88NAN 10323

Institut National
Polytechnique de Lorraine

Centre de Recherche en
Informatique de Nancy
Inria Lorraine

Perception de la distance par mise en correspondance de régions entre des images stéréoscopiques

THÈSE

INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE NANCY

présentée et soutenue publiquement le **8 mars 1988**

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Lorraine

(Spécialité Informatique)

par

Brigitte Wrobel-Dautcourt

Composition du jury :

Président : Jean-Paul HATON

Rapporteurs : Pierre LESCANNE
Henri MAÎTRE

Examinateurs : Radu HORAUD
Roger MOHR

D 136 008497 0

13600 84910

88NAN10323

Institut National
Polytechnique de Lorraine

Centre de Recherche en
Informatique de Nancy
Inria Lorraine

[M]1988 WROBEL-DAUTCOURT, B

Perception de la distance par mise en correspondance de régions entre des images stéréoscopiques

THÈSE

INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE NANCY

présentée et soutenue publiquement le **8 mars 1988**

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Lorraine

(Spécialité Informatique)

par

Brigitte Wrobel-Dautcourt

Composition du jury :

Président : Jean-Paul HATON

Rapporteurs : Pierre LESCANNE
Henri MAÎTRE

Examinateurs : Radu HORAUD
Roger MOHR

Résumé

Notre travail s'inscrit dans le cadre du projet ORASIS dont l'objectif est le contrôle de la navigation d'un robot mobile autonome se déplaçant dans un environnement de bureau ou d'atelier sous des conditions naturelles d'éclairage. Ce robot doit être capable de voir, reconnaître et apprendre l'univers dans lequel il évolue.

Cette thèse présente le résultat de notre travail sur la vision du robot, plus précisément la *perception de la distance par stéréovision*. Le principe de base de la stéréovision est d'apparier des indices visuels des deux images traitées puis de déterminer par un calcul géométrique leur position dans la scène tridimensionnelle. Les indices images considérés sont les régions, puisque celles-ci contiennent une information de surface nécessaire à la construction d'une scène tridimensionnelle la plus complète possible.

Notre travail se scinde en trois étapes : extraction des régions, appariement puis construction des informations tridimensionnelles. Les difficultés de cette approche sont essentiellement dans la mise en correspondance puisque les régions à apparier présentent de grandes différences aussi bien photométriques que morphologiques. De plus, les deux surfaces visualisées d'un même objet sont partitionnées en un nombre inégal de régions. La solution apportée s'appuie sur l'utilisation de plusieurs indices visuels (région, contours, ...) afin de bénéficier d'informations plus riches d'une part, et sur le renforcement de la segmentation à chaque niveau du système d'autre part.

Notre travail montre également les limites d'un traitement de "bas-niveau" sur les images et la nécessité d'inclure la compétence de traitements de plus "haut-niveau" ; la méthode de stéréovision que nous avons développée en est un exemple.

Mots clés :

- robot mobile - stéréovision - extraction de régions -
- appariement - segments tridimensionnels - surfaces tridimensionnelles -
- interaction entre processus -

Abstract

Our study is part of the ORASIS project. The aim of this project is to control the navigation of an autonomous mobile robot which moves in an office environment under natural light. The robot is supposed to be able to see and learn the surrounding world.

This thesis presents the result of our work on the robot's vision system, especially on the perception of depth by means of stereovision. The fundamental principle of stereovision is to match image primitives between two views in order to compute their three-dimensional position. The image primitives chosen are regions, as they carry a surface information which allows to compute a complete three-dimensional map.

Our work is composed of three steps: region growing, matching and construction of the three-dimensional information. The difficulties of this approach are essentially in the matching: the regions to match are different which respect to their photometric and morphologic characteristics, and the surface of the same object is partitioned into non equivalent numbers of regions. Our solution is based first of all on the use of several kinds of image primitives (regions, edge segments, ...) instead of just one; this yields more information to deal with. Secondly, we enhance the segmentation at each step of the system.

Our work also shows the limit of "low-level" vision and the need for including "high-level" processing; our stereovision method stands as a good example.

Key words :

- mobile robot - stereovision - region growing -
- matching - three-dimensional edge segments - three-dimensional surfaces -
- processus interaction -

à mes parents,

à Thierry,

Cette thèse est pour moi l'occasion d'exprimer des remerciements à tous ceux qui m'ont conseillée et aidée à réaliser ce projet dans de bonnes conditions, et je suis heureuse aujourd'hui de remercier plus particulièrement :

Monsieur Roger Mohr, Professeur à l'Institut National Polytechnique de Lorraine, qui m'a proposé un sujet de recherche passionnant, et m'a encadrée et orientée tout au long de ces années de travail. Je lui suis particulièrement reconnaissante pour ses conseils, ses encouragements et également pour son enthousiasme.

Monsieur Jean-Paul Haton, Professeur à l'Université de Nancy 1 et Directeur de Recherches INRIA au CRIN, qui m'a accueillie dans le laboratoire Reconnaissance des Formes et Intelligence Artificielle du CRIN, et qui me fait l'honneur de présider ce jury. Je le remercie de sa confiance et pour le grand intérêt qu'il a toujours porté à mes travaux.

Monsieur Pierre Lescanne, Directeur de Recherches CNRS au CRIN, et Monsieur Henri Maître, Professeur à l'Ecole Nationale Supérieure des Télécommunications à Paris, qui ont acceptés d'être rapporteurs de ce mémoire et de siéger à ce jury. Je les remercie pour leurs remarques et conseils complémentaires qui m'ont permis d'apporter plusieurs améliorations à la qualité de cette thèse. Qu'ils trouvent ici l'expression de ma gratitude pour l'intérêt dont ils ont fait preuve à l'égard de ce travail.

Monsieur Radu Horaud, Chargé de Recherches CNRS au LIFIA à Grenoble qui a accepté de faire partie du jury de soutenance. Il m'a fait un grand plaisir en ayant accepté d'examiner mon travail et de le juger.

Je veux aussi remercier les personnes des différents laboratoires impliqués dans le projet ORASIS *avec lesquelles j'ai collaboré à un moment ou un autre. Je tiens plus particulièrement à remercier Nicholas Ayache et Giorgio Toscani pour leurs discussions et leur contribution matérielle ; Olivier Monga pour la confrontation de nos idées et logiciels concernant la segmentation d'images, pour les très nombreuses discussions ayant donné naissance à plusieurs collaborations fructueuses.*

Je dois des remerciement chaleureux à toute l'équipe Vision du CRIN, plus spécialement à Djemel Ziou et Eric Thirion pour avoir mis à ma disposition les outils logiciels qu'ils ont développés ; Yolande Belaïd qui a eu la gentillesse de développer un logiciel indispensable à mon module de stéréovision. Je remercie également Karl Tombre pour sa disponibilité, sa compétence, et pour ses très nombreux moments consacrés à la relecture de mes différents articles et de ce mémoire.

Merci également à Anne Boyer pour ses utiles conseils de présentation et de rédaction de cette thèse ; à Pierre Mousel et Azim Roussanaly qui partagent mon bureau mais aussi ma "vie de recherche". C'est avec grand plaisir que je remercie Claude Wrobel qui m'a conseillée lors de mes choix concernant mon cursus universitaire ; et François Schwaab pour l'aide et le soutien moral qu'il m'a prodigués, et sans lesquels ce travail ne serait pas.

Enfin je tiens à ce que tous mes collègues et amis du CRIN trouvent ici mes remerciements pour l'atmosphère de travail sympathique dont j'ai bénéficié, merci à tous ceux qui ont pris de leur temps pour discuter avec moi et me conseiller.

Table des matières

1 PERCEPTION DE L'ENVIRONNEMENT **5**
1 Percevoir et comprendre l'environnement 6
2 Percevoir l'environnement 7
3 Méthodes actives de perception 7
4 Méthodes passives de perception 10
4.1 La vision artificielle 10
4.2 Perception de l'environnement par la vision 12
4.3 Notre système de perception de l'environnement 13

2 STÉRÉOVISION **17**
1 Généralités 18
2 Acquisition des images 18
2.1 Utiliser une seule caméra 19
2.2 Utiliser plusieurs caméras 20
3 Mise en correspondance 20
4 Géométrie du système optique de stéréovision 23
4.1 Construction d'une image 23
4.2 Géométrie épipolaire 23
4.3 Problème de l'appariement 26
4.4 Lois d'unicité et de continuité 26
4.5 Calibration 28
5 Calcul des coordonnées tridimensionnelles 31
6 Stéréovision multiple 32
6.1 Limites de la vision binoculaire 32
6.2 Solutions 35
6.3 Choix du système de stéréovision 36

3 CHOIX DES PRIMITIVES IMAGES **39**
1 Introduction 40
2 Points images 41

2.1 Points images 42
2.2 Points d'intérêt 44
2.3 Conclusion 46
3 Segments de contour 46
3.1 Appariement de segments isolés 49
3.2 Appariement d'ensembles de segments 52
4 Arcs de cercle 53
5 Régions 54
6 Utilisation conjointe de plusieurs primitives 56
7 Interaction entre processus 58

4 SEGMENTATION D'IMAGES EN RÉGIONS 61
1 Introduction 62
2 Choisir entre une "approche contour" et une "approche région" . . . 63
2.1 Approche "contour" 64
2.2 Approche "région" 66
2.3 Notre approche 66
3 Définition formelle de la segmentation 67
4 Algorithme 69
4.1 Structure de l'algorithme 70
4.2 Structure de données et implantation de l'algorithme 71
4.3 Complexité de l'algorithme 73
4.4 Partition initiale 76
4.5 Résumé 78
5 Prédicats d'homogénéité 79
5.1 Utilisation d'un prédicat d'homogénéité 80
5.2 Utilisation de plusieurs prédicats 82
6 "Coopération" régions-contours 85
7 Conclusion 87

5 MISE EN CORRESPONDANCE DES RÉGIONS 95
1 Introduction 96
2 Prise en compte des différences de segmentation 97
2.1 Position du problème 97
2.2 Notre approche 97
2.3 Implantation 100
3 Phase 1 : génération des hypothèses 101
3.1 Mise en correspondance des frontières 102
3.2 Couples de droites épipolaires associées 103
3.3 Recherche de chemin optimal dans un graphe 105
3.4 Prise en compte des lignes environnantes 109
3.5 Génération des hypothèses 111
4 Phase 2 : Validation des hypothèses 116
4.1 Structure générale de l'algorithme 116
4.2 Stratégie d'examen des hypothèses 116
4.3 Evaluation de la qualité de correspondance 117
5 Résultats 120
6 Conclusion 124

6 CONSTRUCTION DES INFORMATIONS TRIDIMENSIONNELLES 127
1 Introduction 128
1.1 Transformation perspective 128
1.2 Transformation perspective inverse 129
2 Différentes méthodes de construction des surfaces tridimensionnelles 131
2.1 L'ensemble des points de la région 131
2.2 L'ensemble des points formant le contour de la région 133
2.3 L'ensemble des points caractéristiques du contour d'une région 135
2.4 Tableau des caractéristiques des régions 137
3 Notre approche 137
3.1 Approximation polygonale des frontières des régions 138
3.2 Approximation polygonale des lignes de contraste 142
3.3 Appariement des segments 144
3.4 Méthode de construction des segments tridimensionnels . . . 147
3.5 Construction des segments tridimensionnels 150
4 Construction des surfaces tridimensionnelles 158
4.1 Position du problème 158
4.2 Méthode de construction 159
5 Résultats et conclusion 168

7 CONCLUSION 175
1 Stéréovision par appariement de régions 175
2 Points importants 181
3 Perspectives 182

A Partition initiale par "quad-tree" 185

B Exemple de déroulement de la segmentation 187

C Calibration du banc stéréoscopique **195**
1 Matrices de transformation perspective 195
1.1 Paires stéréoscopiques *bureau1*, *bureau2* et *couloir* 195
1.2 Paire stéréoscopique *pièce* 195
2 Relation géométrique entre les caméras 196
2.1 Paires stéréoscopiques *bureau1*, *bureau2* et *couloir* 196
2.2 Paire stéréoscopique *pièce* 197

D Résultats **199**
1 Scène *bureau1* 200
2 Scène *bureau2* 205
3 Scène *couloir* 210
4 Scène *pièce* 215

Liste des Figures

1.1 Télémètre laser 8
1.2 Projection de plans lumineux sur une scène (1) 9
1.3 Projection de plans lumineux sur une scène (2) 9

2.1 Capteur tridimensionnel 19
2.2 Robot mobile 20
2.3 Acquisition d'une image (1) 24
2.4 Acquisition d'une image (2) 24
2.5 Géométrie épipolaire 25
2.6 Loi d'ordre 27
2.7 Transformation perspective 29
2.8 Mire de calibration du banc stéréoscopique 30
2.9 Couples de lignes épipolaires associées 31
2.10 Construction d'un point tridimensionnel 32
2.11 Répercution d'une erreur sur le calcul d'un point tridimensionnel . . 33
2.12 Intersection de plans 34
2.13 Répercution d'une erreur sur le calcul d'un segment tridimensionnel (1) 34
2.14 Répercution d'une erreur sur le calcul d'un segment tridimensionnel (2) 35
2.15 Système de vision à trois caméras 37

3.1 Stéréogramme aléatoire 43
3.2 Points de contraste d'une image 45
3.3 Différentes résolutions d'une image 47
3.4 Lignes de contraste 48
3.5 Approximation polygonale 49
3.6 Appariement de segments d'une scène d'intérieur (1) 51
3.7 Appariement de segments d'une scène d'intérieur (2) 52
3.8 Appariement d'arcs de cercle 54
3.9 Résultats d'un algorithme de stéréovision couplé à l'extraction et la construction de la surface tridimensionnelle 59

4.1 Détection de contours (1) 65
4.2 Détection de contours (2) 65
4.3 Division récursive de régions 77
4.4 Image originale 77
4.5 Partition initiale 78
4.6 Partition initiale - balayage de l'image 79
4.7 Segmentation finale (1) 80
4.8 Segmentation finale (2) 81
4.9 Segmentation finale (3) 81
4.10 Représentation de la frontière entre deux régions 84
4.11 Segmentation finale (4) 85
4.12 Contrôle de la croissance des régions par les contours 86
4.13 Carte de points de contraste 87
4.14 Segmentation contrôlée par les contours 88
4.15 Schéma récapitulatif des traitements pour la segmentation des images stéréoscopiques 91
4.16 Couple stéréoscopique (1) 92
4.17 Segmentation des images du couple stéréoscopique (1) 92
4.18 Couple stéréoscopique (2) 93
4.19 Segmentations des images du couple stéréoscopique (2) 93

5.1 Différence de segmentation entre les deux images - cas général 97
5.2 Appariement des intersections des régions 98
5.3 Appariement des unions des régions 98
5.4 Appariement des intersections des régions et divisions de régions . . 99
5.5 Appariement des unions des régions et fusions de régions 100
5.6 Appariement des régions 101
5.7 Appariement des frontières 102
5.8 Droite épipolaire 103
5.9 Couple de droites épipolaires associées 104
5.10 Graphe de recherche des points homologues 105
5.11 Scène artificielle 106
5.12 Chemin optimal "artificiel" 106
5.13 Construction d'un chemin optimal 107
5.14 Arcs du graphe pour la construction du chemin optimal 108
5.15 Construction d'un chemin optimal avec prise en compte des informations contextuelles 110
5.16 Recherche d'un ensemble optimal de chemins 112
5.17 Lignes épipolaires 113
5.18 Chemin optimal (1) 113
5.19 Chemin optimal (2) 114
5.20 Génération des hypothèses d'appariement de régions 115
5.21 Appariement des points frontières des régions 115
5.22 Couple stéréoscopique - images originales 121
5.23 Partitions stéréoscopiques à apparier 121
5.24 Fusion de régions par le processus d'appariement 122
5.25 Appariement des régions 122
5.26 Partitions finales 123
5.27 Schéma récapitulatif des traitements pour l'appariement des régions 124

6.1 Projection perspective 128
6.2 Projection perspective inverse 130
6.3 Estimation de la position d'un point tridimensionnel connaissant ses deux projections sur chacune des vues (1) 130
6.4 Construction d'une surface tridimensionnelle (1) 132
6.5 Construction d'une surface tridimensionnelle (2) 134
6.6 Construction d'un segment tridimensionnel (1) 136
6.7 Construction d'une surface tridimensionnelle (3) 136
6.8 Principe de l'approximation polygonale d'une chaîne de contour . . . 139
6.9 Partitions en régions 139
6.10 Approximation polygonale des frontières des régions 140
6.11 Superposition des segments de contours à l'image originale (1) . . . 141
6.12 Extraction des chaînes de points de contraste des images stéréoscopiques 142
6.13 Approximation polygonale des chaînes de points de contraste extraites des images stéréoscopiques 143
6.14 Superposition des segments de contours à l'image originale (2) . . . 143
6.15 Segments à apparier 144
6.16 Chemin optimal représentant la mise en correspondance des segments 145
6.17 Utilisation de des régions appariées pour guider l'appariement des régions 146
6.18 Recalage des segments de contours aux régions 146
6.19 Appariement des segments 147
6.20 Construction d'un segment tridimensionnel (2) 148
6.21 Détermination des zones des segments pouvant s'associer 148
6.22 Construction d'un segment tridimensionnel (3) 149
6.23 Zones des segments pouvant s'associer en respect des contraintes épipolaires 149
6.24 Segments tridimensionnels de la scène (1) 151

6.25 "Vue de face" des segments tridimensionnels (lignes de contraste) . . 151
6.26 "Vue de côté" des segments tridimensionnels (lignes de contraste) . . 152
6.27 "Vue de dessus" des segments tridimensionnels (lignes de contraste) 152
6.28 "Vue de dessus" des segments tridimensionnels (frontières des régions) 153
6.29 Incertitude des segments 3d : "vue de dessus" (lignes de contraste) . 156
6.30 Incertitude des segments 3d : "vue de face" (lignes de contraste) . . 157
6.31 Problème des surfaces cachantes 158
6.32 Construction d'une surface plane par une méthode d'approximation au sens des moindres carrés. 160
6.33 Appartenance d'un segment aux surfaces qu'il délimite (Règle 1) . . 163
6.34 Appartenance d'un segment aux surfaces qu'il délimite (Règle 2) . . 164
6.35 Appartenance d'un segment aux surfaces qu'il délimite (Règle 3) . . 165
6.36 Construction des surfaces tridimensionnelles - problème des surfaces trouées 167
6.37 Schéma récapitulatif des traitements pour le calcul des surfaces tridimensionnelles 168
6.38 Construction de certaines surfaces tridimensionnelles 169
6.39 Construction de toutes les surfaces tridimensionnelles 170
6.40 Projection des segments "corrigés" sur le sol de la pièce 171

7.1 Schéma de tous les traitements nécessaires au module de stéréovision 180

A.1 Quad-tree 186

B.1 Exemple de partition à fusionner 188
B.2 Graphe d'adjacence valué associé à la partition 188
B.3 Arbre binaire de recherche associé à la partition 189
B.4 Nouvelle partition après fusion de deux régions 189
B.5 Arbre binaire de recherche associé à la partition 190
B.6 Nouvelle partition après fusion de deux régions 190
B.7 Arbre binaire de recherche associé à la partition 191
B.8 Nouvelle partition après fusion de deux régions 191
B.9 Arbre binaire de recherche associé à la partition 192
B.10 Nouvelle partition après fusion de deux régions 193
B.11 Arbre binaire de recherche associé à la partition 193

C.1 Vecteur directeur d'une ligne épipolaire dans l'image droite 196
C.2 Vecteur directeur d'une ligne épipolaire dans l'image gauche 197
C.3 Couples de lignes épipolaires associées 198
C.4 Couples de lignes épipolaires associées 198

D.1 Couple stéréoscopique (2) 200
D.2 Partitions initiales en régions 200
D.3 Appariement des régions 201
D.4 Partitions finales en régions 201
D.5 "Vue de face" des segments tridimensionnels (lignes de contraste) . . 202
D.6 "Vue de dessus" des segments tridimensionnels (lignes de contraste) 203
D.7 Surfaces tridimensionnelles de la scène - Premiers résultats 204
D.8 Couple stéréoscopique (2) 205
D.9 Partitions initiales en régions 205
D.10 Appariement des régions 206
D.11 Partitions finales en régions 206
D.12 "Vue de face" des segments tridimensionnels (lignes de contraste) . . 207
D.13 "Vue de dessus" des segments tridimensionnels (lignes de contraste) 208
D.14 Surfaces tridimensionnelles de la scène - Premiers résultats 209
D.15 Couple stéréoscopique (2) 210
D.16 Partitions initiales en régions 210
D.17 Appariement des régions 211
D.18 Partitions finales en régions 211
D.19 "Vue de face" des segments tridimensionnels (lignes de contraste) . . 212
D.20 "Vue de dessus" des segments tridimensionnels (lignes de contraste) 213
D.21 Surfaces tridimensionnelles de la scène - Premiers résultats 214
D.22 Couple stéréoscopique (2) 215
D.23 Partitions initiales en régions 215
D.24 Appariement des régions 216
D.25 Partitions finales en régions 216
D.26 "Vue de face" des segments tridimensionnels (lignes de contraste) . . 217
D.27 "Vue de dessus" des segments tridimensionnels (lignes de contraste) 218
D.28 Surfaces tridimensionnelles de la scène - Premiers résultats 219

INTRODUCTION

"D'une image d'un objet, le cerveau humain retire une multitude d'informations, est-il possible de doter une machine ou un robot d'une telle capacité d'analyse ?" (Jean Serra) [Serra 83].

L'étude présentée dans cette thèse s'inscrit dans le cadre des recherches effectuées pour le projet ORASIS[1] intéressant six laboratoires français :

- INRIA, Unité de Recherche de Rocquencourt, LE CHESNAY,
- INRIA, Unité de Recherche de Sophia-Antipolis, VALBONNE,
- IRISA, Institut de Recherche en Informatique et Systèmes Aléatoires, INRIA Unité de Recherche de RENNES,
- LIFIA-IMAG, Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle, GRENOBLE,
- CERFIA, Cybernétique des Entreprises, Reconnaissance des Formes, Intelligence Artificielle, TOULOUSE,
- et CRIN INRIA-Lorraine, Centre de Recherche en Informatique de Nancy.

L'objectif principal est le développement de la méthodologie de la vision haut niveau ; l'application support est la vision d'un *robot mobile autonome*, capable de voir, reconnaître et apprendre l'univers dans lequel il évolue. Les points techniques la concernant sont la segmentation d'images, la perception de la distance par vision stéréoscopique, la représentation des connaissances, l'identification, la localisation et la navigation.

[1] Objectif Robot Autonome et Système d'Intelligence Sensorielle ; ORASIS signifie également vision en grec.

Robot mobile :

Un robot mobile est appelé à se déplacer dans un environnement particulier. L'autonomie du robot se caractérise par sa capacité à acquérir par lui-même les données nécessaires et à raisonner sur ces données pour effectuer la tâche demandée.

Les premiers travaux concernant les robots mobiles ont eut lieu dans les années soixante avec la réalisation de SHAKEY au SRI[2]. L'objectif était alors d'*"étudier les processus de contrôle en temps réel d'un système robotique interagissant avec un environnement complexe"* [Nilsonn 69].

Du fait des progrès de l'informatique tant au point de vue logiciel que matériel, la recherche sur les robots mobiles se poursuit actuellement de manière intensive. Les robots mobiles représentent en fait un terrain d'essai formidable pour confronter les techniques de Reconnaissance des Formes et d'Intelligence Artificielle avec la complexité du monde réel et les contraintes de la robotique. Afin de situer le projet ORASIS, nous citons ci-après quelques projets de robots mobiles en les classant selon leur technique d'acquisition de l'information tridimensionnelle.

Capteurs à ultra-sons :

NEPTUNE : développé à l'Université de Carnegie-Mellon [Moravec 85] ;

IMP : (Intelligent Mobile Platform) projet de l'Université de Carnegie-Mellon pour définir une plateforme intelligente, capable d'apprendre de façon semi-automatique (guidé par un opérateur) [Crowley 85] ;

Vision monoculaire :

AVL : (Autonomous Land Vehicle) projet de navigation d'un véhicule autonome destiné à faire du suivi de routes, développé à l'Université du Maryland [Waxman 86].

Analyse d'une séquence d'images :

HERO : l'idée de base est d'utiliser la continuité de la séquence d'images pour faciliter la mesure du déplacement apparent des caractéristiques visuelles (flot optique) [Tsuji 85].

Stéréovision active :

VESA : ce robot mobile a été développé à Rennes [Zhao 84] avec une source d'éclairement lumineuse et un capteur d'images.

Stéréovision passive :

CART : ce robot, développé à l'Université de Stanford [Moravec 80] est le premier doté d'un système de perception par stéréovision ;

[2]Stanford Research Institute

FIDO : développé à l'Université de Carnegie-Mellon [Thorpe 85] ;

HILARE : ce robot, développé au LAAS[3], utilise plusieurs capteurs de distance : capteurs à ultra-sons, odomètre, télémètre laser et stéréovision. Cette redondance permet d'améliorer la qualité des données acquises [RobertdeSaintVincent 86] ;

ORASIS : le robot construit par l'Inria Rocquencourt est muni de deux caméras pour permettre l'acquisition de la distance par stéréovision. Ce robot s'est vu attribuer une caméra supplémentaire pour améliorer la qualité des résultats. D'autres capteurs : de proximité (ultra-sons) et de déplacement (capteur sur les roues) interviennent également.

Des détails concernant les caractéristiques particulières de ces robots mobiles peuvent être trouvés dans [RobertdeSaintVincent 86].

Stéréovision :

Au sein du vaste projet ORASIS, notre recherche concerne *l'acquisition de l'information de distance par stéréovision.* L'étude et la réalisation sont présentées dans cette thèse. Le plan que nous suivons se présente ainsi :

Le *premier chapitre* introduit les problèmes de la perception et de la compréhension automatiques de l'environnement. Nous discutons des différentes méthodes généralement utilisées pour percevoir l'environnement d'un robot et plus particulièrement celles concernant l'*acquisition de l'information tridimensionnelle.* Nous situons notre choix pour une méthode de stéréovision et présentons les problèmes généraux liés à la vision artificielle.

Nous présentons au *chapitre 2* les généralités de la *vision stéréoscopique* et montrons l'importance et les difficultés de la phase d'appariement. Nous détaillons l'acquisition des images, les différentes approches pour effectuer la mise en correspondance, la géométrie et les contraintes d'un tel système de vision.

Au *chapitre 3*, nous présentons une revue des *primitives images* généralement utilisées, ainsi que les techniques d'appariement mises en œuvre. Ceci nous permet d'expliquer notre choix pour la primitive région et de situer notre approche pour résoudre problème de la vision stéréoscopique. Celle-ci est détaillée page 56.

L'objectif du *chapitre 4* est de présenter la méthode de *segmentation d'images* développée pour obtenir les partitions en régions utilisées ultérieurement lors de la mise en correspondance. Nous exposons tout d'abord le problème de la segmentation et présentons ensuite, d'une part, l'outil algorithmique général mis en œuvre, et d'autre part, les prédicats d'homogénéité de régions utilisés. Nous terminons en détaillant une technique simple de contrôle de la segmentation par une carte de points de contraste pré-calculée.

Le *chapitre 5* est consacré à la présentation de l'algorithme d'*appariement des régions.* La prise en compte des différences de segmentation entre les images – fondamentale pour notre problème – a influencé l'approche que nous avons choisie.

[3]Laboratoire d'Automatique et d'Analyse des Systèmes du CNRS, TOULOUSE

Ces différences interviennent au niveau des caractéristiques photométriques et morphologiques des régions ; de plus, les surfaces des mêmes objets sont partitionnée en un nombre inégal de régions. Notre méthode se partage en deux phases : génération des hypothèses puis validation de ces hypothèses. La première phase est réalisée par la mise en correspondance des points formant la frontière des régions à l'aide d'une méthode de programmation dynamique. La seconde phase concerne la validation des hypothèses d'appariement des régions avec des modifications locales des partitions en régions.

Le propos du *chapitre 6* est de présenter la méthode utilisée pour la *construction des informations tridimensionnelles* d'une scène à partir des couples de régions homologues déterminés. Notre approche s'appuie sur la construction des segments tridimensionnels représentant les frontières des régions permettant ensuite la construction des surfaces tridimensionnelles.

Nous concluons sur les points forts de notre approche : d'une part utilisation de plusieurs indices visuels (région, contours, ...) afin de bénéficier d'informations plus riches, et d'autre part renforcement de la segmentation à chaque niveau du système. Notre travail montre les limites d'un traitement de "bas-niveau" sur les images et la nécessité d'inclure la compétence de traitements de plus "haut-niveau" ; la méthode de stéréovision que nous avons développée en est un exemple.

Chapitre 1

PERCEPTION DE L'ENVIRONNEMENT

Ce chapitre introduit le problème de la perception et de la compréhension automatiques de l'environnement.

Nous discutons des différentes méthodes généralement utilisées pour percevoir l'environnement, essentiellement celles concernant l'acquisition de l'information tridimensionnelle, et présentons les problèmes inhérents à la vision artificielle. Nous exposons ensuite notre choix d'une méthode de stéréovision pour notre application particulière d'aide à la navigation d'un robot mobile, ce robot évoluant dans un environnement de bureau ou d'atelier.

1 Percevoir et comprendre l'environnement

L'utilité de la perception et de la reconnaissance de l'environnement par une machine est particulièrement importante dans le domaine de la robotique.

Actuellement dans l'industrie, de nombreux systèmes automatiques fonctionnant de façon "aveugle" sont implantés. Ils ne possèdent aucune perception, de quelque nature que ce soit, de leur environnement. Ce sont des robots manipulateurs destinés à effectuer une tâche répétitive dans un environnement bien défini et immuable. Toutefois, si l'on désire qu'ils soient capables de s'adapter à leur monde local, d'interagir sur les éléments qui les entourent, il est indispensable qu'ils soient aussi capables de percevoir et de comprendre cet environnement.

Cette compréhension automatique du monde est fondamentale en robotique et a donné naissance à de très nombreuses études, assemblage de pièces industrielles avec contrôle [Rogala 87], arrimage d'un satellite dans l'espace, ...

Les principaux problèmes posés par la perception et la compréhension de l'environnement sont les suivants :

- acquisition des informations, ces informations sont de type couleur, texture, ... et plus particulièrement de type coordonnées tridimensionnelles puisque l'espace qui est perçut possède une structure à trois dimensions ;
- représentation de ces informations ;
- reconnaissance des objets de la scène.

L'acquisition est à la base de tout système de compréhension de l'environnement et apparaît donc comme une phase très importante ; elle doit être capable de fournir les données les plus correctes et les plus fiables possibles.

Les informations ainsi obtenues sont généralement exprimées dans un repère tridimensionnel centré sur le système d'acquisition (caméra, télémètre, ...). Elles peuvent ensuite être transformées dans le repère de l'objet, afin de donner une **représentation** stable des objets. Le centre de gravité de l'objet peut, par exemple, être le centre du repère utilisé. Les objets sont alors décrits par un ensemble de primitives de type surfacique ou volumique. Nous trouvons par exemple des primitives "plan", "cône", "sphère" ou définies par des surfaces quadratiques [Shihari 81], [MullerBelaid 84].

La dernière phase consiste à **décrire** la scène observée, c'est-à-dire donner une étiquette sémantique aux différents objets présents. Un ensemble composé d'un plan et de quatre cylindres peut, par exemple, être instancié par l'objet "table" si quelques contraintes supplémentaires sont vérifiées [Bajcsy 80].

La réalisation de chacune de ces trois phases est extrêmement délicate et aucun système actuel ne les résoud parfaitement. C'est un véritable défi que d'atteindre le niveau de la perception visuelle de l'homme : lorsqu'on demande à une personne de décrire ce qu'elle voit, elle n'éprouve aucune difficulté à nommer les objets qui l'entourent : table, téléphone, chaise, ...

Le cadre de notre recherche se limite à la perception de l'environnement, soit l'acquisition des informations de la scène, aussi nous ne détaillerons pas les problèmes liés à la description et à la reconnaissance des objets.

2 Percevoir l'environnement

Les informations concernant le monde réel que l'on veut extraire sont essentiellement de deux natures :

- *photométriques* : la quantité de lumière (intensité lumineuse), la couleur, ... peuvent être pertinentes à connaître ;
- *spatiales* : les coordonnées tridimensionnelles des éléments composant l'environnement sont fondamentales pour certaines applications.

Ces informations servent ensuite de données à d'autres traitements : assemblage, navigation, interprétation ...

Les sources d'inspiration pour la création des systèmes perceptifs sont variées, mais un grand nombre sont issues des "modèles" existant chez l'homme ou chez les animaux. Nous pouvons les regrouper en deux catégories :

- les capteurs qui ont une action directe sur l'environnement étudié : systèmes acoustiques, tactiles, télémètres, ... de cette caractéristique fondamentale découle leur nom : **capteurs actifs**. Ils acquièrent généralement l'information spatiale.
- les capteurs qualifiés de **passifs** qui n'ont aucune action sur leur environnement. Ces capteurs utilisent généralement la *vision* comme mode d'acquisition des informations élémentaires (images), desquelles sont extraites les informations photométriques et spatiales.

Le terme *capteur*, quoique courant dans la littérature, n'est cependant guère adapté. Il sous-entend un élément technologique alors que les informations "élémentaires" acquises doivent être traitées, quelque fois de façon assez complexe, pour obtenir les informations désirées. Nous parlons donc dans la suite de ce chapitre, de méthodes ou de systèmes plutôt que de capteurs.

Puisque nous sommes plus particulièrement intéressés par l'acquisition de l'information spatiale d'une scène, nous présentons ci-après les méthodes actives et passives permettant de la déterminer.

3 Méthodes actives de perception

Certains systèmes sont **acoustiques**. Leur principe est d'utiliser la vitesse de propagation des ultra-sons. Le temps mis par une onde pour parcourir la distance aller-retour entre la source et un point de l'objet de la scène permet de calculer la

position tridimensionnelle de ce point. Un système de détection de l'environnement pour guider un robot mobile par des capteurs à ultra-sons a été développé au LAAS[1] [Bauzil 81].

D'autres systèmes utilisent la perception **tactile** pour reconnaître les objets. Citons dans ce domaine les travaux de S.A. Stansfield [Stansfield 86]. Le capteur tactile utilisé se présente sous la forme d'une matrice (10×16) de surfaces sensitives protégées d'élastomère. La géométrie du capteur est planaire. Il est placé en contact avec l'objet à étudier et il effectue la mesure des forces résultantes de la pression appliquée sur la surface. Le capteur est monté sur le poignet d'un bras manipulateur à 6 degrés de liberté.

Les **télémètres** sont un autre type de systèmes actifs. Le principe est de mesurer soit le temps de propagation de faisceaux lumineux entre l'émetteur, l'objet de la scène et le récepteur, soit le déphasage en amplitude entre le rayon émis et le rayon réfléchi dans le cas d'un faisceau laser modulé en amplitude. Le schéma de principe du télémètre laser est donné figure 1.1.

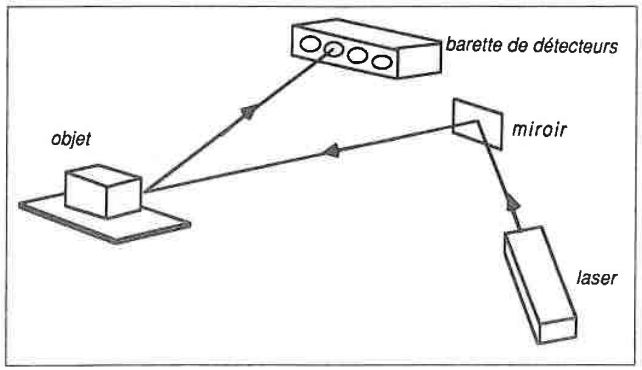

Figure 1.1: *Télémètre laser.*

Certains systèmes actifs intègrent un système de vision et font alors généralement intervenir une **source lumineuse** supplémentaire. On utilise une lumière structurée donnant un sur-éclairement de la scène et permettant alors une détection plus facile des points ainsi marqués ; les lasers sont le plus couramment employés. Citons les travaux de [Borianne 84] et [Lelandais 84] qui procèdent par projection de plans lumineux. La figure 1.2 illustrant ce principe est issue de l'article de E.W. Kent [Kent 86].

