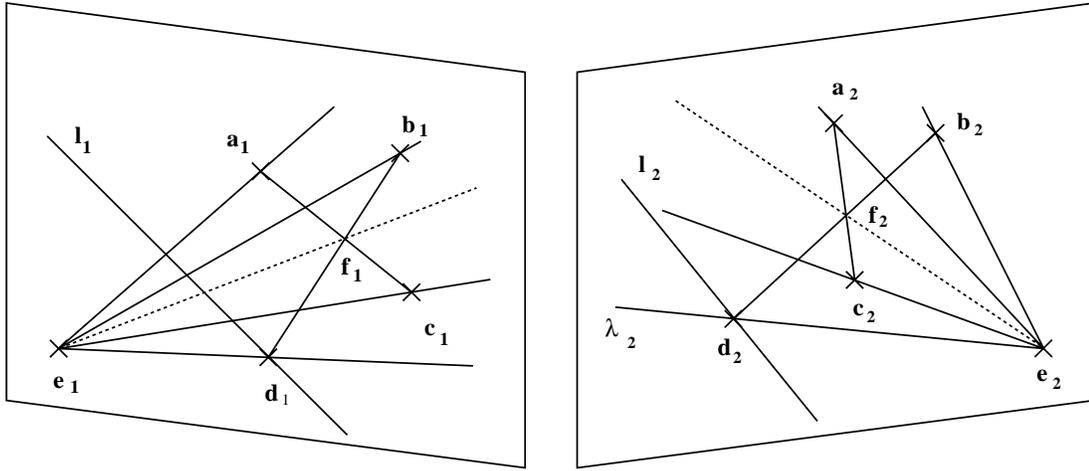


FIG. A.1 - Construction de points coplanaires : première méthode.

4. D'après le test de coplanarité,  $d_1$  et  $d_2$  sont les projections d'un point  $D$  de  $L$  coplanaire avec  $A$ ,  $B$  et  $C$ , si et seulement si les deux faisceaux  $(e_1a_1, e_1b_1, e_1c_1, e_1f_1)$  et  $(e_2a_2, e_2b_2, e_2c_2, e_2f_2)$  ont les mêmes birapports.

### A.1.2 Calculs

Pour simplifier les calculs, on applique au préalable aux images une technique de rectification. Il s'agit de transformer chacune des images afin que les épipoles soient à l'infini et que les épipolaires soient parallèles. On peut, de plus, faire en sorte que ces épipolaires soient parallèles à l'axe des abscisses.

#### Rectification des images

La rectification des images est un algorithme classique. Déjà les photogrammètres utilisaient des miroirs pour reprojeter leurs images et les rectifier. On trouvera diverses méthodes dans [Lus 87, Aya 89, Moh 92a, Hor 93].

Etant données deux images appariées  $I_1$  et  $I_2$  contenant des points, il s'agit de trouver deux transformations projectives  $T_1$  et  $T_2$  telles que les transformées des deux images,  $I'_1$  et  $I'_2$ , vérifient les conditions suivantes :

1. les épipoles  $e_1$  et  $e_2$  ont pour coordonnées dans chaque image  $(1, 0, 0)$  (c'est à dire qu'ils sont à l'infini dans la direction de l'axe des abscisses) ;
2. Les droites épipolaires sont parallèles à l'axe des abscisses ;
3. l'épipolaire associée à un point a la même ordonnée que lui.

La première condition se traduit ainsi :

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = T_1 \begin{pmatrix} x_{e_1} \\ y_{e_1} \\ t_{e_1} \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = T_2 \begin{pmatrix} x_{e_2} \\ y_{e_2} \\ t_{e_2} \end{pmatrix}$$

Pour satisfaire les deux dernières conditions, on choisit trois points de la première image  $p_1, q_1$  et  $r_1$ . On impose qu'il soient des points fixes de  $T_1$ . Cette homographie est alors déterminée complètement, car les images de quatre points ont été choisies. On impose de plus que les points qui leur correspondent dans la deuxième image,  $p_2, q_2$  et  $r_2$  soient transformés de telle manière qu'ils aient les mêmes ordonnées que  $p_1, q_1$  et  $r_1$ .

Cela ne détermine pas complètement  $T_2$ . Les degrés de liberté restant peuvent être utilisés pour que les points de la deuxième image gardent des abscisses compris entre 0 et 256 ou 512 par exemple.

La matrice fondamentale des deux nouvelles images,  $I'_1$  et  $I'_2$ , est alors telle que le point de coordonnées  $(x, y, t)$  est associé à la droite de l'autre image de même ordonnée, donc de coordonnées  $(0, t, -y)$ . Cette matrice est alors :

$$F = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

### Reconstruction

On suppose qu'on a rectifié les images comme indiqué précédemment, et on travaille donc désormais dans les nouvelles images obtenues. Cette rectification n'a pour but que de simplifier les calculs. Ceux-ci sont toutefois possibles directement avec les images d'origine, mais l'emploi d'un système de calcul formel est alors fortement souhaitable, vue la longueur des expressions.

On suppose que  $l_1$  et  $l_2$  sont connues par leurs coordonnées homogènes  $(u_i, v_i, w_i)$  avec  $i \in \{1, 2\}$ . Un point quelconque de  $l_1$ ,  $d_1$  par exemple, peut être défini en fonction d'un paramètre  $\mu$  : on considère pour cela les deux points  $m_1$  et  $n_1$  de coordonnées homogènes respectives  $(-w_{l_1}, 0, u_{l_1})$  et  $(0, -w_{l_1}, v_{l_1})$ , qui appartiennent à  $l_1$ , et on écrit  $d_1 = \mu m_1 + (1-\mu)n_1$ . Les coordonnées de  $d_1$  sont alors :  $(-\mu v_{l_1} w_{l_1}, -(1-\mu)u_{l_1} w_{l_1}, u_{l_1} v_{l_1})$ .

On peut alors calculer les coordonnées de  $f_1$  :

$$\begin{aligned} x_{f_1} &= (y_{a_1} x_{c_1} - x_{a_1} y_{c_1})(-\mu v_{l_1} w_{l_1} - x_{b_1} u_{l_1} v_{l_1}) \\ &\quad - (x_{b_1}(1-\mu)u_{l_1} w_{l_1} - y_{b_1} \mu v_{l_1} w_{l_1})(x_{c_1} - x_{a_1}) \\ y_{f_1} &= (x_{a_1} y_{c_1} - y_{a_1} x_{c_1})((1-\mu)u_{l_1} w_{l_1} + y_{b_1} u_{l_1} v_{l_1}) \\ &\quad - (x_{b_1}(1-\mu)u_{l_1} w_{l_1} - y_{b_1} \mu v_{l_1} w_{l_1})(y_{c_1} - y_{a_1}) \\ t_{f_1} &= (y_{c_1} - y_{a_1})(\mu v_{l_1} w_{l_1} + x_{b_1} u_{l_1} v_{l_1}) - ((1-\mu)u_{l_1} w_{l_1} + y_{b_1} u_{l_1} v_{l_1})(x_{c_1} - x_{a_1}) \end{aligned}$$

Le point  $d_2$  est l'intersection de  $l_2$  et de la droite épipolaire  $\lambda_2$  associée à  $d_1$ . Notons  $F$  la matrice fondamentale.  $\lambda_2$  a pour coordonnées homogènes :

$$\lambda_2 = Fd_1 = (0, u_{l_1} v_{l_1}, (1-\mu)u_{l_1} w_{l_1})$$

Celles de  $d_2$  sont donc :

$$\begin{aligned} x_{d_2} &= w_{l_2} u_{l_1} v_{l_1} - v_{l_2}(1-\mu)u_{l_1} w_{l_1} \\ y_{d_2} &= u_{l_2}(1-\mu)u_{l_1} w_{l_1} \\ t_{d_2} &= -u_{l_2} u_{l_1} v_{l_1} \end{aligned}$$

et celles de  $f_2$  :

$$\begin{aligned}
x_{f_2} &= (y_{a_2}x_{c_2} - x_{a_2}y_{c_2}) [w_{l_2}u_{l_1}v_{l_1} - v_{l_2}(1 - \mu)u_{l_1}w_{l_1} + x_{b_2}u_{l_2}u_{l_1}v_{l_1}] \\
&\quad - (x_{c_2} - x_{a_2}) [y_{b_2}(w_{l_2}u_{l_1}v_{l_1} - v_{l_2}(1 - \mu)u_{l_1}w_{l_1}) - x_{b_2}u_{l_2}(1 - \mu)u_{l_1}w_{l_1}] \\
y_{f_2} &= (y_{a_2}x_{c_2} - x_{a_2}y_{c_2}) [u_{l_2}(1 - \mu)u_{l_1}w_{l_1} + y_{b_2}u_{l_2}u_{l_1}v_{l_1}] \\
&\quad - (y_{c_2} - y_{a_2}) [y_{b_2}(w_{l_2}u_{l_1}v_{l_1} - v_{l_2}(1 - \mu)u_{l_1}w_{l_1}) - x_{b_2}u_{l_2}(1 - \mu)u_{l_1}w_{l_1}] \\
t_{f_2} &= (x_{c_2} - x_{a_2}) [u_{l_2}(1 - \mu)u_{l_1}w_{l_1} + y_{b_2}u_{l_2}u_{l_1}v_{l_1}] \\
&\quad - (y_{c_2} - y_{a_2}) [w_{l_2}u_{l_1}v_{l_1} - v_{l_2}(1 - \mu)u_{l_1}w_{l_1} + x_{b_2}u_{l_2}u_{l_1}v_{l_1}]
\end{aligned}$$

On exprime alors la condition de coplanarité :

$$\frac{|e_1a_1c_1||e_1b_1f_1|}{|e_1a_1f_1||e_1b_1c_1|} = \frac{|e_2a_2c_2||e_2b_2f_2|}{|e_2a_2f_2||e_2b_2c_2|}$$