Il est alors possible de calculer les coordonnées tridimensionnelles des points de la scène ainsi éclairés par simple triangulation à condition de connaître avec précision la position de la source de lumière et le modèle de la caméra utilisée. L'ensemble

[1] LAAS : Laboratoire d'Automatique et d'Analyse des Systèmes du C.N.R.S., TOULOUSE

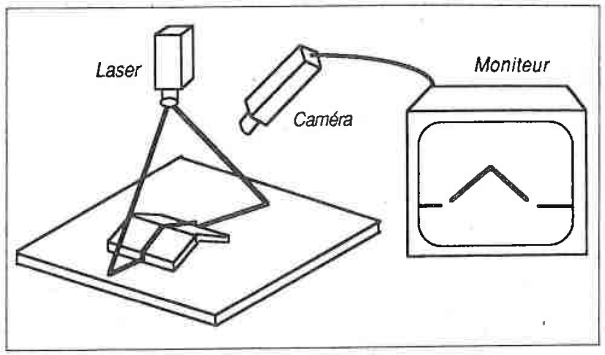

Figure 1.2: *Projection de plans lumineux sur une scène.*

de la scène peut être étudié si l'on peut faire bouger soit la source de lumière, soit la scène. Le mouvement de la source de lumière est généralement obtenu en faisant pivoter un miroir réfléchissant les rayons lumineux. Déplacer la scène revient par exemple à placer les objets à étudier sur un plateau tournant (figure 1.3).

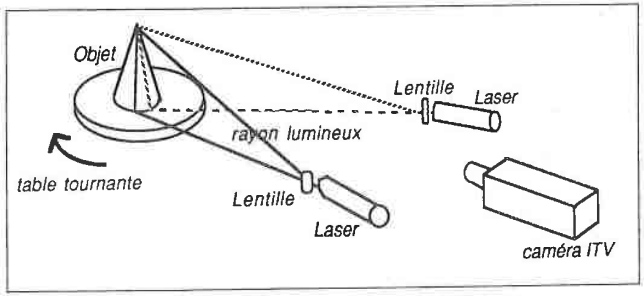

Figure 1.3: *Projection de plans lumineux sur une scène. L'objet est placé sur une table tournante contrôlée par ordinateur et éclairée par deux raies lumineuses.*

Les méthodes actives fournissent des mesures précises et peuvent s'avérer très efficaces, mais la description du monde obtenue est généralement très locale. Les systèmes utilisant des sources lumineuses supplémentaires imposent en outre de travailler sous des conditions d'éclairage très contraignantes et ne conviennent pas aux surfaces réfléchissantes. Ils ont également l'inconvénient d'interférer avec le milieu et d'être limités en distance.

Néanmoins, les méthodes actives sont très utilisées dans de nombreuses appli-

cations, et particulièrement dans le domaine de la robotique. On peut, par exemple, commander un bras manipulateur pour saisir les objets présents dans la scène [Jarvis 83], ou reconnaître des pièces usinées et leurs positions [Faugeras 86a].

Il est cependant des cas où certains systèmes actifs sont inutilisables du fait soit de leur interférence avec le milieu étudié (scène d'intérieur : le laser peut s'avérer dangereux d'utilisation, les ultra-sons sont nocifs aux animaux, ...), soit de la profondeur de la scène observée (images satellites).

4 Méthodes passives de perception

Les méthodes passives n'ont aucune influence sur l'environnement étudié ; leur rôle est simplement de recevoir les signaux lumineux émis ou réfléchis par la scène. Les capteurs utilisés sont principalement les **caméras**. L'information directement acquise n'est pas tridimensionnelle, mais seulement bidimensionnelle. Une ou plusieurs images sont traitées afin d'en extraire l'information tridimensionnelle souhaitée.

De telles méthodes de perception entrent alors dans le domaine de recherche en *vision artificielle*, domaine qui a posé et continue de poser de formidables problèmes et défis. Aussi, avant détailler le sujet de notre étude, nous exposons brièvement les problèmes généraux de la *vision*.

4.1 La vision artificielle

La vision est un processus de traitement de l'information, en l'occurence des images. Ces images peuvent être de types variés et le résultat de la vision est une interprétation de l'environnement, interprétation également variable[2].

Concernant la vision humaine, l'information qui est réellement disponible sur la rétine des yeux n'est ni plus ni moins qu'une collection de points (environ 1 million). En chaque point, il y a tout simplement une information qui donne une indication quant à la quantité de lumière et de couleur qui proviennent de l'espace environnant et qui ont été projetés en cet endroit de la rétine. Et pourtant l'homme voit un téléphone, une lampe, un bureau et il en perçoit les distances ...

Une des ambitions des chercheurs dans ce domaine est de construire un *modèle computationnel* [Horaud 87a] pour la perception visuelle. Pour l'homme, la vision est une expérience permanente et d'une extrême facilité. Cette facilité n'est cependant qu'une apparence, la vision humaine faisant partie d'un système très perfectionné mais aussi très complexe. On sait maintenant qu'une douzaine de centres cérébraux, soit environ 60% du cortex, participent à l'activité de la vision. La vue est donc d'une importance capitale pour l'intelligence humaine et du fait de cette liaison intime entre l'œil et le cerveau, il est difficile, sinon impossible, de séparer la vision des autres activités intellectuelles.

La vision est liée à l'intelligence, ceci ne signifie nullement que des animaux moins intelligents que l'homme ne "voient" pas, mais ils ont une perception et une

[2] La mouche, la poule et l'homme "voient" et pourtant ils n'ont pas les mêmes rétines, donc les mêmes images, et ils n'ont certainement pas la même interprétation des objets vus.

interprétation différentes, du moins pouvons nous le penser. La vision a suscité l'intérêt de nombreux scientifiques et philosophes depuis fort longtemps. Parmi ceux-ci, les neuro-biologistes mènent des recherches théoriques et expérimentales afin d'essayer de comprendre l'anatomie et le fonctionnement du cerveau, qui est loin cependant de leur avoir révélé tous ses secrets.

La **vision par ordinateur** s'intègre alors dans le domaine de l'**intelligence artificielle**. Elle en constitue aujourd'hui une discipline que nous pouvons qualifier de délicate du fait de la difficulté de s'inspirer d'un acte instinctif et de le simuler, dans la mesure où nous n'avons que peu de renseignements sur cet acte. Les actes raisonnés sont plus faciles à modéliser en intelligence artificielle que les actes instinctifs comme la vision, à laquelle nous nous intéressons.

Aussi, le but n'est pas de fournir une explication du fonctionnement de la vision biologique, mais de créer un modèle qui, vu de l'extérieur, possède des propriétés semblables.

Les applications de la vision par ordinateur sont nombreuses et variées. Nous pouvons citer, sans vouloir être exhaustifs :

- l'inspection des images médicales : étude des chromosomes [Bonton 87], de tumeurs, de lésions osseuses, ..., par le traitement d'images d'électrophorèses, de scanner, de résonance magnétique, d'échographie, ... [Geographic 87] ;
- l'inspection d'images géologiques : reconstitution et étude des strates par analyse d'images sismiques [Keskes 82], ...
- la télédétection : cartographie, interprétation automatique des images fournies par satellite [Berthod 87], ...
- la communication entre l'homme et la machine : lecture optique, reconnaissance de l'écriture [Belaid 85], analyse et lecture automatique de documents [Shihari 86], [Tombre 87], ...
- l'automatisation de processus industriels : reconnaissance d'objets isolés ou en vrac, saisie et assemblage de pièces, reconnaissance d'objets défectueux [Kent 86], ...
- le contrôle et la navigation de robots mobiles : le but généralement poursuivi est alors de construire un robot remplaçant au mieux l'homme pour des tâches répétitives : convoyage, surveillance, nettoyage industriel, suivi de routes [Wallace 85], ou devant évoluer dans un milieu hostile et/ou dangereux : mines, milieu sous-marin (trajectographie [Montanvert 87]), sites nucléaires contaminés, ...

  On peut aussi vouloir automatiser la perception et la modélisation de l'environnement visualisé par le robot [RobertdeSaintVincent 86], ...
- l'étude du mouvement : suivi de cible [Stelmaszyk 85], analyse de flux, trajectométrie [Coulon 85], estimation du déplacement [Toscani 87b], analyse de séquences d'images [Guilloux 84], [Bouthemy 87] ...

4.2 Perception de l'environnement par la vision

Dans ce domaine complexe de la vision, certaines méthodes ont été développées dans le but de percevoir l'environnement et plus particulièrement de déterminer les informations tridimensionnelles des scènes observées, aussi nous en présentons certaines ci-après. Mais avant de poursuivre, mentionnons que l'abréviation couramment employée pour le mot *trois dimensions* ou l'adjectif *tridimensionnel* est *3d*. Nous en faisons également usage dans cette thèse pour une simplicité et une commodité de lecture. Pour les mêmes raisons nous utilisons les initiales *2d* pour *bidimensionnel*.

Certains systèmes passifs de perception de l'environnement sont fondés sur l'utilisation des **relations d'occlusion** des objets pour détecter leurs distances relatives [Rosenberg 78]. La principale difficulté de cette méthode est de reconnaître les objets partiellement observés.

Citons également les méthodes utilisant l'aspect de la **texture** selon le principe suivant : l'élément de texture sera d'autant plus petit que l'objet est éloigné de la caméra [Witkin 81], [Bajcsy 76].

L'utilisation de la **photométrie** permet également la mesure de la distance en déterminant l'orientation des différentes surfaces visualisées. Le monde ordinaire est composé pour une grande part d'objets tridimensionnels opaques : le niveau de gris d'un point dépend de la direction et de l'intensité de la source lumineuse, des matériaux composant l'objet et également de sa position géométrique. Il est alors possible de déterminer la normale aux surfaces visualisées en s'appuyant sur leurs propriétés de réflectance, à condition d'avoir une mesure très précise des niveaux de gris des points et d'étudier des objets possédant des surfaces homogènes. Parmi les nombreux travaux effectués en photométrie, nous pouvons citer ceux fondés sur l'étude des surfaces mates (surfaces lambertiennes[3]) [Horn 70], ceux utilisant la photométrie à quatre sources lumineuses [Coleman 82] pour éliminer l'effet de surbrillance de certaines surfaces, et ceux étudiant la propriété miroir[4] des surfaces [Castan 84b].

Cependant, ces méthodes ne fournissent que des indications concernant l'orientation des surfaces et non des informations tridimensionelles absolues. Ceci permet donc de construire, sous de bonnes conditions, les surfaces à une "homothétie 3d" près. Beaucoup de progrès ont été fait dans le domaine de la photométrie ; cependant, l'inconvénient principal de ces méthodes est d'utiliser un modèle de réflexion approximatif, alors que la réalité est bien plus complexe. De ce fait, aucun système de vision par ordinateur opérationnel ne les utilise actuellement.

Une autre façon d'acquérir passivement l'information 3d d'une seule image est d'**identifier** les formes bidimensionnelles visualisées avec un modèle objet tridimensionnel connu. C'est de cette manière que procèdent, entre autres, R.A. Brooks et R. Horaud. Le premier, dans le système ACRONYM [Brooks 78], définit les apparences 2d des objets 3d du modèle. Ces prédictions sont ensuite vérifiées sur les images [Brooks 81]. A l'inverse, le second a développé une stratégie de recherche permettant d'établir de manière unique une correspondance entre un ensemble d'indices images et un modèle d'objet 3d spécifique [Horaud 88]. Cette méthode utilise une projection inverse des indices images dans l'espace 3d et combine des contraintes géométriques à des heuristiques pour trouver une solution. Le résultat est la position spatiale et l'orientation (6 degrés de liberté) de chaque objet reconnu. La méthode a l'avantage d'être valide pour une grande variété d'images perspectives et ne requiert pas une segmentation de bas-niveau parfaite. Cette technique de reconnaissance est cependant limitée aux objets contenant suffisamment de lignes droites et de surfaces planes. Le domaine d'application concerne donc essentiellement les objets manufacturés.

[3]Les surfaces lambertiennes possèdent une finition mate idéale et leur fonction de réflexion, très simple, est seulement proportionnelle au cosinus de l'angle incident de la lumière. De telles surfaces ont la propriété de paraître sous la même intensité lumineuse, quelle que soit la position de l'observateur.

[4]Les surfaces miroir, comme les surfaces lisses métalliques, réfléchissent la lumière selon un angle de réflexion égal à l'angle d'incidence et selon une direction telle que les vecteurs d'incidence, de réflexion et de normale à la surface soient coplanaires.

Avant de poursuivre, remarquons que la vision humaine procède également de manière passive, en accord avec la définition donnée précédemment[5]. L'homme perçoit un espace tridimensionnel bien que chacune des images sur la rétine soit bidimensionnelle. La vision humaine fait alors coopérer plusieurs sources d'informations, telles que la taille et la projection perspective des objets, les recouvrements partiels, la variation d'éclairage des surfaces continues, le mouvement et la focalisation des yeux sur un endroit intéressant du champ visuel, l'identification de certaines formes avec un modèle connu similaire, et la **stéréovision**. Cette dernière méthode peut être automatisée et permet, en utilisant deux images, de percevoir le relief et de fournir une description 3d absolue de la scène.

Précisons qu'une simple transformation géométrique (projection) permet de calculer la position dans un plan image des points 3d réels connus, mais que la transformation inverse fournit pour chaque point de l'image, une infinité de points 3d situés sur une droite dans le monde réel et vérifiant la projection précédente. L'idée de base de la stéréovision est d'utiliser une seconde image de la même scène prise sous un angle de vue différent, afin de lever cette indétermination.

Les principes et méthodes de la stéréovision sont développés au chapitre suivant, aussi nous ne nous étendons pas davantage ici sur ce sujet.

4.3 Notre système de perception de l'environnement

Les différentes méthodes de perception sont généralement comparées sur les caractéristiques suivantes :

- vitesse d'acquisition,
- densité des points acquis,
- précision obtenue sur les résultats,
- type des résultats : coordonnées 3d avec ou sans informations de type photométrique.

[5]Sans modification, lumineuse, acoustique, ..., de l'environnement observé.

Le choix du système d'acquisition 3d dépend généralement de l'application considérée et de la qualité de l'information demandée.

Les systèmes actifs sont utilisés avec succès pour des applications :

- où les objets à étudier sont proches : les capteurs laser ou à ultras-sons sont limités en distance,
- où la taille de la scène est relativement petite afin d'obtenir une densité correcte de points 3d : le temps d'acquisition est fonction du temps de balayage de la scène,
- et lorsque les conditions d'environnement le permettent : éclairement particulier respecté, aucune interférence avec le milieu étudié, bonnes caractéristiques des surfaces étudiées.

Par exemple, il est particulièrement efficace d'utiliser un capteur actif, de type laser, pour la reconnaissance et la manipulation d'objets manufacturés, où les conditions précédemment énoncées sont respectées. Dans les autres cas, étude du relief à l'aide d'images aériennes, navigation d'un robot mobile, ..., il semble plus naturel d'utiliser une méthode passive.

Cependant, cette classification "type de systèmes" en fonction du type d'images n'est qu'indicative. En effet, certains robots mobiles sont équipés exclusivement de capteurs actifs [Moravec 85], le relief peut être perçu avec des images aériennes radar [Pettengill 81], et la stéréovision est utilisées dans le domaine médical [Camillerapp 87] ou pour certaines pièces manufacturés [Demazeau 85], ...

Remarquons que certains systèmes de vision combinent les deux types de systèmes pour en conserver les avantages respectifs. Un système passif est utilisé pour obtenir des informations globales sur le monde observé, celles-ci servent ensuite à guider un capteur actif dans une exploration plus locale, mais également plus complète et quelques fois plus précise. Par exemple, lors de la navigation d'un robot, il est possible d'utiliser une méthode de stéréovision pour déterminer une carte tridimensionnelle vaste et un capteur de proximité de type acoustique (dans la mesure du possible) pour contourner les obstacles proches ou déterminer plus complétement un objet particulier de la scène.

Le cadre de notre étude est plus particulièrement la description tridimensionnelle d'une scène observée par un robot mobile évoluant dans un environnement de bureau, d'appartement ou d'atelier, sous des conditions naturelles d'éclairage. Cette étude est intégrée au projet ORASIS. Les limitations inhérentes aux méthodes actives d'acquisition de l'information 3d nous ont fait préférer une méthode d'acquisition passive, en l'occurence la **stéréovision**, car elle est la seule à fournir l'information de distance absolue nécessaire pour connaître la position du robot mobile.

Le résultat doit être une carte tridimensionnelle la plus dense possible de la scène observée. Cette carte doit aider le robot à se localiser dans la pièce, à y naviguer correctement, et à enrichir ses modèles de scène par apprentissage. De plus, il est envisagé de faire une interprétation des objets vus, soit par appariement de la scène

avec un modèle 3d connu et/ou appris par le robot[6], soit par interprétation directe des données à l'aide, par exemple, de règles et d'un moteur d'inférence[7].

Les méthodes d'interprétation des données ne sont pas décrites car elles dépassent le cadre de cette thèse et sont seulement en phase d'étude pour le projet ORASIS. Nous les aborderons seulement dans la conclusion.

[6] Approche choisie au CRIN.
[7] Méthode testée à l'INRIA-Sophia-Antipolis.

Chapitre 2

STÉRÉOVISION

Ce chapitre est consacré aux généralités de la vision stéréoscopique. Nous présentons les différentes façons d'acquérir des images stéréoscopiques, les méthodes de mise en correspondance. Nous décrivons également la géométrie et les contraintes d'un système de vision binoculaire et exposons les points délicats de ce type de vision.

1 Généralités

Le principe de base de la stéréovision est de comparer deux images d'une même scène prise sous deux angles différents. La première étape consiste à **apparier** des indices visuels homologues extraits des deux images. Ces indices visuels, aussi appelés primitives images, peuvent être des points, des segments de contours, des surfaces d'objet et/ou des objets .... Deux primitives images, une dans chaque image, mises en correspondance représentent les deux projections de la même primitive réelle dans la scène.

La seconde étape est la **construction tridimensionnelle** de la scène : connaissant les paramètres du système optique de prise de vue obtenus lors d'une étape de calibration, on calcule pour chaque paire de primitives images homologues les coordonnées réelles 3d de la primitive dans la scène. La connaissance de la disparité entre les deux images stéréoscopiques permet, par triangulation, de calculer la profondeur. Une bonne introduction à la vision binoculaire peut être trouvée dans [LimozinLong 86].

Nous nous proposons de présenter ci-après les différents problèmes inhérents à la vision stéréoscopique, c'est-à-dire :

- l'acquisition des images,
- la mise en correspondance (ou appariement),
- la détermination de la géométrie du système optique,
- le calcul des coordonnées tridimensionnelles.

Nous devons accorder une attention particulière au problème de l'appariement : il représente la phase cruciale et délicate de tout système de vision stéréoscopique, et permet ensuite de calculer les coordonnées tridimensionnelles attendues. La mise en correspondance doit être traitée avec soin, c'est d'elle dont dépend la majeure partie des résultats finaux. Un grand nombre de travaux trouvés dans la littérature ont été entièrement consacrés à ce problème fondamental de l'appariement.

Mentionnons que certains systèmes de stéréovision utilisent trois images ou plus, essentiellement pour supprimer les ambiguïtés d'appariement qui subsistent ou pour accélérer le temps de traitement. La dernière partie de ce chapitre est consacrée à cette vision, que nous qualifions de "stéréovision multiple".

2 Acquisition des images

L'acquisition peut être effectuée de deux manières différentes : soit en utilisant une seule caméra enregistrant successivement des images, soit par l'emploi d'autant de caméras situées à des emplacements différents que d'images désirées.

2.1 Utiliser une seule caméra

Cette première solution est moins coûteuse, ne demande la calibration que d'une seule caméra et permet d'enregistrer séquentiellement les images. Elle impose cependant soit de déplacer la caméra, soit de la munir d'un miroir pivotant et nécessite donc de connaître avec précision la position de la caméra après chaque mouvement ce qui est très difficile dans le cas d'un déplacement. Cette méthode est toutefois utilisée quand la profondeur de la scène demande une grande distance entre les deux endroits de prise de vues des images : images aériennes par exemple.

Pour s'affranchir de ces inconvénients, il est possible d'utiliser une seule caméra enregistrant en même temps deux images. Un tel capteur a été conçu et développé par Y. Demazeau (LIFIA) dans le but d'allier la simplicité du matériel avec la facilité d'utilisation. Nous en donnons ci-après le principe, mais les détails concernant la structure matérielle, le principe de la calibration et une application de localisation de fils électriques peut être trouvée dans [Demazeau 85].

Une "image gauche" et une "image droite" sont envoyées par un système optique simple composé d'un miroir gauche, d'un miroir droit et d'un prisme central, vers l'objectif d'une caméra (figure 2.1). L'image reçue est alors coupée verticalement en deux moitiés correspondant à l'image gauche et à l'image droite de l'ensemble de la scène. La résolution stéréoscopique n'est alors que la moitié de la résolution de la caméra.

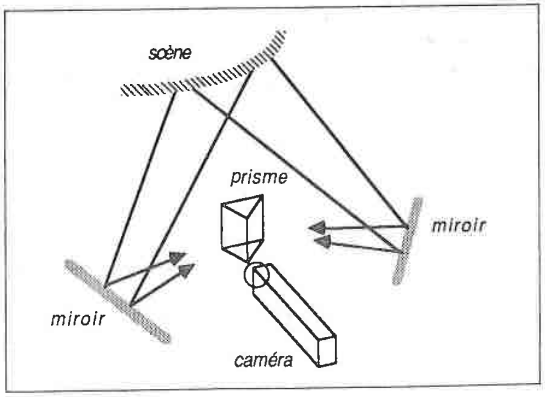

Figure 2.1: *Capteur tridimensionnel du LIFIA.*

Ce système a l'avantage de simplifier la calibration du capteur puisqu'une seule caméra participe à l'acquisition, la géométrie du système est unique. De plus la simultanéité d'acquisition des vues droite et gauche permet de capter des scènes mobiles.

2.2 Utiliser plusieurs caméras

L'inconvénient d'un système qui utilise plusieurs caméras est de devoir connaître d'une part le réglage de chacune des caméras et d'autre part la géométrie du système d'acquisition[1] ; mais ces mesures restent bien sur valables pour toutes les prises de vues ultérieures.

Le montage du capteur est simplifié : aucun prisme ni miroir. De plus, si le système de vision reste rigide, il peut être facilement embarqué. C'est un capteur de ce type que nous utilisons pour notre projet : le robot mobile construit à l'INRIA-Rocquencourt est muni de deux caméras. La figure 2.2 montre le robot et le support horizontal des caméras.

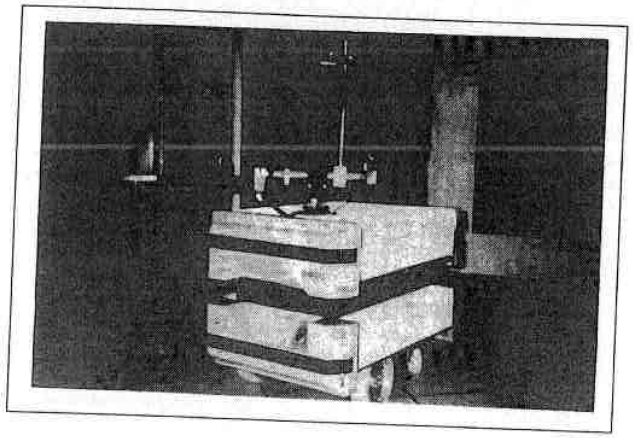

Figure 2.2: *Robot mobile et support des caméras (banc stéréoscopique) INRIA-Rocquencourt.*

3 Mise en correspondance

Le problème fondamental est de trouver les **appariements** entre les deux images, c'est-à-dire de mettre en correspondance les primitives images homologues. Nous appelons **primitives** des indices images représentant des ensembles caractéristiques de points. De telles primitives peuvent être, par exemple, des lignes de contraste, des segments de contours, des régions, ou tout simplement des points. Le troisième chapitre est entièrement consacré à une présentation détaillée des primitives généralement employées.

Avant de présenter diverses méthodes de mise en correspondance en stéréovision, remarquons que l'appariement est une technique fondamentale en analyse d'images

[1] les relations de position entre les caméras

et que de nombreuses méthodes ont été développées. Celles-ci se regroupent en deux grandes classes :

- appariements scène - modèle,
- appariements scène - scène.

Les méthodes de la première catégorie sont généralement utilisées pour résoudre les problèmes d'interprétation de scènes. Une mise en correspondance "grossière" est suffisante dans la plupart des cas. La description du modèle sert ensuite de référence pour interpréter les différentes primitives extraites de la scène. Dans ce domaine, citons les systèmes ACRONYM [Brooks 78], PVV [Lux 84], HYPER [Ayache 86] et le système de vision tridimensionnelle TRIDENT implanté dans notre laboratoire [Belaid 86].

TRIDENT est inspiré d'ACRONYM ; il est constitué d'un ensemble d'outils de perception réalisant l'extraction d'indices visuels 3d [MullerBelaid 84] et d'une étape d'interprétation. Les objets utilisés pour l'interprétation sont définis par composition de primitives et le type de modèle utilisé atteint rapidement un haut degré de complexité. Les différents modèles décrivant de façon générique les scènes en fonction des objets qui les composent, étant actuellement entrés "à la main", il convient de les valider afin de les débarasser de toute les incohérences possibles. La validation des modèles employés dans TRIDENT est assurée par le système MEPHISTO [Camonin 87].

La seconde catégorie d'appariement s'intéresse aux problèmes de comparaison d'images. De telles méthodes s'appliquent au domaine de la stéréovision, mais également à ceux de l'analyse du mouvement, de la détection de cibles, de la classification d'images, ... [Rosenfeld 82], où l'on utilise des séquences d'images ou des systèmes de vision à caméras multiples. Les attributs des primitives sont de type géométrique et statistique (taille, forme, couleur, ...), plus rarement de type symbolique (table, bureau, ...).

Les différentes images utilisées pour des appariements de type scène - scène sont généralement liées par de fortes **contraintes géométriques**, et les méthodes d'appariements, quelles qu'elles soient, prennent en compte les particularités du système de prise de vue utilisé.

Pour résoudre le problème de l'appariement des primitives homologues en stéréovision, plusieurs approches ont été explorées. Citons, entre autres :

- la **modélisation de la vision humaine**. Les travaux de B. Julesz [Julesz 71] lui ont permis de tester des phénomènes se produisant au niveau de la rétine ou au niveau du cortex de manière quasi-indépendante. Une "structure" de vision en relief a été proposée où la vision binoculaire s'opère sans qu'aucune reconnaissance de formes de hiérarchie supérieure (cortex) ne se fasse au niveau monoculaire.

  Une équipe pluri-disciplinaire composée de biologistes, de psychologues, d'informaticiens, et de mathématiciens s'est intéressée à la fois aux problèmes de connaissance et de modélisation de la vision binoculaire. Le résultat de cette

étude a été de proposer une théorie de la stéréovision chez l'homme, de la modéliser sur le plan algorithmique [Marr 76] et de l'implanter sur ordinateur [Grimson 81].

Les recherches concernant cette approche se poursuivent, comme le confirment, entre autres, les récents travaux de N.C. Griswold [GRISWOLD86.

- la **corrélation** [Moravec 77], [Castan 84a]. L'idée est de construire une correspondance entre les images, généralement point par point. On s'efforce donc d'appliquer une partie d'image sur l'autre image en la "déformant" le moins possible. Le principe est d'extraire une fenêtre autour d'une primitive à apparier, puis de chercher à réaliser la meilleure superposition de cette fenêtre "sur" l'autre image. La taille des fenêtres peut être fixe : 10×10 points par exemple [Henderson 78], ou variable [Yakimovsky 78]. On estime ensuite la disparité qui en résulte. Cette approche fournit d'excellents résultats pour certaines applications où les images stéréoscopiques sont bien texturées ; mais les déformations de perspective sont généralement mal traitées. Cette mise en correspondance rigide limite cette technique, aussi certaines méthodes ont vu le jour pour rendre l'appariement plus élastique. Signalons pour terminer que les méthodes de corrélation sont très gourmandes en temps de calcul mais sont assez facilement réalisables en "matériel".

- la **relaxation** [Berthod 84], [Medioni 85]. Le principe est d'affecter un coefficient de vraissemblance à chaque couple potentiel de primitives. L'influence du voisinage intervient ensuite pour faire évoluer ce poids. La notion de voisinage permet de modéliser facilement les contraintes imposées par le problème.

- les techniques de **programmation dynamique** [Benard 83], [Ohta 83], [Mohr 84]. Les algorithmes mis en œuvre sont fondés sur une comparaison dynamique des primitives des deux images. Ils ont l'avantage d'être tolérants avec les petites distorsions, donc de bien traiter les parties cachées des objets. Cette caractéristique est relativement importante en stéréovision.

- les méthodes utilisant dès **contraintes structurelles** pour s'affranchir de la rotation ou du changement d'échelle du système de coordonnées [Hwang 82]. Les différentes primitives (segments, régions, points d'intersections de segments) sont étiquetées et les relations structurelles entre primitives voisines sont déterminées. Il s'agit ensuite, connaissant toutes les modifications possibles des relations entre deux vues, de trouver les meilleurs appariements de primitives.

  Dans les récents travaux de R. Horaud, chaque primitive image est caractérisée par sa position, son orientation et également par les types de relations géométriques avec les primitives voisines [Horaud 87b]. L'appariement tient compte des similarités d'ordre topographique entre les deux images traitées.

- la **prédiction-vérification d'hypothèses**. Le principe est de faire des hypothèses de mise en correspondance de certaines primitives, de construire des

appariements respectant les contraintes données par les appariements précédents, puis de sélectionner les appariements issus des hypothèses ayant fournies le plus de couples de primitives compatibles [Ayache 85].

Signalons dès à présent que l'algorithme que nous avons développé s'intègre dans cette dernière catégorie : prédiction, puis vérification d'hypothèses, tout en incluant une technique de programmation dynamique lors de la génération des hypothèses.

Nous reviendrons plus en détail sur certaines de ces méthodes dans les chapitres suivants lors de la présentation des différentes primitives images pouvant être utilisées. Mais précisons tout d'abord la **géométrie** des systèmes de vision binoculaire, géométrie respectée par l'ensemble des méthodes de stéréovision.

4 Géométrie du système optique de stéréovision

Nous présentons dans cette partie la géométrie d'un système de vision binoculaire, appelée géométrie épipolaire, et les contraintes qui lui sont liées. La connaissance du système optique utilisé permet d'une part, de faciliter la mise en correspondance et d'autre part, de calculer les coordonnées des points 3d de la scène. Pour ce faire, il faut également connaître les modèles des caméras utilisées, ceux-ci sont déterminées lors d'une étape de calibration.

Nous exposons successivement les trois parties suivantes :

- la géométrie épipolaire,
- les lois d'unicité et de continuité qui contraignent la mise en correspondance,
- la calibration.

4.1 Construction d'une image

Du point de vue physique, le centre focal (ou centre optique) d'une caméra est situé entre la scène et le plan image. Ceci produit une image inversée sur le plan de vue comme l'illustre la figure 2.3.

Cependant, il revient au même de raisonner en plaçant le centre optique en arrière du plan image. Cette représentation permet de simplifier les calculs et le raisonnement (figure 2.4).

Ceci est la raison pour laquelle tous les schémas d'illustration de la géométrie du système optique présentés dans la suite de ce document situent le centre optique en arrière du plan image. Ceci n'a, bien évidemment, aucune influence sur le traitement informatique réalisé.

4.2 Géométrie épipolaire

La connaissance de la géométrie épipolaire permet de définir les zones de correspondance privilégiées de deux images. Ces zones ne dépendent que de la position relative des deux caméras comme cela est expliqué ci-après.

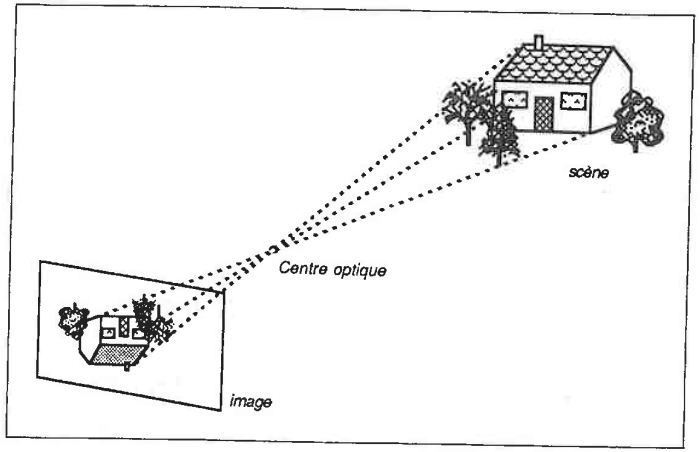

Figure 2.3: *Acquisition d'une vue d'une scène par un système optique. Le centre optique est situé entre la scène et le plan sur lequel se forme l'image.*

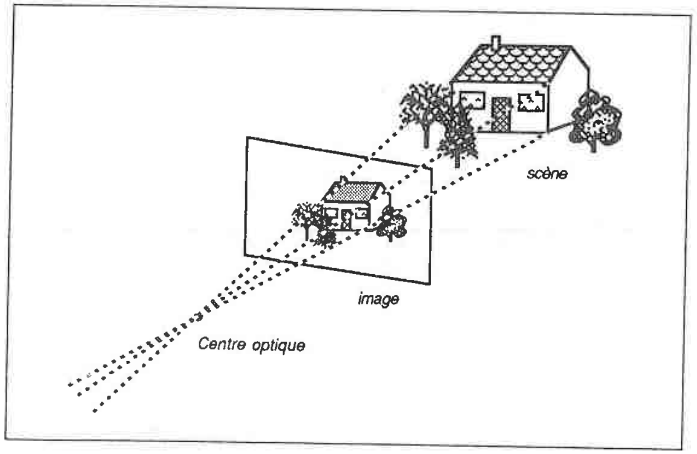

Figure 2.4: *Acquisition d'une vue d'une scène par un système optique. Le centre optique est représenté en arrière du plan sur lequel se forme l'image.*

Soient C_1 et C_2 les *centres optiques* des deux caméras 1 et 2 ; ils sont également appelés les centres focaux, ou encore les "foyers" de chacune des caméras (figure 2.5).

Soient $Proj_1$ et $Proj_2$ les deux *plans de vues* ou *plans images* associés aux caméras.

Soit P un point dans scène tridimensionnelle.

Le point P_1 est défini par l'intersection de la demi-droite (P, C_1) avec le plan $Proj_1$; on l'appelle point image puisqu'il appartient à l'image de la caméra 1, il représente également la *projection* du point P sur $Proj_1$.

Nous avons la définition symétrique pour P_2.

Puisque les points P_1 et P_2 sont les projections d'un même point P de la scène, ils sont qualifiés de points *homologues*, ou synonynement de points *correspondants*.

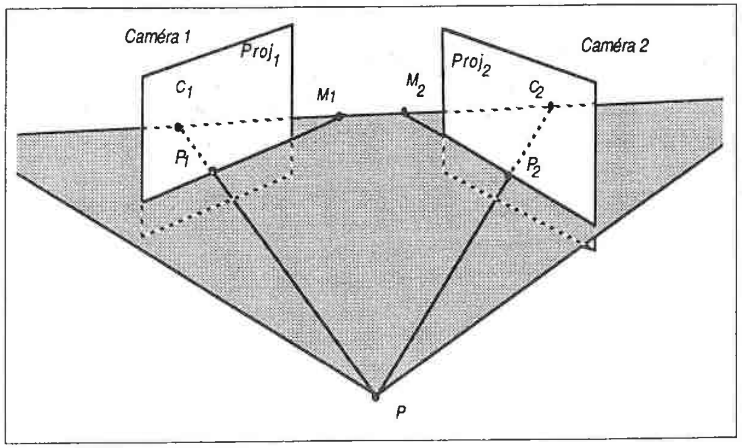

Figure 2.5: *Géométrie épipolaire.*

On appelle *plan épipolaire*, le plan défini par les deux centres optiques C_1 et C_2 et le point P.

On appelle *droite épipolaire*, l'intersection d'un plan image avec le plan épipolaire, cette intersection existe toujours. Les deux droites épipolaires portant chacune les projections du point P sont appelées droites épipolaires *associées* ou *conjuguées*.

Les deux centres optiques sont fixes alors que le point P peut représenter tout point dans la scène. Ceci implique que tous les plans épipolaires sont sécants ; leur intersection est la droite passant par C_1 et C_2. De ce fait, toutes les droites épipolaires appartenant à un même plan image sont sécantes en un point de la droite (C_1, C_2). Ce point est appelé *centre épipolaire* ou *épipôle*. M_1 et M_2 représentent respectivement les centres épipolaires des plans images $Proj_1$ et $Proj_2$.

Le centre épipolaire M_1 correspond également à la projection du centre optique

C_2 de l'autre caméra sur le plan $Proj_1$. Réciproquement pour M_2.

4.3 Problème de l'appariement

Les données dont nous disposons pour effectuer l'appariement sont les suivantes :

- les deux images $Proj_1$ et $Proj_2$,
- les couples de droites épipolaires associées.

Le problème de l'appariement est de trouver les couples de points homologues. Il s'énonce de la façon suivante :

> *Etant donné un point dans l'une des deux images, trouver son homologue dans l'autre image.*

Du fait de la géométrie épipolaire, les points images homologues sont situés sur un couple de droites épipolaires associées, le problème se réduit alors à :

> *Etant donné un point dans l'une des deux images, trouver son homologue sur la droite épipolaire associée dans l'autre image.*

Cette droite épipolaire est le lieu géographique de tous les homologues potentiels du point considéré.

La géométrie épipolaire permet donc de réduire fortement la dimension du problème de mise en correspondance : la recherche des appariements peut se faire dans un espace à une dimension (droite épipolaire) au lieu de deux (plan image).

4.4 Lois d'unicité et de continuité

Des contraintes, liées au monde réel observé, doivent également être prises en compte :

- Les points images proviennent d'une scène ayant une certaine consistance et cohérente. Plus précisément des points voisins dans une image appartiennent vraissemblablement à la même surface d'un objet et sont donc situés à une profondeur voisine.
- De plus, nombre de surfaces sont opaques et cachent les objets situés derrière elles, en fonction de l'axe de visée des caméras. Un point image est donc la projection d'un et d'un seul point de la scène.

Ces deux contraintes ont été formalisées par D. Marr [Marr 77] lors de ses travaux sur la vision binoculaire et sont énoncées selon les deux "lois" suivantes :

Loi d'unicité : Chaque point de la scène possède une projection unique sur chacune des images. Chaque point d'une vue a donc au plus un correspondant unique dans l'autre vue.

Il existe donc des points images qui n'ont aucun correspondant dans l'autre vue, leur homologue se trouve sur la surface cachée d'un objet.

Loi de continuité : La variation relative de la scène est faible ; les surfaces sont localement continues.

Cette dernière loi exclut donc toute variation brutale de la disparité, c'est-à-dire de l'écart entre les représentations dans les images des primitives homologues. Les fortes variations de distance situées à la frontière des objets représentent un cas d'exception à cette loi.

Ces règles ne sont pas absolues, mais fortement probables ; par exemple, des objets de la scène produisant des reflets comme des vitres ou des miroirs ne respectent pas la première loi. Tout le problème est de satisfaire ces règles **parfaitement presque partout** et non partout presque parfaitement.

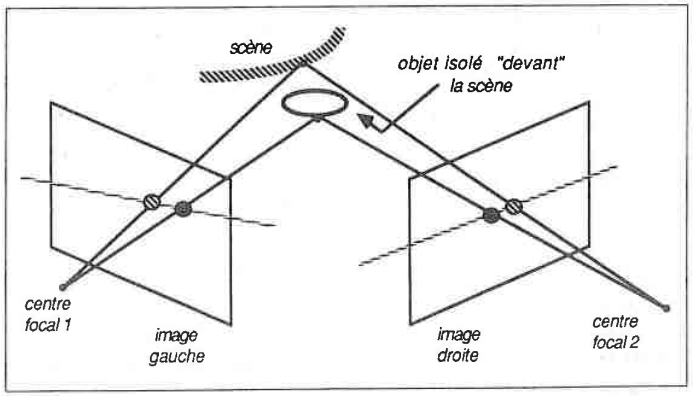

Figure 2.6: *Inversion des projections perspectives le long des épipolaires pour un objet se situant en avant de la scène.*

Il est à noter que, pour une majorité d'images, les projections des objets apparaissent dans le même ordre dans les images le long des épipolaires. Une primitive A, perçue à droite d'une primitive B dans l'une des images, a la même disposition relative dans l'autre image. Cette hypothèse conduit à une troisième "loi" :

Loi d'ordre : Les projections de points de la scène apparaisssent dans le même ordre sur les couples d'épipolaires associées.

Cette troisième règle peut simplifier la mise en correspondance le long des lignes épipolaires et permettre ainsi d'utiliser des techniques de programmation dynamique pour résoudre le problème de l'appariement.

Cette règle est cependant mise en défaut pour les objets isolés et en avant de la scène comme les nuages dans les images satellites ou les pattes de chaises dans les scènes d'intérieur. L'hypothèse communément admise est que l'existence de tels objets est limitée, environ 5 % des primitives, sauf pour des cas particuliers comme celui de la localisation de fils électriques [Demazeau 85].

4.5 Calibration

Rappelons que nous travaillons avec des scènes naturelles, captées par des caméras, et qu'il est indispensable de tenir compte des **paramètres des capteurs** utilisés et de leur **position relative**, afin d'obtenir les résultats les plus corrects possibles.

En pratique, il est difficile de mesurer directement de telles valeurs avec suffisamment de précision. Elles sont donc généralement calculées par une étape de **calibration**. Le terme **étalonnage** est également très utilisé dans la littérature.

La calibration comporte en fait deux phases :

- calibrer indépendamment chacune des caméras,
- trouver les relations entre les caméras.

Calibrer une caméra

Calibrer une caméra, c'est mesurer la transformation qui fait passer d'un point 3d du monde réel à sa projection sur le plan d'une image (figure 2.7). Cette transformation est liée aux quatre paramètres optiques intrinsèques de la caméra : la distance focale, les deux coordonnées du centre optique et le facteur d'échelle. Nous avons utilisé pour le projet ORASIS des caméras CCD, aussi il faut ajouter un second facteur d'échelle aux paramètres des caméras.

La méthode généralement implantée pour déterminer la transformation se fait par la résolution d'un système d'équations : on utilise un certain nombre de points de la scène dont on connaît les coordonnées 3d, et dont on localise chacune des projections sur les plans images.

La scène utilisée pour calibrer notre système de vision est un plan sur lequel a été dessiné, par table traçante, un quadrillage de lignes noires sur fond blanc. Cette mire est représentée figure 2.8. Les points utilisés pour la calibration sont les intersections de ces droites. Les coordonnées réelles de ces points sont donc connues très précisément, car tracées de manière automatique avec une précision supérieure au $1/10^{ème}$ de millimètre.

A partir de l'image de cette mire fournie par la caméra, on extrait et on chaîne les points de contraste représentant les lignes. On effectue ensuite une approximation polygonale des chaînes obtenues afin de retrouver les points d'intersection des droites.

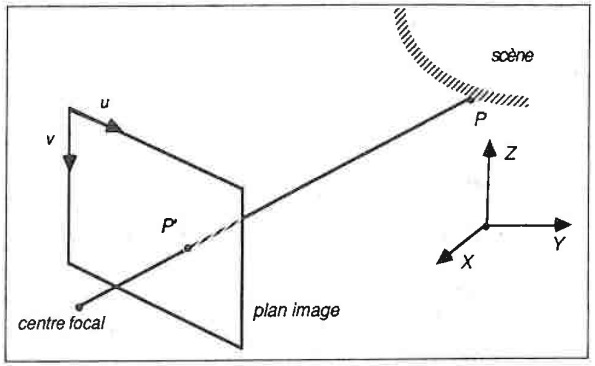

Figure 2.7: *Transformation perspective. Mesure de la transformation qui fait passer d'un point $P(x, y, z)$ de la scène à sa projection $P'(u, v)$ sur le plan image de la caméra.*

Pour chaque point considéré, nous avons à notre disposition ses coordonnées 3d (en centimètres) dans le repère de la scène et les coordonnées images (en pixels) de sa projection. En théorie, et selon la méthode employée, seuls 3 ou 5 points suffisent pour déterminer la transformation. Mais, pour augmenter les précisions obtenues, les 113 points de la mire sont effectivement utilisés. Une méthode d'approximation au sens des moindres carrés permet d'extraire les coefficients des matrices de transformation, mais l'utilisation du filtre de Kalmann permet de les estimer de façon optimale [Toscani 87a].

Calibrer un couple de caméras

Le but est de trouver les relations entre les deux images, c'est-à-dire de déterminer les lignes épipolaires.

Par définition, le centre épipolaire de la caméra gauche est la projection du centre focal de la caméra droite sur le plan image gauche (et réciproquement).

La matrice de projection de chacune des caméras est connue de par l'étape précédente, les coordonnées des centres épipolaires peuvent donc être calculés. Il est ensuite possible, pour un point donné dans une image, de déterminer l'équation de la ligne épipolaire dans l'autre image sur laquelle se trouve son homologue.

Réalisation de la calibration

La calibration s'avère indispensable pour résoudre correctement le problème de la vision stéréoscopique. Elle permet de déterminer les paramètres optiques de chaque caméra (projection perspective) et la relation géométrique reliant ces deux mêmes

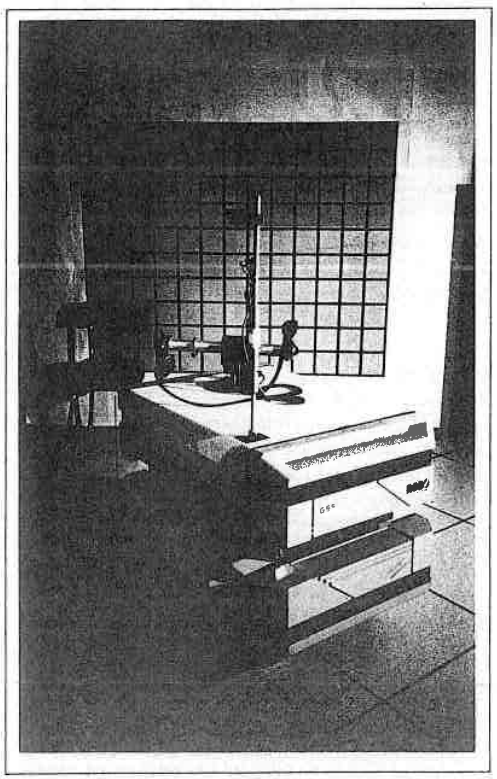

Figure 2.8: *Mire utilisée pour la calibration du banc stéréoscopique [Faugeras 87]. Robot et support des caméras au premier plan.*

caméras.

Ces données sont utilisées pour calculer les couples de lignes épipolaires d'une part (figure 2.9) et pour construire les points 3d par perspective inverse d'autre part, quand on connaît leur deux projections images.

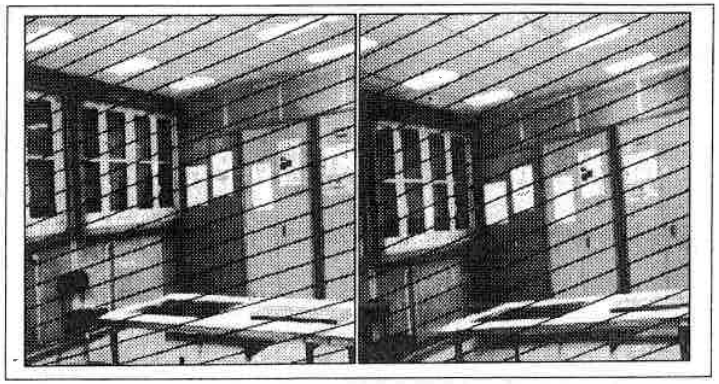

Figure 2.9: *Couples de lignes épipolaires associées.*

La calibration de notre système a été réalisée à l'INRIA-Rocquencourt par G. Toscani, et l'on pourra consulter [Faugeras 86b], [Faugeras 87] et [Toscani 87a] pour plus de détails. Ce laboratoire nous a fournit les images stéréoscopiques accompagnées de leur calibration. Les paramètres du banc stéréoscopique se présentent plus précisément sous la forme suivante :

- deux matrices de transformation perspective, une pour chaque image,
- une matrice permettant, connaissant un point image d'une vue de trouver l'équation de la droite épipolaire portant son homologue dans l'autre vue et une seconde matrice pour le problème symétrique.

Des exemples de telles matrices sont présentés Annexe C.

5 Calcul des coordonnées tridimensionnelles

Cette étape permet de passer des couples de primitives images à leur représentant tridimensionnel dans la scène par perspective inverse.

Soit un appariement de points P_1 et P_2 entre les deux images utilisées. Il est possible, pour chacun de ces deux points, de calculer la droite 3d sur laquelle se trouve le point réel 3d P puisque nous connaissons la matrice de projection associée à chaque caméra. La droite Δ_1 est issue du centre focal de la caméra gauche et passe par la projection P_1. La définition symétrique convient à la droite Δ_2.

Le point 3d P cherché se situe à l'intersection de ces deux demi-droites (figure 2.10).

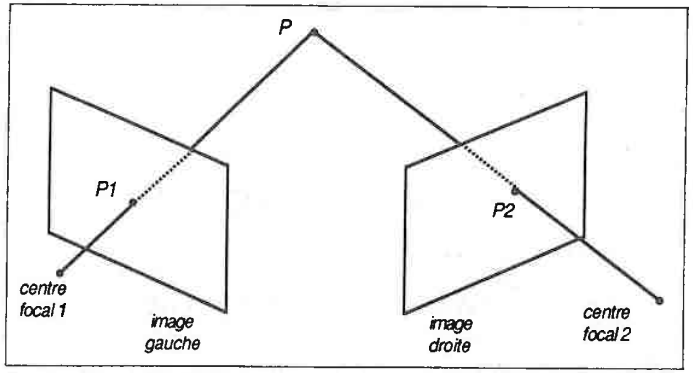

Figure 2.10: *Construction du point tridimensionnel P à partir de ses deux projections images P1 et P2.*

Les deux vues utilisées sont très peu différentes, et de ce fait la précision sur la mesure de la distance est bien moins bonne que sur les autres coordonnées. De plus, les coordonnées réelles calculées par perspective inverse sont très sensibles aux erreurs commises sur les positions des primitives appariées (figure 2.11), et ceci d'autant plus que les vues sont rapprochées.

Rappelons donc que les résultats généralement obtenus par une méthode de stéréovision n'ont pas toujours une très bonne précision et peuvent s'avérer très coûteux en temps de calcul. Ils couvrent cependant une large partie du champ visuel. En ce sens, ils sont complémentaires des capteurs actifs et peuvent leur être couplés de manière efficace. La stéréovision effectuée dans le cadre du projet ORASIS n'est actuellement couplée à aucun autre capteur.

De plus, il est à remarquer que la construction 3d de primitives de type points ou lignes s'effectue facilement, alors que le calcul d'éléments de type surface pose certains problèmes. Ceux-ci seront abordés dans le dernier chapitre de cette thèse.

6 Stéréovision multiple

6.1 Limites de la vision binoculaire

La pratique habituelle en matière de stéréovision est de copier, ou tout au moins de s'inspirer, de la vision humaine, et d'utiliser deux caméras, de préférence alignées horizontalement. Un tel système souffre cependant particulièrement de l'incapacité de déterminer correctement les structures parallèles aux lignes épipolaires, soit le plus souvent les horizontales.

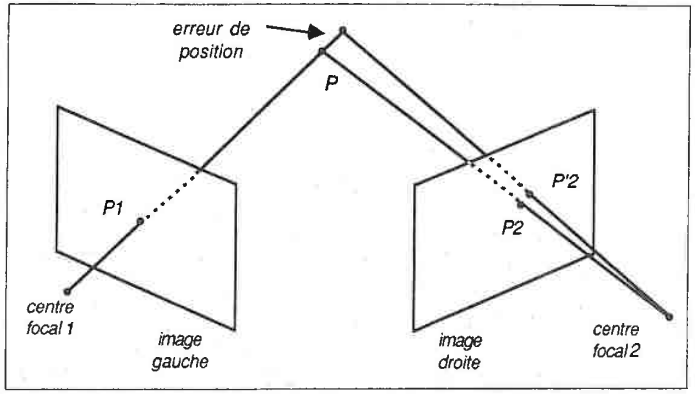

Figure 2.11: *Répercution d'une erreur d'appariement de deux points sur la construction 3d du point résultant. P_1 est apparié à P'_2 au lieu de P_2.*

Prenons l'exemple des primitives images **segments de droite**. Soient deux segments S_1 et S_2 appariés. Le segment tridimensionnel S construit est le segment maximal qui admet S_1 et S_2 comme projections sur les deux plans images. La droite 3d Δ portant le segment S est déterminée par l'intersection dans l'espace des deux faisceaux de plans $Plan_1$ et $Plan_2$ (figure 2.12). $Plan_1$ passe par le centre focal de la caméra gauche et par le segment S_1 de l'image gauche. $Plan_2$ passe par le centre focal de la caméra droite et par le segment S_2 de l'image droite.

Les extrémités du segment 3d S sont calculées à partir des zones des segments S_1 et S_2 pouvant s'associer en respect des contraintes épipolaires. Ceci correspond à l'intersection maximale des faisceaux et protège donc le segment 3d des variations de longueur de S_1 et S_2.

Une incertitude de localisation d'une extrémité d'un segment peut influer, dans certains cas seulement, sur les zones des segments pouvant s'associer et donc sur la longueur du segment 3d S construit (figure 2.13).

Plaçons nous à présent dans le cas où les deux segments appariés sont portés par les lignes épipolaires. Les plans $Plan_1$ et $Plan_2$ définis précédemment sont alors confondus (du fait de la géométrie épipolaire). L'intersection des plans est le plan épipolaire tout entier et ne définit aucune direction pour le segment à construire. De plus, les zones des segments S_1 et S_2 pouvant se correspondre sont les segments tout entiers.

En conséquence, toute incertitude sur les extrémités des segments provoque des variations sur les extrémités du segment 3d mais également, et ceci est beaucoup plus grave, sur son orientation dans le plan (figure 2.14).

L'incertitude provoquée sur le segment 3d étant beaucoup trop importante, les

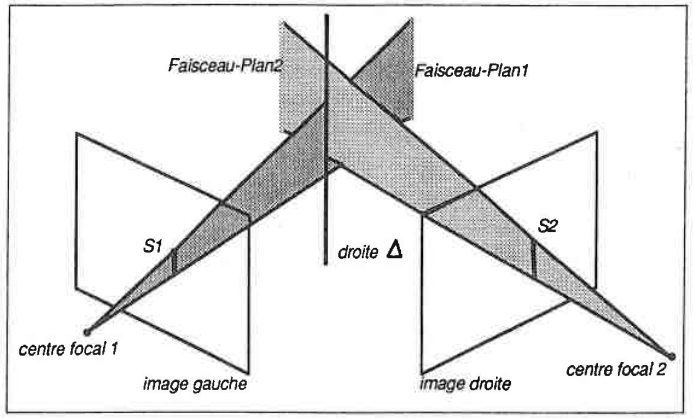

Figure 2.12: *La droite 3d Δ est définie par l'intersection des faisceaux de plans $Plan_1$ et $Plan_2$ issus des segments appariés et contient le segment 3d S à construire.*

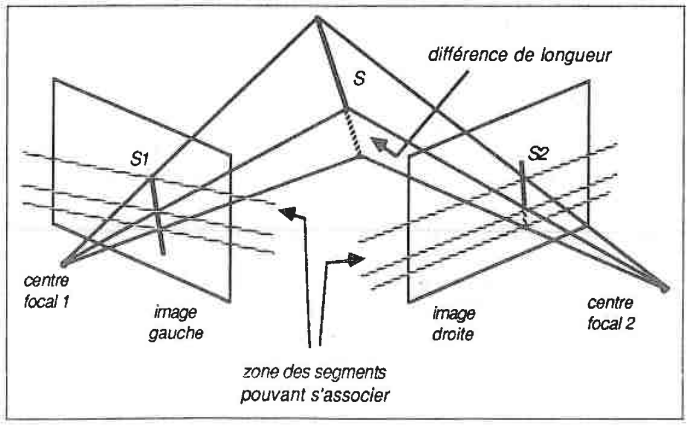

Figure 2.13: *Les segments S_1 et S_2 ne sont pas portés par les épipolaires. Une incertitude sur la longueur d'un segment (ici S_2) provoque une incertitude sur la longueur seulement du segment 3d S construit.*

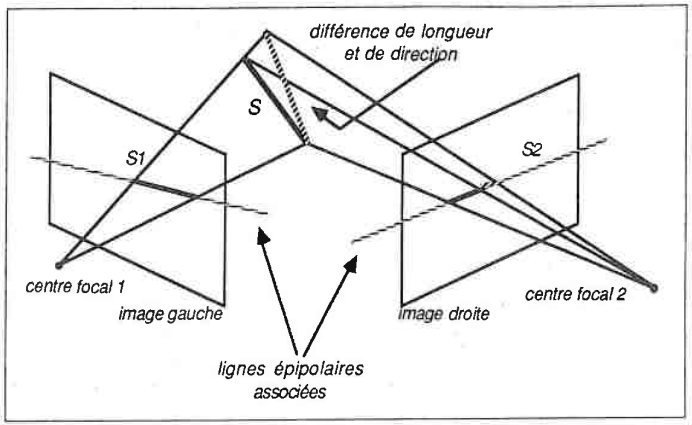

Figure 2.14: *Les segments S_1 et S_2 sont portés par les épipolaires. Une incertitude sur la longueur d'un segment (ici S_2) provoque une incertitude sur la longueur et sur l'orientation du segment 3d S construit.*

segments portés par les lignes épipolaires, ou plus généralement, ceux quasi-parallèles à ces lignes, ne peuvent pas être utilisés lors de la construction de la scène tridimensionnelle. Ceci représente un inconvénient important lorsque les images traitées contiennent de nombreuses primitives de cette sorte.

6.2 Solutions

Reprenons l'exemple des primitives segments lors de l'étude de scènes d'intérieur. Le monde visualisé est essentiellement composé de lignes horizontales et verticales. Un système de vision stéréoscopique "horizontal" fait donc "perdre" la moitié environ des primitives utilisées.

Face à ce problème, signalons que certains chercheurs placent leurs caméras alignées sur un support incliné. Les primitives "perdues" portées par les épipolaires sont obliques et apparaissent donc en nombre limité quand les scènes réelles étudiées sont à prédominantes horizontales et verticales.

D'autres chercheurs, dans le but de d'augmenter le nombre d'appariements et également de restreindre les ambiguïtés d'association, préfèrent utiliser une troisième caméra. Le système de vision "trinoculaire" permet alors de traiter les cas qui présentent des problèmes.

Les méthodes utilisées se partagent généralement en trois catégories :

- Privilégier l'un des couples et utiliser une troisième caméra en contrôle des appariements effectués Ceci s'avère particulièrement utile lorsqu'on utilise une

méthode de prédiction-vérification d'hypothèses : la troisième image sert alors au cours de la phase de génération d'hypothèses. Le temps perdu à rechercher dans la troisième image des appariements sûrs est très largement compensé par le temps gagné lors de la phase de propagation : les hypothèses de mise en correspondance sur lesquelles s'appuie la suite du traitement sont fiables, les retours arrières et/ou les calculs inutiles concernant de fausses hypothèses sont réduits. La phase de propagation peut être simplement supprimée : la troisième caméra offre un moyen de vérification puisque la connaissance d'un appariement entre deux primitives permet de prédire sa position (et son orientation dans le cas d'un segment) de la primitive correspondance dans la troisième vue. C'est de cette manière que procèdent N. Ayache et F. Lustman [Ayache 87], [Lustman 87] pour accélérer leur algorithme de prédiction-vérification d'hypothèses [Ayache 85].

- Travailler avec les trois couples stéréoscopiques issus des trois vues et confronter les appariements effectués [Ohta 86b].
- Rechercher des triplets vérifiant l'ensemble des propriétés géométriques [Pietikainen 86].

Il existe différentes manières pour positionner trois caméras ; par exemple les placer, avec une même orientation de l'axe optique, aux angles d'un triangle rectangle [Ohta 86b] (figure 2.15) ou d'un triangle isocèle [Gerhard 86]. Dans ces deux cas, les lignes épipolaires sont déterminées par la position géométrique des caméras et non par la calibration. Par exemple, les lignes épipolaires des images U et L (figure 2.15) sont définies comme parallèles à l'axe des Y. L'utilisation d'une bonne méthode de calibration, comme celle développée à l'INRIA-Rocquencourt [Faugeras 86b] permet un calcul précis de la géométrie épipolaire et autorise alors de placer les caméras dans n'importe quelle position, généralement avec des axes optiques convergents pour une meilleure restitution des parties cachées.

Plusieurs avantages se dégagent de la stéréovision trinoculaire par rapport à la vision binoculaire :

- réduction importante du temps de calcul pour effectuer les appariements,
- suppression du problème des lignes parallèles aux épipolaires,
- construction 3d plus précise.

6.3 Choix du système de stéréovision

Les systèmes de stéréovision binoculaire sont encore les plus courants car les plus simples à mettre en œuvre et les moins coûteux, mais les systèmes de vision trinoculaire se développent de plus en plus, pour les raisons évoquées précédemment.

Il existe également des systèmes utilisant une succession d'images captées lors du déplacement d'un satellite ou d'un robot mobile par exemple.

Pour notre projet, nous utilisons deux images captées par des caméras placées sur le robot mobile. Pour des raisons de simplicité, ces caméras ont tout d'abord

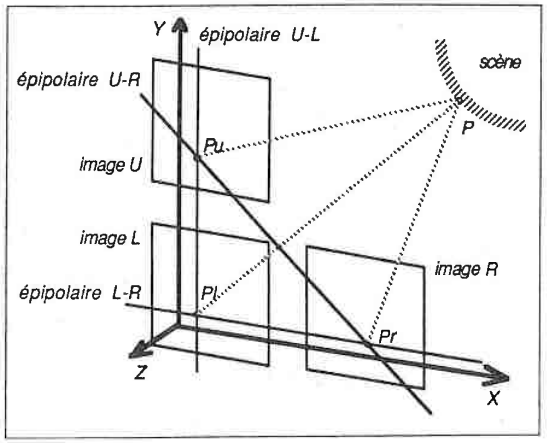

Figure 2.15: *Système de vision trinoculaire où les caméras sont disposées sur un triangle rectangle.*

été posées sur un support horizontal et la majeure partie des résultats présentés au cours de cette thèse sont issus d'images prises par un tel système binoculaire.

Cependant, suite aux inconvénients précédemment cités lors de la construction tridimensionnelle de la scène, nous avons également utilisé des images captées par des caméras situées sur un support oblique. La comparaison entre ces deux types d'acquisition est présentée à la fin de ce mémoire.

Avant de terminer cette partie consacrée à la stéréovision multiple, remarquons que la vision humaine est "seulement" binoculaire. Le système utilisé est cependant continuellement en mouvement et possède une direction épipolaire variable dans le temps, et combine d'autres techniques : étude du recouvrement, identification des objets, ...

Chapitre 3

CHOIX DES PRIMITIVES IMAGES

Dans ce chapitre, nous présentons brièvement notre approche du problème de la vision stéréoscopique et plus particulièrement de l'étape de mise en correspondance. Nous situons cette approche par rapport à d'autres méthodes existantes. L'originalité de notre méthode réside pour une grande part dans le type de primitives images utilisées : les régions, dont nous justifions le choix par comparaison avec les autres primitives images les plus fréquemment employées.

1 Introduction

Il existe deux grandes approches pour résoudre le problème de la stéréovision. La première consiste à construire la scène tridimensionnelle grâce à de nombreux appariements de primitives images simples, généralement des **points**. Dans cette scène composée de points 3d, on cherche ensuite à former des structures : plans, surfaces courbes, etc. L'objectif est d'obtenir une description plus symbolique et de préparer la tâche d'un système d'interprétation de scène.

La seconde approche, que l'on peut qualifier de duale de la première, est de chercher dans les images des primitives plus complexes, c'est-à-dire des regroupements structurés de points. Ces primitives peuvent être de type **segments de contour**, **régions**, etc. Elles ont alors une signification essentiellement physique, comme le contour d'un objet, ou sa surface. Il est également possible de travailler avec des primitives de niveau encore supérieur, comme des ensembles de segments : on s'intéresse par exemple aux **jonctions de segments**, ces jonctions représentent alors un coin de pièce, un angle de bureau, etc.

Avec cette approche, on peut choisir des primitives de plus en plus complexes et leurs donner des caractéristiques de type sémantique. Le cas extrême est d'apparier les descriptions des objets présents dans les images. Ceci nécessite une phase d'interprétation de chacune des images avant d'effectuer leur comparaison. Dans ce cas, l'étape de mise en correspondance est simplifiée : on cherche à apparier des **objets** avec des **objets**, par exemple une chaise avec une chaise, un bureau avec un bureau, etc. Il s'agit ensuite de calculer les coordonnées tridimensionnelles des objets ainsi mis en correspondance. On peut trouver les motivations d'une telle approche dans [Demazeau 86].

La première approche, fondée sur une mise en correspondance de points, comporte donc deux phases délicates :

- l'appariement, du fait du grand nombre de primitives utilisées,
- l'interprétation de la scène à l'aide des points 3d obtenus.

La seconde approche, par appariement de primitives présente également deux problèmes qui sont complémentaires des deux précédents :

- l'extraction des primitives : points d'intérêt, segments, régions, objets étiquetés sémantiquement, etc.
- la construction des primitives 3d, puis de la scène.

Ces deux approches de la stéréovision ne peuvent s'appliquer dans tous les cas, et le choix des primitives images utilisées dépend du type des images traitées. Par exemple, les primitives de type **segment de contour** issues d'images aériennes ou satellites sont peu présentes et donc peu utiles. En revanche, ces mêmes primitives sont significatives lorsque l'on travaille avec des images dans lesquelles on peut détecter des structures rectilignes : vrac de pièces industrielles, vue aérienne de réseaux de canaux ou de routes, scènes d'intérieur, etc.

Notre méthode s'apparente à la seconde approche présentée, sans aller toutefois trop loin dans la complexité des primitives utilisées. Nous manipulons des primitives formées par des ensembles de points, mais nous ne cherchons pas à donner une étiquette sémantique aux structures ainsi extraites ; nous ne tentons pas de les reconnaître. Cette interprétation de scène sera effectuée ultérieurement sur les données tridimensionnelles fournies par notre système de stéréovision. Rappelons que cette phase de plus "haut niveau" ne fait pas partie du cadre de notre projet.

L'approche de stéréovision choisie comporte alors les trois étapes classiques suivantes :

- extraction des primitives,
- appariement de ces primitives,
- calcul des informations tridimensionnelles.

Nous pouvons alors constater que la manière de résoudre le problème de la stéréovision dépend généralement des primitives choisies pour l'appariement, qui dépendent elles-mêmes du domaine d'application considéré. En d'autres termes, "comment apparier" dépend souvent de "quoi apparier".

Nous présentons donc, dans la suite de ce chapitre, les différentes primitives qu'il est possible d'extraire, les méthodes de stéréovision les utilisant actuellement et leur(s) domaine(s) d'application privilégié(s). Cette comparaison va nous permettre d'exposer nos motivations pour un choix de primitives de type **région**, compte tenu du problème particulier que nous avons à résoudre. Nous exposons successivement les primitives images suivantes :

- points,
- segments de contour,
- arcs de cercle,
- régions.

Quelles que soient les primitives images sélectionnées, remarquons que leur extraction doit être faite avec attention pour détecter les mêmes primitives dans les deux images à apparier. Ce traitement, comme nous l'expliquons ci-après, n'est ni facile, ni garantit de succès.

2 Points images

On peut séparer les primitives "points" en deux catégories selon que la mise en correspondance s'effectue sur **tous les points des images** ou seulement sur certains points particuliers, communément appelés **points d'intérêt**. Ces deux catégories sont présentées ci-après.

2.1 Points images

On peut choisir d'apparier les points des **images initiales**, ce qui évite alors de perdre du temps pour extraire des primitives, ou de mettre en correspondance les points des **images traitées** (seuillées, lissées, etc).

Dans les deux cas, si l'appariement des points entre les deux images est convenablement réalisé et fournit un très grand nombre de couples de points homologues, nous sommes en droit d'espérer obtenir une carte dense lors de la construction des informations tridimensionnelles.

Points des images initiales

Les **attributs** des points utilisés pour évaluer les appariements effectués sont le plus souvent locaux, comme l'**intensité lumineuse**.

Les images idéalement traitées possèdent les caractéristiques suivantes :

- Deux points images, projections du même point physique, ont **exactement** la même intensité lumineuse.
- Les zones composées de points voisins de même intensité lumineuse sont souvent **très petites**, ce qui réduit l'ambiguïté lors de la mise en correspondance.

Les méthodes d'appariement de points travaillant sur des images respectant les deux contraintes précédentes fournissent généralement de bons résultats. Il convient de citer les travaux de Marr et Poggio [Marr 77] qui ont conçu le premier et sans doute le plus simple algorithme de vision stéréoscopique, lequel accepte en donnée un stéréogramme aléatoire, et est fondé sur une méthode de relaxation.

Nous avons également choisi d'utiliser une primitive **point image** lors de l'implantation d'une méthode de stéréovision traitant des images d'une scène tridimensionnelle artificielle, l'appariement de points étant alors fondé sur une technique de **programmation dynamique** par recherche de chemin optimal dans un graphe [Mohr 84]. Dans ce cas précis, les points des deux vues artificielles sont colorés de manière aléatoire et les images possèdent les deux caractéristiques citées précédemment.

Il n'est cependant pas viable d'étendre ces méthodes au traitement de scènes naturelles car les deux hypothèses citées préalablement ne sont plus vérifiées :

- L'intensité lumineuse de deux points images homologues peut être assez différente du fait du bruit des capteurs, de la qualité imparfaite des images, et essentiellement de la différence d'angle de prise de vue.
- Les scènes naturelles comportent des zones importantes de points d'intensité lumineuse quasi-identique (surface d'un bureau, mur, sol, etc) rendant impossible de tels appariements.

Les algorithmes de stéréovision fondés sur l'appariement de tous les points des images initiales utilisent généralement :

- soit des données artificielles dans le dessein de comprendre la vision humaine [Marr 77], [Grimson 81] (figure 3.1) ou de justifier une nouvelle approche du problème [Mohr 84],
- soit des données réelles bruitées artificiellement, par exemple, en projetant une lumière artificielle aléatoire sur la scène comme le propose H.K. Nishihara [Nishihara 84].

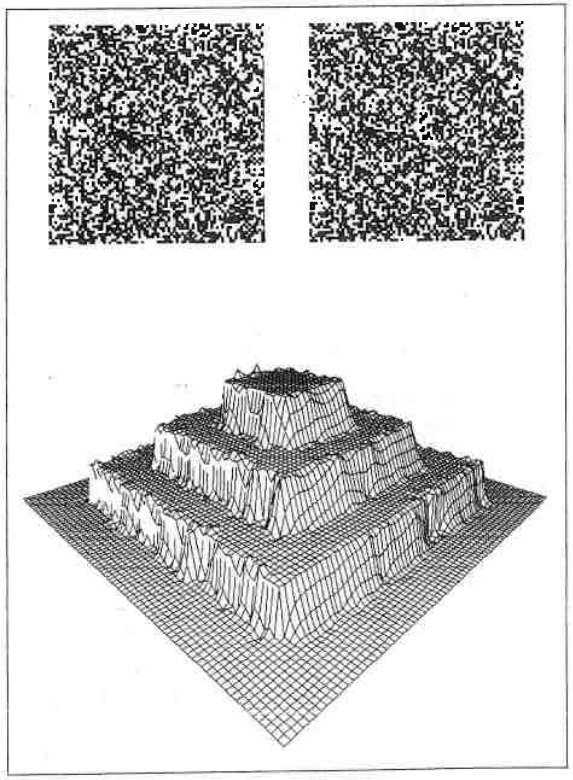

Figure 3.1: *Stéréogramme aléatoire et construction tridimensionnelle de la surface observée. [Grimson 81].*

Points des images traitées

Pour illustrer ce cas, il convient de citer la méthode d'appariement de J. Shen [Shen 87]. L'idée est de mettre en correspondance les points des **images binaires de polarité**. Chaque image de polarité résulte de la soustraction de deux lissages différents de chacune des images initiales. La mise en correspondance est effectuée par une méthode de corrélation, accélérée par une observation astucieuse du codage d'un vecteur binaire.

Cette approche convient bien à des images très texturées.

2.2 Points d'intérêt

Afin de limiter le nombre de candidats à l'appariement et de diminuer ainsi la combinatoire des algorithmes, nous pouvons sélectionner seulement certains points des images. Citons par exemple les **points d'intérêt** extraits par l'opérateur de Moravec [Moravec 77] ; les points sélectionnés sont définis comme étant *"les maxima locaux de l'image des minima des variances directionnelles"* et ont la propriété de bien se distinguer de leurs voisins.