En développant en fonction des coordonnées, on obtient :

$$\begin{aligned}
0 &= (y_{a_1} - y_{c_1})(y_{b_2} - y_{c_2}) \\
&\quad (y_{b_1} ((y_{c_1} - y_{a_1})(\mu v_{l_1}w_{l_1} + x_{b_1}u_{l_1}v_{l_1}) \\
&\quad\quad - ((1 - \mu)u_{l_1}w_{l_1} + y_{b_1}u_{l_1}v_{l_1})(x_{c_1} - x_{a_1})) \\
&\quad\quad - (x_{a_1}y_{c_1} - y_{a_1}x_{c_1})((1 - \mu)u_{l_1}w_{l_1} + y_{b_1}u_{l_1}v_{l_1}) \\
&\quad\quad + (x_{b_1}(1 - \mu)u_{l_1}w_{l_1} - y_{b_1}\mu v_{l_1}w_{l_1})(y_{c_1} - y_{a_1})) \\
&\quad (y_{a_2} ((u_{l_2}(1 - \mu)u_{l_1}w_{l_1} + y_{b_2}u_{l_2}u_{l_1}v_{l_1})(x_{c_2} - x_{a_2}) \\
&\quad\quad - (y_{c_2} - y_{a_2})(w_{l_2}u_{l_1}v_{l_1} - v_{l_2}(1 - \mu)u_{l_1}w_{l_1} + x_{b_2}u_{l_2}u_{l_1}v_{l_1})) \\
&\quad\quad - (y_{a_2}x_{c_2} - x_{a_2}y_{c_2})(u_{l_2}(1 - \mu)u_{l_1}w_{l_1} + y_{b_2}u_{l_2}u_{l_1}v_{l_1}) \\
&\quad\quad + (y_{b_2}(w_{l_2}u_{l_1}v_{l_1} - v_{l_2}(1 - \mu)u_{l_1}w_{l_1}) \\
&\quad\quad\quad - x_{b_2}u_{l_2}(1 - \mu)u_{l_1}w_{l_1})(y_{c_2} - y_{a_2})) \\
&\quad - (y_{a_2} - y_{c_2})(y_{b_1} - y_{c_1}) \\
&\quad (y_{b_2} ((u_{l_2}(1 - \mu)u_{l_1}w_{l_1} + y_{b_2}u_{l_2}u_{l_1}v_{l_1})(x_{c_2} - x_{a_2}) \\
&\quad\quad - (y_{c_2} - y_{a_2})(w_{l_2}u_{l_1}v_{l_1} - v_{l_2}(1 - \mu)u_{l_1}w_{l_1} + x_{b_2}u_{l_2}u_{l_1}v_{l_1})) \\
&\quad\quad - (y_{a_2}x_{c_2} - x_{a_2}y_{c_2})(u_{l_2}(1 - \mu)u_{l_1}w_{l_1} + y_{b_2}u_{l_2}u_{l_1}v_{l_1}) \\
&\quad\quad + (y_{b_2}(w_{l_2}u_{l_1}v_{l_1} - v_{l_2}(1 - \mu)u_{l_1}w_{l_1}) \\
&\quad\quad\quad - x_{b_2}u_{l_2}(1 - \mu)u_{l_1}w_{l_1})(y_{c_2} - y_{a_2})) \\
&\quad (y_{a_1} ((y_{c_1} - y_{a_1})(\mu v_{l_1}w_{l_1} + x_{b_1}u_{l_1}v_{l_1}) \\
&\quad\quad - ((1 - \mu)u_{l_1}w_{l_1} + y_{b_1}u_{l_1}v_{l_1})(x_{c_1} - x_{a_1})) \\
&\quad\quad - (x_{a_1}y_{c_1} - y_{a_1}x_{c_1})((1 - \mu)u_{l_1}w_{l_1} + y_{b_1}u_{l_1}v_{l_1}) \\
&\quad\quad + (x_{b_1}(1 - \mu)u_{l_1}w_{l_1} - y_{b_1}\mu v_{l_1}w_{l_1})(y_{c_1} - y_{a_1}))
\end{aligned}$$

C'est une équation du second degré en  $\mu$ . On peut la simplifier en tenant compte du fait que, grâce à la rectification des images, les points  $a_1$  et  $a_2$  ont la même ordonnée, ainsi que les points  $b_1$  et  $b_2$ , et  $c_1$  et  $c_2$  respectivement.

Pour résoudre cette équation, l'emploi d'un système de calcul formel est indispensable. Le discriminant de cette équation, lorsqu'il est développé est un polynôme contenant plus de 5300 monômes, ceux-ci étant des facteurs de degré total 23 ou 24. L'équation peut toutefois être résolue en quelques secondes. Les deux solutions trouvées sont :

$$\mu_1 = \frac{y_{b_1}v_{l_1} + w_{l_1}}{w_{l_1}}$$

et :

$$\mu_2 = -\frac{u_{l_1} D}{w_{l_1} N}$$

avec :

$$D = \begin{array}{cccc} -y_{c_1} x_{a_1} x_{b_2} u_{l_2} v_{l_1} & -y_{c_1} x_{a_1} w_{l_2} v_{l_1} & +y_{c_1} x_{b_1} w_{l_2} v_{l_1} & +y_{c_1} x_{a_2} u_{l_2} x_{b_1} v_{l_1} \\ -y_{c_1} x_{b_1} w_{l_1} v_{l_2} & +y_{c_1} x_{a_1} w_{l_1} v_{l_2} & +x_{b_1} w_{l_1} y_{a_1} v_{l_2} & -x_{b_1} y_{a_1} w_{l_2} v_{l_1} \\ +y_{a_1} x_{c_1} x_{b_2} u_{l_2} v_{l_1} & -y_{a_1} u_{l_2} x_{c_2} x_{b_1} v_{l_1} & -y_{a_1} x_{c_1} w_{l_1} v_{l_2} & +y_{a_1} x_{c_1} w_{l_2} v_{l_1} \\ +x_{b_1} w_{l_1} u_{l_2} x_{a_2} & -x_{c_1} y_{b_1} v_{l_1} x_{a_2} u_{l_2} & -u_{l_2} w_{l_1} x_{c_2} x_{b_1} & +u_{l_2} w_{l_1} x_{c_2} x_{a_1} \\ -x_{c_1} w_{l_1} x_{a_2} u_{l_2} & +u_{l_2} x_{c_2} x_{a_1} y_{b_1} v_{l_1} & +x_{c_1} w_{l_1} v_{l_2} y_{b_1} & -x_{a_1} w_{l_1} y_{b_1} v_{l_2} \\ -y_{b_1} x_{c_1} w_{l_2} v_{l_1} & -x_{a_1} w_{l_1} x_{b_2} u_{l_2} & +x_{c_1} w_{l_1} x_{b_2} u_{l_2} & +x_{a_1} y_{b_1} w_{l_2} v_{l_1} \end{array}$$

$$N = \begin{array}{cccc} y_{c_1} x_{b_1} u_{l_1} v_{l_2} & -y_{c_1} x_{a_1} u_{l_1} v_{l_2} & +y_{c_1} x_{a_2} u_{l_2} v_{l_1} & -y_{c_1} x_{b_2} u_{l_2} v_{l_1} \\ -y_{a_1} u_{l_2} x_{c_2} v_{l_1} & -x_{b_1} u_{l_1} y_{a_1} v_{l_2} & +y_{a_1} x_{c_1} u_{l_1} v_{l_2} & +x_{b_2} u_{l_2} y_{a_1} v_{l_1} \\ +x_{a_1} u_{l_1} x_{b_2} u_{l_2} & +u_{l_2} u_{l_1} x_{c_2} x_{b_1} & -x_{b_1} u_{l_1} u_{l_2} x_{a_2} & +u_{l_2} x_{c_2} y_{b_1} v_{l_1} \\ +x_{c_1} u_{l_1} x_{a_2} u_{l_2} & +x_{a_1} u_{l_1} y_{b_1} v_{l_2} & -u_{l_2} u_{l_1} x_{c_2} x_{a_1} & -x_{c_1} u_{l_1} x_{b_2} u_{l_2} \\ -x_{a_2} u_{l_2} y_{b_1} v_{l_1} & -x_{c_1} u_{l_1} v_{l_2} y_{b_1} & & \end{array}$$

La première solution correspond au cas dégénéré où  $d_1$  se trouve sur la droite épipolaire passant par  $b_1$ , ainsi que  $f_1$ .  $d_2$  et  $f_2$  se trouvent alors sur la droite épipolaire passant par  $b_2$  et les faisceaux de droites que l'on considère pour calculer les birapports ont deux droites confondues. Ils définissent alors les mêmes birapports.

Mathématiquement, l'appartenance de  $d_1$  et  $b_1$  à la même droite épipolaire se traduit par l'égalité de leurs ordonnées respectives. Soit :

$$-(1 - \mu)u_{l_1}w_{l_1} = y_{b_1}u_{l_1}v_{l_1}$$

ce qui donne :

$$\mu = \frac{y_{b_1}v_{l_1} + w_{l_1}}{w_{l_1}}$$

Il ne reste plus alors qu'à reporter l'expression de  $\mu_2$  dans les formules des coordonnées de  $d_1$  et  $d_2$ , ce qui est laissé à la charge du lecteur !

## A.2 Deuxième méthode

Cette deuxième méthode est une méthode constructive : on utilise des constructions géométriques linéaires pour déterminer les projections que l'on cherche. Ces constructions sont principalement la construction de la droite contenant deux points, l'intersection de deux droites, et la construction de la droite épipolaire associée à un point.

La méthode a été présentée de manière géométrique dans le corps du texte. Cette présentation est rappelée pour mémoire ici. Ce paragraphe en reprend ensuite les étapes en faisant les calculs de manière plus analytique.

### A.2.1 Principe de la méthode

Etant données les projections de cinq points  $A, B, C, D$  et  $E$  dans deux images, on cherche dans ces images les projections du point  $J$  qui est l'intersection du plan  $(ABC)$  avec la droite  $(DE)$  (cf. FIG. A.2). On suppose que la géométrie épipolaire entre les deux images est connue.