P. Limozin-Long [LimozinLong 86] propose d'apparier des points de Moravec extraits d'une image de scène d'intérieur par une technique de relaxation. Cette méthode est décrite au paragraphe suivant, lors de la présentation de la primitive segment de contour, qu'elle accepte également d'apparier.

On peut aussi choisir d'apparier les **points de contraste** pour lesquels il existe actuellement de nombreuses méthodes d'extraction ; ces méthodes fournissent des résultats sensiblement différents. L'inconvénient majeur provient de sensibilité des extracteurs à la texture des surfaces des objets : celle-ci engendre des points non significatifs et différents sur les deux vues (figure 3.2).

Nous trouvons dans la littérature un grand nombre de travaux portant sur l'appariement des points de contraste, nous pouvons en citer quelques uns :

- appariement des points de contraste déterminés par les passages par zéro des deuxièmes dérivées directionnelles [Marr 77], ou déterminés par les passages par zéro du Laplacien [Marr 80] ;
- appariement des points de contraste par une méthode de type corrélation [Shirai 79] ;
- utilisation d'une méthode de programmation dynamique [Ohta 86a].

Les **attributs** des primitives points sont généralement locaux : valeur de l'opérateur de Moravec ou module du gradient, mais ils sont suffisamment discriminants. L'étape d'appariement est également facilitée de par le faible nombre des points à traiter.

L'inconvénient de cette approche se manifeste lors de la construction de l'information tridimensionnelle. Les points ne peuvent contenir qu'une quantité assez faible d'information : intensité lumineuse et voisinage local, ils ne sont donc pas assez nombreux pour fournir une carte de coordonnées 3d satisfaisante.

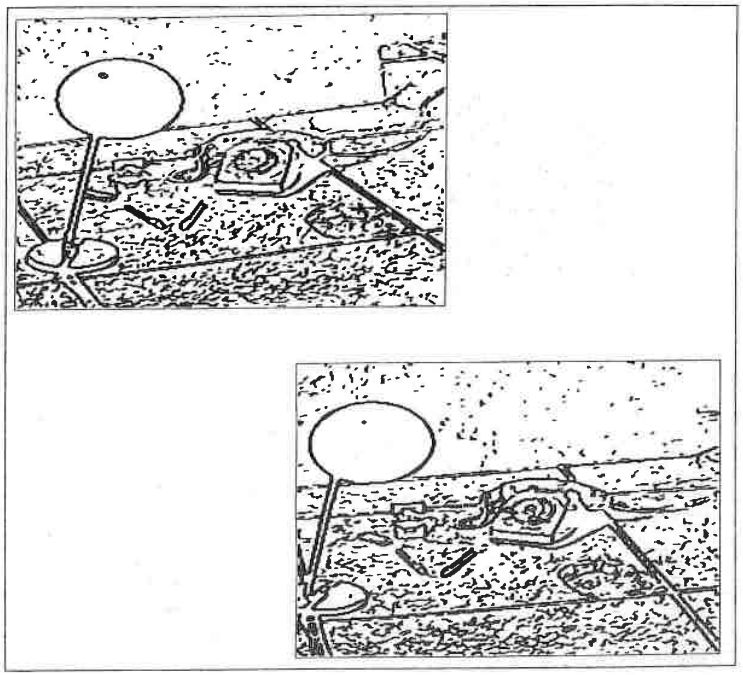

Figure 3.2: *Paire stéréo d'une scène de bureau. Extraction des points de contraste par seuillage d'un* **gradient morphologique normalisé** *[Tombre 85].*

2.3 Conclusion

La complexité algorithmique inhérente aux méthodes utilisant tous les points comme primitives images est un fort handicap. Cette complexité est fortement réduite lors de l'appariement des points d'intérêt mais dans ce cas la carte 3d en résultant est très pauvre.

L'idée adoptée par plusieurs chercheurs est alors d'utiliser ces appariements de points particuliers comme "points d'ancrage", qui servent ensuite à guider la mise en correspondance des points situés dans leur voisinage. La diminution de la complexité algorithmique ainsi obtenue permet de réduire le temps de calcul. Le nombre important des résultats tridimensionnels permet ensuite leur exploitation. Ceci s'applique avec succès aux images où les points sont différents de leurs voisins de manière significative : images très texturées comme les images aériennes traitées par M.J. Hannah [Hannah 84]. Cette approche a également été utilisée par H.H. Backer et T.O. Binford [Backer 81] où la mise en correspondance des contours le long de chaque épipolaire guide l'appariement des zones situées entre ces contours.

Une autre méthode ayant le même objectif, et pouvant être combinée à l'utilisation de certains appariements comme points d'ancrage [Hannah 84], est d'apparier des points en travaillant sur des couples d'images identiques, mais de résolution différente. Chaque image de la paire stéréoscopique est réduite, et l'appariement commence sur les images les plus petites. A chaque étape, le résultat de la mise en correspondance guide l'appariement de l'étape suivante sur des images moins réduites. Cette approche peut être combinée avec l'utilisation de certains appariements servant de points d'ancrage [Hannah 84], [Backer 81].

Ceci est également l'idée clef de W.E.L. Grimson [Grimson 81] (figure 3.3). Pour certaines images, il utilise en plus un artefact pour augmenter de manière fiable le nombre de points de contraste avec lesquels il travaille : par projection de peinture noire sur la scène pour formées des taches (bien contrastées). Ceci permet de se rapprocher des stéréogrammes aléatoires (figure 3.1).

La caractéristique de toutes ces approches est de traiter des primitives ne contenant qu'une information très locale, généralement de type photométrique. Si ces méthodes conviennent bien aux images très texturées, elles sont en revanche à éviter pour les applications où les images contiennent des structures comme des lignes ou des surfaces homogènes nécessitant de corréler les informations photométriques et géométriques des images. Ces méthodes sont de plus très gourmandes en temps de calcul, et demandent à être implantées de manière astucieuse.

3 Segments de contour

Les deux objectifs de l'appariement sont les suivants :

- *"réduire suffisamment le nombre de primitives à associer"* afin d'avoir une mise en correspondance rapide,
- *"obtenir une carte tridimensionnelle relativement dense"* dans le but d'être exploitable dans la suite du projet.

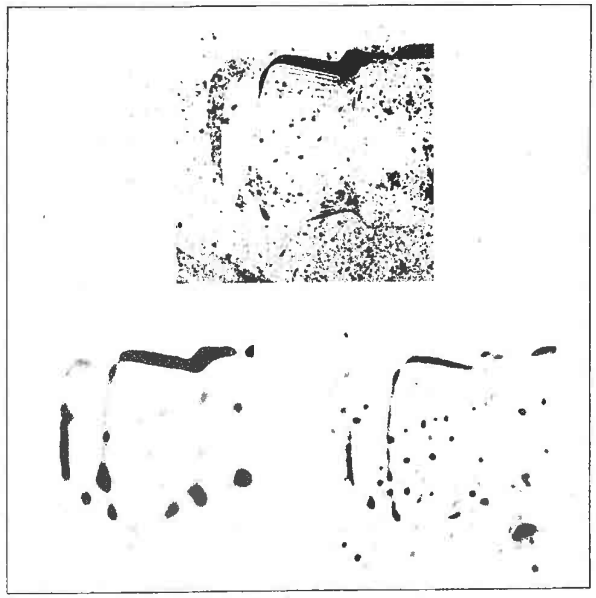

Figure 3.3: *Différentes résolutions d'une même image [Grimson 81].*

Un compromis peut être réalisé en choisissant une primitive apparaissant peu souvent dans les images mais contenant beaucoup d'information. Le **segment de contour** possède ces deux caractéristiques.

Les primitives segments peuvent être facilement extraites des deux images, généralement par approximation polygonale des chaînes construites par suivi de points de contour d'épaisseur un. On pourra consulter les travaux de N. Keskes [Keskes 82] et plus récemment ceux de G. Giraudon [Giraudon 87], où le chaînage et l'approximation sont réalisés en une seule passe sur l'image ; un tel traitement peut donc être très rapide.

Cette fois encore, le même traitement est appliqué à chacune des deux images de la paire de manière indépendante, et inévitablement des différences de segmentation apparaissent. Ces différences proviennent essentiellement de l'instabilité de l'approximation polygonale et non de l'extraction de ces chaînes de contour.

En effet, une ligne de contour dépend faiblement de la variation d'éclairage et, par conséquent, de l'angle de prise de vue des caméras. Les chaînes de contour résultant de l'extraction sont donc relativement similaires (figure 3.4). En revanche, ces chaînes sont approchées par un nombre différent de segments (figure 3.5), caractéristique dont doit tenir compte le processus d'appariement.

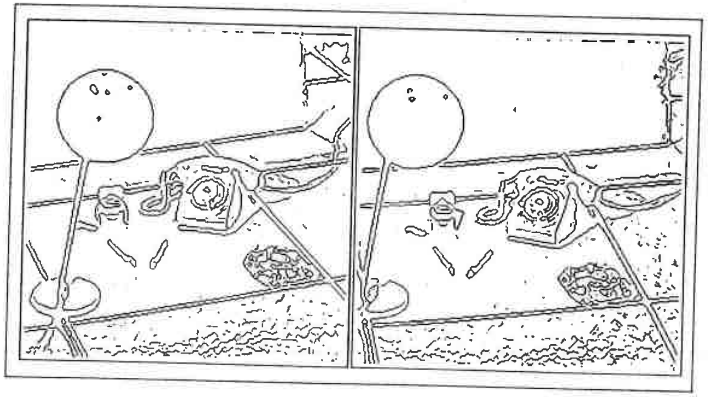

Figure 3.4: *Extraction des lignes de contraste sur chacune des images d'une paire stéréoscopique selon la méthode proposée par R. Deriche [Deriche 87a].*

L'utilisation de la primitive **segment** permet de simplifier la tâche délicate de la mise en correspondance : le nombre de possibilités d'association de primitives est fortement diminué. Ceci explique le grand nombre de méthodes de stéréovision fondées sur l'utilisation de cette primitive. Nous pouvons séparer les techniques mises en œuvre en deux classes :

- celles qui apparient des segments,

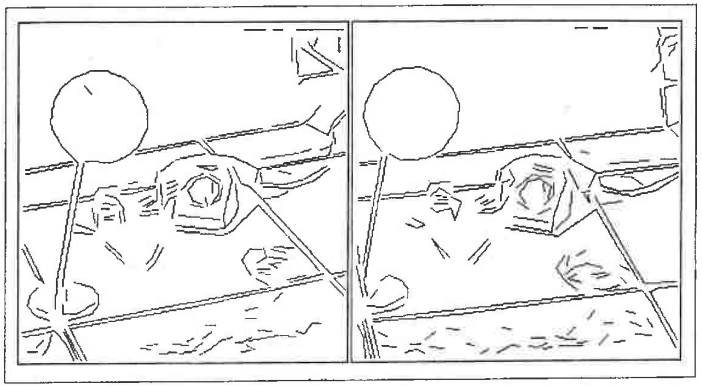

Figure 3.5: *Approximation polygonale des lignes de contraste précédemment détectées.*

- celles qui apparient des ensembles de segments ; un tel ensemble étant considéré comme une seule primitive image.

L'appariement de segments isolés peut être vu comme un cas particulier de l'appariement d'ensembles de segments. La différence essentielle est que la seconde catégorie regroupe des méthodes utilisant de très fortes contraintes de voisinage. Néanmoins, cette différence tend à fortement diminuer, certaines contraintes de type topographique sont utilisées pour l'appariement de segments, comme nous le décrivons ci-après.

3.1 Appariement de segments isolés

Citons dans ce domaine, sans être exhaustif, les travaux de :

G. Medioni [Medioni 85]. La méthode mise en œuvre s'apparente aux méthodes de relaxation. Une particularité de cet algorithme est de prendre en compte les appariements des segments voisins pour évaluer la qualité de mise en correspondance de chaque couple de primitives.

N. Ayache et B. Faverjon [Ayache 85]. Ils utilisent une méthode de prédiction-vérification récursive d'hypothèses. Le principe est de sélectionner quelques bons appariements hypothétiques et d'essayer de les propager à leur voisinage. Seules les hypothèses de meilleur score sont validées. L'algorithme comporte trois grandes étapes :

- *prédiction d'hypothèses* : on sélectionne un certain nombre d'appariements de segments, en fonction de la similarité de leurs attributs et des contraintes épipolaires.

- *propagation récursive d'hypothèses* : on cherche dans le voisinage de chaque couple de segments homologues d'autres appariements qui conservent la disparité proposée, en respect avec l'hypothèse de continuité de Marr. La qualité de chaque hypothèse est calculée en fonction du nombre de mises en correspondance engendrées par chaque appariement initial.
- *résolution des conflits et sélection des résultats* : il s'agit essentiellement de respecter l'hypothèse d'unicité de Marr et de valider les appariements correspondants aux meilleurs scores.

P. Limozin-Long [LimozinLong 86]. L'algorithme de mise en correspondance est fondé sur une méthode de relaxation continue, cette méthode est vue comme une optimisation d'étiquetage sous contraintes. L'étiquetage solution est celui qui optimise un critère de mesure de consistance. L'algorithme comporte les trois étapes suivantes :

- *initialisation* : on définit un étiquetage initial conforme au formalisme proposé ; les **objets** manipulés sont tous les couples de primitives potentiels, et les **étiquettes** sont "présent" ou "absent" pour noter la présence ou l'absence des objets utilisés.
- *optimisation d'étiquetage* : l'influence du contexte est prise en compte pour renforcer ou inhiber les objets (respect de l'hypothèse de continuité). A la convergence, ce système rend un étiquetage consistant, mais ambigu : les objets sont plus ou moins "présents" et plus ou moins "absents".
- *détermination des résultats* : il est nécessaire de trouver un étiquetage discret non ambigu le plus proche possible de celui proposé à l'étape précédente (respect de l'hypothèse d'unicité). Les objets doivent alors prendre soit l'étiquette "présent", soit l'étiquette "absent".

R. Horaud, T. Skordas [Horaud 88]. L'appariement est réalisé en tenant compte d'une part des contraintes géométriques et d'autre part des similarités topographiques. L'idée est d'utiliser des relations de voisinage qui sont invariantes par projection perspective, comme la colinéarité, et de mettre en correspondance des groupements de segments structurellement cohérents. La méthode se partage en quatre parties :

- *recherche des attributs des primitives* : chaque segment est caractérisé par sa position, son orientation et ses relations de voisinage avec les autres segments. La relation de voisinage est très locale : deux segments sont voisins si il n'existe aucun autre segment entre eux et cette relation possède quatre attributs : "à droite de", "à gauche de", "colinéaire à", "de même intersection que".
- *formation des hypothèses* : les hypothèses sont créées à partir de la similitude des caractéristiques locales des segments. La méthode étant complétement ascendante, toutes les hypothèses doivent être formées à cette étape.

- *construction des relations de compatibilité entre les hypothèses* : l'idée est d'utiliser les relations de voisinage entre les segments pour définir les relations de compatibilité entre les hypothèses précédemment établies. Les relations de voisinage étant locales, il devient nécessaire de construire une règle de compatibilité transitive pour les étendre. Ceci pose le problème de la propagation de relations dont la définition comporte des incertitudes. Par exemple, la fermeture transitive de la relation "est colinéaire à" définie à ϵ près peut trouver colinéaires des segments qui ne le sont pas.
- *appariement* : la meilleure correspondance se traduit par le grand nombre de nœuds mutuellement compatibles. De par la différence entre les deux images, dues aux occlusions, aux alignements accidentels et/ou aux erreurs d'extraction de primitives, la recherche d'une correspondance parfaite est exclue. L'approche se propose donc de trouver la meilleure correspondance possible, c'est-à-dire de trouver les cliques maximales en terme de recherche dans un graphe.

Les résultats de l'appariement de segments sont présentés successivement figures 3.6 et 3.7. Ils concernent les deux dernières méthodes présentées.

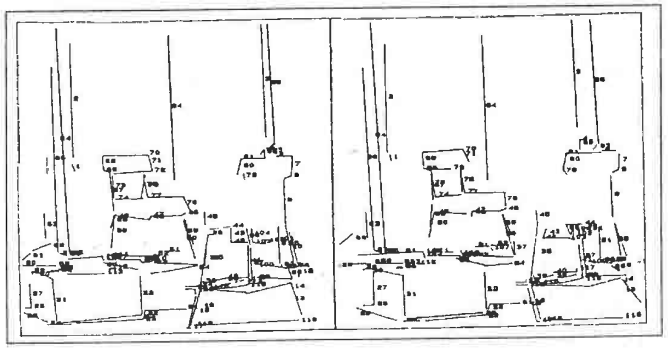

Figure 3.6: *Appariement de segments d'une scène d'intérieur [LimozinLong 86].*

Ces différentes méthodes, conçues pour traiter des images contenant des lignes significatives, donnent de bons résultats : les segments sont correctement situés dans les images et l'erreur de construction des segments tridimensionnels est minimisée. Le temps de traitement est faible : en moyenne 15 secondes pour des images 256 × 256 sur un Perkin-Elmer [Ayache 85], sauf pour la dernière méthode présentée où la recherche des cliques maximales effectuée par des algorithmes classiques n'est pas très rapide. Cette dernière approche est cependant très intéressante car elle permet de prendre en compte des groupements de segments et de respecter les structures trouvées dans les images.

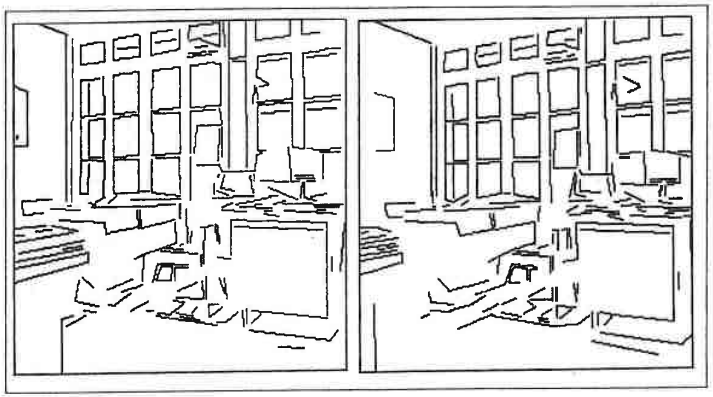

Figure 3.7: *Segments appariés d'une scène d'intérieur [Horaud 88].*

Le premier inconvénient des méthodes d'appariement de segments est la "perte" d'une assez forte proportion des segments : typiquement 50 % des segments ne sont pas appariés [Ayache 85].

Le second inconvénient est plus contraignant : lors de la construction des informations 3d, nous obtenons un "nuage" de segments 3d dans l'espace, sans connaissance de ce qui se trouve de part et d'autre de ces segments. Il devient alors nécessaire d'effectuer un traitement ultérieur afin de compléter la carte 3d obtenue car, pour notre projet, il est important d'avoir une représentation contenant à la fois des volumes pour déterminer l'espace libre dans lequel le robot peut se déplacer sans danger et les surfaces pour les tâches de navigation ou de reconnaissance.

Face à ce problème, il est possible d'effectuer une triangulation de Delaunay sur les segments 3d qui respecte la visibilité des indices [Boissonnat 87]. Une autre approche est de trouver les structures principales de la scène (plans) et de les identifier à l'aide de règles et d'un moteur d'inférence [Thonnat 86].

3.2 Appariement d'ensembles de segments

Les ensembles de segments le plus souvent utilisés pour ce type d'appariement sont les **jonctions de segments**. On utilise généralement des tables de transition pour recenser les différentes vues possibles d'une même primitive. Ceci permet de prédire comment une forme dans une image peut apparaître dans l'autre image. En exemple, citons les travaux de :

C. Thorpe et S. Shafer [Thorpe 83]. Les images à traiter sont segmentées et les primitives étiquetées selon le dictionnaire proposé par D.A. Huffman [Huffman 71] et M.B. Clowes [Clowes 71]. L'appariement doit respecter les trois contraintes suivantes :

- conservation des points d'intersection (points anguleux) : puisqu'un même objet doit apparaître sur les deux vues, nous devons dans la majorité des cas observer les points anguleux dans chacune des images ;
- conservation des types des segments : un segment concave, respectivement convexe, le reste quel que soit son angle de prise de vue ;
- conservation des adjacences.

L'appariement est alors vu comme la recherche de cliques dans un graphe où chaque nœud représente un appariement potentiel entre une jonction de chaque image. Les arcs répertorient les relations d'adjacence.

Si les résultats proposés sont excellents sur des données artificielles ou des images réelles d'objets polyhédriques isolés ou partiellement cachés et par là même tendent à confirmer cette approche, il manque néanmoins des essais sur des scènes plus complexes. La méthode semble trop contraignante pour être appliquée telle quelle sur des images réelles où les primitives images sont généralement incomplètes. Les idées concernant la conservation des liens de voisinage entre les primitives sont pertinentes mais leur mise en œuvre est assez lourde ; ces idées devraient être intégrées à un processus d'appariement travaillant sur des primitives plus simples. Ceci rejoint alors la méthode développée au paragraphe précédant concernant l'appariement des segments de contour en respectant les "groupements" formés [Horaud 88].

4 Arcs de cercle

Une approximation des chaînes de contour par des arcs de cercle permet un gain de précision par rapport à une approximation rectiligne et une diminution du nombre de primitives.

La principale hypothèse simplificatrice utilisée est que l'arc de cercle qui approche au mieux les points d'une chaîne de contour, est centré sur la médiatrice du segment trouvé par approximation polygonale. Les arcs de cercle sont caractérisés par leur centre et leurs deux extrémités. Chaque arc est déterminé par minimisation de la distance des points de la chaîne de contour à l'arc approchant.

Une méthode d'extraction et d'appariement de telles primitives a été mise au point par M. Alami [Alami 87]. Nous présentons figure 3.8 un exemple de résultats d'appariement.

La méthode de mise en correspondance s'avère efficace pour des scènes d'environnement robotique malgré une augmentation du volume de codage nécessaire aux primitives.

On peut cependant s'interroger sur l'avenir d'une telle primitive pour la stéréovision car il n'existe pas, à notre connaissance, de méthode de construction des informations tridimensionnelles. De plus, les inconvénients de cette méthode rejoignent ceux cités pour l'appariement des segments de contour : la carte 3d ne contient pas d'informations surfaciques, seuls les contours des objets peuvent être déterminés dans l'espace.

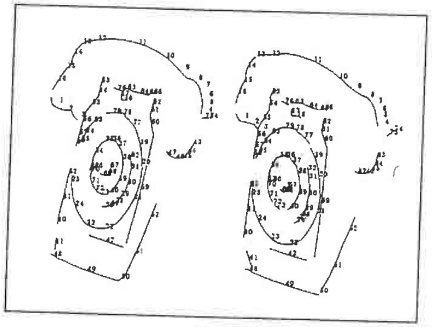

Figure 3.8: *Appariement de primitives de type arc de cercle [Alami 87].*

5 Régions

En choisissant les segments de contour pour réaliser l'appariement, nous venons de constater qu'un problème se pose lors de la construction de la scène : connaître l'état de la scène "entre" les segments. Il manque une information concernant les **surfaces**.

Cette information de surface se trouve dans la primitive **région**. Une région est définie comme un ensemble de points connexes et homogènes selon une ou plusieurs caractéristiques : intensité lumineuse, variance de l'intensité lumineuse, etc, et elle correspond idéalement à un objet ou une partie d'objet dans l'image.

La mise en correspondance se fait sur un petit nombre de primitives : une moyenne de 250 régions par image 256 × 256 points (image d'une scène d'intérieur).

L'utilisation de cette primitive est encore peu courante car elle possède trois inconvénients qu'il faut limiter ou supprimer : le temps de calcul (segmentation et mise en correspondance), l'instabilité des partitions à apparier et la localisation peu précise des régions.

En effet, l'extraction des régions est fortement influencée par le niveau et les variations faibles ou systématiques de l'éclairage et le bruit dû au capteur. Les partitions en régions obtenues sont alors quelque peu différentes lorsque le même processus de segmentation est appliqué indépendamment sur chaque des images de la paire stéréoscopique. Ce phénomène doit être pris en compte lors de la mise en correspondance.

Cette primitive, longtemps tenue à l'écart des préoccupations les chercheurs, commence à apparaître dans l'actualité. Les récents travaux cités ci-après en témoignent :

- La méthode de recalage stéréoscopique développé à l'ENST[1] [Remion 87] est fondée sur l'appariement des régions. L'idée est d'intégrer au processus de mise

[1] ENST : Ecole Nationale Supérieure des Télécommunications, Laboratoire Image, PARIS

en correspondance la contrainte géométrique liant les primitives qui représentent les projections d'un même objet plan, c'est-à-dire les primitives régions. Cette contrainte s'énonce sous la forme suivante :

Si deux régions R_g et R_d, exemptes de parties cachées, sont les projections de la même surface plane, alors elles doivent vérifier l'équation :

$$T_p(R_g) = R_d$$

où T_p est la transformation géométrique faisant passer des coordonnées des points de l'image gauche d'un même plan 3d P aux coordonnées de leurs points dans l'image droite. Une telle contrainte inclut la contrainte épipolaire.

Pour être en mesure de résoudre une telle équation, il convient de modéliser numériquement la géométrie des régions. Les six éléments inertiels : la surface, les coordonnées du centre de gravité et les moments d'inertie, en forment une bonne description, quoique incomplète.

La méthode d'appariement consiste à sélectionner de manière itérative le couple de régions dont les six paramètres vérifient au mieux la contrainte géométrique.

Cette approche donne des résultats satisfaisants et à l'avantage d'être très tolérante pour les différences photométriques entre régions homologues des images. Il convient cependant d'observer qu'une proportion significative des erreurs est imputable à la segmentation.

- L'approche conçue à l'INRIA-Rocquencourt [Cocquerez 87] fait appel à la notion de graphe d'association, l'idée de base est d'apparier les deux graphes issus des deux images stéréoscopiques. Les nœuds des graphes représentent les régions et les arcs caractérisent les relations d'adjacence.

  L'objectif est de déterminer une injection qui optimise la similitude des attributs des nœuds (régions) mis en correspondance, ainsi que l'isomorphie des deux graphes. Etant donné que dans la plupart des cas, une isomorphie complète ne peut être déterminée, le terme ϵ-isomorphie convient mieux. La méthode s'effectue en deux temps :

  - *mise en correspondance des régions* en utilisant simplement leurs attributs (photométriques et géométriques). Cette stratégie revient à optimiser une fonction de qualité de correspondance sans se préoccuper de l'existence d'une fonction de mise en correspondance des arcs du graphe. L'appariement est effectué de manière itérative pour les couples de régions de meilleur qualité de ressemblance.
  - *optimisation de l'ϵ-isomorphie des graphes* : l'idée est de propager un ensemble de mises en correspondance initiales à l'aide de relation de voisinage existant entre les deux graphes : une expansion en parallèle des appariements des régions est ainsi réalisée.

Si les attributs des régions sont discriminants et stables d'une image à l'autre, cette méthode donne des résultats satisfaisants. Cela est souvent le cas quand les points de vue associés aux deux images sont très proches. Cette approche est cependant fortement perturbée par les différences de segmentation existant entre les deux images utilisées.

Nous avons également choisi de travailler avec la primitive **région** car l'information de surface qu'elle contient doit permettre de construire une scène tridimensionnelle dense comportant des surfaces. L'enjeu de notre travail est alors de résoudre le problème de l'instabilité des primitives régions. Ceci est réalisé, d'une part en faisant usage d'un processus de segmentation qui soit ancré sur les lignes de contraste préalablement détectées, et d'autre part en renforcant la segmentation en régions à chaque niveau du système.

- Nous décrivons au chapitre suivant (chapitre 4) la **segmentation** des images en régions et l'aide apportée par l'extraction préalable des points de contraste.
- Le chapitre 5 est consacré à l'**appariement** des régions : nous y décrivons la mise en correspondance des contours des régions par une technique de programmation dynamique qui sert à guider la mise en correspondance des régions et à améliorer les partitions initiales. Notre méthode procède par prédiction-vérification d'hypothèses.

  Si nous représentons les deux partitions à apparier sous forme de deux graphes où les nœuds représentent les régions (avec leurs attributs photométriques et géométriques) et les arcs les relations d'adjacence (avec leurs attributs tels que la longueur de la frontière, la valeur du gradient moyen, etc), nous pouvons résumer notre méthode de la façon suivante :

  - "pseudo-appariement" des **arcs** des graphes en tenant compte des relations de voisinage des régions essentiellement selon la direction des lignes épipolaires ; chaque appariement "vote" pour l'appariement des nœuds (régions) concernés par les relations mises en jeu ;
  - les votes précédemment comptabilisés et la similitude des attributs des **régions** permettent leur mise en correspondance itérative : appariements successifs des couples de régions de meilleure qualité de correspondance.
- le chapitre 6 concerne la **construction des informations tridimensionnelles**. Elle s'effectue en deux phases :
  - construction des segments 3d correspondant au bord des régions,
  - construction des surfaces 3d.

6 Utilisation conjointe de plusieurs primitives

Il est à remarquer que l'une des directions actuelles dans le domaine de la stéréovision est l'emploi, non plus d'un type de primitives, mais de plusieurs. Cette

façon de procéder permet aux appariements de se renforcer mutuellement, de lever certaines ambiguïtés ou d'éviter des erreurs.

Ceci peut se faire de plusieurs manières. Par exemple :

- en incluant des caractéristiques d'autres primitives aux primitives effectivement appariées. L'appariement de primitives mixtes **lignes-régions** a été réalisé par P. Long [LimozinLong 86].

  Chaque primitive segment possède des informations concernant une des régions de l'image que ce segment borne. Dans ce cas, les régions restent ouvertes et le terme "zone" leur conviendrait peut être mieux. Les attributs des segments : "longueur", "orientation dans l'image", "position du point milieu", sont enrichis de l'attribut "moyenne des niveaux de gris de la région" sur laquelle se situe le segment considéré.
- en appariant **simultanément** plusieurs types de primitives. J.J. Hwang et E.L. Hall [Hwang 82] utilisent conjointement des informations géométriques et structurelles pour effectuer l'appariement. On extrait de chaque image les régions, les segments de contour et les points anguleux (points d'intersection des segments). Ces informations, liées par des relations topographiques, sont répertoriées dans une table. L'appariement consiste à trouver les étiquetages consistants entre les deux tables de relation associées aux deux images par un processus de relaxation. Puisque les étiquettes dans les deux tables doivent se correspondre une à une, le problème de l'appariement des deux images est réduit à un problème d'appariement de régions, d'appariement de segments et d'appariement de points anguleux.
- en appariant un type de primitives pour **guider** ensuite l'appariement d'un autre type de primitives. Par exemple, l'appariement des régions peut guider l'appariement des segments : la qualité de correspondance entre deux segments est alors fonction de leurs attributs et éventuellement de ceux de leurs voisins, mais également du résultat de l'appariement des régions dans le voisinage des segments considérés. La méthode réciproque peut également être mise en œuvre ; c'est de cette manière que nous avons choisi de procéder, aussi cette technique est elle plus amplement développée au chapitre 5 consacré à l'appariement.

La tendance actuelle est d'aller au-delà des appariements de primitives bien particulières et de tirer profit de la richesse de l'information contenue dans chacun des types d'indices visuels traités pour mener à bien la tâche de mise en correspondance et d'être capable de construire au mieux l'information 3d résultat.

Cette idée doit cependant être modulée par le fait suivant : lors d'une application en temps réel, comme la navigation d'un robot mobile, il doit aussi être tenu compte du temps de traitement, et donc du temps d'extraction et de manipulation des différentes primitives. Nous aurons l'occasion de revenir sur ce point.

Mentionnons également qu'une des principales difficultés d'une telle approche est de gérer des informations provenant de primitives différentes.

Notre choix de primitives à apparier s'est porté sur les régions, essentiellement dans le dessein d'obtenir des informations 3d sur les surfaces. Il s'est cependant trouvé indispensable d'utiliser d'autres primitives images (les segments de contour) pour effectuer correctement la mise en correspondance, mais également pour obtenir une meilleure précision lors de la reconstruction : les segments de contour sont bien localisés dans les images et représentent de ce fait un point de passage obligé. Le problème de l'utilisation conjointe de plusieurs types de primitives est donc abordé plusieurs fois dans cette thèse.

7 Interaction entre processus

Le problème de la stéréovision se décompose généralement en trois phases successives : extraction des primitives, mise en correspondance et construction des informations tridimensionnelles. Ces étapes sont le plus souvent implantées de manière séquentielle, car elles sont complètement indépendantes. Il s'avère cependant que de notables améliorations de résultats peuvent s'obtenir soit en intégrant plusieurs traitements, soit en leurs demandant une étroite coopération.

Les travaux de Hoff et Ahuja [Hoff 87] mêlent la détection des points de contraste, leur mise en correspondance et l'interpolation des surfaces tridimensionnelles. A chaque niveau de résolution de l'image, le traitement est le suivant :

- *appariement des points de contraste* correspondant aux passages par zéro du Laplacien guidé par la profondeur de la surface résultat calculée à l'étape précédente,
- *construction de petites surfaces 3d* carrées correspondant aux couples de points 3d trouvés (représentant une petite surface du fait de la résolution de l'image),
- *interpolation* de ces nouvelles surfaces avec les surfaces déjà construites,
- *prédiction* des emplacements des futurs points de contraste à apparier lors de l'étape suivante, et extraction de ces points.

Cette approche est indépendante du domaine d'application car la seule contrainte utilisée est la continuité locale des surfaces, ce qui est le cas des scènes réelles où le monde est localement continu. Elle permet de plus de distribuer relativement uniformément les calculs sur chaque processus au lieu de les concentrer sur l'étape d'appariement, dans le cas de la stéréovision par mise en correspondance de points. Les résultats sont assez bons : seules les frontières sont occasionnellement mal localisées de part leur détection (figure 3.9).

Lors de l'appariement des régions, nous constatons que la plupart des erreurs de mise en correspondance proviennent de la segmentation des images [Cocquerez 87], [Remion 87]. Cet inconvénient peut être considérablement réduit lors de l'intégration d'un module de segmentation au module d'appariement des régions : nous avons mis en œuvre une méthode d'appariement qui poursuit la segmentation initiale en régions [Wrobel 87a].

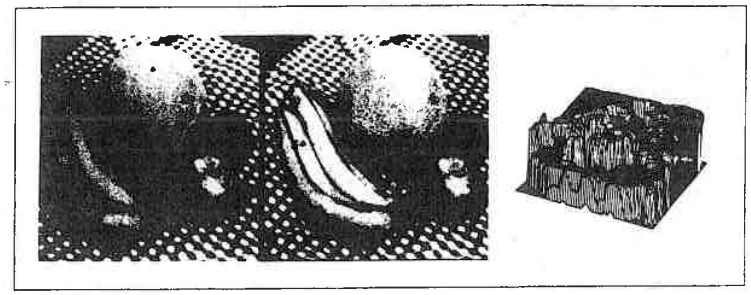

Figure 3.9: *Résultats d'un algorithme de stéréovision couplé à l'extraction et à la construction des surfaces tridimensionnelles [Hoff 87].*

L'interaction entre les traitements ou plus généralement la coopération entre les traitements, est une approche qu'il convient de suivre et de développer. Chaque processus doit pouvoir travailler de manière isolée pour fournir des résultats "de base", puis être guidé par les autres processus qui disposent d'autres informations pour affiner ses résultats.

Cette interaction n'est pas propre à la stéréovision et se développe dans d'autres domaines de vision : citons le système PVV [Lux 84] où l'on décide de l'endroit où il faut extraire les points de contraste, ou la coopération entre l'extraction de contours et la reconnaissance de formes [Gandelin 87], par exemple.

La stéréovision peut aussi être intégrée à des traitements de niveau supérieur comme la tâche de navigation et/ou de localisation d'un robot mobile ; cette étude est actuellement menée dans notre laboratoire dans le cadre du projet ORASIS. D'autres coopérations peuvent donc être envisagées.

Chapitre 4

SEGMENTATION D'IMAGES EN RÉGIONS

L'objectif de ce chapitre est de présenter la méthode de segmentation d'images développée pour obtenir les partitions en régions utilisées lors de la mise en correspondance.

Dans les premiers paragraphes, nous exposons le problème de la segmentation et les deux grandes approches (extraction de contours ou de régions) permettant de le résoudre, puis nous précisons la méthode que nous avons choisie. Nous donnons ensuite une définition formelle de la segmentation, et présentons, d'une part, l'outil algorithmique général mis en œuvre, et d'autre part, les prédicats d'homogénéité de régions utilisés. Ces prédicats dépendent du type des images traitées.