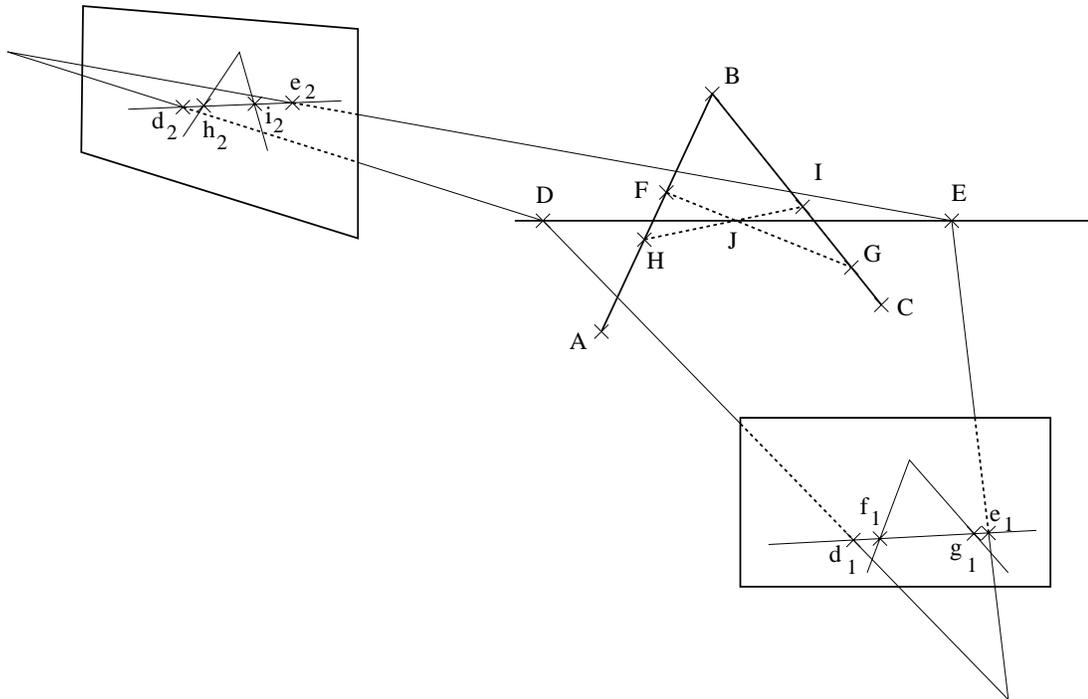


FIG. A.2 - Deuxième méthode de construction de points coplanaires : vue 3D.

Le plan d'interprétation de la droite  $(DE)$  par rapport à la première image, c'est à dire le plan qui contient cette droite et le centre de projection de la première image, coupe respectivement  $(AB)$  et  $(BC)$  en  $F$  et  $G$ . Les droites  $(DE)$  et  $(FG)$  se coupent en un point qui est coplanaire avec  $A, B$  et  $C$  : c'est le point  $J$ . Dans la première image, la construction des projections de  $F$  et  $G$ ,  $f_1$  et  $g_1$  respectivement, est immédiate (cf. FIG. A.3).

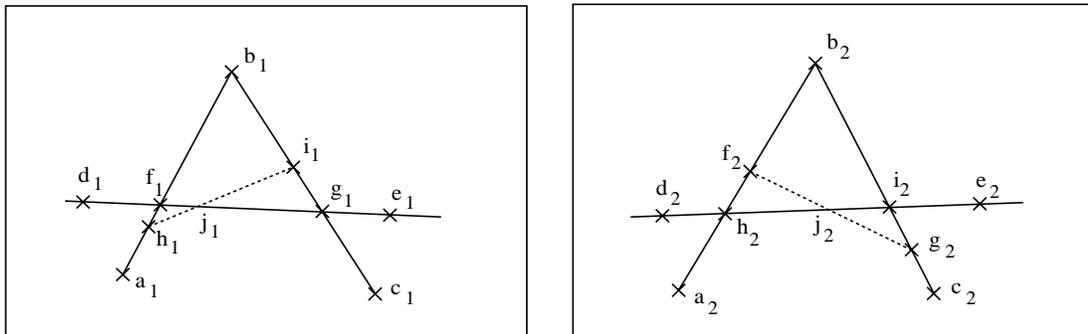


FIG. A.3 - Deuxième méthode de construction de points coplanaires : les deux images.

On peut construire de manière similaire  $H_2$  et  $I_2$ , les projections dans la deuxième image des points  $H$  et  $I$ , intersections de  $(AB)$  et  $(BC)$  avec le plan d'interprétation de  $(DE)$  par rapport à cette deuxième image. A l'aide de la géométrie épipolaire, on peut alors construire  $h_1, i_1, f_2$  et  $g_2$ . Par exemple,  $h_1$  est l'intersection de la droite épipolaire associée à  $h_2$  dans la première image avec la droite  $(a_1b_1)$ . Comme  $J = (FG) \cap (HI)$ ,

on a  $j_1 = (f_1 g_1) \cap (h_1 i_1)$  et  $j_2 = (f_2 g_2) \cap (h_2 i_2)$ . Ce sont les points que nous cherchons.

### A.2.2 Calcul vectoriel

Les calculs correspondants peuvent être effectués très simplement dans un formalisme vectoriel si l'on confond les points et les droites avec les vecteurs tridimensionnels de leurs coordonnées projectives. La droite définie par deux points ou le point intersection de deux droites se calculent alors par un simple produit vectoriel. Par exemple, la droite  $(d_1 e_1)$  a pour coordonnées  $d_1 \wedge e_1$ .

On obtient alors successivement, en notant  $F$  la matrice fondamentale :

$$\begin{aligned} f_1 &= (d_1 \wedge e_1) \wedge (b_1 \wedge a_1) & \text{et} & & f_2 &= (b_2 \wedge a_2) \wedge F.f_1 \\ g_1 &= (d_1 \wedge e_1) \wedge (b_1 \wedge c_1) & \text{et} & & g_2 &= (b_2 \wedge c_2) \wedge F.g_1 \\ h_2 &= (d_2 \wedge e_2) \wedge (b_2 \wedge a_2) & \text{et} & & h_1 &= (b_1 \wedge a_1) \wedge {}^t F.h_2 \\ i_2 &= (d_2 \wedge e_2) \wedge (b_2 \wedge c_2) & \text{et} & & i_1 &= (b_1 \wedge c_1) \wedge {}^t F.i_2 \end{aligned}$$

et finalement :

$$\begin{aligned} j_1 &= (h_1 \wedge i_1) \wedge (f_1 \wedge g_1) = (h_1 \wedge i_1) \wedge (d_1 \wedge e_1) \\ j_2 &= (f_2 \wedge g_2) \wedge (h_2 \wedge i_2) = (f_2 \wedge g_2) \wedge (d_2 \wedge e_2) \end{aligned} \quad (\text{A.20})$$

### Une variante

On peut mener les calculs de manière légèrement différente en remarquant que, par l'intermédiaire du plan  $(ABC)$ , il existe une homographie  $T$  entre les deux images. Cette homographie met en correspondance les points  $a_1, b_1, c_1$  et  $e_1$  avec les points  $a_2, b_2, c_2$  et  $e_2$ . Soit  $L$  et  $X$  une droite et un point du plan  $(ABC)$ , et  $l_1, l_2, x_1$  et  $x_2$  leurs projections dans chacune des deux images. On a :

$$\forall X \in L \quad {}^t_2 x_2 = {}^t_2 T.x_1 = 0$$

On en déduit  ${}^t_1 = {}^t_2 T$  et  $l_2 = {}^t T^{-1} l_1$ .

En reprenant nos notations, posons :  $l_1 = (d_1 e_1)$  et  $l_2 = (d_2 e_2)$ . On a alors, dans la deuxième image :

$$j_2 = l_2 \wedge ({}^t T^{-1} l_1)$$

Cette expression est, en fait analogue à celle que nous avons déjà obtenu. En effet, on a :

$$\begin{aligned} l_2 &= d_2 \wedge e_2 \\ l_1 &= d_1 \wedge e_1 \end{aligned}$$

et

$$f_2 \wedge g_2 = (T f_1) \wedge (T g_1) = (\det T). {}^t T^{-1} (f_1 \wedge g_1) = (\det T). {}^t T^{-1} l_1$$

Les coordonnées projectives étant définies à un facteur multiplicatif près, on a retrouvé l'expression de l'équation (A.20). La seule différence que peut introduire cette variante vient de la manière dont est évaluée l'homographie  $T$  : évaluation implicite dans le premier calcul, évaluation explicite dans cette variante.

### A.2.3 Calcul analytique

Comme précédemment, on suppose que les images sont rectifiées. La matrice fondamentale est donc :

$$F = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

On prend des notations semblables au paragraphe précédent pour les coordonnées des points. Le calcul suit exactement les étapes de la méthode. On construit d'abord  $f_1 = (a_1b_1) \cap (d_1e_1)$  et  $g_1 = (c_1b_1) \cap (d_1e_1)$  :

$$\begin{aligned} x_{f_1} &= (y_{a_1}x_{b_1} - x_{a_1}y_{b_1})(x_{e_1} - x_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(x_{b_1} - x_{a_1}) \\ y_{f_1} &= (y_{a_1}x_{b_1} - x_{a_1}y_{b_1})(y_{e_1} - y_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(y_{b_1} - y_{a_1}) \\ t_{f_1} &= (x_{b_1} - x_{a_1})(y_{e_1} - y_{d_1}) - (x_{e_1} - x_{d_1})(y_{b_1} - y_{a_1}) \\ \\ x_{g_1} &= (y_{c_1}x_{b_1} - x_{c_1}y_{b_1})(x_{e_1} - x_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(x_{b_1} - x_{c_1}) \\ y_{g_1} &= (y_{c_1}x_{b_1} - x_{c_1}y_{b_1})(y_{e_1} - y_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(y_{b_1} - y_{c_1}) \\ t_{g_1} &= (x_{b_1} - x_{c_1})(y_{e_1} - y_{d_1}) - (x_{e_1} - x_{d_1})(y_{b_1} - y_{c_1}) \end{aligned}$$

Les coordonnées de  $h_2$  et  $i_2$  sont obtenues par simple changement de l'indice 1 en 2 dans les formules précédentes.