En dernière partie, nous montrons une façon simple de contrôler notre processus de croissance de régions par une carte de points de contraste pré-calculée. Ceci représente un essai, très encourageant, d'utilisation conjointe de deux informations duales : les discontinuités locales dans l'image (points de contraste) et les zones homogènes (régions).

1 Introduction

Segmenter une image en régions, c'est réaliser une partition de l'image en un nombre de zones isolées, appelées **régions**, qui possèdent une certaine uniformité pour une ou plusieurs caractéristiques (intensité lumineuse, teinte, ...) et qui soient différentes, pour au moins une de ces caractéristiques, des régions voisines.

Que devrait être une bonne segmentation ?

Il est actuellement difficile de qualifier la qualité d'une partition en régions car le résultat d'une segmentation est généralement jugé par l'homme en fonction de critères sémantiques difficilement implantables. De plus, cette qualité dépend souvent du traitement ultérieur choisi, elle diffère selon que l'objectif est, par exemple, la réduction du volume des données stockées ou l'interprétation de scènes.

Nous pouvons toutefois essayer de donner une définition générale d'une "bonne segmentation" :

- les régions devraient être uniformes et homogènes en fonction de caractéristiques comme le niveau de gris ou l'histogramme de répartition des niveaux de gris ;
- les régions adjacentes devraient prendre des valeurs relativement différentes du point de vue des critères en fonction desquels elles sont uniformes ;
- les frontières des régions devraient vérifier une certaine régularité ; on souhaite que les régions ne soient pas déchiquetées, qu'elles ne comportent pas de petits trous et que les frontières soient lisses, non bruitées et correctement positionnées.

Effectuer une segmentation qui atteint de tels objectifs est très difficile pour les deux raisons suivantes :

- Des régions uniformes et homogènes comportent justement des contours bruités :
  - La zone de frontière entre deux surfaces dans une image réelle s'étend généralement sur une largeur de plusieurs pixels, alors qu'elle devrait "idéalement" être filiforme. Cette frontière ne peut donc être déterminée qu'approximativement.
  - Les valeurs des pixels situés dans la zone de la frontière sont relativement différentes de celles composant les deux régions à séparer :
    * les intégrer aux régions rendent celles-ci beaucoup moins homogènes et uniformes ;
    * ne pas les intégrer crée des contours bruités ou même une région "fictive" car n'existant pas dans la réalité et représentant toute la "zone" de la frontière.
- Insister sur le fait que les régions adjacentes doivent comporter des différences significatives entre les valeurs de leurs caractéristiques a pour effet la fusion intempestive de ces régions :

  - Il est naturel de trouver, dans une image réelle, des zones adjacentes de caractéristiques similaires et représentant pourtant des surfaces différentes.

D'une manière générale, nous pouvons dire qu'il n'est pas simple d'extraire des indices visuels sans ambiguïté et sans qu'ils soient entachés d'erreurs, une partie de ces erreurs provenant notamment des incertitudes dues au capteur. Cette difficulté explique le nombre très important des travaux menés dans le domaine de la segmentation d'images. Elle explique également que les systèmes de plus haut niveau, comme la stéréovision ou l'interprétation de scènes, préfèrent travailler avec des primitives plus faciles à extraire : les contours.

2 Choisir entre une "approche contour" et une "approche région"

Dans notre cas, nous souhaitons que la segmentation nous fournisse une "bonne" partition dans laquelle les régions correspondent aux projections d'objets ou parties d'objets présents dans la scène d'où est issue l'image. Le problème est de trouver la fonction "qualité" à optimiser afin d'obtenir cette partition.

Par exemple, nous pouvons définir la qualité d'une partition comme étant la différence de chaque point avec la moyenne de la région à laquelle il appartient. Ceci est décrit par la formule suivante :

$$C_2(S) = \sum_{k,l \in I} (I(k,l) - M_{kl})^2$$

avec C : fonction de la qualité de la partition à optimiser,
S : partition dont on calcule la qualité,
I : image à segmenter,
$I(k,l)$: valeur du point image à la position (k,l),
M_{kl} : moyenne arithmétique des valeurs des points de la région à laquelle appartient le point (k,l).

Optimiser la qualité d'une partition revient dans ce cas à minimiser cette différence, c'est-à-dire à tenter de créer des régions où les points ont sensiblement les mêmes valeurs. Nous détaillons page 82 la fonction de qualité utilisée.

Les techniques de segmentation existantes sont nombreuses, mais elles sont généralement fondées sur l'un des deux principes de base : **la discontinuité** ou **la similarité**.

- Les approches s'appuyant sur la recherche des discontinuités locales de la fonction de niveau de gris correspondent aux techniques de recherche de **contours**. Par définition, un point de contour est un lieu de forte variation de l'intensité dans l'image, ou une "crête" dans l'image gradient. Ces points de fort gradient sont souvent attachés à la présence d'un objet ou d'un reflet. Ils correspondent

généralement à des transitions entre deux zones de caractéristiques différentes dans l'image. L'objectif est alors de décrire l'image avec un ensemble de points constituant les frontières des différentes entités physiques présentes dans la scène.

- Les approches fondées sur le principe de similarité correspondent aux techniques de recherche de **régions** homogènes dans l'image ; elles font intervenir la recherche de la continuité de certaines caractéristiques sur des groupes de points adjacents.

Ces deux approches sont **duales** en ce sens qu'une région définit une ligne par son contour et qu'une ligne fermée définit une région. Elles mènent cependant à des algorithmes différents et ne fournissent pas les mêmes résultats.

Pour parvenir à segmenter une image en régions nous avons donc le choix entre ces deux approches.

2.1 Approche "contour"

Les avantages d'une méthode par recherche des lignes de contours sont multiples [Lux 84] :

- Les points de contour sont physiquement bien définis puisqu'ils correspondent aux endroits de l'image gradient où le rapport signal/bruit est maximal.
- Les crêtes de l'image gradient sont indépendantes du niveau d'éclairage, et également des variations faibles ou systématiques de l'éclairage[1]. De plus, de par leur définition, les crêtes sont précisément les endroits où l'image gradient possède le meilleur rapport signal/bruit si l'on suppose que le bruit est uniforme et indépendant du niveau du signal.

L'opération d'extraction de points de contours se fait généralement par filtrage de l'image. Cette extraction fait apparaître des discontinuités, les séquences de points voisins regroupés en lignes sont interrompues. La cause de ces discontinuités est double :

- le gradient est trop faible à certains endroits de la ligne de contours pour donner naissance à un point de contour.
- les jonctions de plusieurs lignes de crêtes de gradient sont mal détectées : les opérateurs privilégient généralement une ligne. Ce problème provient des filtres utilisés pour détecter les contours "simples" entre seulement deux zones différentes. Ce principe est illustré figure 4.1. Ces mêmes filtres ne sont pas adaptés au traitement des jonctions des lignes de contours, c'est-à-dire où un contour "multiple" passe par un point (figure 4.2).

  Ce problème peut être partiellement résolu en utilisant des opérateurs directionnels qui ne détectent les contours que dans une seule direction sur l'image [Binford 81].

[1] le gradient est constant ou faible

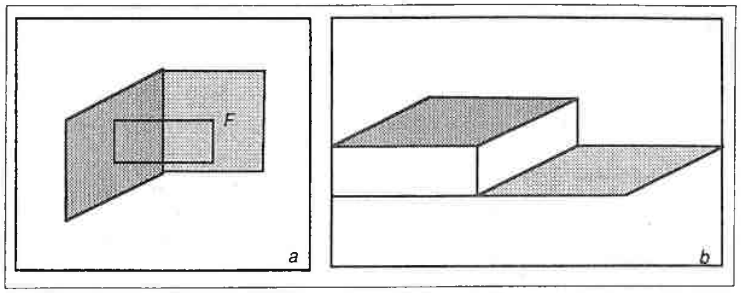

Figure 4.1: **a** : *Vue d'une scène artificielle.* **b** : *Coupe et agrandissement de la fenêtre F : la "hauteur" des surfaces correspond à leur niveau de gris. Les filtres sont généralement écrits pour détecter de tels contours.*

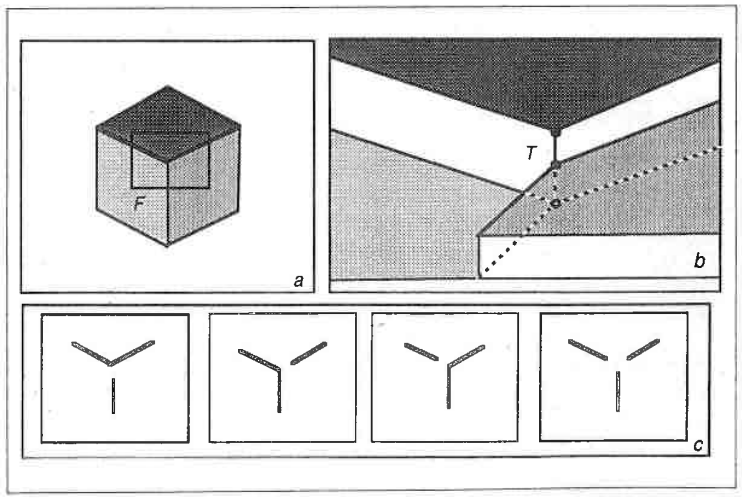

Figure 4.2: **a** : *Vue d'une scène artificielle.* **b** : *Coupe et agrandissement de la fenêtre F. Les filtres sont souvent inopérants au voisinage du point triple T.* **c** : *Exemples de résultats : mauvaise détection de la jonction des lignes.*

Puisque les lignes de contours sont discontinues, il devient nécessaire d'utiliser un algorithme de fermeture de contours. Certaines méthodes réalisent l'extraction et la fermeture en même temps. La difficulté est qu'une ligne de contraste peut traverser l'image en sautant d'un objet à un autre, passer par des ombres et des reflets. Ceci signifie que son interprétation en terme bord d'objet, ou d'élément de la scène, peut varier le long de la ligne.

Un processus de fermeture de contours doit donc résoudre les problèmes suivants :

- déterminer le seuil de qualité afin d'interrompre le suivi d'une "mauvaise ligne",
- trouver les critères de sélection du point par lequel on poursuit le suivi : module du gradient, plus forte dérivée dans une direction, ...

Il s'avère que ces algorithmes fournissent de bons résultats sur des images bien contrastées et sont utilisés avec succès en robotique. Mais il faut souligner que cette approche s'accomode mal de la présence de texture dans les images. Les images de scène d'intérieur que nous traitons se situent entre ces deux types d'images en ce sens où les images ne pas très bien contrastées car l'éclairage n'est pas contrôlé, et les surfaces des objets ont une texture relativement "fine" donc discrète. Cependant, la méthode d'extraction et de fermeture de contours développée par R. Deriche et J.P. Cocquerez et testée sur des images de scène d'intérieur fournit de très bons résultats [Deriche 87b].

2.2 Approche "région"

La notion de région n'a pas les propriétés que nous venons d'énoncer. Au contraire, la segmentation d'une image en régions dépend souvent des conditions d'éclairage. Le niveau et la direction de l'éclairage influent directement sur la valeur de l'intensité lumineuse de chaque point. Les attributs utilisés pour caractériser les régions sont calculés directement à partir de ces valeurs ; l'influence du niveau du signal et du bruit n'est pas annulée par l'utilisation d'un opérateur tel que le gradient.

Plus précisément, dans les zones à forte variation d'intensité, la nature même des frontières entre zones homogènes ne nous permet pas de les détecter de façon sûre ; de tels contours ne sont pas parfaits et leur localisation par recherche de zones homogènes n'est pas garanti.

2.3 Notre approche

Plutôt que de choisir entre l'une ou l'autre de ces deux approches, nous avons décidé de les faire *"coopérer"*, c'est-à-dire d'utiliser conjointement des propriétés d'homogénéité et des discontinuités locales dans l'image. L'idée est de contrôler un algorithme de recherche de régions par une carte de points de contours pré-calculée.

La suite de ce chapitre est consacrée à la définition formelle de la segmentation d'image en régions, puis à la présentation d'un algorithme de recherche d'une

partition par optimisations successives de sa qualité globale. Nous montrons ensuite comment caractériser les propriétés d'homogénéité des régions et la manière d'améliorer les résultats par l'utilisation d'une carte de points de contours.

3 Définition formelle de la segmentation

Une segmentation d'une image I utilisant le prédicat d'homogénéité P est généralement définie comme un ensemble $S = \{R_1, ..., R_n\}$ de I telle que : [Zucker 76]

(a) $I = \bigcup_{i=1,n} R_i$

(b) $\forall i \in [1,n]$, R_i *est connexe*,

(c) $\forall i \in [1,n]$, $P(R_i) = vrai$

(d) $\forall i \neq j \in [1,n]$, R_i *et* R_j *étant deux ensembles voisins*, $P(R_i \cup R_j) = faux$

- La première condition indique simplement que la segmentation est complète : tout point de l'image doit appartenir à une région et une seule. Un algorithme ne doit pas s'arrêter avant d'avoir traité tous les points de l'image.
- La seconde condition indique que les régions sont des ensembles de points connexes. C'est la raison pour laquelle les algorithmes prennent en compte le voisinage des points.
- Le critère de segmentation est défini à la troisième condition. Une région doit vérifier un prédicat d'homogénéité, lequel dépend généralement de la nature de l'image traitée. A titre d'exemple, citons un prédicat P imposant à des régions de posséder des points de niveaux de gris quasi-identiques :

$$P(R_i) = [Max_i - Min_i < Seuil]$$

Max_i et Min_i : niveau de gris maximum et minimum des points de la région R_i ;

$Seuil$ est le seuil accepté pour qualifier l'homogénéité des régions. Par exemple, $Seuil = 10$ pour des niveaux de gris variant de 0 à 255 pour l'image.

- La quatrième condition est une façon d'exprimer la maximalité de chaque région.

La vérification de ces quatre conditions est une **condition nécessaire et suffisante** pour qu'une partition d'une image soit une segmentation ; rien, toutefois, n'implique son unicité. En particulier, les résultats de la segmentation dépendent de l'ordre dans lequel les régions sont traitées et non uniquement du contenu de l'image comme cela devrait être le cas. L'exemple donné ci-après illustre cet inconvénient :

Soit une image 2×2 où chacun des quatre points est représenté par la valeur de son intensité lumineuse :

1	2
4	3

Si nous prenons le critère de ressemblance suivant :

> *"deux points voisins appartiennent à la même région si l'écart total entre leurs valeurs est inférieur ou égal à 1",*

nous obtenons plusieurs partitions qui dépendent de l'algorithme utilisé pour segmenter :

		ou		
1	2		1	2
4	3		4	3

Pour les mêmes raisons, une segmentation effectuée sur la transposée d'une image ne donne généralement pas la transposée de la segmentation de l'image originale.

Il est cependant possible de réduire cette indétermination en ajoutant l'optimisation d'une fonction C caractérisant la qualité globale d'une segmentation. Nous ajoutons alors aux quatre conditions précédentes, la condition **(e)** suivante :

(e) De toutes les segmentations S possibles qui vérifient les quatre conditions (a), (b), (c) et (d), nous cherchons la (ou une) segmentation S^* qui optimise la fonction de qualité globale C, c'est-à-dire :

$\forall S \in S_p(I),\ C(S^*) \geq C(S)$

où $S_p(I)$ est l'ensemble de toutes les partitions possibles de I ;
et C est une fonction positive, évaluant la qualité globale de la partition S.

Cette condition n'est pas suffisante pour obtenir une segmentation unique puisque plusieurs partitions de même valeur minimale pour la fonction C peuvent exister. Cependant, cette condition nous fournit une définition plus précise de la segmentation.

L'objectif est alors de trouver une *partition optimale* vérifiant des *contraintes de voisinage*. Si un tel problème peut être formulé de la façon suivante :

> *"Soit G un graphe valué, trouver une partition en sous-graphes connexes optimisant un critère donné",*

alors le problème est *NP-difficile* [Horowitz 78].

En effet, soit $I = \{i_1, i_2, \ldots i_n\}$ un ensemble de n valeurs ;
soit G le graphe complet de n sommets valués par les n valeurs $i_1, i_2, \ldots i_n$;
soit C le critère sur la partition $S = \{R_1, R_2, \ldots R_k\}$, défini par :

$$C(S) = \begin{cases} \infty & si\ k > 2 \\ Max_i\left(\sum_{j \in R_i}(i_j)\right) - Min_i\left(\sum_{j \in R_i}(i_j)\right) & sinon \end{cases}$$

S'il existe une partition S telle que $C(S) = 0$ alors il existe une partition de I en deux sous-ensembles tels que la somme des entiers de ces deux sous-ensembles soit égale.

Or le problème de l'existence d'une partition d'un ensemble d'entiers en deux sous-ensembles de sommes égales est *NP-complet*. Optimiser le critère $C(S)$ permettrait de répondre à ce problème d'existence. Cependant, notre propos n'est pas de décider de l'existence d'une telle partition, mais d'en déterminer une ; ce problème est donc *NP-difficile*.

Pour notre problème, notons ici que ni le critère, ni le graphe, ne correspondent exactement à notre représentation. Notons que notre algorithme travaille avec un graphe planaire représentant une partition image. Or, à aucun moment, la planaérité du graphe n'est prise en compte. L'utilisation de cette caractéristique permet peut-être au problème de ne pas être *NP-difficile*, mais nous n'avons pas cherché à le vérifier. Nous pouvons donc seulement conclure que la recherche d'une partition optimale est probablement un problème *NP-difficile*.

Nous nous proposons donc de résoudre ce problème de façon sous-optimale. Notre idée est d'adopter une **stratégie d'optimisation locale** au cours des étapes de l'algorithme, dont on espère que la solution ne soit pas trop distante de l'optimum réel.

Nous pouvons alors décomposer le problème de la segmentation d'images en régions en deux parties :

- établir une stratégie d'utilisation du prédicat d'homogénéité pour optimiser la qualité globale de la segmentation ; cette optimisation s'effectue au fur et à mesure du déroulement de l'algorithme ;
- définir le prédicat d'homogénéité à utiliser ; ce prédicat caractérise les régions à extraire et dépend donc de la nature de l'image traitée.

Cette première partie correspond à la définition de l'*outil algorithmique* à utiliser pour optimiser la qualité de la segmentation ; la seconde partie est la définition de l'*outil mathématique*. Nous présentons ces deux outils ci-après, leur définition et leur formalisation sont le résultat de la synthèse de nos travaux et de ceux d'O. Monga de l'INRIA, et de plus amples détails peuvent être trouvés dans [Gambotto 85], [Gagalowicz 85] et [Monga 87].

4 Algorithme

Il existe de nombreuses techniques de segmentation d'images, mais notre propos n'est pas d'en discuter dans ce document, et l'on pourra se reporter à la comparaison de ces méthodes par R. Haralick [Haralick 85] et à leur critiques par T. Pavlidis [Pavlidis 86]. Les trois principales classes des techniques de segmentation sont les suivantes :

- techniques de fusion,

- techniques de division,
- techniques de fusion/division.

Les méthodes fondées sur des fusions de régions[2] procèdent généralement par fusions successives jusqu'à trouver une partition qui satisfasse la définition donnée précédemment. La difficulté de ces méthodes est de trouver le critère de regroupement à utiliser.

Les méthodes procédant par division de régions[3] possèdent deux difficultés : trouver d'une part le critère de division et d'autre part la façon dont on divise une région non homogène.

Les méthodes mixtes combinent à la fois la nature ascendante des techniques de fusion et la nature descendante des techniques de division. Le principe est de partitionner arbitraitement l'image initiale, puis, à chaque étape, soit de diviser les régions si elles ne sont pas homogènes, soit de les fusionner si deux régions voisines sont similaires. Ces techniques cumulent les difficultés des deux techniques précédentes et posent également le problème de la stratégie à adopter.

Une solution simple consiste à diviser tout d'abord l'image initiale en régions très homogènes en appliquant un découpage simple (chacune des zones est partagée en quatre sous-régions par exemple), puis à fusionner des régions adjacentes vérifiant un prédicat d'homogénéité [Pavlidis 77].

Nous avons choisi de mettre en œuvre une technique de croissance de régions, mais pour des raisons de coût en temps de calcul, l'algorithme de fusion débute, non pas sur l'image originale, mais sur une partition initiale obtenue par divisions successives jusqu'à obtenir de petites régions très homogènes. Notre algorithme s'apparente alors à la troisième classe de technique : division/fusion.

4.1 Structure de l'algorithme

L'idée de base de l'algorithme est d'optimiser la qualité globale C de la segmentation par fusions successives de régions vérifiant le prédicat P [Morris 86]. Parmi toutes les fusions de régions adjacentes permises à une étape, seule la meilleure va être réalisée. Ceci revient à chercher, dans toute l'image, les deux régions adjacentes dont la réunion optimise le critère de fusion Q.

Nous avons donc la procédure *Croissance* suivante :

Procédure Croissance (Prédicat P, Qualité Q) :

- *tantqu'il existe un couple de régions adjacentes dont la réunion vérifie le prédicat d'homogénéité P faire*
  - *choisir parmi tous les couples (R_i, R_j) vérifiant $P(R_i \cup R_j)$, celui pour lequel la qualité locale de fusion $Q(R_i \cup R_j)$ est optimale*

[2] Ces méthodes sont également appelées "croissance de régions" ou "regroupement de régions", qui se rassemblent sous les dénominations "merge" ou "region growing" en anglais.
[3] Traditionnellement appelées "split" en anglais.

  - *mettre à jour la partition S :*

$$S = S - \{R_i\} - \{R_j\} \cup \{R_i \cup R_j\}$$

avec S : partition en cours de création ;

L'algorithme de segmentation en régions a alors la structure suivante :

- *données :*
  I : image initiale ;
  P : prédicat d'homogénéité ;
  Q : fonction de qualité locale.
- $S = \{\{(k,l)\}/(k,l) \in I\}$
- *Croissance (P,Q)*

4.2 Structure de données et implantation de l'algorithme

La détermination, à chaque étape, de tous les couples de régions candidats à une fusion, sous-tend une complexité algorithmique importante. Des structures de données adaptées permettent cependant d'aboutir à une implantation algorithmique de coût réduit.

Graphe image :

La partition en cours de création se présente sous la forme d'un ensemble de régions, chacune de ces régions est caractérisée par un ensemble de valeurs (attributs) et possède un lien avec ses voisins ; ce lien est également caractérisé. Ceci nous conduit naturellement à représenter une partition en régions sous forme d'un graphe d'adjacence valué :

- à chaque nœud du graphe, nous associons une région et ses attributs (nombre de points, somme des niveaux de gris des points, ...) ;
- à chaque arc du graphe, nous associons deux régions et les attributs qui caractérisent la relation existant entre les deux nœuds concernés. Nous ne conservons bien évidemment que les arcs entre les paires de nœuds qui vérifient certaines propriétés comme la connexité (images 2d) ou le recouvrement (images 3d).

L'implantation de cette structure de données doit essentiellement favoriser les deux traitements suivants :

- accès rapide au meilleur arc du graphe ; c'est-à-dire le meilleur couple de régions adjacentes qui vérifient le prédicat d'homogénéité P ;
- mise à jour facile des propriétés locales (attributs) des nœuds et des arcs modifiés par la fusion.

Accès rapide au meilleur arc :

L'accès facile au meilleur arc du graphe impose une gestion triée de ces arcs. Nous utilisons un arbre binaire de recherche [Aho 87] pour stocker les couples de régions proposés à la fusion : nous plaçons chaque pointeur d'arc liant deux régions qui vérifient le prédicat d'homogénéité P dans un arbre. Cet arbre stocke les arcs triés en fonction des valeurs de la qualité locale de fusion Q calculée sur les couples de nœuds concernés. A la feuille d'une branche située d'un côté de l'arbre se trouve le pointeur sur le couple de régions à fusionner. L'extraction de cette valeur se fait en $\mathcal{O}(log(n))$. Remarquons que nous ne nous somme pas préoccupés de l'équilibrage de l'arbre, les nœuds qui servent à la construction de l'arbre sont insérés selon un ordre quelconque. Nous espérons qu'en moyenne l'arbre soit proche de l'équilibre.

Lors de la fusion de deux nœuds N_1 et N_2, nous créons un nouveau nœud N_3 représentant l'union des deux anciens nœuds et autant de nouveaux arcs que de régions adjacentes aux deux régions N_1 et N_2. Chaque nouvel arc, liant deux nœuds et vérifiant le prédicat d'homogénéité P, est ajouté dans l'arbre binaire de recherche en fonction de la qualité de fusion Q des deux régions concernées.

Dans ce cas, la recherche d'un arc de coût minimal dans l'arbre se poursuit jusqu'à trouver un arc entre deux nœuds qui existent encore.

En Annexe B, nous présentons un exemple de représentation d'un graphe d'adjacence et d'un arbre binaire de recherche associés à une image et leur évolution au cours du processus de segmentation.

Mise à jour facile des attributs du graphe image :

La mise à jour des attributs des nœuds et des arcs, pour être rapide, ne doit évaluer les nouveaux attributs qu'en fonction des anciens et éviter ainsi tout retour coûteux sur les données de l'image initiale. Il faut également que le prédicat d'homogénéité P et la fonction de qualité locale de fusion Q d'un couple de régions puissent être calculés directement en fonction des attributs des deux régions concernées. Ceci se traduit par les conditions **(1)** et **(2)** suivantes :

Posons S : *partition de l'image ;*
R_i *et* R_j : *deux régions de la partition ;*
P : *prédicat d'homogénéité ;*
Q : *qualité de la fonction locale de fusion ;*
N : *vecteur d'attributs d'une région (nœud du graphe d'adjacence) ;*
A : *vecteur d'attributs d'une relation entre deux régions (arc du graphe d'adjacence).*

(1) $\forall R_i, R_j \in S$,
il existe deux fonctions F_p *et* F_q *d'évaluation du prédicat d'homogénéité* P *et de la qualité locale* Q *des deux régions, telles que :*
$P(R_i \cup R_j) = F_p(N(R_i), N(R_j), A(R_i, R_j))$
$Q(R_i \cup R_j) = F_q(N(R_i), N(R_j), A(R_i, R_j))$

(2) $\forall R_i, R_j \in S$,
il existe deux fonctions F_n *et* F_a *de mise à jour des attributs des nœuds et des arcs, telles que :*
$N(R_i \cup R_j) = F_n(N(R_i), N(R_j))$
$A(R_i, R_j \cup R_k) = F_a(A(R_i, R_j), A(R_i, R_k))$

Des détails supplémentaires peuvent être trouvés dans [Monga 87].

L'algorithme prend alors la structure suivante :

- *<u>données :</u>*
  F_p *et* F_q *: les fonctions d'évaluation des prédicats et de la qualité locale ;*
  F_n *et* F_a *: les fonctions de mise à jour des attributs des nœuds et des arcs du graphe d'adjacence ;*
  P *: prédicat d'homogénéité ;*
  Q *: fonction de qualité locale.*
- *construire le graphe image d'adjacence*
- *mettre dans l'arbre binaire de recherche tous les couples régions adjacentes* (R_i, R_j) *vérifiant* $P(R_i, R_j)$ *en fonction de la qualité locale de fusion* $Q(R_i, R_j)$
- *<u>tantque</u> arbre non vide <u>faire</u>*
  - *rechercher un arc de coût minimal dans l'arbre jusqu'à trouver un arc entre deux nœuds qui existent encore*
  - *effectuer la fusion par une mise-à-jour des nœuds et des arcs à l'aide des deux fonctions* F_n *et* F_a

Le résultat de la segmentation se trouve dans le graphe image d'adjacence.

4.3 Complexité de l'algorithme

Etudions la complexité de l'algorithme proposé :

Soit N le nombre de nœuds du graphe initial.

Soit A le nombre d'arcs du graphe initial.

Chaque nœud du graphe est connecté en moyenne à $2A/N$ nœuds. Posons $V = 2A / N$. Cette valeur reste sensiblement identique pendant toute la durée du processus de segmentation. En effet, d'après la relation d'Euler, nous avons la relation suivante pour un graphe connexe :

$$N + F - A = 2 \tag{4.1}$$

avec N : nombre de sommets (nombre de nœuds dans notre cas) ;
A : nombre d'arcs ;
F : nombre de facettes (une facette est définie par les arcs qui la délimitent) ;

Cette formule est également vraie dans le cas particulier d'un graphe planaire, comme celui avec lequel nous travaillons. Or,

- une facette est délimitée par un ensemble d'arcs, une facette est donc limitée par au moins 3 arcs ;
- chaque arc délimite au plus 2 facettes ;

nous pouvons donc majorer le nombre de facettes par :

$$F \leq \frac{A \times 2}{3} \tag{4.2}$$

De (4.1) et (4.2), nous déduisons un majorant du nombre d'arcs :

$$A \leq 3 \times (N - 2)$$

Ceci signifie que le nombre d'arcs dans un graphe est toujours inférieur à 3 fois le nombre de facettes :

$$\frac{A}{N} \leq 3 \tag{4.3}$$

De plus :

$$\frac{N}{2} \leq A \tag{4.4}$$

De (4.3) et (4.4), nous concluons que :

$$0,5 \leq \frac{A}{N} \leq 3$$

d'où :

$$1 \leq \frac{2A}{N} \leq 6$$

$$1 \leq V \leq 6$$

La valeur V peut varier dans un intervalle très restraint. Le nombre moyen de nœuds connectés à un autre nœud reste donc sensiblement constant pendant la segmentation comme le montre le tableau de valeurs présenté à la fin de ce paragraphe. Pour cet exemple V varie entre 5.153 et 5.430.

Soit F le nombre de fusions réalisées pendant tout le traitement.

Soit T_n le nombre d'opérations effectuées par la fonction F_n de mise à jour des attributs des nœuds après une fusion.

Soit T_a le nombre d'opérations effectuées par la fonction F_a de mise à jour des attributs des arcs après une fusion.

Soit T_x le nombre d'opérations effectuées pour l'évaluation d'un prédicat d'homogénéité et de la fonction de qualité locale entre deux régions.

Soit M le nombre de fusions potentielles de chaque nœud avec l'un de ses voisins qui vérifient le prédicat d'homogénéité.

- Le nombre d'opérations nécessaire pour mettre à jour les attributs des nœuds et des arcs dans le graphe image est de :

$$F \times (T_n + T_a \times V)$$

- Après la fusion de deux nœuds, il faut calculer V nouvelles valeurs du prédicat d'homogénéité puisque ce nouveau nœud est connecté à une moyenne de V autres nœuds, le nombre d'opérations demandé est de :

$$F \times (V \times T_x)$$

- Seuls M arcs parmi V nouveaux arcs vérifient le prédicat P et doivent donc être insérés dans l'arbre binaire de recherche, le nombre d'opérations est de :

$$F \times (M \times log(M \times N))$$

puisque le nombre maximal d'éléments dans cette structure est de : $M \times N$.

- La complexité de tout l'algorithme s'écrit alors :

$$\mathcal{O}(F \times (T_n + V \times T_a + V \times T_x + M \times log(M \times N)))$$

Le nombre F de fusions est directement lié au nombre de régions à traiter, en revanche, les autres coefficients intervenant dans le calcul de la complexité dépendent seulement de la "densité d'arc" de ce graphe. Il est donc très coûteux d'appliquer cet algorithme sur l'image originale : le graphe est alors composé de tous les points de l'image comme nœuds et de tous les points en relation[4] comme arcs.

L'idée est alors d'effectuer une partition initiale de l'image avant d'utiliser notre algorithme de regroupement.

Nous présentons ci-après le tableau de valeurs illustrant le fait que chaque région de la partition à un nombre quasi constant de régions voisines, compris entre 5 et 6 pour notre exemple. Les étapes de l'algorithme sont arbitraires, elles correspondent à différents arrêts lors du déroulement du processus et sont déterminées aléatoirement.

[4] 4 ou 8 voisins dans le cas d'une image 2d par exemple, si l'on travaille en 4 ou 8-connexité.

Nombre moyen de régions voisines par région Image de la lampe présentée figure 4.4, page 77 de taille 256 × 256 points			
étapes du processus de segmentation	nombre de régions (nœuds) N	nombre de relations de voisinage (arcs) A	nombre moyen de régions voisines par région $V = 2A/N$
1	6259	16128	5.153
2	5622	14584	5.188
3	4781	12483	5.220
4	3965	10368	5.229
5	3191	8579	5.210
6	2772	7181	5.181
7	2643	6819	5.160
8	393	1074	5.460
9	295	803	5.440
10	288	782	5.430

4.4 Partition initiale

Cette première partition est réalisée par une technique classique de division de régions : la procédure SPLIT de Pavlidis [Pavlidis 77]. La méthode consiste à diviser récursivement l'image en quatre parties égales jusqu'à ce que chaque zone obtenue vérifie un prédicat d'homogénéité (figure 4.3). Pour implanter ceci, nous utilisons une description de l'image sous forme de "quad-tree", aussi appelé arbre quaternaire. Des détails concernant cette structure sont donnés dans l'Annexe A.

Le prédicat utilisé peut être identique à celui utilisé dans l'algorithme de fusion, mais en diminuant sa tolérance. Ceci permet d'obtenir des régions petites mais très homogènes, et de réduire le volume des données d'environ 75 %, ce qui est nettement appréciable.

Par exemple, le prédicat d'homogénéité P suivant peut être utilisé pour effectuer une partition initiale :

$$P(R_i) = [Max_i - Min_i < Seuil]$$

avec R_i : région traitée ;
Max_i, respectivement Min_i : niveau de gris maximum, respectivement minimum, des points de la région R_i ;
$Seuil$: seuil accepté pour qualifier l'homogénéité des régions.

Pour une valeur de $Seuil$ de 30 sur une image 256 × 256 de niveaux de gris variant de 0 à 255 (figure 4.4), nous obtenons les résultats présentés figure 4.5.

Une première version de l'algorithme de création de la partition initiale a été de construire les régions par agglomération de points au fur et à mesure du balayage séquentiel de l'image [Wrobel 85]. Mais il s'est avéré que le sens de parcours des

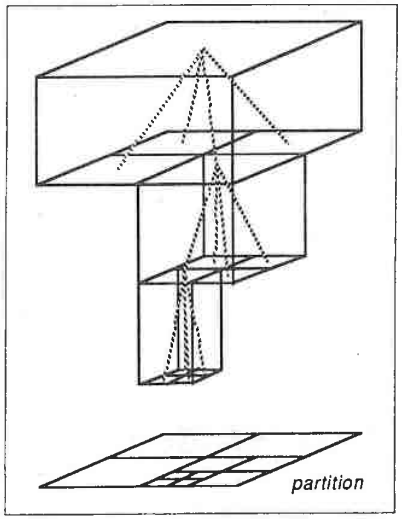

Figure 4.3: *Division récursive de régions suivant un critère d'homogénéité.*

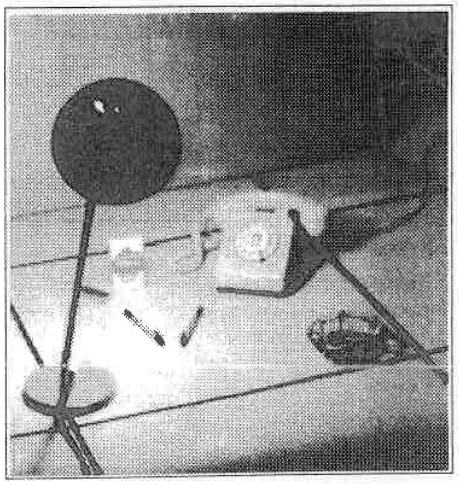

Figure 4.4: *Image originale. 65536 points de 256 niveaux de gris.*

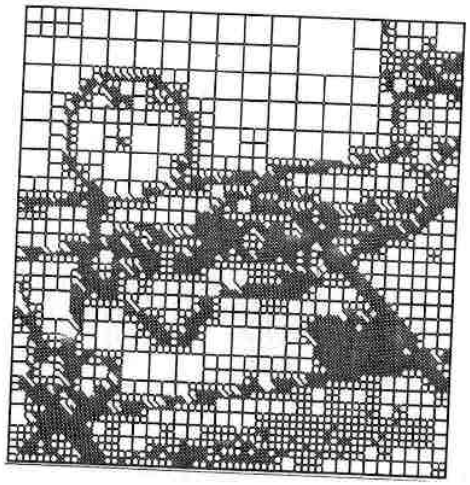

Figure 4.5: *Partition initiale (11119 régions).*

données avait une influence non négligeable sur la forme des régions créées, qui se retrouvait ensuite dans la partition finale. L'image 4.6 illustre la déformation des régions de la partition initiale, le balayage de l'image étant effectué de la gauche vers la droite et de haut en bas. Cette technique de création de la partition initiale a donc été abandonnée au profit de la méthode de division présentée ci-avant.

4.5 Résumé

Nous venons de définir une stratégie de segmentation par :

- un algorithme itératif de fusions de régions qui optimise, localement au fur et à mesure de son déroulement, la qualité globale de la segmentation ;
- un graphe image d'adjacence valué et deux fonctions de mise à jour des attributs des nœuds et des arcs de ce graphe ;
- des structures de données adaptées aux traitements effectués.

L'outil algorithmique étant défini, nous pouvons passer à l'étude de la segmentation d'images de scènes naturelles et définir le (ou les) prédicat(s) d'homogénéité à utiliser qui, eux, dépendent des caractéristiques de l'image traitée.

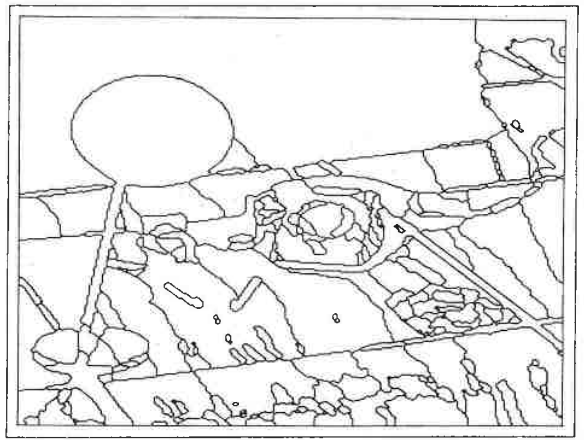

Figure 4.6: *Partition initiale obtenue par balayage séquentiel de l'image. Le sens de traitement des données influe sur la forme des régions.*

5 Prédicats d'homogénéité

L'homogénéité des régions dépend de la nature de l'image traitée. Les régions extraites d'une image satellite ne peuvent pas posséder les mêmes propriétés d'homogénéité que les régions extraites d'images de scènes d'intérieur.

L'objectif de la segmentation est de déterminer la projection des différents objets présents dans la scène observée. Or une image naturelle dépend des diverses caractéristiques de la scène, telles que l'orientation et les matériaux des objets, l'éclairage, l'angle de prise de vue de la caméra, etc. La conjonction de tous ces paramètres fournit une image ne contenant qu'un seul type d'information : le **pixel**, c'est-à-dire l'association d'une *intensité lumineuse* à une *position* dans l'image. Il est évidemment plutôt difficile de séparer l'influence de ces différents facteurs afin de trouver les projections des objets.

La qualité de la segmentation est maximale si chaque région correspond à un objet et réciproquement. Pour arriver à une telle segmentation, il faut connaître précisément les propriétés d'homogénéité des régions ce qui n'est généralement pas le cas ; il est en effet difficile de trouver une caractérisation de l'homogénéité globale et adaptée à toutes les régions.

Nous montrons ci-après que l'utilisation d'un prédicat d'homogénéité simple est insuffisant pour obtenir une bonne segmentation. Il s'avère nécessaire :

- soit d'utiliser un prédicat plus complexe, combinaison de prédicats "élémentaire".

- soit, plus simplement, d'utiliser plusieurs prédicats "élémentaires" de manière séquentielle.

5.1 Utilisation d'un prédicat d'homogénéité

Il semble assez naturel de travailler avec les moyennes de niveaux de gris des régions, par exemple, et de fusionner des régions adjacentes de moyennes quasi-identiques. Le prédicat d'homogénéité s'écrit alors comme suit :

$$P(R_i \cup R_j) = [\mid Moy_i - Moy_j \mid < Seuil]$$

avec R_i et R_j : les deux régions qui sont proposées au fusionnement si le prédicat est vrai :
Moy_i et Moy_j : moyennes respectives des niveaux de gris des points des régions R_i et R_j ;
Seuil : seuil de fusionnement.

Seuil est à déterminer ; expérimentalement, aucune valeur n'est satisfaisante. Prenons par exemple $Seuil = 10$ pour une image où les niveaux de gris des points varient entre 0 et 255 ; la partition résultat est sur-segmentée (figure 4.7).

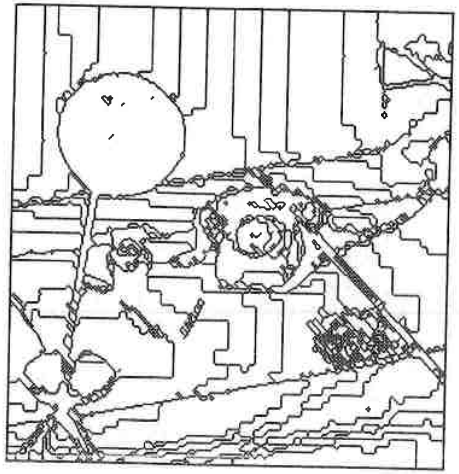

Figure 4.7: *Segmentation finale, Seuil = 10 (3578 régions).*

L'augmentation de la valeur de *Seuil*, pour s'affranchir des problèmes de sur-segmentation essentiellement dus aux dégradés de lumière, provoque des fusions intempestives de régions adjacentes appartenant à des objets différents. Ceci est

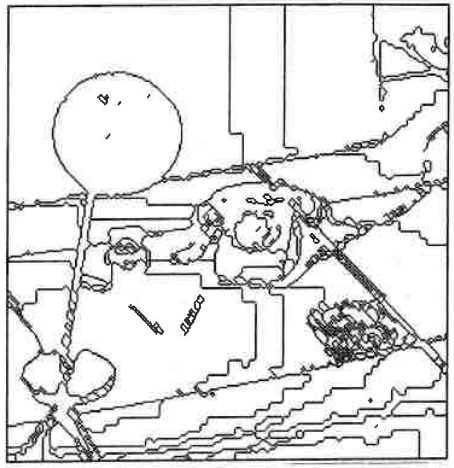

Figure 4.8: *Segmentation finale, Seuil = 15 (2786 régions).*

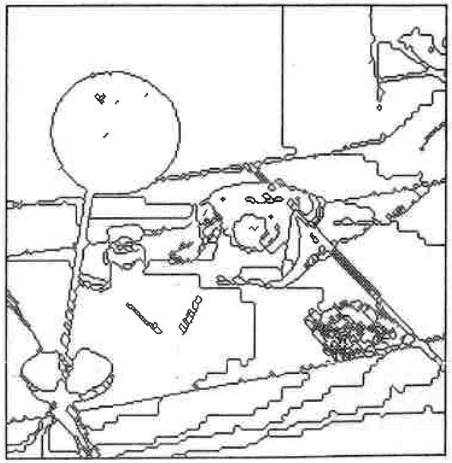

Figure 4.9: *Segmentation finale, Seuil = 20 (2162 régions).*

illustré (figures 4.8 et 4.9) par la fusion des régions correspondantes au téléphone et au support sur lequel il est posé.

La conclusion de ces essais est qu'il n'est pas suffisant de caractériser l'homogénéité de régions par la moyenne des niveaux de gris de leurs points. L'utilisation isolée d'autres prédicats aboutit aux mêmes constatations. lorsque la scène étudiée est aussi complexe que celle présentée.

5.2 Utilisation de plusieurs prédicats

La solution que nous proposons pour obtenir de meilleurs résultats est d'utiliser de façon optimale plusieurs propriétés d'homogénéité, ces propriétés devant être décomposées de manière hiérarchique. A chacune de ces propriétés est associé un prédicat ; chaque prédicat est utilisé séquentiellement par l'algorithme sous-optimal décrit précédemment. Cet algorithme prend alors la structure suivante :

- *données :*
  (P_1, Q_1), ..., (P_n, Q_n) : *séquence de prédicats et de fonctions de qualité locale associées.*
- *construire la partition initiale (généralement à l'aide du prédicat P_1)*
- *pour chacune de ces n paires (P_k, Q_k) faire*
  - *Croissance (P_k, Q_k)*

Dans le cas de la segmentation d'images naturelles à niveaux de gris, les propriétés souhaitées pour chaque région sont les suivantes :

(a) La moyenne des gradients calculée sur les points d'une région est faible, c'est-à-dire qu'aucun contour ne traverse la région.

(b) Une région peut être décomposée en petites zones de moyennes quasi-identiques.

(c) Chacune de ces zones peut être décomposée en zones de points de niveaux de gris quasi-identiques.

Ces trois propriétés servent à définir les trois prédicats d'homogénéité ; il est évident que le premier prédicat P_1 de la séquence travaille au niveau de l'image initiale, le second prédicat P_2 travaille sur les régions obtenues à l'étape précédente par le prédicat P_1, etc.

Soit I une image, $I(k,l)$ un point de cette image, et $S = \{R_1, ...R_n\}$ une partition de I, les trois prédicats issus des trois propriétés sont les suivants :

(c) Prédicat P_1 :

$$P_1(R_i \cup R_j) = [Max_{ij} - Min_{ij} < Seuil_1]$$

avec Max_{ij}, respectivement Min_{ij} : niveau de gris maximum, respectivement minimum, des points de la "nouvelle" région $R_i \cup R_j$;
$Seuil_1$: le seuil accepté pour l'homogénéité des régions.

- La fonction de qualité globale à optimiser s'écrit sous la forme :

$$C_1(S) = \sum_{i=1}^{n} \sum_{k,l \in R_i} \left((I(k,l) - Max_i)^2 - (I(k,l) - Min_i)^2 \right)$$

- La fonction de qualité locale s'écrit :

$$Q_1(R_i \cup R_j) = Max_{ij} - Min_{ij}$$

(b) Prédicat P_2 :

$$P_2(R_i \cup R_j) = [\mid Moy_i - Moy_j \mid < Seuil_2]$$

avec R_i, R_j : les régions candidates à une fusion ;
Moy_i, Moy_j : moyennes respectives des niveaux de gris des points des régions R_i et R_j ;
$Seuil_2$: seuil de fusion.

- La fonction de qualité globale à optimiser s'écrit sous la forme :

$$C_2(S) = \sum_{k,l \in I} (I(k,l) - M_{kl})^2$$

où M_{kl} est la moyenne des niveaux de gris des points de la région à laquelle appartient le point (k,l).

- La fonction de qualité locale s'écrit :

$$Q_2(R_i \cup R_j) = \mid Moy_i - Moy_j \mid$$

(a) Prédicat P_3 :

$$P_3(R_i \cup R_j) = \left[\frac{\sum_{((i,j),(k,l)) \in F(R_i,R_j)} \mid I(i,j) - I(k,l) \mid}{Card(F(R_i,R_j))} < Seuil_3 \right]$$

avec $F(R_i, R_j)$: l'ensemble des couples de points situés le long de la frontière entre les régions R_i et R_j (figure 4.10) ;
$Card(F(R_i, R_j))$: longueur de la frontière entre R_i et R_j ;
$Seuil_3$: seuil de fusion.

- La fonction de qualité globale s'écrit sous la forme :

$$C_3(S) = \sum_{((i,j),(k,l)) \in X_S(I)} | I(i,j) - I(k,l) |$$

avec X_S : l'ensemble des couples de points connexes de I appartenant à une même région, et ce, pour toutes les régions de la partition S.

- La fonction de qualité locale s'écrit :

$$Q_3(R_i \cup R_j) = \frac{\sum_{((i,j),(k,l)) \in F(R_i,R_j)} | I(i,j) - I(k,l) |}{Card(F(R_i, R_j))}$$

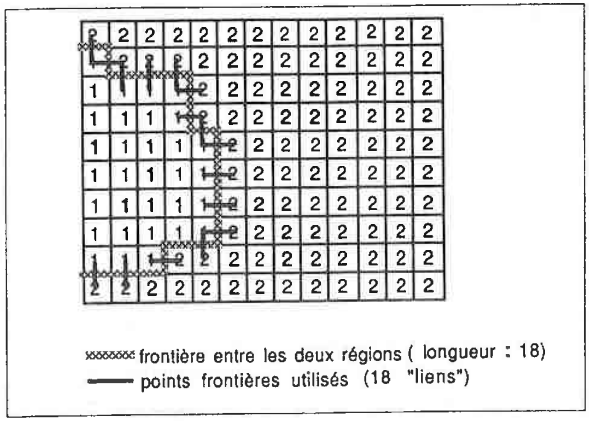

Figure 4.10: *Représentation de la frontière entre les deux régions R_1 et R_2.*

Nous donnons figure 4.11, un exemple de résultat obtenu avec l'utilisation séquentielle de ces trois prédicats d'homogénéité. Les améliorations apportées à la segmentation sont évidentes. Une adaptation plus fine des différents seuils utilisés permet d'obtenir une meilleure partition, mais puisqu'une détection automatique de tels paramètres en fonction de l'image traitée n'est pas réalisée, nous préférons travailler avec un jeu de valeurs identiques pour tous les types d'images traités.

L'utilisation optimale d'une suite de critères de regroupement permet de mieux utiliser les propriétés d'homogénéité des régions que l'on recherche. Ces critères dépendent des caractéristiques des images traitées et doivent être redéfinis pour chaque type d'application.

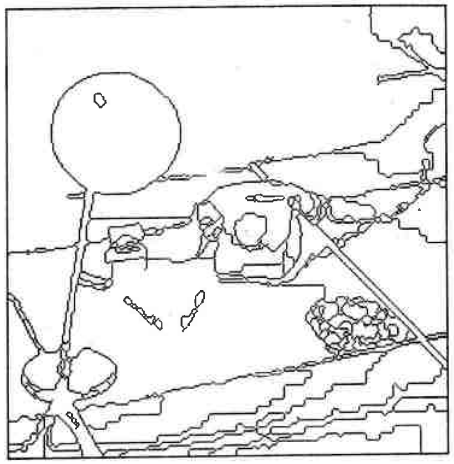

Figure 4.11: *Partition obtenue par l'utilisation optimale de trois prédicats d'homogénéité successifs (162 régions).*

6 "Coopération" régions-contours

L'idée implicite est d'utiliser conjointement les caractéristiques d'homogénéité des régions et les discontinuités locales (points de contours).

Avant de poursuivre, notons que nous préférons utiliser le terme **"point de contraste"** plutôt que **"point de contour"** qui est pourtant plus fréquemment utilisé dans la littérature. Cependant, il faut remarquer que le terme point de contraste est plus adapté car il se réfère seulement aux propriétés de l'image et non aux formes qui peuvent être présentes dans la scène [Demazeau 86]. Par exemple, les points de contraste dûs à la texture d'une surface ne correspondent à aucun contour d'objet.

Les points de contraste, préalablement calculés, sont utilisés pour guider le processus de segmentation ; ils contrôlent la croissance des régions comme l'illustre la figure 4.12.

Ce contrôle de la croissance des régions est effectué en interdisant toute fusion entre régions adjacentes pour lesquelles la proportion de points de contraste situés sur leur frontière commune dépasse un certain seuil. Ce seuil est fixé à 0 pour tous les prédicats sauf celui évaluant les "forces" des frontières (prédicat P_3). Les trois prédicats précédemment cités sont donc modifiés de la façon suivante :

$$P_1(R_i) = [Max_i - Min_i < Seuil_1] \wedge [Ct(R_i, R_j) = 0]$$

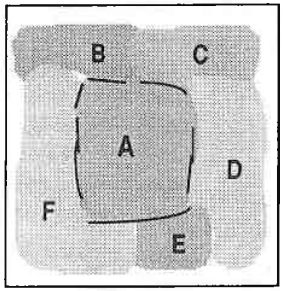

Figure 4.12: *Superposition de la carte de points de contraste et de la partition en régions en cours de création. La région A est suffisamment entourée de points de contraste, sa croissance est arrêtée.*

$$P_2(R_i \cup R_j) = [\mid Moy_i - Moy_j \mid < Seuil_2] \wedge [Ct(R_i, R_j) = 0]$$

$$P_3(R_i \cup R_j) = \left\{ \begin{array}{ll} \left[\frac{\sum_{(i,j),(k,l) \in F(R_i,R_j)} |I(i,j) - I(k,l)|}{Card(F(R_i,R_j))} < Seuil_3 \right] & \underline{\text{si}} \left(\frac{Ct(R_i,R_j)}{Card(F(R_i,R_j))} < Seuil_4 \right) \\ faux & \underline{\text{sinon}} \end{array} \right\}$$

pour refuser toute fusion comportant une proportion
trop importante de points de contraste sur la frontière
commune des deux régions concernées.

avec $Ct(R_i, R_j)$: nombre de points de contraste sur la frontière entre les deux régions R_i et R_j.

Cette "coopération" régions-contours permet d'utiliser des seuils plus tolérants lors de l'évaluation des prédicats d'homogénéité, donc d'autoriser plus de fusions fondées sur une comparaison des attributs des régions, sans dégrader la partition obtenue.

Notre algorithme apparaît comme un moyen très naturel de concilier l'information contours et l'information régions [Wrobel 87b]. Ceci est utile en particulier dans les zones de l'image où aucune de ces deux informations prise séparément ne permet de déterminer correctement les contours des objets. Dans ce cas, assez fréquent pour les images de type scène d'intérieur, la projection d'un objet est entourée de lignes de contraste brisées et est voisine de régions homogènes de mêmes caractéristiques. La hiérarchie du processus de regroupement permet la formation de la région correspondant à la projection de l'objet. Cette région n'est ensuite plus fusionnée, car elle est entourée de points de contraste.

Notre méthode peut être résumée de la façon suivante : nous effectuons une première segmentation (extraction de points de contraste) avec des critères stables ;

puis une seconde segmentation où les critères sont plus tolérants car portant sur des valeurs peu stables.

Les régions grandissent, soit jusqu'à buter sur les lignes de contraste, soit jusqu'à atteindre le seuil maximal définissent leur homogénéité. De cette manière, la localisation des frontières des régions est correcte ; ces frontières sont généralement situées sur les lignes de contraste pré-calculées.

Cette approche nous permet de tirer profit de la richesse des informations contenues dans les deux types de primitives, et également de leur dualité.

Nous présentons ci-après une carte de contraste (figure 4.13) et la segmentation obtenue (figure 4.14) avec les prédicats indiqués précédemment. L'extraction des points de contraste est effectuée selon la méthode développée par R. Deriche à l'Inria-Rocquencourt [Deriche 87a]. Son implantation dans notre laboratoire a été réalisée par D. Ziou. Nous en utilisons donc ici seulement les résultats.

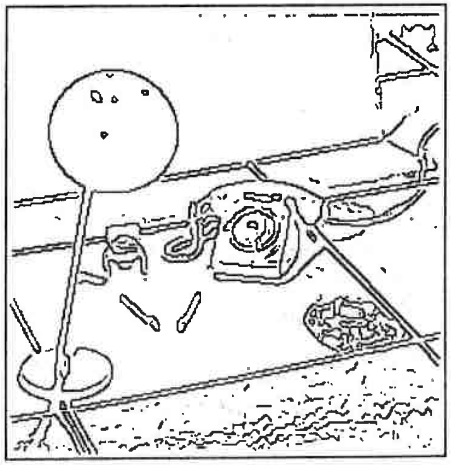

Figure 4.13: *Extraction des points de contraste par la méthode développée par R. Deriche [Deriche 87a].*

Remarquons que n'importe quelle carte de points de contraste définissant des points d'arrêt peut, en fait, être utilisée.

7 Conclusion

Le choix des primitives à apparier entre les deux images stéréoscopiques s'étant porté sur les régions, nous avons été confrontés au problème de leur extraction. Les nombreuses techniques de segmentation d'images existantes ne nous ont pas

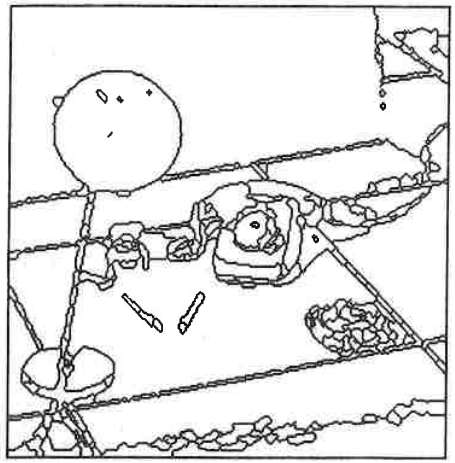

Figure 4.14: *Résultat de la segmentation où la croissance de régions est contrôlée par une carte de points de contraste pré-calculée.*

semblé satisfaisantes dans la mesure où elle dépendent trop souvent du domaine d'application considéré, mais nous nous en sommes cependant fortement inspirés pour construire notre approche.

Discussion de notre approche :

La méthode développée possède quatre points en sa faveur :

- Elle sépare l'algorithme de la définition des prédicats d'homogénéité qui caractérisent les régions recherchées. Cette séparation confère à notre algorithme sa **généralité**.
- Aucun sens d'examen des données n'est défini a priori, l'algorithme est essentiellement guidé par les données.
- L'algorithme **optimise** au fur et à mesure de son déroulement la qualité globale de la partition. Ceci nous permet d'obtenir une segmentation que nous espérons proche de l'optimum réel.
- L'utilisation de **plusieurs prédicats d'homogénéité** autorise une meilleure définition des caractéristiques des régions. Ils sont utilisés de manière séquentielle par l'algorithme sous-optimal.

Cependant, avec l'approche que nous avons déterminée, nous ne proposons aucun moyen pour définir les seuils apparaissant dans les prédicats. Les seuils utilisés

prennent des valeurs trouvées empiriquement et qui sont employées pour toutes les images de type scène d'intérieur. La détermination automatique des paramètres est un objectif que nous avons tenté d'atteindre, sans toutefois aboutir à une solution satisfaisante. Cette phase de test d'adaptation des seuils correspond au premier inconvénient de notre méthode.

Le deuxième inconvénient est de déterminer les prédicats à utiliser selon le type d'images traitées : nous ne proposons pas de manière pour les construire, ce travail est laissé au soin de l'utilisateur. Il faut donc décrire l'image sous forme d'une suite de propriétés que les régions doivent vérifier. De là sont déduits les prédicats d'homogénéité. Leur utilisation séquentielle dépend de la façon dont les propriétés sont "emboîtées". Les propriétés peuvent être de toutes catégories, comme par exemple de type sémantique : quand une région a une signification sémantique, elle est reconnue et sa croissance peut être arrêtée.

Le travail qui reste à faire pour l'utilisateur se résume de la façon suivante :

- décrire la suite des propriétés que doivent vérifier les régions à extraire ;
- définir la suite des prédicats d'homogénéité et les fonctions locales de qualité à optimiser ;
- déterminer les différents seuils utilisés dans les prédicats.

Remarquons que les prédicats d'homogénéité peuvent se combiner pour ne former qu'un seul prédicat contrôlant la fusion des régions. Ceci pose cependant les problèmes de déterminer la fonction de combinaison des prédicats et de trouver leurs coefficients de pondération, et risque également d'allourdir le traitement par des calculs inutiles.

L'idée d'utiliser plusieurs critères de regroupement apparaît aussi dans le travail de F. Zaroli [Zaroli 83] qui propose une approche liant à la fois les techniques de classification et de fusion. Cet outil a été créé pour un module d'interprétation de scènes et utilise plusieurs images de la scène étudiée : images à niveaux de gris, images de distance, etc.

Applications :

Pour illustrer la généralité de cette approche, citons son utilisation pour la segmentation d'images 3d sismiques [Monga 86], ou dans le domaine du mouvement. Dans ce domaine, la segmentation représente une étape intermédiaire importante lors du processus de traitement de séquences d'images et d'analyse de scène dynamique, et l'un des objectifs est d'obtenir une segmentation selon des critères de mouvement. L'outil algorithmique est identique, les prédicats d'homogénéité au sens du mouvement représentent des modèles de vitesse. Deux prédicats sont actuellement utilisés, le premier pour obtenir une partition en régions de "mouvement constant", le second pour déterminer les régions de "mouvement linéaire" [Bouthemy 87]. Des prédicats qualifiant des mouvements plus complexes peuvent ensuite être utilisés.

Utilisation de l'information "contour" :

Nous avons également montré comment réaliser la "coopération" entre un détecteur-contour et un détecteur-région : l'utilisation conjointe de deux informations duales permet d'obtenir de meilleures partitions. Une carte de points de contraste pré-calculée permet de contrôler la croissance des régions ; ceci joue le rôle d'une adaptation locale des valeurs des prédicats d'homogénéité. En effet, les fusions de régions se font plus facilement dans les zones de l'image où il n'y a que peu de points de contraste, donc représentant certainement une seule surface, même si les caractéristiques locales des régions sont assez différentes. Ce cas est illustré par les régions de la partition initiale situées sur une surface où la lumière est dégradée.

L'inconvénient de la segmentation en régions "classique" est ici corrigé : les frontières des régions sont bien localisées, car elles s'appuyent sur des lignes de contraste fiables. De plus l'homogénéité des régions, généralement assez délicate à définir, n'est plus la seule caractéristique prise en compte.

Application à la stéréovision :

Nous donnons figure 4.15 le schéma récapitulatif des traitements à effectuer sur chaque image du couple stéréoscopique et figures 4.17 et 4.19 les résultats obtenus sur les images originales présentées figures 4.16 et 4.18.

Il faut cependant mentionner que, si la segmentation de chaque image d'une paire stéréoscopique est satisfaisante, il subsiste des différences, quelquefois très importantes, entre les deux partitions à apparier comme nous pouvons le constater sur les images présentées. Ces différences sont essentiellement dûes à l'écart entre les angles d'observation de la scène.

Notre idée est de tenir compte de ces différences entre les partitions lors de l'appariement. Le processus d'appariement accepte en données les deux partitions et cette redondance d'information peut lui permettre de jouer le rôle, dans certaines parties de l'image et lorsque le besoin s'en fait sentir, d'un processus de segmentation. L'objectif est de remettre en cause certaines zones de la partition pour améliorer la qualité de correspondance entre les deux images. Ceci fait l'objet du chapitre suivant.

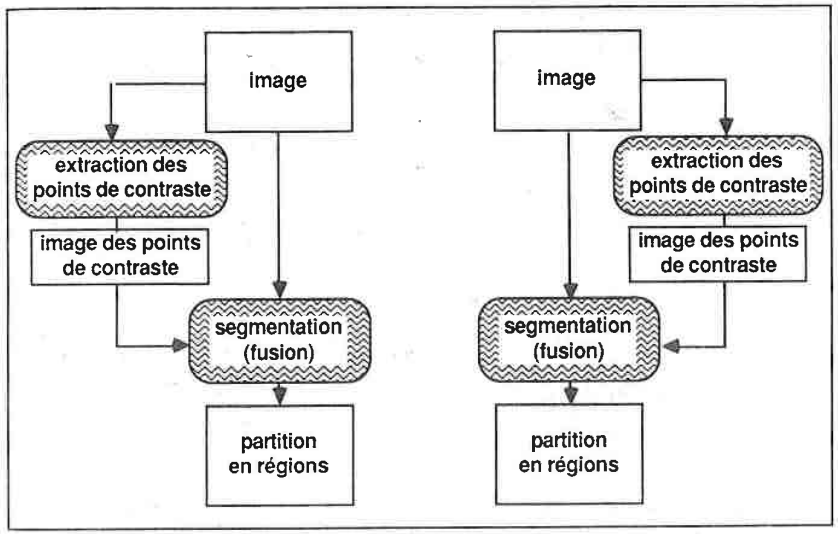

Figure 4.15: *Schéma récapitulatif des traitements pour la segmentation en régions des images stéréoscopiques.*

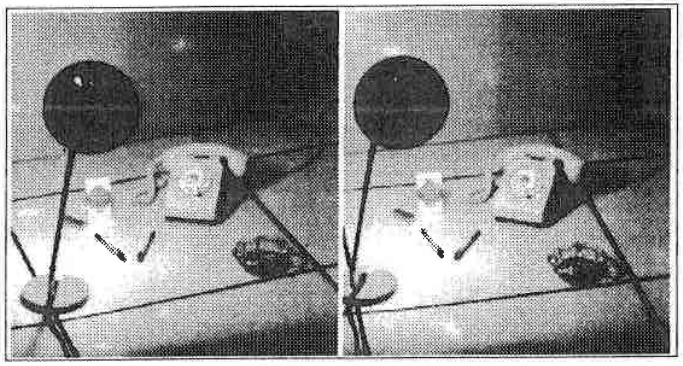

Figure 4.16: *Couple stéréoscopique. Images initiales.*

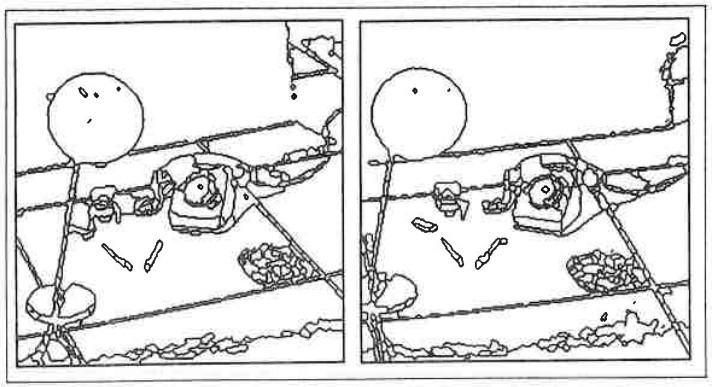

Figure 4.17: *Couple stéréoscopique. Segmentations finales.*

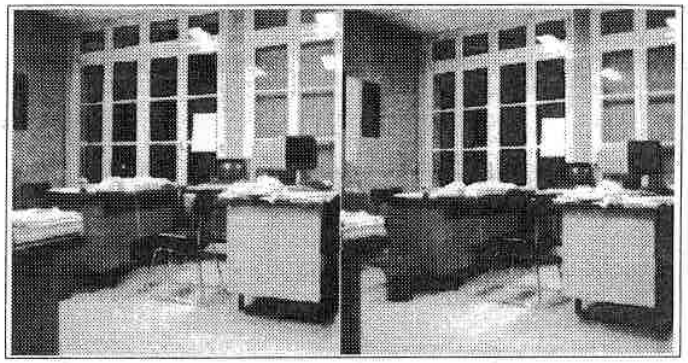

Figure 4.18: *Couple stéréoscopique. Images initiales.*

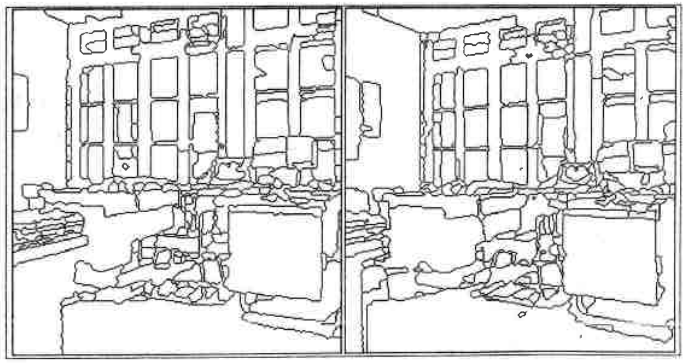

Figure 4.19: *Couple stéréoscopique. Segmentations finales.*

Chapitre 5

MISE EN CORRESPONDANCE DES RÉGIONS

Ce chapitre est consacré à la présentation de l'algorithme d'appariement des régions. La prise en compte des différences de segmentation, fondamentale pour notre problème, a influencé l'approche que nous avons choisie. Notre méthode se partage en deux phases : génération des hypothèses puis validation de ces hypothèses.

La première phase est réalisée par la mise en correspondance des points formant la frontière des régions à l'aide d'une méthode de programmation dynamique.

La seconde phase concerne la validation des hypothèses d'appariement des régions avec remise en cause de la segmentation en régions : certaines fusions sont effectuées entre des régions voisines pour améliorer la qualité de correspondance.

1 Introduction

Rappelons brièvement que le principe de base de la stéréovision est la mise en correspondance de deux images d'une même scène prise sous deux angles différents : c'est-à-dire la formation de couples de primitives homologues, constitués d'une primitive de l'image gauche et d'une primitive de l'image droite. Cette étape d'appariement fait l'objet de ce chapitre.

Résoudre le problème de la mise en correspondance en stéréovision, c'est choisir un sous-ensemble cohérent de couples de primitives homologues parmi l'ensemble de tous les couples potentiels. Cette "cohérence" est définie par le respect des hypothèses d'unicité et de continuité présentées au chapitre 2. De plus, cet ensemble doit optimiser la similitude des attributs de ces régions, attributs de quelque nature que ce soit : photométrique, morphologique ou topographique.

Puisque les partitions à apparier ne sont pas similaires, il faut rester très prudent lorsque deux régions appartenant chacune à une vue ont des attributs suffisamment similaires pour répondre au critère d'appariement : il est nécessaire de s'assurer que ces régions ne sont pas segmentées incorrectement. Ceci signifie qu'il n'est pas raisonnable de vouloir résoudre le problème de l'appariement des régions en comparant seulement leurs attributs : il faut bien évidemment tenir compte de leurs relations, mais également refaire la segmentation de la partie d'image concernée quand celle-ci n'est pas jugée satisfaisante.

Ceci nous conduit à effectuer une phase de "pré-appariement", pour guider ensuite la phase de mise en correspondance des régions. Notre méthode s'apparente aux techniques de "prédiction-vérification" d'hypothèses, et comporte donc les deux phases suivantes :

Phase 1 - Génération des hypothèses : Cette étape effectue la construction de toutes les hypothèses d'appariement de régions en tenant compte de leurs attributs, mais surtout de leurs relations de voisinage. Nous construisons les mises en correspondance des arcs entre les deux graphes d'adjacence et engendrons les hypothèses d'appariement entre les régions liées par ces arcs.

Phase 2 - Validation des hypothèses : Chaque hypothèse d'appariement est évaluée en fonction de la similitude des attributs des régions concernées, et est validée si elle est cohérente[1] avec les autres couples de régions homologues. Des modifications des régions initiales sont effectuées si leur extraction n'est pas jugée suffisamment correcte.

Cette méthode est strictement ascendante : toutes les hypothèses d'appariement doivent être formulées lors de la première phase. Ceci signifie que les différences de segmentation des images à apparier doivent être prises en compte au cours des deux phases, comme nous l'expliquons au paragraphe suivant.

Si nous représentons chacune des partitions sous forme d'un graphe d'adjacence où les nœuds sont les régions et les arcs représentent les relations de voisinage

[1] Respect de la loi d'unicité seulement. Une région représente une surface continue. Les discontinuités des surfaces se présentent donc très certainement à leur frontière, la loi de continuité n'est donc pas utilisée ici.

entre les nœuds, notre méthode consiste en un "pseudo-appariement" des arcs des deux graphes pour engendrer des hypothèses de mise en correspondance des régions concernées, puis en l'examen et en la validation de certaines de ces hypothèses.

La construction des informations tridimensionnelles à partir de ces appariements fait l'objet du chapitre 6.

2 Prise en compte des différences de segmentation

2.1 Position du problème

Le cas le plus général de la différence entre deux partitions à apparier est celui où deux zones images doivent être homologues mais se trouvent inégalement partitionnées, l'une en n régions et l'autre en m régions. La figure 5.1 illustre le cas où une zone A est partitionnée en cinq régions A_1, A_2, A_3, A_4 et A_5 dans la vue gauche et une zone B en trois régions B_1, B_2 et B_3 dans la vue droite.

Notre objectif est d'apparier ces deux zones dans le cas où elles représentent la même partie d'objet, et de modifier en conséquence les partitions initiales en régions.

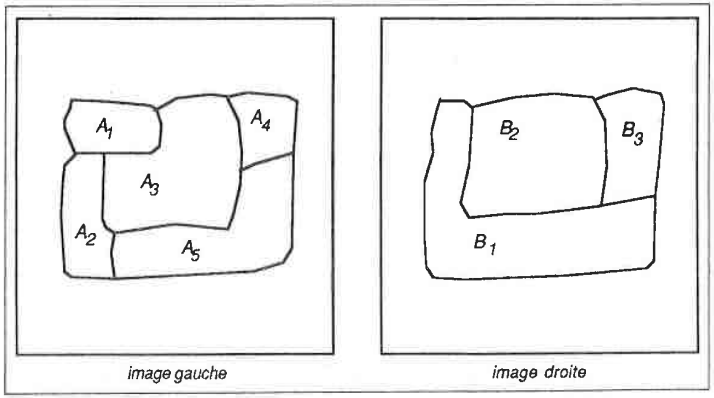

Figure 5.1: *Les deux zones A et B inégalement partitionnées représentent la même partie d'objet. Elles doivent être reconnues homologues.*

2.2 Notre approche

Face à ce problème de différence de segmentation, nous pouvons adopter deux attitudes :

- soit mettre en correspondance l'"intersection" des régions candidates à l'appariement, c'est-à-dire rechercher les sous-parties des régions qui se correspondent

effectivement, comme l'illustre la figure 5.2 ;

- soit apparier l'"union" de ces régions, c'est-à-dire trouver les deux zones maximales qui se correspondent (figure 5.3).