Le point  $f_2$  est le point d'intersection de la droite  $(a_2b_2)$  avec la droite épipolaire associée au point  $f_1$ . Les coordonnées de cette dernière droite sont obtenues en faisant le produit de la matrice fondamentale par le vecteur des coordonnées de  $f_1$ . On obtient alors :

$$\begin{aligned} x_{f_2} &= (y_{a_2}x_{b_2} - x_{a_2}y_{b_2})((x_{b_1} - x_{a_1})(y_{e_1} - y_{d_1}) - (x_{e_1} - x_{d_1})(y_{b_1} - y_{a_1})) \\ &\quad - (x_{a_2} - x_{b_2})((y_{a_1}x_{b_1} - x_{a_1}y_{b_1})(y_{d_1} - y_{e_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(y_{a_1} - y_{b_1})) \\ y_{f_2} &= (y_{b_2} - y_{a_2})((y_{a_1}x_{b_1} - x_{a_1}y_{b_1})(y_{d_1} - y_{e_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(y_{a_1} - y_{b_1})) \\ t_{f_2} &= -(y_{b_2} - y_{a_2})((x_{b_1} - x_{a_1})(y_{e_1} - y_{d_1}) - (x_{e_1} - x_{d_1})(y_{b_1} - y_{a_1})) \end{aligned}$$

De même pour  $g_2$  :

$$\begin{aligned} x_{g_2} &= (y_{c_2}x_{b_2} - x_{c_2}y_{b_2})((x_{b_1} - x_{c_1})(y_{e_1} - y_{d_1}) - (x_{e_1} - x_{d_1})(y_{b_1} - y_{c_1})) \\ &\quad - (x_{c_2} - x_{b_2})((y_{c_1}x_{b_1} - x_{c_1}y_{b_1})(y_{d_1} - y_{e_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(y_{c_1} - y_{b_1})) \\ y_{g_2} &= (y_{b_2} - y_{c_2})((y_{c_1}x_{b_1} - x_{c_1}y_{b_1})(y_{d_1} - y_{e_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(y_{c_1} - y_{b_1})) \\ t_{g_2} &= -(y_{b_2} - y_{c_2})((x_{b_1} - x_{c_1})(y_{e_1} - y_{d_1}) - (x_{e_1} - x_{d_1})(y_{b_1} - y_{c_1})) \end{aligned}$$

Pour les points  $h_1$  et  $i_1$ , on utilise le même procédé, mais en remarquant qu'il faut utiliser la transposée de la matrice fondamentale pour obtenir les coordonnées des droites épipolaires dans la première image des points de la deuxième image. Cela donne pour  $h_1$  :

$$\begin{aligned} x_{h_1} &= (y_{a_1}x_{b_1} - x_{a_1}y_{b_1})((x_{b_2} - x_{a_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{a_2})) \\ &\quad - (x_{a_1} - x_{b_1})((y_{a_2}x_{b_2} - x_{a_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{a_2})) \\ y_{h_1} &= (y_{b_1} - y_{a_1})((y_{a_2}x_{b_2} - x_{a_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{a_2})) \\ t_{h_1} &= -(y_{b_1} - y_{a_1})((x_{b_2} - x_{a_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{a_2})) \end{aligned}$$

et pour  $i_1$  :

$$\begin{aligned} x_{i_1} &= (y_{c_1}x_{b_1} - x_{c_1}y_{b_1})((x_{b_2} - x_{c_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{c_2})) \\ &\quad - (x_{c_1} - x_{b_1})((y_{c_2}x_{b_2} - x_{c_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{c_2})) \\ y_{i_1} &= (y_{b_1} - y_{c_1})((y_{c_2}x_{b_2} - x_{c_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{c_2})) \\ t_{i_1} &= -(y_{b_1} - y_{c_1})((x_{b_2} - x_{c_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{c_2})) \end{aligned}$$

Les points  $j_1$  et  $j_2$  sont alors définis par intersection de droites :

$$j_1 = (f_1g_1) \cap (h_1i_1) \quad \text{et} \quad j_2 = (f_2g_2) \cap (h_2i_2)$$

Cela donne :

$$\begin{aligned} x_{j_1} &= (x_{g_1}y_{f_1} - x_{f_1}y_{g_1})(x_{i_1}t_{h_1} - x_{h_1}t_{i_1}) - (y_{h_1}x_{i_1} - x_{h_1}y_{i_1})(x_{g_1}t_{f_1} - x_{f_1}t_{g_1}) \\ &= (((y_{a_1}x_{b_1} - x_{a_1}y_{b_1})(y_{e_1} - y_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(y_{b_1} - y_{a_1})) \\ &\quad ((y_{c_1}x_{b_1} - x_{c_1}y_{b_1})(x_{e_1} - x_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(x_{b_1} - x_{c_1})) \\ &\quad - ((y_{a_1}x_{b_1} - x_{a_1}y_{b_1})(x_{e_1} - x_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(x_{b_1} - x_{a_1})) \\ &\quad ((y_{c_1}x_{b_1} - x_{c_1}y_{b_1})(y_{e_1} - y_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(y_{b_1} - y_{c_1}))) \\ &\quad (((y_{a_1}x_{b_1} - x_{a_1}y_{b_1})((x_{b_2} - x_{a_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{a_2})) \\ &\quad - (x_{a_1} - x_{b_1})((y_{a_2}x_{b_2} - x_{a_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{a_2}))) \\ &\quad (y_{b_1} - y_{c_1})((x_{b_2} - x_{c_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{c_2})) \\ &\quad - ((y_{c_1}x_{b_1} - x_{c_1}y_{b_1})((x_{b_2} - x_{c_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{c_2})) \\ &\quad - (x_{c_1} - x_{b_1})((y_{c_2}x_{b_2} - x_{c_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{c_2}))) \\ &\quad (y_{b_1} - y_{a_1})(((x_{b_2} - x_{a_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{a_2}))) \\ &\quad - ((y_{b_1} - y_{a_1})((y_{a_2}x_{b_2} - x_{a_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{a_2}))) \\ &\quad ((y_{c_1}x_{b_1} - x_{c_1}y_{b_1})((x_{b_2} - x_{c_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{c_2})) \\ &\quad - (x_{c_1} - x_{b_1})((y_{c_2}x_{b_2} - x_{c_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{c_2}))) \\ &\quad - ((y_{a_1}x_{b_1} - x_{a_1}y_{b_1})((x_{b_2} - x_{a_2})(y_{d_2} - y_{e_2}) - (x_{d_2} - x_{e_2})(y_{b_2} - y_{a_2})) \\ &\quad - (x_{a_1} - x_{b_1})((y_{a_2}x_{b_2} - x_{a_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{a_2}))) \\ &\quad (y_{b_1} - y_{c_1})((y_{c_2}x_{b_2} - x_{c_2}y_{b_2})(y_{e_2} - y_{d_2}) - (y_{d_2}x_{e_2} - x_{d_2}y_{e_2})(y_{b_2} - y_{c_2}))) \\ &\quad (((y_{c_1}x_{b_1} - x_{c_1}y_{b_1})(x_{e_1} - x_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(x_{b_1} - x_{c_1})) \\ &\quad ((x_{b_1} - x_{a_1})(y_{e_1} - y_{d_1}) - (x_{e_1} - x_{d_1})(y_{b_1} - y_{a_1})) \\ &\quad - ((y_{a_1}x_{b_1} - x_{a_1}y_{b_1})(x_{e_1} - x_{d_1}) - (y_{d_1}x_{e_1} - x_{d_1}y_{e_1})(x_{b_1} - x_{a_1})) \\ &\quad ((x_{b_1} - x_{c_1})(y_{e_1} - y_{d_1}) - (x_{e_1} - x_{d_1})(y_{b_1} - y_{c_1}))) \end{aligned}$$

$$y_{j_1} = (y_{f_1}t_{g_1} - y_{g_1}t_{f_1})(y_{h_1}x_{i_1} - x_{h_1}y_{i_1}) - (y_{h_1}t_{i_1} - y_{i_1}t_{h_1})(y_{f_1}x_{g_1} - x_{f_1}y_{g_1})$$

$$t_{j_1} = (y_{f_1}t_{g_1} - y_{g_1}t_{f_1})(x_{i_1}t_{h_1} - x_{h_1}t_{i_1}) - (y_{h_1}t_{i_1} - y_{i_1}t_{h_1})(x_{g_1}t_{f_1} - x_{f_1}t_{g_1})$$

$$x_{j_2} = (x_{g_2}y_{f_2} - x_{f_2}y_{g_2})(x_{i_2}t_{h_2} - x_{h_2}t_{i_2}) - (y_{h_2}x_{i_2} - x_{h_2}y_{i_2})(x_{g_2}t_{f_2} - x_{f_2}t_{g_2})$$

$$y_{j_2} = (y_{f_2}t_{g_2} - y_{g_2}t_{f_2})(y_{h_2}x_{i_2} - x_{h_2}y_{i_2}) - (y_{h_2}t_{i_2} - y_{i_2}t_{h_2})(y_{f_2}x_{g_2} - x_{f_2}y_{g_2})$$

$$t_{j_2} = (y_{f_2}t_{g_2} - y_{g_2}t_{f_2})(x_{i_2}t_{h_2} - x_{h_2}t_{i_2}) - (y_{h_2}t_{i_2} - y_{i_2}t_{h_2})(x_{g_2}t_{f_2} - x_{f_2}t_{g_2})$$

Les dernières substitutions sont laissées au lecteur... Notons que, si les formules sont assez longues, les calculs effectués sont très simples : multiplication d'une matrice  $3 \times 3$  par un vecteur de dimension 3, et intersection de deux droites du plan.