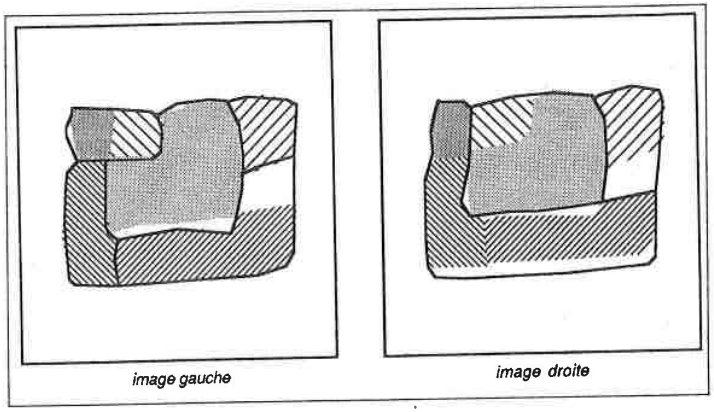

Figure 5.2: *Mettre seulement en correspondance chaque "intersection" des régions, représentée sur cet exemple par les différentes zones hachurées.*

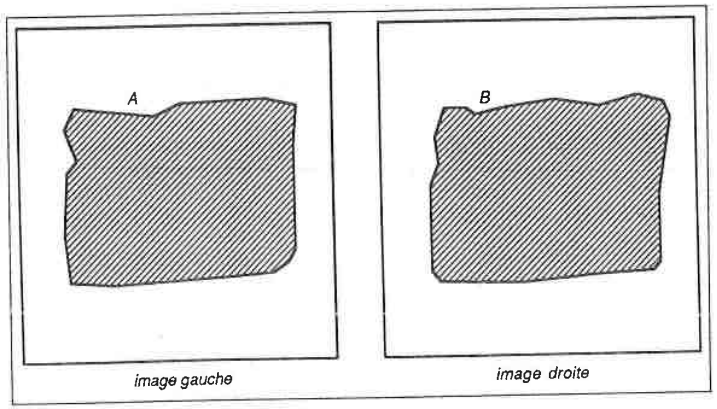

Figure 5.3: *Mettre en correspondance "l'union" des régions, représentée sur cet exemple par les zones hachurées.*

Définition de l'intersection : Nous appelons *intersection* la partie d'une région délimitée partiellement par les "portions" de sa ligne de frontière appariées à des "portions" de la ligne de frontière d'une seule région dans l'autre image.

De l'exemple figure 5.2, nous montrons figure 5.4 deux couples de lignes frontières homologues qui délimitent deux intersections homologues. Un problème crucial apparaît alors : connaissant les portions de frontières qui subsistent, *diviser* les régions initiales pour trouver les intersections de régions homologues.

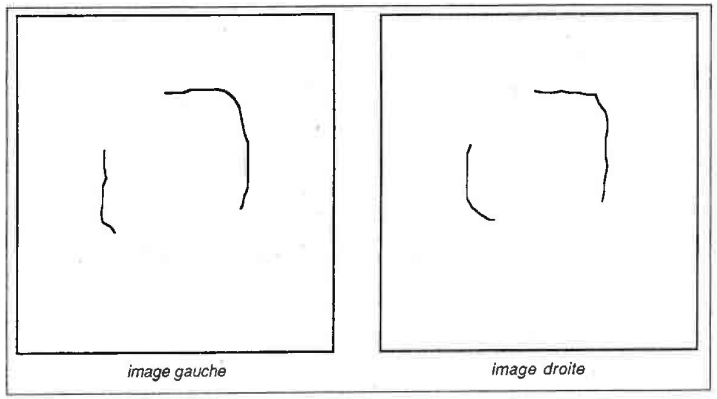

Figure 5.4: *Les lignes de frontières appariées ne permettent pas facilement de déterminer les intersections des régions homologues.*

Définition de l'union : Nous appelons *union* la région délimitée par la plus grande frontière fermée formée de "portions" de lignes appariées. Elle est la région résultant de la fusion de toutes les régions intérieures à cette frontière.

La figure 5.5 représente les lignes de frontières appariées, les lignes extérieures représentent les frontières des régions homologues. La localisation de la frontière externe déterminée, il suffit de fusionner les régions qui lui sont "internes".

Cette approche est donc beaucoup plus simple que l'appariement des intersections, aussi, nous avons choisi de l'implanter.

Il faut toutefois remarquer que l'exemple de différence de segmentation figure 5.1, quoique complet, est relativement sympathique, car seules des fusions peuvent améliorer l'appariement. En réalité, il faut *fusionner* certaines régions et en *diviser* d'autres, en fonction de critères qu'il n'est pas facile de formaliser.

Pour simplifier la résolution d'un tel problème, nous supposons que les partitions sont sur-segmentées. De ce fait, lors de la phase d'appariement, seules des

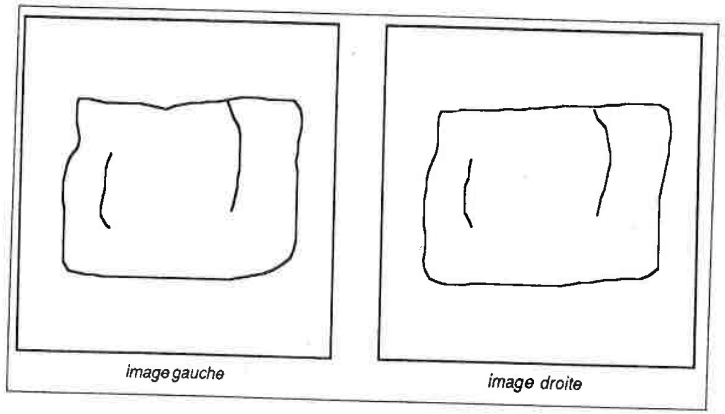

Figure 5.5: *Les lignes de frontières appariées délimitent une région interne, celle-ci est construite par la fusion de toutes les régions internes à la ligne frontière extérieure.*

fusions sont requises pour trouver les unions à mettre en correspondance. Cette hypothèse de "sur-segmentation" est réalisée par l'arrêt précoce de l'algorithme de segmentation.

En d'autres termes, nous avons choisi une méthode d'appariement qui poursuit en fait la croissance des régions pour améliorer la qualité de correspondance entre les deux partitions [Wrobel 87b].

2.3 Implantation

Pour être capable d'apparier correctement deux zones homologues, nous avons besoin d'engendrer, lors de la première phase du processus, toutes les hypothèses d'appariement possibles, soit pour l'exemple illustré figure 5.1 les huit hypothèses suivantes :

$$\mathcal{H} = \begin{cases} (A_1, B_1) & (A_1, B_2) \\ (A_2, B_1) & \\ (A_3, B_2) & (A_3, B_3) \\ (A_4, B_3) & \\ (A_5, B_2) & (A_5, B_3) \end{cases}$$

Cet ensemble d'hypothèses est ensuite examiné lors de la seconde phase du processus : le résultat doit être de corriger les différences de segmentation et d'apparier correctement les deux zones homologues.

La prise en compte des différences de segmentation est donc réalisée au niveau des deux phases du processus d'appariement de la façon suivante :

Phase 1 - Génération des hypothèses : Nous devons construire les hypothèses

d'appariement entre les régions qui représentent la même partie d'objet, c'est-à-dire entre toutes les régions qui ont une intersection.

Phase 2 - Validation des hypothèses : A chaque hypothèse d'appariement, nous cherchons à améliorer la qualité de correspondance des deux régions concernées, en réalisant la fusion de leurs régions voisines. Ceci est effectué en examinant les hypothèses d'appariement engendrées dans le voisinage des deux régions.

3 Phase 1 : génération des hypothèses

Pour illustrer notre démarche, prenons un exemple simple de différence de segmentation entre deux zones à apparier. Cet exemple est représenté figure 5.6.

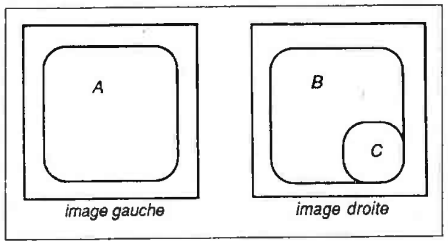

Figure 5.6: *Résultat souhaité : la région A est l'homologue de la région obtenue en fusionnant B avec C.*

La génération des deux hypothèses :

$$\begin{cases} (A, B) \\ (A, C) \end{cases}$$

peut être obtenue en travaillant sur les frontières des régions. L'idée est de trouver que (figure 5.7) :

- les frontières gA_1 et dB sont homologues,
- les frontières gA_2 et dC sont homologues.

Cela revient à trouver les frontières homologues des régions des deux partitions, c'est-à-dire d'apparier les relations de voisinage entre régions[2].

A partir des couples de frontières homologues, nous pouvons engendrer les hypothèses de mise en correspondance des régions délimitées par ces frontières.

[2]Ce qui revient à apparier les arcs des deux graphes images représentant les deux partitions.

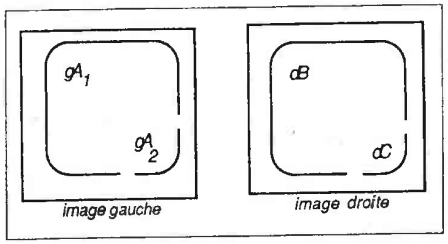

Figure 5.7: *Résultat souhaité : les frontières gA_1 et dB sont homologues et les frontières gA_2 et dC sont également homologues.*

3.1 Mise en correspondance des frontières

Plutôt que d'apparier globalement les frontières des régions, nous avons décidé d'apparier les points formant ces frontières.

La géométrie épipolaire, présentée au chapitre 2, permet de réduire fortement la dimension du problème de mise en correspondance : les primitives images homologues sont situées sur un couple de droite épipolaires associées. Le banc stéréoscopique avec lequel ont été prises les paires d'images est calibré [Faugeras 86b], nous sommes donc en mesure de définir les couples d'épipolaires.

Puisque l'on travaille avec des éléments de la taille du point,

> *"le problème de la mise en correspondance des points de n lignes épipolaires associées"*

se ramène à :

> *"n problèmes de mise en correspondance des points de deux lignes épipolaires associées."*

Chacun de ces problèmes peut être traité comme une *recherche de chemin optimal dans un graphe.*

Cette méthode a été testée dans un contexte différent lors de la mise en correspondance de tous les points images de deux vues artificielles [Mohr 84]. Elle s'était avérée performante, aussi nous l'avons conservée et adaptée à notre cas particulier : appariement des points frontières seulement.

Il est assez classique de transformer les images initiales afin que les lignes épipolaires deviennent les lignes horizontales de la trame de l'image [Benard 83] ; cependant, nous n'effectuons pas cette "rectification".

Nous présentons ci-après la manière de déterminer les couples de droites épipolaires associées, puis la technique utilisée pour apparier les points homologues situés sur un couple de droites épipolaires. Le traitement décrit doit donc être appliqué à tous les couples de droites épipolaires des deux vues.

3.2 Couples de droites épipolaires associées

Déterminer une ligne épipolaire

L'étape de calibration permet, connaissant la position (u, v) d'un point P dans une image, de calculer l'épipolaire Δ dans l'autre image sur laquelle se trouve l'homologue de ce point. Cette droite passe par le centre épipolaire E associé à l'image et a pour vecteur directeur $\vec{A}$.

Ce vecteur $\vec{A}$ s'écrit en fonction de la position du point $P(u, v)$ de la façon suivante :

$$\vec{A} = \alpha \times P$$

$$\vec{A} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \end{pmatrix} \times \begin{pmatrix} u \\ v \end{pmatrix}$$

En pratique :

- Soit $P_g(u_g, u_g)$ un point de l'image gauche. L'épipolaire droite Δ_d associée à ce point passe par le centre épipolaire E_d et a pour vecteur directeur $\vec{A_d}$ défini par :

$$\vec{A_d} = \begin{cases} \alpha_{g11} + \alpha_{g12} \times u_g + \alpha_{g13} \times v_g \\ \alpha_{g21} + \alpha_{g22} \times u_g + \alpha_{g23} \times v_g \end{cases}$$

La figure 5.8 illustre ce calcul géométrique.

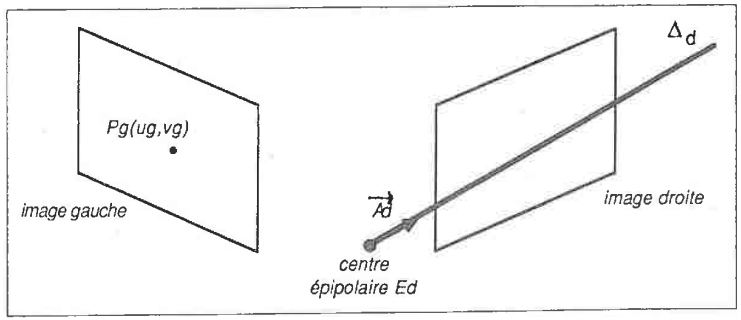

Figure 5.8: *La droite épipolaire Δ_d de vecteur directeur $\vec{A_d}$ est déterminée à partir de la position du point $P_g(u_g, u_g)$ dans l'image gauche et du centre épipolaire droit E_d.*

- Nous avons le schéma symétrique pour l'image droite. Soit $P_d(u_d, v_d)$ un point de l'image droite. L'épipolaire gauche Δ_g associée à ce point passe par le centre épipolaire E_d et a pour vecteur directeur $\vec{A_g}$ défini par :

$$\vec{A_g} = \begin{cases} \alpha_{d11} + \alpha_{d12} \times u_d + \alpha_{d13} \times v_d \\ \alpha_{d21} + \alpha_{d22} \times u_d + \alpha_{d23} \times v_d \end{cases}$$

La position des deux centres épipolaires et les douze coefficients des matrices α_d et α_g sont déterminés lors de l'étape de calibration. Les coordonnées des centres épipolaires droit et gauche, sont déterminés en pixels respectivement dans le repère de l'image droite et gauche.

Déterminer un couple de droites épipolaires associées

Nous savons calculer à partir d'un point P dans une image la droite épipolaire dans l'autre image sur laquelle se situe son homologue. L'intersection du plan épipolaire, passant par le point P et les deux centres épipolaires E_d et E_g, et du plan image auquel appartient P, détermine la droite épipolaire associée.

En pratique :

Soit $P_g(u_g, v_g)$ un point dans l'image gauche. La droite épipolaire Δ_d et l'intersection du plan (P_g, E_g, E_d) avec le plan image gauche, c'est-à-dire la droite Δ_g (P_g, E_g) sur l'image, forment un couple de droites épipolaires associées (figure 5.9).

Les points de la droite Δ_d ont tous leur homologue sur la droite Δ_g, et réciproquement.

Remarquons que pour effectuer ces calculs, nous avons besoin de connaître les coordonnées des deux centres épipolaires E_g et E_d et les six coefficients d'une seule matrice α.

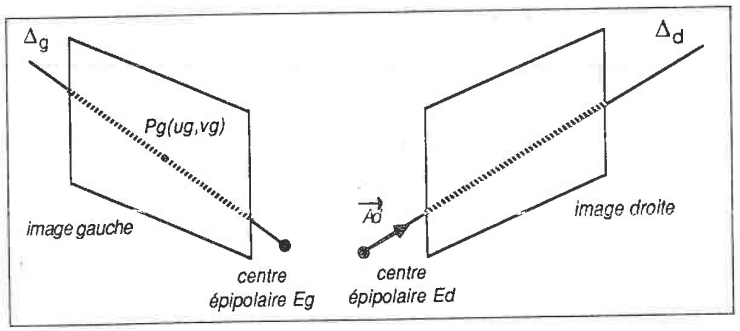

Figure 5.9: *Les deux droites Δ_g et Δ_d sont deux droites épipolaires associées.*

3.3 Recherche de chemin optimal dans un graphe

Construction du graphe :

Tout graphe est formé de nœuds et d'arcs. Nous avons choisi de définir un nœud comme un appariement potentiel entre les points des images. Les points du graphe représentent alors tous les couples de correspondance possibles. Un tel graphe est représenté figure 5.10 où les points de la ligne épipolaire gauche sont en abscisse, ceux de la ligne épipolaire droite en ordonnée. La recherche des couples de correspondance s'effectue dans ce graphe, la combinatoire du problème peut être fortement diminuée si l'on exclut de ce graphe les nœuds engendrant des points 3d trop proches ou trop éloignés des caméras.

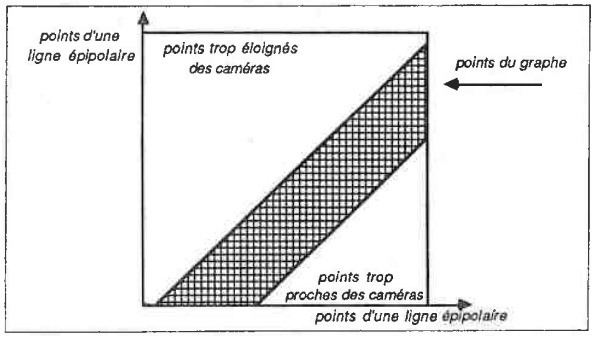

Figure 5.10: *Graphe de recherche des points homologues.*

Les arcs du graphe sont définis au paragraphe suivant.

La figure 5.12 représente le chemin optimal "artificiel" obtenu en reliant les nœuds sélectionnés. Ces nœuds indiquent les points appariés entre les deux lignes épipolaires de la scène artificielle visualisée figure 5.11. La ligne épipolaire de l'image gauche est située sur l'axe horizontal, la ligne épipolaire de l'image droite est sitée sur l'axe vertical. Notons que cette correspondance laisse des points sans équivalent : il s'agit des points vus par une caméra seulement.

Effectuer un bon appariement, c'est choisir un sous-ensemble de couples de points homologues parmi l'ensemble de tous les couples possibles. Un chemin idéal reliant toutes les bonnes correspondances comme celui présenté figure 5.12 possède deux caractéristiques :

- si l'on somme, le long du chemin, la fonction qualifiant la similarité des points mis en correspondance, alors ce chemin a un coût maximal,
- le chemin est formé d'une majorité de *segments parallèles à la diagonale* joints par des *segments parallèles aux axes.*

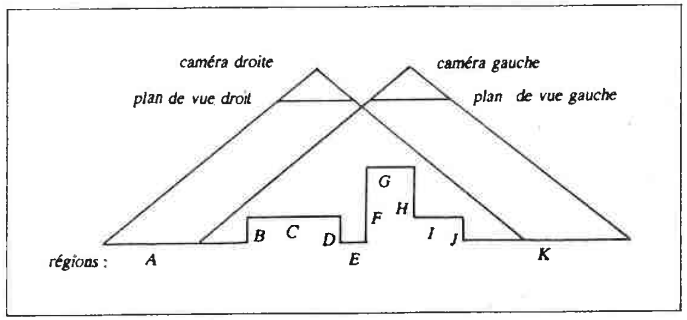

Figure 5.11: *Coupe d'une scène artificielle observée par deux caméras.*

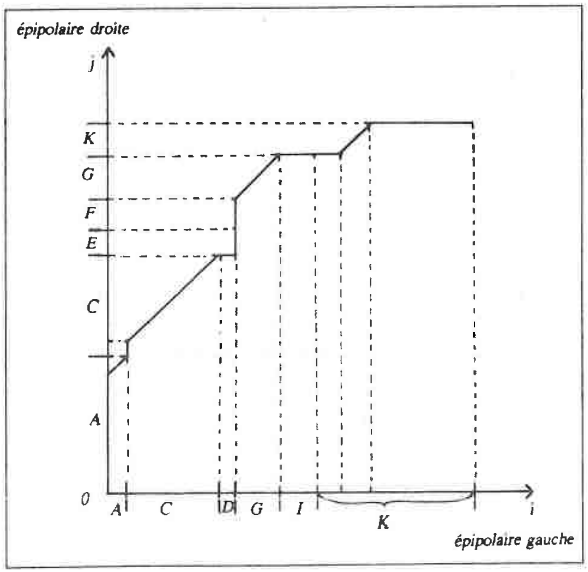

Figure 5.12: *Chemin optimal reliant les couples de correspondance sélectionnés dans le graphe.*

Déterminer un tel chemin de coût optimal revient alors à faire un compromis entre *maximiser le nombre d'appariements* et optimiser la qualité de correspondance des points appariés.

Nous disons alors réciproquement :

> *Si l'on sait construire un chemin du coin inférieur gauche au coin supérieur droit du graphe mettant en correspondance des points d'attributs similaires, alors on effectue une bonne correspondance.*

La recherche d'un chemin optimal dans un graphe conjugue :

- la recherche d'un score global de mise en correspondance, ce score est fonction de la qualité de similitude des attributs des points examinés ;
- le respect de l'hypothèse de continuité dans les mises en correspondance ;
- le respect de l'hypothèse d'unicité ;
- le respect de l'hypothèse d'ordre. Cette propriété est vérifiée à 95 % par les objets des scènes que nous traitons. Ceci représente cependant une limitation, donc une faiblesse, de notre approche.

Le graphe étant sans circuit, le calcul du chemin optimal se fait par propagation depuis une extrémité : partant de l'origine, on cherche en chaque point le chemin optimal y menant et son coût.

Soit I_x un point de l'axe horizontal (figure 5.13).

Soit I_{x-1} le point prédécesseur de I_x.

Si le calcul du chemin le plus "court" de l'origine à I_{x-1} est effectué pour tous les couples (I_{x-1}, j) où j varie de 0 à n, alors le calcul du chemin le plus court de l'origine à I_x se fait en une seule itération.

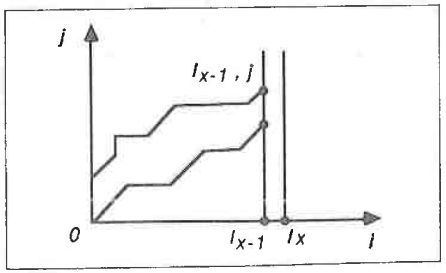

Figure 5.13: *Le calcul du chemin optimal arrivant à I_x se fait en une seule itération.*

Ceci s'écrit sous la forme algorithmique suivante :

- *pour chaque point i de l'épipolaire gauche faire*
  - *pour chaque point j de l'épipolaire droite compris dans l'intervalle de disparité autorisé (pour supprimer les points physiques trop proches ou trop éloignés des caméras) faire*
    - * *Remarque : à ce point on connaît le coût des meilleurs chemins aboutissant aux prédécesseurs de (i,j)*
    - * *prolonger les chemins aboutissant à (i,j) et retenir l'arc du meilleur chemin*
    - * *retenir l'arc sélectionné et donner à (i,j) le coût du chemin*
- *rechercher le meilleur coût aboutissant au "coin supérieur droit" du graphe et reconstruire le chemin en reprenant les arcs sélectionnés*

Valuation des arcs du graphe :

Nous avons besoin de seulement trois types d'arcs élémentaires représentés figure 5.14.

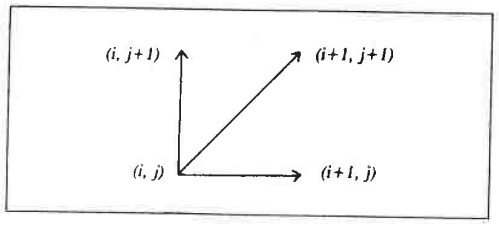

Figure 5.14: *Les trois types d'arcs élémentaires joignant les points du graphe.*

Intuitivement, on favorise les arcs diagonaux qui nous fournissent des couples de correspondance, et on pénalise les arcs horizontaux et verticaux représentant les quelques discontinuités des surfaces de la scène ; ces arcs ne font pas intervenir de correspondance et ne sont donc empruntés que lorsqu'aucun couplage (i,j) n'est possible.

Si la correspondance (i,j) est acceptée, alors :

- les correspondances $(i+1,j)$ et $(i,j+1)$ sont interdites du fait de la contrainte d'unicité ;
- la correspondance $(i+1,j+1)$ est possible.

Les arcs sont donc valués de la manière suivante :

arcs verticaux et horizontaux : Pénalité.

arcs diagonaux : Calcul de la qualité de correspondance Q entre les deux points considérés ; cette mesure de qualité est locale et dépend des attributs de chaque point frontière considéré. Cette qualité est fonction de la similitude des attributs suivants :

- attributs de la ligne de contour à laquelle il appartient : longueur et direction de cette ligne ;
- attributs des régions dont il est frontière : moyenne des niveaux de gris, ...

Programmation dynamique :

Le moyen le plus classique de déterminer de tels chemins est la programmation dynamique. Cette technique est largement utilisée dans le domaine de l'image quand il s'agit de mettre en correspondance des chaînes comprenant des dissemblances, par exemple :

- en stéréovision : M. Benard [Benard 83] utilise l'algorithme de Viterbi [Forney 73] que l'on peut considérer comme une variation de la programmation dynamique ;
- pour la détection du mouvement : [Guilloux 84] .

De nombreux problèmes de reconnaissance de la parole sont également traités par la programmation dynamique, technique introduite par T. Vintsyuk [Vintsyuk 68]. Cette technique aide à la correction des distorsions essentiellement non linéaires dues aux variations du rythme d'élocution de la parole. Le travail d'A. Boyer, dans notre laboratoire, se propose de reconnaître des mots enchaînés et isolés à l'aide de la programmation dynamique [Boyer 87].

Signalons aussi que la programmation dynamique est une technique classique de la recherche opérationnelle dont le principe de base a été formalisé par J. Bellman [Bellman 57], dans un cadre beaucoup plus large :

> *"Dans une séquence optimale de décisions, quelle que soit la première décision prise, les décisions subséquentes forment une sous-séquence optimale, compte tenu des résultats de la première décision"*

Ce que nous traduisons, dans notre cas par : *"Tout sous-chemin est lui-même optimal".*

3.4 Prise en compte des lignes environnantes

La méthode précédente (recherche de chemin optimal dans un graphe) ne tient compte que des informations sur une ligne, il semble alors naturel de prendre en compte les informations des lignes environnantes. En effet, les surfaces détectées sont continues presque partout et les disparités dans le voisinage d'un point sont presque toujours proches de la disparité du point considéré.

Ceci peut être mis en œuvre par une technique s'inspirant des méthodes générales de relaxation. L'algorithme précédent fournit les valeurs initiales des disparités. Puis, pour chaque ligne, on calcule la mesure de qualité Q' pour la correspondance (i,j) de la façon suivante :

$$Q'(i,j) = Q(i,j) + C(i,j)$$

avec $Q(i,j)$: mesure de la qualité de la correspondance (i,j) définie précédemment,
$C(i,j)$: compatibilité de la correspondance (i,j) avec les correspondances des points des lignes "supérieures" et "inférieures" dans les images.

Cette correspondance favorise les disparités comprises entre celles des lignes voisines (figure 5.15). De plus, elle pénalise de façon quadratique les autres disparités : les petits écarts sont faiblement pondérés, alors que la pondération des grands écarts est très forte.

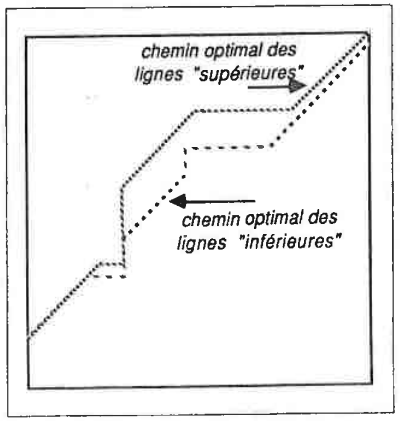

Figure 5.15: *On favorise la construction du chemin de la mise en correspondance des deux lignes épipolaires courantes dans l'intervalle de disparité délimité par les deux chemins trouvés lors de l'appariement des lignes "supérieures" et "inférieures".*

Il suffit alors d'appliquer le même algorithme avec Q' remplaçant Q et d'itérer.

Cette méthode a été implantée avec succès pour une autre application de stéréovision [Mohr 84], et expérimentalement 3 ou 4 itérations sont suffisantes pour obtenir des résultats satisfaisants.

En fait, le but de la mise en correspondance des points frontières des régions est d'engendrer des hypothèses d'appariement de régions, et non de calculer immédiatement des coordonnées 3d. Ceci ne justifie guère la mise en place de cette méthode quelque peu coûteuse en temps machine pour éliminer les irrégularités obtenues.

Nous reviendrons cependant sur cette décision pour améliorer les résultats, si cela s'avère nécessaire.

La même préoccupation a été traitée par Y. Ohta et T. Kanade [Ohta 83] : la mise en correspondance des points appartenant à une ligne de contraste se fait également par programmation dynamique sur les couples de lignes épipolaires. Puisque cette recherche, appelée *intra-scanline*, ne prend pas en compte la dépendance mutuelle entre les différentes lignes de l'image, il est nécessaire d'effectuer une recherche, appelée *inter-scanline*, qui établit une consistence entre les lignes. L'approche suivie consiste à rechercher l'ensemble optimal de chemins dans un espace à trois dimensions composé de l'empilement des graphes précédemment définis (figure 5.16). Les côtés de cet espace représentent les images stéréoscopiques.

Le coût d'un ensemble de chemins est défini par la somme des coûts de chaque chemin qui le compose. La recherche d'un ensemble optimal de chemins est effectuée au fur et à mesure de la croissance parallèle des chemins optimaux dans chaque graphe.

3.5 Génération des hypothèses

Pour chaque couple de lignes épipolaires associées, nous obtenons un ensemble cohérent de couples de points frontières appariés, plus ou moins correct selon la qualité des partitions en régions des images.

Nous présentons figure 5.17 un couple d'images de frontières à apparier, avec visualisation de deux couples de lignes épipolaires associées. Les chemins optimaux trouvés sont représentés figures 5.18 et 5.19.

Ces deux chemins sont à composantes essentiellement diagonales. On peut cependant constater quelques écarts de disparité que nous expliquons ci-après :

- L'écart de disparité A du chemin figure 5.18 correspond à la différence de segmentation des montants des fenêtres à l'extrême gauche.

  L'écart B est dû à la sur-segmentation du carreau dans l'image gauche : la segmentation de ce carreau a été perturbée par l'existence de plusieurs reflets que l'on peut constater dans l'image originale figure 5.17.

  Le troisième écart C provient de la difficulté de segmenter correctement l'affiche blanche posée sur la fenêtre, et ce, dans les deux vues.

  L'écart D correspond à la présence d'une région supplémentaire dans le carreau le plus à droite de la vue droite. Celle-ci est due à un reflet plus perceptible dans la vue de droite.

- L'écart E du second chemin optimal (figure 5.19) est dû à une problème de segmentation et à la création d'une région supplémentaire dans l'image de droite entre le bureau (à gauche dans la scène) et le sol.

  La différence de segmentation de l'ombre du bureau de droite dans la scène provoque l'écart de disparité F dans le chemin.

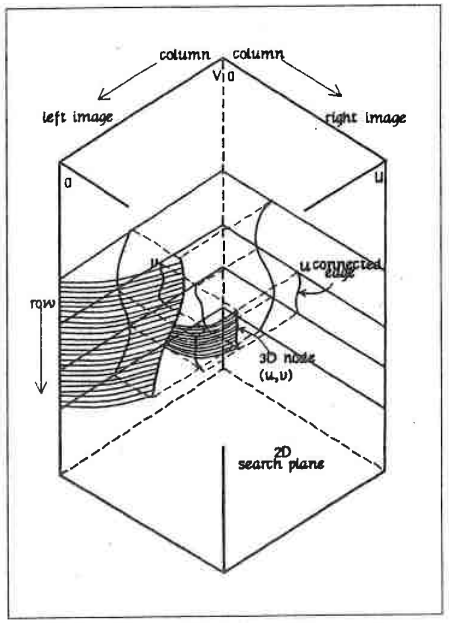

Figure 5.16: *L'espace de recherche 3d pour la méthode de programmation dynamique se fait en tenant compte des lignes environnantes [Ohta 83]. Les lignes de contraste forment des ensembles de nœuds 3d. La sélection d'un ensemble optimal de chemins est effectuée à chaque nœud 3d.*

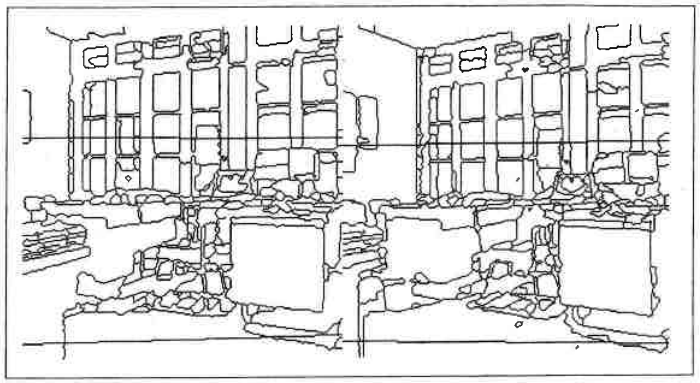

Figure 5.17: *Couple stéréoscopique d'images des frontières de régions. Visualisation de deux couples de lignes épipolaires associées.*

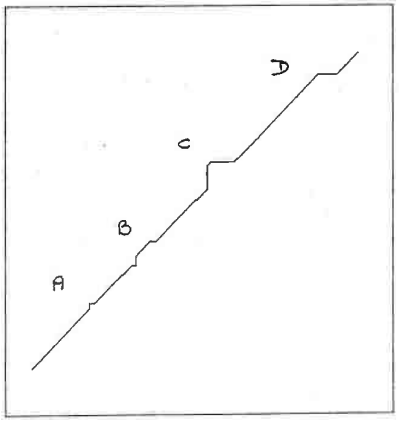

Figure 5.18: *Chemin optimal résultat de la mise en correspondance des points frontières des deux lignes épipolaires "hautes" associées. La ligne épipolaire de la vue gauche est en abscisse, la ligne épipolaire de la vue droite est en ordonnée.*

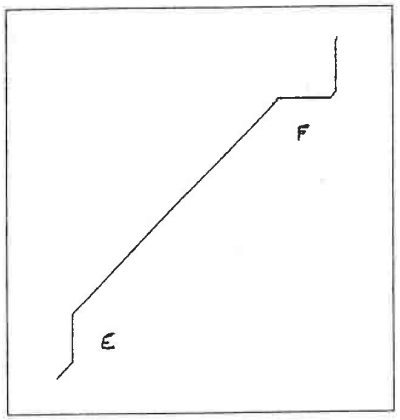

Figure 5.19: *Chemin optimal résultat de la mise en correspondance des points frontières des deux lignes épipolaires "basses" associées. La ligne épipolaire de la vue gauche est en abscisse, la ligne épipolaire de la vue droite est en ordonnée.*

Remarquons qu'en présence d'une répétition de motifs (comme les bords des carreaux des fenêtres), la technique de recherche d'un chemin optimal dans un graphe tend à maximiser le nombre d'appariements effectués. Ceci permet donc de lever l'ambiguïté de mise en correspondance de certains points ayant plusieurs homologues potentiels (de caractéristiques similaires) dans l'autre vue.

Chaque couple de points homologues engendre des hypothèses quant à la mise en correspondance des régions que ces points délimitent. Par exemple, soit (P_g, P_d) un couple de points homologues, alors d'après le cas illustré figure 5.20 nous engendrons les trois hypothèses d'appariement suivantes :

$$\begin{cases} (A_g, A_d) \\ (B_g, B_d) \\ (C_g, C_d) \end{cases}$$

Cette phase d'appariement de points frontières a donc pour résultat un ensemble d'hypothèses concernant la mise en correspondance des régions. Chaque hypothèse est valuée par un <u>score</u> représentant le nombre de fois que cette hypothèse a été engendrée par un couple de points frontières homologues.

Les hypothèses correctes possèdent le score le plus élevé. En effet, prenons l'exemple des points formant la frontière des carreaux des fenêtres. La majeure partie de ces points est correctement appariée et engendre donc des hypothèses correctes. Ces hypothèses ont un score élevé.