### **A.3 Conclusion sur ces méthodes**

La deuxième méthode possède plusieurs avantages par rapport à la première.

- Tout d'abord, on obtient une expression polynomiale et non rationnelle, ce qui est un atout de stabilité numérique.
- Deuxièmement, elle est constructive. Les points recherchés sont construits géométriquement : cela permet de vérifier à chaque étape la stabilité des calculs effectués. Dans la première méthode, toutes les difficultés sont concentrées dans l'équation à résoudre.
- Enfin, elle ne donne qu'une solution. Dans la première méthode, si on ne mène pas tous les calculs de manière symbolique, les deux solutions peuvent être difficiles à différencier, à cause du bruit qui affecte la position des points par exemple.

Notons toutefois que si les formules diffèrent, c'est aussi parce que les hypothèses de départ ne sont pas exactement les mêmes. Dans un cas la droite est définie par ses coordonnées, dans l'autre cas par deux points qu'elle contient.

# Bibliographie de l'auteur

## Articles de revues

- P. Gros. Modélisation d'images en vision par ordinateur. Dans *Techniques et Sciences Informatiques*, volume 11, numéro 4, pages 43 à 68, 1992.
- L. Quan, P. Gros et R. Mohr. Invariants of a pair of conics revisited. Dans *International Journal "Image and Vision Computing"*, volume 10, numéro 5, pages 319 à 323, 1992.

## Chapitre de livre

- J.L. Mundy, D. Kapur, S.J. Maybank, P. Gros et L. Quan. Geometric interpretation of joint conic invariant. Dans *Geometric Invariance in Computer Vision*. J. Mundy et A. Zisserman éditeurs, pages 77–86, MIT Press 1992.

## Communications dans des congrès internationaux

- P. Gros, R. Horaud et R. Mohr. How to recover line drawings from raw data. Dans *Application of Invariance in Computer Vision*, Colloque DARPA - CCE, pages 259–271, Reykjavik, Islande, mars 1991.
- P. Gros et R. Mohr. Automatic object modelization in computer vision. Dans *Proceedings of the workshop "Advances in Structural and Syntactic Pattern Recognition"*, Berne, Suisse, août 1992, dans *Series on Machine Perception and Artificial Intelligence*, volume 5, pages 385–400, H. Bunke éditeur.
- P. Gros et L. Quan. Projective invariants for vision. A paraître dans *Geometric Methods in Computer Vision II - SPIE 1993 International Symposium on Optical Applied Science and Engineering*, San Diego, Californie, USA, juillet 1993.

**Communications dans des congrès nationaux**

- P. Gros et P. Rolin. Gestion d'un agenda électronique par mail. Dans *Actes des Conférences Techniques de la Convention Unix 90*, Association Française des Utilisateurs d'Unix, pages 63–75, CNIT La défense, mars 1990.
- L. Quan, P. Gros et R. Mohr. Invariants of a pair of conics revisited. Dans *Proceedings of the British Machine Vision Conference*, pages 71–77, Glasgow, Ecosse, septembre 1991.
- P. Gros. Modélisation d'images par appariement et correction. Dans *Actes du congrès Reconnaissance des Formes et Intelligence Artificielle*, AFCET-RFIA, Tome 3, pages 1335–1344, Villeurbanne, novembre 1991.

**Rapports techniques**

- P. Gros et L. Quan. Présentation de la théorie des invariants sous une forme utilisable en vision par ordinateur. Rapport technique RT 69 IMAG - 7 LIFIA, juillet 1991.
- P. Gros. Modélisation d'images par appariement et correction. Rapport technique RT 72 IMAG - 8 LIFIA, septembre 1991.
- F. Veillon, E. Arbogast et P. Gros. Curve Matching using Cubic B-Splines. Rapport technique RT 85 IMAG - 13 LIFIA, juin 1992.
- P. Gros et L. Quan. Projective Invariants for Vision. Rapport technique RT 90 IMAG - 15 LIFIA, décembre 1992.

**Exposés lors de séminaires**

- Séminaire Robotique et vision du LIFIA, Grenoble, juillet 1990.
- Séminaire LIFIA Grenoble – CRIN Nancy sur la géométrie, Chamrousse, avril 1991.
- Journées nationales du projet ORASIS du PRC Communication Homme–Machine, Nancy, mai 1991.
- ESPRIT-BRA No. 3274 “FIRST”, annual review, Bruxelles, octobre 1991.
- Séminaire du projet MOVI de l'INRIA, Saint Pierre de Chartreuse, Septembre 1992.

**Film**

- P. Gros et J. Rey. Hélioséismologie. Film produit par le Service Audio-visuel de l'Ecole Polytechnique, juillet 1988.

# Bibliographie

- [Alb 11] L.B. Alberti. *Trattato della Pittura*. Nuremberg, 1511.
- [Alo 88] J. Aloimonos, I. Weiss, et A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, pages 333–356, 1988.
- [Ana 89] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.
- [Arb 91a] E. Arbogast. La représentation des contours et leur segmentation. *Journées Nationales du PRC Communication Homme Machine*, pages 453–464, Janvier 1991.
- [Arb 91b] E. Arbogast. *Modélisation automatique d’objets non polyédriques par observation monoculaire*. Thèse de doctorat, Institut National Polytechnique de Grenoble, France, 1991.
- [Arb 91c] E. Arbogast et R. Mohr. 3D structures inference from images sequences. *International Journal of Pattern Recognition and Artificial Intelligence*, 5(5):749, 1991.
- [Åst 92] K. Åström et L. Morin. Random cross ratios. Rapport no. RT 88 IMAG - 14 LIFIA, IRIMAG-LIFIA, Octobre 1992.
- [Aya 85] N. Ayache et B. Faverjon. Un algorithme rapide de stéréoscopie passive utilisant la prédiction et la vérification d’hypothèses. *Actes du 5ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Grenoble, France*, pages 1217–1232, 1985.
- [Aya 89] N. Ayache. *Vision Stéréoscopique et Perception Multisensorielle: Application à la Robotique Mobile*. Science Informatique. InterEditions, 1989.

- [Bam 86] B. Bamieh et R.J.P. de Figueiredo. A general moment-invariants/attributed-graph method for three-dimensional object-recognition from a single image. *IEEE Journal of Robotics and Automation*, 2(1):31–41, 1986.
- [Bar 86] D. Barba et J.F. Bertrand. Automatic region construction by edge detection and contour following in image segmentation. *Proceedings of the 8th International Conference on Pattern Recognition, Paris, France*, pages 681–683, 1986.
- [Ber 90] M. Berger. *Géométrie*. Nathan, 1990.
- [Bes 88] P.J. Besl. Active optical range imaging sensors. Springer-Verlag, New York, USA, 1988.
- [Bey 92] H.A. Beyer. Accurate calibration of CCD cameras. *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 96–101, 1992.
- [Bha 87] B. Bhanu. Guest editor's introduction. *Computer (Special Issue on CAD-Based Robot Vision)*, Août 1987.
- [Bol 86] R.C. Bolles et R. Horaud. 3DPO: A three-Dimensional Part Orientation system. *The International Journal of Robotics Research*, 5(3):3–26, 1986.
- [Bow 91] K. Bowyer. Why aspect graphs are not (yet) practical for computer vision. *Proceedings of the IEEE workshop on Direction on automated CAD-based Vision, Maui, Hawaii, USA*, pages 97–104, 1991.
- [Bra 92] P. Brand. Corrections des Distorsions. Rapport à paraître, IRIMAG–LIFIA, Grenoble, France, 1992.
- [Bri 83] M.H. Brill et E.B. Barret. Closed-form extension of the anharmonic ratio to N-space. *Computer Vision, Graphics and Image Processing*, 23:92–98, 1983.
- [Bro 71] D.C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, 1971.
- [Bur 88] J.B. Burns et L.J. Kitchen. Rapid object recognition from a large model base using prediction hierarchies. *Proceedings of the 1988 DARPA Image Understanding Workshop*, pages 711–719, 1988.
- [Bur 92] J.B. Burns, R.S. Weiss, et E.M. Riseman. The non-existence of general-case view-invariants. J.L. Mundy et A. Zisserman, éditeurs, *Geometric Invariance in Computer Vision*, chapitre 6, pages 120–131. MIT Press, 1992.
- [Can 86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [Cha 91] H. Chabbi et G. Masini. A combined use of regions and segments to construct facets. *Proceedings of the 6th International Conference on Image Analysis and Processing, Como, Italie, Septembre 1991*.

- [Che 91] C.H. Chen et P.G. Mulgaonkar. CAD-based feature-utility measures for automatic vision programming. *Direction in automated CAD-based vision*, pages 106–114. IEEE computer society, 1991.
- [Cip 92] R. Cipolla et A. Blake. Surface orientation and time to contact image divergence and deformation. G. Sandini, éditeur, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italie*, pages 187–202. Springer-Verlag, Mai 1992.
- [Coc 87] J.P. Cocquerez et A. Gagalowicz. Mise en correspondance de régions dans une paire d'images stéréo. *Actes du Congrès MARI 87, Paris, France*, pages 165–173, Tome 2, 1987.
- [Coc 90] S.D. Cochran. Surface description from binocular stereo. Rapport, University of Southern California, Los Angeles, Californie, USA, Novembre 1990.
- [Com 41] H. Commissaire et G. Cagnac. *Cours de Mathématiques Spéciales*. Masson, 1941.
- [Con 87] J.H. Connell et M. Brady. Generating and generalizing models of visual objects. *Artificial Intelligence*, 31:159–183, 1987.
- [Con 90] P. Constant, S. Matwin, et F. Oppacher. LEW: Learning by Watching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):294–308, Mai 1990.
- [Cro 90] J.L. Crowley et P. Stelmazyk. Measurement and integration of 3D structures by tracking edges lines. O. Faugeras, éditeur, *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 269–280. Springer-Verlag, Avril 1990.
- [DD 86] A. Dahan-Dalmedico et J. Peiffer. *Une Histoire des Mathématiques. Routes et Dédales*. Points Sciences. Editions du seuil, 1986.
- [Dem 87] M. Demazure. *Géométrie: Catastrophes et Bifurcations*. Ecole Polytechnique, 1987.
- [Der 87a] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.
- [Der 87b] R. Deriche et J.P. Cocquerez. Extraction de composantes connexes basée sur une détection optimale des contours. *Actes du Congrès MARI 87, Paris, France*, pages 1–9, Tome 2, 1987.
- [Der 90a] R. Deriche et O. Faugeras. Tracking line segments. *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 259–267. Springer-Verlag, Avril 1990.
- [Der 90b] R. Deriche et G. Giraudon. Accurate corner detection: an analytical study. *Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japon*, 1990.

- [Des 37] R. Descartes. *Géométrie*, 1637.
- [Des 39] G. Desargues. Brouillon Project d'une Atteinte aux Evénemens des Rencontres du Cône avec un Plan, 1639.
- [Did 76] E. Diday et J.C. Simon. Clustering Analysis. K.S. Fu, éditeur, *Communication and Cybernetics*. Springer-Verlag, 1976.
- [Dix 90] J. Dixmier. Quelques aspects de la théorie des invariants. *Gazette des Mathématiciens*, 43:39–64, Janvier 1990.
- [Ett 88] G.J. Ettinger. Large hierarchical object recognition using librairies of parameterized model sub-parts. *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, Californie, USA*, pages 32–41, Juin 1988.
- [Fau 92a] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? G. Sandini, éditeur, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italie*, pages 563–578. Springer-Verlag, Mai 1992.
- [Fau 92b] O.D. Faugeras. *3D Computer Vision*. MIT Press, 1992.
- [Fau 92c] O.D. Faugeras, Q.T. Luong, et S.J. Maybank. Camera Self-Calibration: Theory and Experiments. G. Sandini, éditeur, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italie*, pages 321–334. Springer-Verlag, Mai 1992.
- [Fav 57] J. Favard. *Cours de Géométrie Différentielle Locale*. Gauthiers-Villard, 1957.
- [Fly 91] P.J. Flynn et A.K. Jain. CAD-based computer vision: from CAD models to relational graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):114–132, Février 1991.
- [For 69] J. Forgarty. *Invariant Theory*. Benjamin, New York, USA, 1969.
- [For 82] J. Forgarty et D. Mummford. *Geometric Invariant Theory*. Springer-Verlag, 1982.
- [For 90] D. Forsyth, J. Mundy, et A. Zisserman. Transformational invariance - a primer. *Proceedings of the British Machine Vision Conference, Oxford, Angleterre*, pages 1–6, Septembre 1990.
- [For 91a] D. Forsyth, J.L. Mundy, A. Zisserman, et C. Rothwell. Invariant descriptors for 3D object recognition and pose. *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Reykjavik, Islande*, pages 171–208, Mai 1991.
- [For 91b] D. Forsyth, J.L. Mundy, A. Zisserman, C. Coelho, A. Heller, et C. Rothwell. Invariant descriptors for 3D object recognition and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):971–991, Octobre 1991.

- [Fra 90] E. Francois et P. Bouthemy. The derivation of qualitative information in motion analysis. O. Faugeras, éditeur, *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 226–230. Springer-Verlag, Avril 1990.
- [Fua 90] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 1990.
- [Goa 86] C. Goad. Fast 3D model-based vision. A.P. Pentland, éditeur, *From Pixels to Predicates*, pages 371–391. Ablex, Norwood, NJ, 1986.
- [Goo 91] L. Van Gool, P. Kempenaers, et A. Oosterlinck. Recognition and semi-differential invariants. *Proceedings of the Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, USA*, pages 454–460, 1991.
- [Goo 92a] L. Van Gool, M. Brill, E. Barret, T. Moons, et E. Pauwels. Semi-differential invariants for nonplanar curves. J. Mundy et A. Zisserman, éditeurs, *Geometric Invariance in Computer Vision*, pages 293–309. MIT Press, 1992.
- [Goo 92b] L. Van Gool, L.T. Moons, E. Pauwels, et A. Oosterlinck. Semi-Differential Invariants. J. Mundy et A. Zisserman, éditeurs, *Geometric Invariance in Computer Vision*, pages 157–192. MIT Press, 1992.
- [Gro 92a] P. Gros et R. Mohr. Automatic object modelization in computer vision. H. Bunke, éditeur, *Proceedings of the workshop "Advances in Structural and Syntactic Pattern Recognition", Bern, Switzerland*, volume 5, série *Series on Machine Perception and Artificial Intelligence*, pages 385–400. World Scientific, Août 1992.
- [Gro 92b] P. Gros et L. Quan. Projective Invariants for Vision. Rapport no. RT 90 IMAG - 15 LIFIA, IRIMAG-LIFIA, Grenoble, France, Décembre 1992.
- [Gru 85] A.W. Gruen. Adaptive least squares correlation: a powerful image matching technique. *S. Afr. Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3):175–187, 1985.
- [Har 92a] R. Hartley, R. Gupta, et T. Chang. Stereo from uncalibrated cameras. *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.
- [Har 92b] R.I. Hartley. Estimation of relative camera positions for uncalibrated cameras. G. Sandini, éditeur, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italie*, pages 579–587. Springer-Verlag, 1992.
- [Hér 91] L. Hérault. *Réseaux de neurones récurrents pour l'optimisation combinatoire*. Thèse de doctorat, Institut National Polytechnique de Grenoble, France, 1991.
- [Hor 89] R. Horaud, B. Conio, O. Le Boulleux, et B. Lacolle. An analytic solution for the perspective 4-point problem. *Computer Vision, Graphics and Image Processing*, 47:33–44, 1989.

- [Hor 90] R. Horaud, T. Skordas, et F. Veillon. Finding geometric and relational structures in an image. *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 374–384. Springer-Verlag, Avril 1990.
- [Hor 93] R. Horaud et O. Monga. *Vision par Ordinateur: Outils Fondamentaux*. Editions Hermès, Paris, 1993.
- [Hu 62] M.K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8:179–187, 1962.
- [Ike 88] K. Ikeuchi et T. Kanade. Applying sensor models to automatic generation of object recognition programs. *Proceedings of the 2nd International Conference on Computer Vision, Tampa, Floride, USA*, pages 228–237, 1988.
- [Kak 88] A. Kak, A. Vayda, R. Cromwell, W. Kim, et C. Chen. Knowledge based robotics. *International Journal of Production Research*, 26(5):707–734, 1988.
- [Kal 86] A. Kalvin, E. Schomberg, J.T. Schwartz, et M. Sharir. Two-dimensional model-based boundary matching using footprints. *The International Journal of Robotics Research*, 5(4):38–54, 1986.
- [Kap 91] D. Kapur et J.L. Mundy. Fitting affine invariant conics to curves. *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Reykjavik, Islande*, pages 209–233, Mai 1991.
- [Kle 74] F. Klein. *Le Programme d'Erlangen*. Collection “Discours de la méthode”. Gauthier-Villars, Paris, 1974.
- [Koe 89] J.J. Koenderink et A.J. van Doorn. Affine structure from motion. Rapport, Utrecht University, Pays-Bas, Octobre 1989.
- [Kro 87] E. Krotkov. Focusing. *International Journal of Computer Vision*, 1(3):223–237, Octobre 1987.
- [Lam 88] Y. Lamdan et H.J. Wolfson. Geometric hashing: a general and efficient model-based recognition scheme. *Proceedings of the 2nd International Conference on Computer Vision, Tampa, Floride, USA*, pages 238–249, 1988.
- [Lam 90] Y. Lamdan, J.T. Schwartz, et H.J. Wolfson. Affine invariant model-based object recognition. *IEEE Journal of Robotics and Automation*, 6:578–589, 1990.
- [Lav 91] J.M. Lavest, G. Rives, et M. Dhome. Utilisation d'un objectif à focale variable en vision monoculaire en vue de la reconstruction 3D. *Actes du 8ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Lyon - Villeurbanne, France*, volume 1, pages 293–301. AFCET, Novembre 1991.
- [Lee 90] C.H. Lee et T. Huang. Finding point correspondences and determining motion of a rigid object from two weak perspective views. *Computer Vision, Graphics and Image Processing*, 52:309–327, 1990.

- [Lee 92] S.W. Lee et R. Bajcsy. Detection of specularity using color and multiple views. G. Sandini, éditeur, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italie*, pages 99–114. Springer-Verlag, Mai 1992.
- [Len 87] R.K. Lenz et R.Y. Tsai. Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. *Proceedings of IEEE International Conference on Robotics and Automation, Raleigh, USA*, pages 68–75, 1987.
- [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.
- [Liq 89] M. Liquière et J. Sallantin. INNE : a structural learning algorithm for noisy examples. *Proceedings of the fourth European Working Session on Learning, Montpellier, France*, pages 111–123, Décembre 1989.
- [Liu 90] Y. Liu, T.S. Huang, et O.D. Faugeras. Determination of camera location from 2D to 3D line and point. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):28–37, Janvier 1990.
- [Luo 92] Q.T. Luong. *Matrice Fondamentale et Autocalibration en Vision par Ordinateur*. Thèse de doctorat, Université de Paris-Sud, Orsay, France, Décembre 1992.
- [Lus 87] F. Lustman. *Vision Stéréoscopique et Perception du Mouvement en Vision Artificielle*. Thèse de doctorat, Université Paris-Sud, Orsay, France, 1987.
- [Mar 78] D. Marr. Representing visual information — A computational approach. *Computer Vision Systems*, pages 61–80, 1978.
- [Mar 82] D. Marr. *Vision*. W.H. Freeman and Company, San Francisco, Californie, USA, 1982.
- [Mat 89] L. Matthies, T. Kanade, et R. Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3:209–236, 1989.
- [May 91] S.J. Maybank. The projection of two non-coplanar conics. *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Reykjavik, Islande*, pages 47–53, Mai 1991.
- [May 92] S.J. Maybank et O.D. Faugeras. A theory of self calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.
- [Mne 86] R. Mneimné et F. Testard. *Introduction à la Théorie des Groupes de Lie Classiques*. Collection Méthodes. Hermann, Paris, 1986.
- [Möb 85] A.F. Möbius. *Gesammelte Werke*, volume 1. Hirzel, Leipzig, 1885.
- [Moh 91a] R. Mohr et E. Arbogast. It can be done without camera calibration. *Pattern Recognition Letters*, 12:39–43, Janvier 1991.

- [Moh 91b] R. Mohr, L. Morin, C. Inglebert, et L. Quan. Geometric solutions to some 3D vision problems. R. Storer J.L. Crowley, E. Granum, éditeur, *Integration and Control in Real Time Active Vision*, ESPRIT BRA Series. Springer-Verlag, 1991.
- [Moh 92a] R. Mohr. Projective geometry and computer vision. S.P. Wang C.H. Chen, L.F.Pau, éditeur, *Handbook of Pattern Recognition and Computer Vision*. World Scientific Pub. Co., 1992.
- [Moh 92b] R. Mohr, L. Morin, et E. Grosso. Relative positioning with uncalibrated cameras. A. Zisserman J.L Mundy, éditeur, *Geometric Invariance in Computer Vision*, pages 440–460. MIT Press, 1992.
- [Moh 92c] R. Mohr, L. Quan, F. Veillon, et B. Boufama. Relative 3D reconstruction using multiples uncalibrated images. Rapport no. RT 84-I-IMAG LIFIA 12, IRIMAG–LIFIA, 1992.
- [Mon 87] O. Monga. An optimal region growing algorithm for image segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, 1(3):351–375, 1987.
- [Mon 88] O. Monga, B. Wrobel-Dautcourt, R. Deriche, G. Giraudon, et J. Shen. Segmentations d’images et robotique mobile. EC2, éditeur, *Actes des Journées Nationales du GRECO–PRC Communication Homme-Machine*, pages 203–230, Paris, France, Novembre 1988.
- [Mor 93] L. Morin. *Quelques Contributions des Invariants Projectifs à la Vision par Ordinateur*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, Janvier 1993.
- [Mou 91] B. Mourrain. *Approche Effective de la Théorie des Invariants des Groupes Classiques*. Thèse de doctorat, Ecole Polytechnique, Palaiseau, France, Septembre 1991.
- [Mun 92] J.L. Mundy et A. Zisserman, éditeurs. *Geometric Invariance in Computer Vision*. MIT Press, Cambridge, Massachusetts, USA, 1992.
- [Nay 90] S.K. Nayar, K. Ikeushi, et T. Kanade. Shape from interreflections. *Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japon*, pages 2–11, Décembre 1990.
- [Naz 84] A.M. Nazif et M.D. Levine. Low level image segmentation: an expert system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(5):555–577, 1984.
- [Nev 77] R. Nevatia et T. Binford. Description and recognition of complex curved objects. *Artificial intelligence*, 8:77–98, 1977.
- [Nie 91] L. Nielsen et G. Sparr. Projective area-invariants as an extension of the cross-ratio. *Computer Vision, Graphics and Image Processing: Image Understanding*, 54:145–159, 1991.

- [Nit 88] D. Nitzan. Three-dimensional vision structure for robot applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(3):291–309, 1988.
- [Par 91] S. Paris. Modélisation structurelle par apprentissage. *Actes du 8ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Lyon – Villeurbanne, France*, volume 3, pages 1345–1350. AFCET, 1991.
- [Pen 87] A. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, Juillet 1987.
- [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.
- [Pon 22] J.V. Poncelet. *Traité des Propriétés Projectives des Figures*, 1822.
- [Pon 90] J. Ponce et D. Kriegman. Computing exact aspect graphs of curved objects. Technical Report no. UIUC-DCS-R-90-1579, Dept. of computer science, University of Illinois, Urbana-Champaign, Illinois, USA, 1990.
- [Por 88] J. Porrill. Optimal combination and constraints for geometrical sensor data. *The International Journal of Robotics Research*, 7(6):66–77, Décembre 1988.
- [Pre 88] W.H. Press, B.P. Flannery, S.A. Teukolsky, et W.T. Vetterling W.T. *Numerical Recipes in C*. Cambridge University Press, 1988.
- [Ram 79] E. Ramis, C. Deschamps, et J. Odoux. *Cours de Mathématiques spéciales*. Masson, 1979.
- [Rot 92] C.A. Rothwell, A. Zisserman, D.A. Forsyth, et J.L. Mundy. Canonical frames for planar object recognition. G. Sandini, éditeur, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italie*, pages 757–772. Springer-Verlag, Mai 1992.
- [Rot 93] C.A. Rothwell, D.A. Forsyth, A. Zisserman, et J.L. Mundy. Extracting projective structure from single perspective views of 3D point sets. *Proceedings of the 4th International Conference on Computer Vision, Berlin, RFA*, pages 573–582. Mai 1993.
- [Sal 92] M. Salotti. Réflexions sur la Segmentation. Rapport de Recherche no. RR 873 -I-, IMAG-TIM3, Grenoble, France, Janvier 1992.
- [Sco 90] G.L. Scott et H.C. Longuet-Higgins. Feature grouping by "relocalisation" of eigenvectors of the proximity matrix. *Proceedings of the British Machine Vision Conference, Oxford, Angleterre*, pages 103–108, Septembre 1990.
- [Sem 52] J.G. Semple et G.T. Kneebone. *Algebraic Projective Geometry*. Oxford Science Publication, 1952.

- [Sha 85a] L.G. Shapiro. A fast structural matching algorithm with applications in stereo vision. *Proceedings of the 4th Scandinavian Conference on Image Analysis, Trondheim, Norvège*, pages 159–166, 1985.
- [Sha 85b] L.G. Shapiro et R. Haralick. A metric for comparing relational descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):90–94, 1985.
- [Shn 87] M.O. Shneier, R.Lumia, et M. Herman. Prediction based vision for robot control. *Computer*, 20(8):46–55, Août 1987.
- [Sko 88] T. Skordas. *Mise en correspondance et reconstruction stéréo utilisant une description structurelle des images*. Thèse de doctorat, Institut National Polytechnique de Grenoble, France, 1988.
- [Sos 92] H. Sossa. *Reconnaissance d'Objets Polyédriques dans une Base de Modèles*. Thèse de doctorat, Institut National Polytechnique de Grenoble, France, Décembre 1992.
- [Spa 90] G. Sparr et L. Nielsen. Shape and mutual cross-ratios with applications to the interior, exterior and relative orientation. *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 607–609. Springer-Verlag, Avril 1990.
- [Spa 91] G. Sparr. Projective invariants for affine shapes of point configurations. *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Reykjavik, Islande*, pages 151–170, Mai 1991.
- [Spa 92] G. Sparr. Depth computations from polyhedral images. G. Sandini, éditeur, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italie*, pages 378–386. Springer-Verlag, Mai 1992.
- [Ste 92] F. Stein et G. Medioni. Structural indexing: efficient 3-D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):125–145, 1992.
- [Sua 74] R. Suardet. *Optique*. Collection de Sciences Physiques. Editions J.-B. Baillière, Paris, France, 1974.
- [Tho 87] D.W. Thompson et J.L. Mundy. Three-dimensional model matching from an unconstrained viewpoint. *Proceedings of IEEE International Conference on Robotics and Automation, Raleigh, USA*, pages 208–220, 1987.
- [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.
- [Vai 89] R. Vaillant et O.D. Faugeras. Using occluding contours for recovering shape properties of objects. *IEEE Workshop on Interpretation of 3D scenes*, pages 26–32, Novembre 1989.

- [Vi 90] T. Viéville et O. Faugeras. Feed-forward recovery of motion and structure from a sequence of 2D-lines matches. *Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japon*, pages 517–520, Décembre 1990.
- [Wei 88] I. Weiss. Projective invariants of shapes. *Proceedings of DARPA Image Understanding Workshop, Cambridge, Massachusetts, USA*, pages 1125–1134, 1988.
- [Wei 91] I. Weiss. Noise-resistant invariant of curves. *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Reykjavik, Islande*, pages 319–344, 1991.
- [Wol 90] H.J. Wolfson. Model-based object recognition by geometric hashing. O. Faugeras, éditeur, *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 526–536. Springer-Verlag, Avril 1990.
- [Wro 87] B. Wrobel. Stéréovision : coopération entre l'extraction et la mise en correspondance symbolique des régions. *Actes du Congrès MARI 87, Paris, France*, pages 181–188, Tome 2, 1987.
- [Yua 89] J.S.C. Yuan. A general phogrammetric solution for the determining object position and orientation. *IEEE Transactions on Robotics and Automation*, 5(2):129–142, 1989.
- [Zha 89] J. Zhao. *Extraction d'information tri-dimensionnelle par stéréovision*. Thèse de doctorat, Université Paul Sabatier, Toulouse, Juillet 1989.
- [Zha 90a] Z. Zhang. *Analyse du mouvement d'une séquence stéréoscopique et ses applications*. Thèse de doctorat, Université de Paris XI, France, Octobre 1990.
- [Zha 90b] Z. Zhang et O. Faugeras. Building a 3D world representation with a mobile robot: 3D line segment representation and integration. *Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, USA*, pages 38–42, Juin 1990.
- [Zha 92] Z. Zhang et O. Faugeras. Three-dimensional motion computation and object segmentation in a long sequence of stereo frames. *International Journal of Computer Vision*, 7(3):211–241, 1992.
- [Zis 91] A. Zisserman et J.L. Mundy, éditeurs. *Proceedings of the DARPA – ESPRIT workshop on Applications of Invariants in Computer Vision, Reykjavik, Islande*, Mai 1991.



# Index

- $\overline{AB}$  : mesure algébrique, 42
- $\{a, b, c, d\}$  ; birapport de 4 nombres ou points, 31
- $\{a ; b, c, d, e\}$  : birapport de 5 points, 32
- absolu, 42
- angle euclidien, 43
- angle non euclidien, 46
- appariement d'images, 133
- aspects visuels d'un objet, 149, 162
- axe optique, 73, 75, 77
- axe principal, 77
  
- bas niveau, 14
- base canonique, 30
- base (changement de), 30
- base projective, 29
- birapport, 31
- birapport de 4 droites, 32
- birapport de 4 plans, 33
- birapport de 4 points, 32
- birapport de 5 droites, 33
- birapport de 5 points, 33
- birapports (comparaison des), 93
- birapports (distance entre), 94
- birapports sur une conique, 37
- bruit dans les images, 80, 156
- Burns (théorème de), 61
  
- CAD-based vision, 150
- CAO, 150
- centre optique, 73
  
- changement de base, 30
- Chasles (conique de), 38
- Chasles (théorème de), 37
- codimension, 55
- collinéation, 35
- comparaison des birapports, 93
- conique, 37
- conique de Chasles, 38
- conjugaison harmonique, 32, 38
- construction de points coplanaires, 90, 175
- coordonnées homogènes, 28
- coordonnées projectives, 33
- coplanaires (construction de points), 90, 175
- coplanarité, 84
- coplanarité (test de), 87
- correction d'images, 155, 157
- corrélation, 133
  
- dédoublément, 157
- dénombrément des invariants, 58
- dépendance linéaire, 29
- Desargues (théorème de), 30
- $\dim V$  : dimension d'une sous-variété, 55
- dimension, 55
- distance entre birapports, 94
- distance entre images, 160
- distance euclidienne, 44
- distance focale, 73, 75, 77
- distance non euclidienne, 46

- droite, 30
- droite épipolaire, 85
- dualité, 30
- éléments de l'infini, 29
- enveloppe de seconde classe, 37
- épipolaire (droite), 85
- épipolaire (géométrie), 85
- épipole, 85
- équianharmonique (quadruplet), 32
- Erlangen (programme d'), 25
- espace des paramètres, 134
- espace projectif, 28
- extrémité libre, 157
- faisceau de droites, 32
- faisceau de plans, 33
- focale (distance), 73, 75, 77
- fonction de répartition du birapport, 94
- fondamentale (matrice), 86
- formule de Möbius, 32, 90
- $GA(n, \mathbb{K})$ : groupe affine de l'espace de dimension  $n$  sur  $\mathbb{K}$ , 41
- géométrie affine, 41
- géométrie algébrique, 25, 44
- géométrie différentielle, 45
- géométrie épipolaire, 85
- géométrie euclidienne, 42
- géométrie non euclidienne, 45
- géométrie projective, 22
- géométrie vectorielle, 42
- $GP(n, \mathbb{K})$ : groupe projectif de l'espace de dimension  $n$  sur  $\mathbb{K}$ , 29, 36
- graphe d'aspects, 151
- groupe, 25, 53
- groupe affine, 41
- groupe de Lie, 55
- groupe linéaire, 42
- groupe non connexe, 57
- groupe opérant sur un ensemble, 53
- groupe (ordre d'un), 56
- groupe projectif, 29
- $Gx$ : orbite d'un élément, 54
- $G_x$ : stabilisateur d'un élément, 60
- harmonique (conjugaison), 32
- haut niveau, 14
- homogènes (coordonnées), 28
- homographie, 36
- Hough, 136
- hyperplan, 30
- hyperplan de l'infini, 41
- indépendance fonctionnelle, 59
- indépendance linéaire, 29
- indexation des modèles, 168
- infini (éléments de l'), 29
- infini (hyperplan de l'), 41
- interstice, 75
- invariant, 52
- invariant 2D de 2 points et 2 droites, 65
- invariant 3D de 2 paires de droites coplanaires, 116
- invariant 3D de 4 droites dont 2 coplanaires, 116
- invariant 3D de 4 droites quelconques, 117
- invariant 3D de 4 points et 1 droite, 115
- invariant différentiel, 45
- invariants (dénombrement des), 58
- invariants (théorie des), 49
- invariants 2D de 2 coniques, 64, 67
- invariants 3D, 113
- invariants 3D de 2 coniques, 120
- invariants 3D de 2 droites et 1 conique, 120
- invariants 3D de 2 points et 3 droites, 116
- invariants 3D de 2 points, 1 droite et 1 conique, 120
- invariants 3D de 3 points et 1 conique, 119
- invariants 3D de 3 points et 2 droites, 115
- invariants 3D de 6 points, 114
- isométrie, 42
- isotropie (sous-groupe d'), 60
- jonction, 160
- jonction en T, 158
- $\mathbb{K}$ : corps des réels ou des complexes, 25
- $\mathbb{K}^n$ : ensemble des  $n$ -uplets d'éléments de  $\mathbb{K}$ , 25

- $\mathbb{K}^{n*}$  : ensemble des  $n$ -uplets non nuls  
 d'éléments de  $\mathbb{K}$ , 25
- matrice associée à une conique, 37  
 matrice fondamentale, 86  
 mesure algébrique, 42  
 mesure de ressemblance, 160  
 méthode infinitésimale, 53  
 méthode par généralisation et contrainte, 63  
 méthode symbolique, 53  
 milieu de deux points, 42  
 mise en correspondance d'images, 133  
 Möbius (formule de), 32, 90  
 Möbius (théorème de), 35  
 modèle de projection affine, 78  
 modèle de projection perspective, 77  
 modélisation automatique, 149  
 modélisation des aspects, 167
- netteté (plan de), 74  
 niveau intermédiaire, 14
- $O(x)$  : orbite d'un élément, 54  
 optique, 73  
 optique (axe), 73, 75, 77  
 optique (centre), 73  
 optique (systèmes épais), 75  
 optique (systèmes minces), 73  
 orbite, 54  
 ordre d'un groupe, 56  
 orientation, 43  
 orthogonalité, 43
- $\mathbb{P}^n$  : espace projectif de dimension  $n$  sur  
 $\mathbb{R}$  ou  $\mathbb{C}$ , 28  
 $\mathbb{P}^n(\mathbb{K})$  : espace projectif de dimension  
 $n$  sur  $\mathbb{K}$ , 28
- parallélisme, 42  
 plan, 30  
 plan de netteté, 74  
 plan principal, 75  
 point unité, 30  
 points complexes, 41  
 points cycliques, 42  
 points de référence, 30  
 points réels, 41  
 polaire, 38
- pôle, 38  
 poursuite d'indices, 134  
 programme d'Erlangen, 25  
 projectif (espace), 28  
 projectif (groupe), 29  
 projection, 36  
 projection affine, 78  
 projection orthogonale, 78  
 projection orthographique, 78  
 projection parallèle, 78  
 projection perspective, 77  
 projective (base), 29  
 projective (géométrie), 22  
 projective (transformation), 35  
 projectives (coordonnées), 33
- quadruplet équiharmonique, 32
- $\overline{\mathbb{R}}$  : droite numérique achevée, 27  
 rapport de trois points, 42  
 reconnaissance, 169  
 rectification des images, 176  
 regroupement des aspects, 162  
 représentation, 28  
 ressemblance entre images, 160
- scalaire, 25  
 similitude, 42, 137  
 sous-espace tangent, 58  
 sous-groupe d'isotropie, 60  
 sous-variété, 55  
 stabilisateur, 60  
 système non-dégénéré d'équations locales, 55
- systèmes optiques épais, 75  
 systèmes optiques minces, 73
- test de coplanarité, 87  
 théorème de Burns, 61  
 théorème de Chasles, 37  
 théorème de Desargues, 30  
 théorème de Möbius, 35  
 théorie des invariants, 49  
 topologie, 25, 44  
 tracking, 134  
 transformation affine, 42  
 transformation projective, 35  
 triangle auto-polaire, 39

vision, 73

vision par ordinateur, 13

vue, 149



## Résumé

Cette thèse concerne la modélisation automatique d'objets polyédriques en vue d'une reconnaissance ultérieure de ces objets à partir d'une seule image. Lorsqu'on dispose de plusieurs images d'un objet à modéliser, le problème consiste à déterminer quelle information est nécessaire pour une reconnaissance future de cet objet et comment cette information peut être extraite des images.

Nous envisageons deux types d'information. Tout d'abord, en se plaçant dans le cadre de la géométrie projective et de la théorie des invariants, nous montrons comment on peut extraire une information 3D numérique (coplanarité de points, birapports) caractéristique d'un objet, à partir de deux images appariées de celui-ci.

Ensuite, nous proposons une méthode d'extraction d'information visuelle à partir d'un ensemble d'images. Après avoir apparié les images, on compare leur contenu, ce qui permet de les regrouper selon l'aspect de l'objet qu'elles représentent. Chacun de ces aspects donne alors lieu à l'établissement d'un modèle qui regroupe l'information disponible à son propos.

Les différents algorithmes proposés sont illustrés sur des exemples d'images réelles ou synthétiques.

**Mots clés :** vision par ordinateur, modélisation d'objets polyédriques, appariement d'images, géométrie projective, invariants, birapport, coplanarité.

## Abstract

This thesis is devoted to a procedure for constructing models of polyhedral objects with the goal of subsequently recognizing individual objects from single images. The problem consists of determining which information is necessary for recognition and how to extract it from a set of images representing the object to be modeled.

Two kinds of information are studied. Firstly, as a part of projective geometry and invariant theory, we show how to extract a numerical 3D information (point coplanarity, cross ratios) characteristic of an object from two matched images of it.

Secondly, we propound a method to extract visual information from a set of images. These images are matched, their content is compared, and they are grouped according to the object aspect they represent. Each of the aspects gives rise to a model subpart which collate the available information.

The different algorithms described in the thesis are tested with real and synthetic images.

**Keywords :** computer vision, polyhedral object model building, image matching, projective geometry, invariants, cross ratio, coplanarity.