En contrepartie, il existe des mauvais appariements entre points frontières. Les points qui posent des problèmes sont en général situés sur les "intersections de

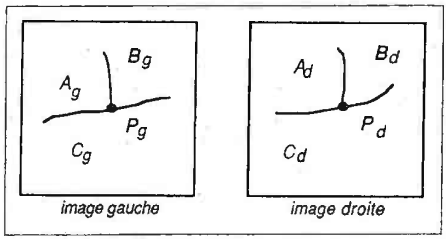

Figure 5.20: *Le couple (P_g, P_d) de points homologues engendre les trois hypothèses : $(A_g, A_d), (B_g, B_d)$ et (C_g, C_d).*

bords" des régions, et apparaissent donc en nombre limité. Les fausses hypothèses engendrées par ces erreurs ont donc des scores beaucoup plus faibles ; elles seront ignorées, la plupart du temps, lors de la phase d'appariement de régions.

Nous présentons figure 5.21 les points frontières des régions pour lesquels un homologue dans l'autre vue a été trouvé et qui fournissent donc des hypothèses d'appariement de régions.

Figure 5.21: *Images stéréoscopiques d'une scène de bureau. Points frontières des régions dont on a trouvé le point homologue dans l'autre vue.*

4 Phase 2 : Validation des hypothèses

4.1 Structure générale de l'algorithme

Nous avons à notre disposition un ensemble d'hypothèses $\mathcal{H}$ d'appariement de régions. A chacune de ces hypothèses est associé un score représentant le cumul des "votes", c'est-à-dire le nombre de fois où chaque hypothèse a été engendrée par les couples de points de contraste créés.

Notre idée est d'examiner ces hypothèses et de les valider au fur et à mesure si certaines conditions sont respectées. La structure de l'algorithme peut être résumée comme suit :

- *tant que l'ensemble des hypothèses $\mathcal{H}$ est non vide faire*
  - *choisir une hypothèse h dans l'ensemble $\mathcal{H}$ des hypothèses à traiter*
  - *retirer cette hypothèse de l'ensemble $\mathcal{H}$:*

$$\mathcal{H} \leftarrow \mathcal{H} - \{h\}$$

  - *calculer la qualité de correspondance entre les deux régions proposées à l'appariement*
  - *valider ou non cette hypothèse, selon la valeur de la qualité*

Dans notre contexte, chaque hypothèse validée sert de "point d'ancrage" pour guider l'appariement des régions non encore traitées. De ce fait, aucun retour arrière n'est autorisé sur les hypothèses validées ; les couples de régions appariées ne sont pas remis en cause. L'ordre d'examen des hypothèses a donc une importance fondamentale.

Notre approche s'apparente à celle utilisée par les algorithmes de mise en correspondance qualifiés de "multi-niveaux" [Hannah 84] [Grimson 81] où le principe est de travailler par affinements successifs, c'est-à-dire de réaliser des correspondances sur des images rendues intentionnellement floues ou lissées pour ne pas tenir compte des détails, puis d'utiliser ces résultats pour guider le traitement sur des images moins floues ou plus grandes.

Nous présentons ci-après les deux principales étapes de notre algorithme :

- choix d'une hypothèse dans l'ensemble $\mathcal{H}$,
- évaluation de la qualité de correspondance de deux régions proposées à l'appariement.

4.2 Stratégie d'examen des hypothèses

Il est nécessaire que les premières hypothèses traitées et validées soient les plus sûres, car elles servent de points d'ancrage pour la suite du traitement.

Il apparaît très naturel de sélectionner en premier lieu les régions les plus "rares", c'est-à-dire ayant des caractéristiques bien particulières, différentes de celles de toutes les autres régions. Par exemple, s'il n'existe que deux régions de couleur rouge, une dans chaque image, il est justifié de commencer par examiner l'hypothèse d'appariement les concernant, si elle existe bien sûr.

Toutes les caractéristiques de régions peuvent être utilisées : photométriques, géométriques ou topographiques pour déterminer les régions "rares". Un tel ordre sur les hypothèses n'est cependant pas facile à réaliser.

Néanmoins, n'oublions pas que chaque hypothèse d'appariement possède un score. Les hypothèses formulées le plus grand nombre de fois correspondent donc aux régions de plus longue frontière, soit généralement aux régions de plus grande surface. Or il faut remarquer que ces "grosses" régions possèdent le plus souvent des caractéristiques assez différentes, alors que les plus petites régions ont fréquemment des attributs similaires[3] et posent des problèmes d'ambiguïté lors de leur mise en correspondance.

Notre méthode est alors d'ordonner les hypothèses d'appariement en fonction de leur score. Ceci revient, de par la façon dont les hypothèses sont construites, à les classer en fonction de la "rareté" des caractéristiques des régions concernées ; les caractéristiques prises en compte qualifiant seulement la taille des régions.

L'ensemble $\mathcal{H}$ de toutes les hypothèses est stocké de manière triée dans une structure de données organisée sous forme d'un arbre binaire de recherche [Aho 87] ; elle présente l'avantage de donner un accès rapide au plus petit ou au plus grand élément stocké.

Choisir une hypothèse d'appariement dans l'ensemble $\mathcal{H}$ revient alors à extraire l'hypothèse de score le plus élevé de la structure de données. La complexité de ce traitement est faible, elle vaut :

$$\mathcal{O}(\mathcal{H}) \times log(\mathcal{C}ard(\mathcal{H})) \textit{ pour chaque hypothèse}$$

et diminue au fur et à mesure du déroulement de l'algorithme puisque l'ensemble $\mathcal{H}$ se vide de ses éléments.

4.3 Evaluation de la qualité de correspondance

Calcul de la qualité :

Typiquement, la qualité de correspondance entre deux régions est locale et fait uniquement intervenir les attributs des régions concernées. Elle est calculée de la façon suivante :

$$Q(R_g, R_d) = \begin{cases} 0 & \text{si l'une des deux régions } R_g \text{ ou } R_d \text{ est déjà} \\ & \text{appariée à une autre région} \\ & \text{(respect de la loi d'unicité)} \\ \sum_{i=1}^{n} C_i \times A_i(R_g, R_d) & \end{cases}$$

[3] Cette remarque peut être illustrée par la paire stéréoscopique de la scène de bureau figure 4.19 où les petites régions représentant les carreaux ont des caractéristiques similaires, alors que les grandes régions : mur, bureau, sol, ... ont des caractéristiques photométriques et géométriques très différentes.

avec $Q(R_g, Rd)$: fonction croissante de la qualité de correspondance entre la région R_g de l'image gauche et la région R_d de l'image droite ;
R_g et R_d : régions proposées à l'appariement ;
$A_i(R_g, R_d)$: fonction qualifiant la similitude de l'attribut numéro i des régions R_g et R_d ;
C_i : coefficient de pondération de la similitude de l'attribut numéro i des deux régions.

Pour notre problème, l'attribut "idéal" d'une région est celui qui caractérise sa géométrie et qui est invariant par projection perspective. Pour l'approcher, nous utilisons la surface et le coefficient de compacité[4] ; mais d'autres attributs comme le centre de gravité ou les moments d'inertie peuvent être choisis. Les attributs que nous avons sélectionnés ont l'avantage de se calculer rapidement, mais il serait certainement plus correct de caractériser également la forme de la frontière des régions : ceci peut être réalisé en déterminant les principales directions des segments de droite approchant la frontière.

Les attributs "classiques" de type photométriques ne sont pas stables du fait de la différence de prise de vue et d'éclairage entre les images ; dans certains cas les différences entre de tels attributs pour deux régions homologues peuvent être relativement élevées. Les attributs photométriques interviennent cependant dans le calcul de la qualité de correspondance, mais leur coefficient de pondération C est faible, alors que ce coefficient est plus élevé pour pondérer la similitude des attributs morphologiques.

En pratique :

L'attribut "principales directions des frontières" n'est pas implanté actuellement pour des raisons de coût : les régions que nous traitons sont également sujettes à des fusions avec leur(s) voisine(s) pour améliorer la qualité de correspondance ; or cet attribut ne se calcule pas à partir des régions à fusionner mais nécessite un retour sur les données initiales. L'utilisation de cet attribut fait cependant partie des améliorations futures de notre méthode.

Le choix des attributs utilisés pour calculer la qualité de correspondance entre les deux régions proposées à l'appariement n'est pas figé, il se résume à la liste suivante :

- attributs de type morphologique :
  - surface de la région (nombre de points) ;
  - coefficient de compacité calculé par $\frac{surface}{périmètre^2}$ où le périmètre est la longueur de la frontière ;
  - longueur et largeur du plus petit rectangle circonscrit à la région.
- attributs de type photométrique :

[4] Le coefficient de compacité est calculé par $\frac{surface}{périmètre^2}$, où le périmètre est la longueur de la frontière.

  - moyenne des niveaux de gris ;
  - plage de variation des niveaux de gris (niveau de gris maximal et minimal).

D'autres attributs de type photométriques peuvent être utilisés comme la variance ou l'histogramme de répartition des niveaux de gris, mais ceux nommés ci-dessus s'avèrent expérimentalement suffisants. L'attention doit être portée sur les attributs morphologiques et non sur les attributs photométriques, aussi les caractéristiques photométriques les plus simples sont considérées comme étant les meilleures.

Nous n'utilisons pas d'attributs topographiques, mais rappelons que les contraintes de voisinage sont prises en compte lors de la phase de génération d'hypothèses. Nous pouvons donc nous autoriser à les "oublier" lors de la phase d'appariement.

En revanche, nous incluons dans le calcul de la qualité de correspondance la "similitude" des bandes épipolaires dans lesquelles se situent les régions. Cela revient à vérifier leur cohérence géométrique.

Amélioration de la qualité :

Pour chaque hypothèse d'appariement, nous calculons la qualité de correspondance entre les deux régions concernées. Il faut cependant se rappeler que ces régions sont peut-être mal segmentées, malgré une assez bonne qualité de correspondance.

L'objectif est d'obtenir un ensemble de couples de la meilleure qualité de correspondance possible. Ceci implique, dans notre cas, d'essayer systématiquement d'augmenter la qualité de chacun des couples de régions, en cherchant à fusionner les deux régions concernées avec une ou plusieurs de leurs voisines. Ces régions voisines ne doivent pas, bien évidemment, être déjà appariées.

La structure de l'algorithme devient alors la suivante :

- *tantque l'ensemble des hypothèses $\mathcal{H}$ est non vide faire*
  - *choisir l'hypothèse h dans l'ensemble $\mathcal{H}$ de score le plus élevé :*

$$h \leftarrow \{h \in \mathcal{H}, score(h) = Max_{\mathcal{H}}(score(h))\}$$

  - *retirer cette hypothèse de l'ensemble $\mathcal{H}$:*

$$\mathcal{H} \leftarrow \mathcal{H} - \{h\}$$

  - *calculer la qualité de correspondance Q entre les deux régions R_g et R_d proposées à l'appariement :*

$$Qual \leftarrow Q(R_g, R_d)$$

  - *tantqu'il existe une région V voisine de R_g (ou R_d), une hypothèse d'appariement (V, R_d) (ou (V, R_g)) dans $\mathcal{H}$, et que la fusion de V avec R_g (ou R_d) améliore la qualité de correspondance $Qual$ faire*

* *sélectionner la région V voisine de R_g (ou R_d) telle que la fusion avec R_g (ou R_d) soit possible et telle que cette fusion produise la meilleure qualité de correspondance $Qual'$, c'est-à-dire :*

$$Qual' \leftarrow Max \begin{pmatrix} Max_{[V_i \in Voisine(R_g)] \wedge [P(V_i \cup R_g)]} Q(R_g \cup V_i, R_d), \\ Max_{[V_i \in Voisine(R_d)] \wedge [P(R_d \cup V_i)]} Q(R_g, R_d \cup V_i) \end{pmatrix}$$

avec $Voisine(R_g)$ et $Voisine(R_d)$: les ensembles de toutes les régions V_i voisines de R_g et R_d ;
$P(V \cup R)$: prédicat d'homogénéité des régions V et R, prend la valeur vrai si la fusion de V avec R est autorisée.

* *effectuer la fusion $(R_g \cup V)$ ou $(R_d \cup V)$ selon l'appartenance de V à $Voisine(R_g)$ ou à $Voisine(R_d)$*
* *mettre à jour la qualité $Qual$:*

$$Qual \leftarrow Qual'$$

– *effectuer ou non l'appariement des deux régions éventuellement modifiées par une fusion*

Le prédicat d'homogénéité P des régions utilisé dans cet algorithme porte sur beaucoup d'attributs de régions, à la différence des prédicats calculés lors de la segmentation des images en régions. Ce prédicat permet de sélectionner l'ensemble des régions voisines qui ont des caractéristiques assez similaires à une région concernée. C'est le processus d'appariement qui choisit dans cet ensemble le couple de régions adjacentes à fusionner en fonction de l'amélioration apportée à la qualité de correspondance.

Les attributs utilisés pour autoriser ou non la fusion de deux régions adjacentes dans une image sont essentiellement photométriques :

- moyenne des niveaux de gris,
- plage de la variation des niveaux de gris (niveau de gris maximal et minimal),
- proportion de points de contraste situés sur la frontière commune des deux régions à fusionner.

5 Résultats

Nous présentons dans cette partie les données et les résultats de la phase d'appariement sur un seul couple d'images stéréoscopiques d'une scène d'intérieur. D'autres résultats, complétés par les informations tridimensionnelles calculées, sont montrés à la fin de ce document.

La scène originale est présentée figure 5.22 et les partitions initiales à apparier figure 5.23. Celles-ci sont issues du processus de segmentation. Nous illustrons ensuite les effets de la remise en cause de la segmentation et les fusions effectuées sur quelques zones particulières dans les images : la figure 5.24 présente quelques fusions

réalisées par le processus afin de montrer son action dynamique sur les partitions initiales en régions. Les figures 5.25 et 5.26 montrent respectivement les régions homologues appariées et les partitions finales des images en régions.

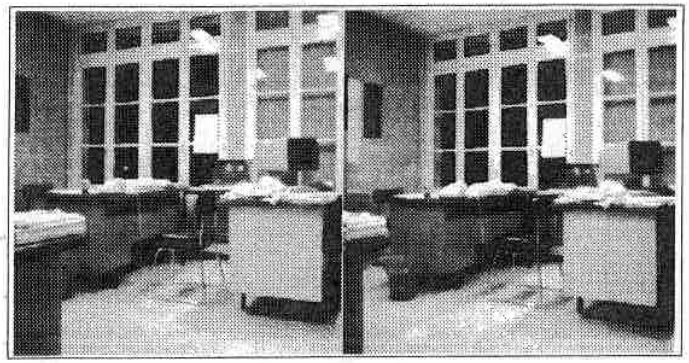

Figure 5.22: *Couple stéréoscopique. Images originales.*

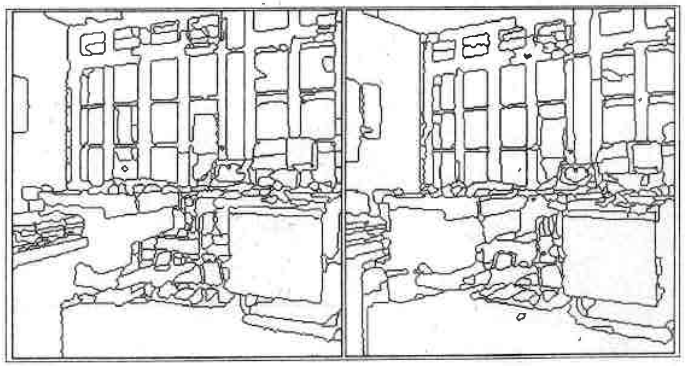

Figure 5.23: *Couple stéréoscopique. Partitions en régions à apparier. (247 régions dans l'image gauche et 254 régions dans l'image droite).*

Les différents paramètres utilisés pour calculer la qualité de correspondance ou de fusion ont été déterminés empiriquement et sont identiques pour toutes les images traitées. Ils sont seulement modulés en fonction des caractéristiques des régions. Par exemple, les régions très découpées sont généralement segmentées de façon totalement différente dans les deux images (voir les montants des fenêtres figure 5.23), et demandent à être traitées avec des paramètres différents.

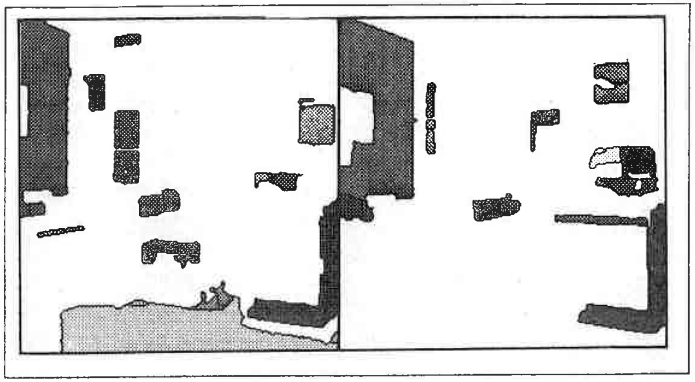

Figure 5.24: *Visualisation de quelques régions adjacentes fusionnées par le processus d'appariement pour améliorer la qualité de correspondance des régions.*

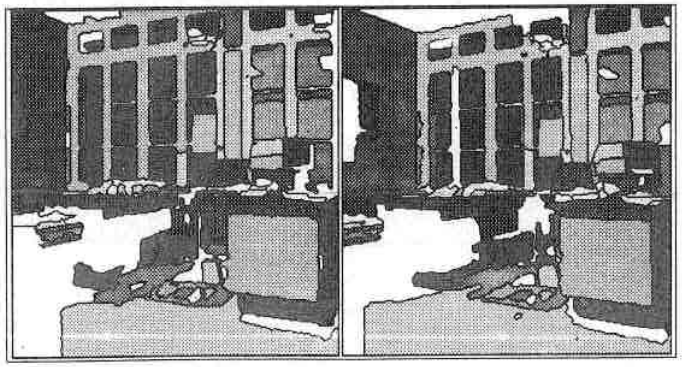

Figure 5.25: *Mise en correspondance des régions (113 régions appariées). Les régions appariées sont visualisées avec le même niveau de gris, les régions laissées en blanc n'ont pas d'homologue.*

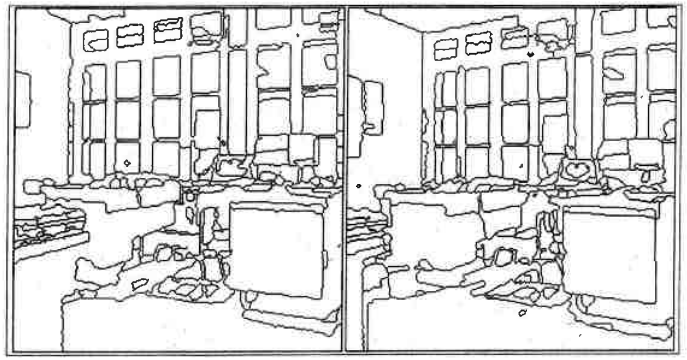

Figure 5.26: *Partitions finales. Le processus d'appariement a réalisé 117 fusions (image gauche : 190 régions, image droite : 194 régions).*

Sur les résultats présentés, nous constatons un bon appariement des principales surfaces des objets de la scène : mur, bureau, sol, fenêtre. Les mises en correspondance des régions de caractéristiques similaires (carreaux des fenêtres) sont réalisées par une méthode de programmation dynamique ; ceci permet de lever l'ambiguïté d'appariement de telles régions.

D'autre part, les fusions effectuées améliorent effectivement les partitions en régions. La sur-segmentation du carreau dans l'image de droite, qui avait provoqué l'écart de disparité B lors de l'étude des chemins optimaux figure 5.18, a été modifiée.

Cependant, il faut mentionner qu'il subsiste des erreurs d'appariement, essentiellement dans les zones des images où la segmentation est très délicate (voir la zone centrale de la scène où les deux chaises sont très mal détectées).

Un autre problème apparaît lorsque certaines zones sont sous-segmentées : la qualité de correspondance ne peut pas être améliorée par des fusions car il est nécessaire d'effectuer des divisions. Ce cas est illustré par le bureau situé à gauche de la scène : ce bureau est bien segmenté à droite, mais il ne forme qu'une seule région avec le bureau central dans l'image de gauche. Les deux régions dans l'image de droite, l'une pour chaque bureau, ne sont pas autorisées à fusionner de par leur trop grande dissimilarité, et la mise en correspondance, de trop mauvaise qualité, n'est pas effectuée.

Bien que notre algorithme de segmentation procède par fusions de régions et soit arrêté précocement, il subsiste des regroupements de régions correspondant à des surfaces différentes. Ce cas de sous-segmentation peut être traité en partie si l'on effectue la division de certaines régions.

Nous avons amélioré la segmentation des régions en utilisant un processus de plus haut niveau : l'appariement. A la vue des résultats de la mise en correspondance, la nécessité d'utiliser là aussi un processus de plus haut niveau pour aider l'appa-

riement commence à se faire sentir. Nous tendons ici vers les limites des possibilités de la mise en correspondance.

6 Conclusion

Appariement :

La segmentation des images stéréoscopiques, opérée grâce à une méthode de fusion assistée par une extraction préalable des points de contraste, fournit deux ensembles de régions distincts. Ceci est essentiellement dû à l'utilisation de critères non stables pour caractériser l'homogénéité des régions.

Au cours de ce chapitre, nous avons montré une façon de prendre en compte les différences de segmentation pour mener à bien l'appariement. Notre méthode se compose de deux phases : génération d'hypothèses, puis validation de ces hypothèses, avec fusion des régions concernées avec leurs voisines si nécessaire. Nous cherchons en fait à apparier l'union des régions candidates à la mise en correspondance. Cette modification des partitions initiales en régions est indispensable pour mener à bien l'appariement des deux vues.

Le traitement effectué peut être résumé par le schéma figure 5.27.

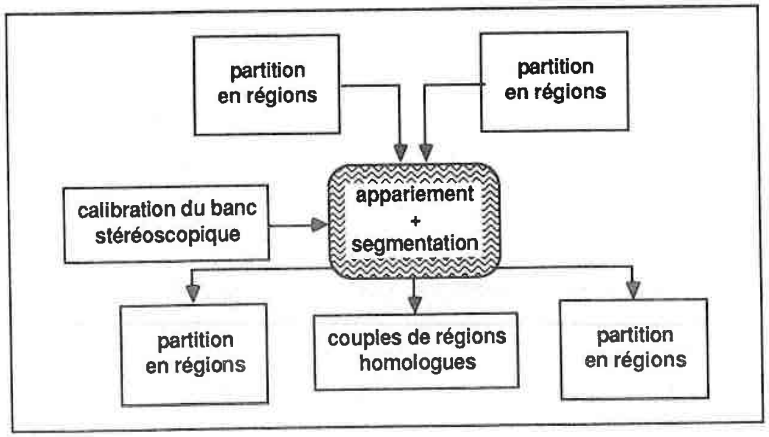

Figure 5.27: *Schéma récapitulatif des traitements effectués pour l'appariement des régions à partir des partitions des images stéréoscopiques.*

Améliorations :

En vue d'améliorer les résultats, il est possible d'ajouter un processus d'appariement procédant par divisions qui traite les régions laissées sans homologue par le

processus d'appariement précédent. Ceci revient à compléter très simplement notre algorithme comme suit :

- *tantque l'ensemble $\mathcal{H}$ des hypothèses est non vide faire*
  - *choisir l'hypothèse dans l'ensemble $\mathcal{H}$ de score le plus élevé :*
    $$h \leftarrow \{h \in \mathcal{H}, score(h) = Max_{\mathcal{H}}(score(h))\}$$
  - *retirer cette hypothèse de l'ensemble $\mathcal{H}$:*
    $$\mathcal{H} \leftarrow \mathcal{H} - \{h\}$$
  - *calculer de la qualité de correspondance Q des deux régions R_g et R_d proposées à l'appariement :*
    $$Qual \leftarrow Q(R_g, R_d)$$
  - *essayer d'améliorer cette qualité Qual par des fusions de régions adjacentes*
  - *si un appariement est possible entre les deux régions éventuellement modifiées par des fusions*
    - *alors effectuer cet appariement*
    - *sinon ajouter l'hypothèse dans l'ensemble $\mathcal{H}'$ des hypothèses non satisfaites :*
      $$\mathcal{H}' \leftarrow \mathcal{H}' \cup \{h\}$$
- *tantque l'ensemble $\mathcal{H}'$ des hypothèses est non vide faire*
  - *choisir l'hypothèse dans l'ensemble $\mathcal{H}'$ de score le plus élevé :*
    $$h \leftarrow \{h \in \mathcal{H}', score(h) = Max_{\mathcal{H}'}(score(h))\}$$
  - *retirer cette hypothèse de l'ensemble $\mathcal{H}'$*
  - *calculer de la qualité de correspondance Q des deux régions R_g et R_d proposées à l'appariement :*
    $$Qual \leftarrow Q(R_g, R_d)$$
  - *essayer d'améliorer cette qualité par des divisions de régions adjacentes*
  - *effectuer ou non l'appariement entre les régions éventuellement modifiées par des divisions*
  - *ajouter dans l'ensemble $\mathcal{H}'$ des nouvelles hypothèses d'appariement entre les régions divisées et non appariées*

Cet algorithme est simple et naturel. La première partie recherche les unions des régions qui peuvent s'apparier, la seconde partie traite les échecs de la première et cherche à apparier les intersections des régions.

Nous ne pouvons cependant garantir la performance de cet algorithme puisque nous ne l'avons pas implanté. Il faut également mentionner que le problème de la division de régions, dans ce cas, n'est pas très simple à résoudre.

Cet algorithme représente cependant une amélioration possible de la version actuelle, et sera implanté si cela s'avère nécessaire. Une version plus complète peut également être envisagée où l'on effectue un traitement itératif d'appariement par fusions et d'appariement par division.

Remarquons cependant que le domaine d'application visé : traitement de scène d'intérieur, a guidé toute notre approche du problème. Notre méthode s'appuie sur l'utilisation de "grands indices", les régions, et est donc mal adapté aux images où les structures sont fines.

Discussion :

Ce chapitre montre qu'un traitement de "bas-niveau" sur une image, en l'occurence l'extraction des régions, ne possède pas suffisamment d'informations pour fournir un résultat satisfaisant dans tous les cas. Il devient indispensable d'utiliser un processus de plus "haut-niveau", ici la comparaison de deux images de la même scène par stéréovision.

Les résultats obtenus s'améliorent au fur et à mesure de leur modification lors de l'enchaînement des traitements : segmentation, puis appariement. Mais là aussi, nous arrivons à un stade où l'information disponible ne suffit plus pour résorber les "doubles" erreurs dues à la fois à la segmentation et à l'appariement. L'amélioration proposée ci-avant, qui se propose d'effectuer des divisions de certaines régions, sera inefficace dans certains cas. Certaines erreurs de bas-niveau ne peuvent plus être résorbées, nous manquons d'éléments pour décider (fusionner, apparier, diviser, ...) et, dans cette ignorance, des erreurs sont commises.

Il nous faut donc davantage d'informations. Certaines peuvent être fournies par une image supplémentaire ; d'autres par le calcul des informations tridimensionnelles : si l'on trouve une incohérence lors de ce traitement, s'il est impossible de construire des surfaces 3d, alors on peut remettre en cause les résultats obtenus lors des deux étapes précédentes.

Nous pouvons également faire intervenir un processus d'interprétation de scène. L'interprétation de certaines parties de l'image pourrait valider ou au contraire annuler l'appariement et/ou la segmentation.

Nous reviendrons sur ces perspectives lors de notre conclusion finale.

Chapitre 6

CONSTRUCTION DES INFORMATIONS TRIDIMENSIONNELLES

Notre propos est de présenter ci-après la méthode utilisée pour la construction des informations tridimensionnelles d'une scène à partir des couples de primitives homologues que nous avons initialement déterminés.

Le plan que nous suivons dans ce chapitre se présente ainsi :

- *La première partie est consacrée à un bref rappel sur la géométrie du système stéréoscopique de prise de vue ; nous y exposons la transformation perspective et son application au calcul des points tridimensionnels.*
- *Dans une deuxième partie, nous présentons les trois principales approches de construction de surfaces tridimensionnelles à partir des couples de régions homologues.*
- *Ceci nous permet d'introduire, en troisième partie, la méthode que nous avons développée. Notre approche s'appuie sur une construction des segments tridimensionnels représentant les frontières des régions, aussi nous présentons les trois étapes nécessaires à ce traitement : l'extraction, l'appariement et le calcul tridimensionnel de tels segments.*
- *Enfin, la construction des surfaces tridimensionnelles est exposée en quatrième partie.*

1 Introduction

L'appariement des images stéréoscopiques, présenté au chapitre précédent, nous fournit un ensemble de couples de régions homologues. Nous pouvons donc à présent envisager de construire des informations tridimensionnelles de la scène visualisée à partir de ces résultats.

Cette étape de construction est composée pour une grande partie de calculs géométriques et demande une connaissance très précise des paramètres du système optique.

L'introduction de ce chapitre est consacrée à la présentation de la projection perspective d'un point de la scène sur un plan image, puis à la projection perspective inverse qui permet de reconstruire un point dans la scène à partir de ses points images.

1.1 Transformation perspective

La calibration de chaque caméra du banc stéréoscopique a été réalisée par G. Toscani à l'INRIA-Rocquencourt [Toscani 87a]. Nous l'avons brièvement présentée au chapitre 2, page 28. Rappelons que calibrer une caméra, c'est déterminer la transformation qui permet de passer de la position d'un point quelconque $P(x, y, z)$ de la scène tridimensionnelle à la position de sa projection $P'(u, v)$ sur le plan image de la caméra (figure 6.1).

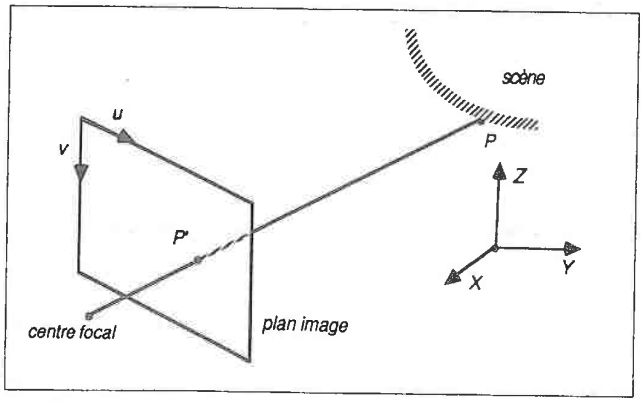

Figure 6.1: *Le point $P(x, y, z)$ de la scène se projette en $P'(u, v)$ sur le plan image. La calibration d'une caméra permet de déterminer cette projection perspective.*

Si les points sont exprimés en coordonnées homogènes, la transformation perspective T s'écrit sous la forme linéaire suivante :

$$P' = T \times P$$

avec P : point de la scène tridimensionnelle, $P(x, y, z)$ s'écrit $(x, y, z, 1)$;
P' : projection de P sur le plan image, $P'(u, v)$ s'écrit $(\mu u, \mu v, \mu)$.

La transformation perspective s'écrit alors :

$$\begin{pmatrix} \mu u \\ \mu v \\ \mu \end{pmatrix} = T \times \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

La matrice T possède 3 lignes et 4 colonnes. La calibration d'une caméra consiste donc à déterminer les 12 coefficients de cette matrice à un coefficient de proportionnalité près. Pour complément, nous donnons en Annexe C, les coefficients d'une telle matrice.

Notons que l'estimation de la transformation perspective s'effectue sous les trois hypothèses suivantes :

- le "plan de vue" sur lequel se forme l'image est effectivement planaire,
- l'axe de visée de la caméra est orthogonal à ce plan de vue,

Les caméras utilisées pour cette application de vision stéréoscopique sont des caméras CCD où les hypothèses citées ci-dessus peuvent être raisonnablement posées.

1.2 Transformation perspective inverse

La construction d'informations tridimensionnelles est l'étape inverse de la projection perspective, c'est-à-dire : connaissant la position d'un point dans l'image et la matrice de projection T, trouver la position du point dans la scène dont il est la projection.

Si l'on ne dispose que d'une seule caméra, cette position est indéterminée puisqu'il existe une infinité de points dans l'espace vérifiant la projection perspective. Ces points 3d appartiennent tous à la demi-droite Δ passant par le centre optique de la caméra et par le point image considéré (figure 6.2).

L'utilisation d'une seconde caméra permet de lever l'indétermination du point P : le point cherché se situe théoriquement à l'intersection des deux demi-droites Δ_g et Δ_d issues des deux projections P_g et P_d homologues.

Les deux points P_g et P_d sont choisis de façon à ce qu'ils se trouvent sur deux lignes épipolaires associées. Il n'y a donc théoriquement aucun problème d'indétermination de P. Cependant, en pratique, du fait de l'incertitude des coefficients des matrices de transformation perspective et des matrices permettant de calculer les droites épipolaires associées, les deux demi-droites Δ_g et Δ_d sont rarement

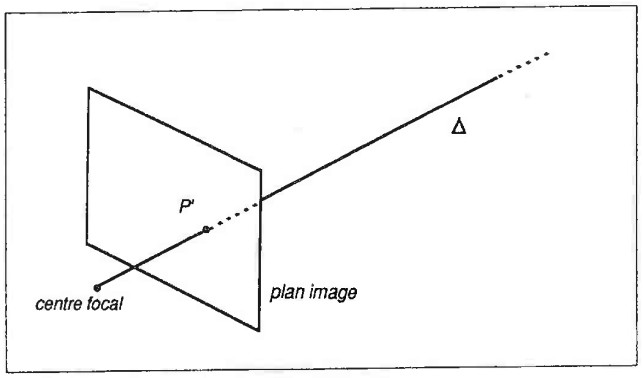

Figure 6.2: *Tous les points de la demi-droite Δ se projettent sur le point image P'.*

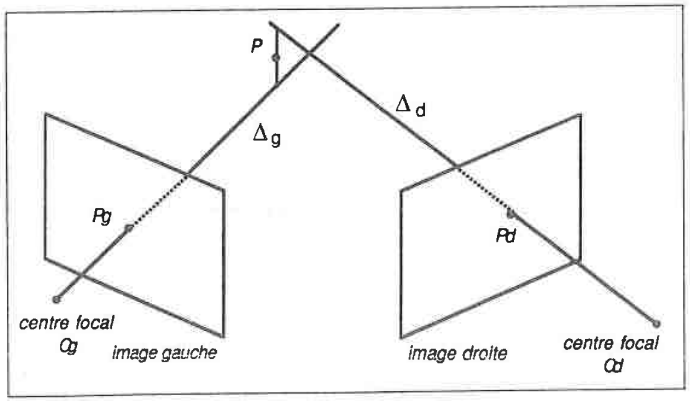

Figure 6.3: *Le point P est situé à équidistance des deux demi-droites Δ_g et Δ_d issues des deux projections P_g et P_d homologues, sur leur perpendiculaire commune.*

coplanaires et ne s'intersectent donc pas. Nous avons choisi d'estimer le point réel 3d par le point situé à équidistance des deux demi-droites Δ_g et Δ_d sur leur perpendiculaire commune (figure 6.3).

La connaissance de la matrice T et d'un couple de points images homologues, nous permet donc de calculer les coordonnées tridimensionnelles d'un point dans la scène. Nous pouvons à présent envisager la construction des surfaces 3d et nous en exposons ci-après les principales approches.

2 Différentes méthodes de construction des surfaces tridimensionnelles

Dans le problème traité, la primitive image utilisée pour effectuer la mise en correspondance des images stéréoscopiques est la région. Or l'appariement de ces primitives ne permet pas une association géométrique directe entre les différentes vues. Nous savons construire des points 3d à partir des couples de points images homologues, mais non immédiatement des surfaces 3d à partir des couples de régions homologues.

Le point apparaît alors comme un *intermédiaire obligatoire* pour effectuer la construction. De ce fait, il semble naturel de caractériser les régions par un ensemble de points. La construction peut alors être réalisée par l'enchaînement des traitements suivants :

1. appariement des points images homologues,
2. calcul des points 3d issus des couples de points,
3. construction des surfaces 3d caractérisées par ces ensembles de points 3d.

Il existe plusieurs façons de caractériser une région par un ensemble de points. Un tel ensemble peut être, par exemple :

- l'ensemble des points images qui constituent la région,
- l'ensemble des points images qui constituent le(s) contour(s)[1] de la région,
- l'ensemble des points particuliers situés sur le(s) contour(s) de la région.

L'utilisation de chacun de ces trois ensembles pour construire les informations tridimensionnelles est étudiée ci-après.

2.1 L'ensemble des points de la région

a. Caractérisation des régions

Par définition, une région est un ensemble connexe de points images. L'ensemble de points le plus simple possible caractérisant une région est celui qui forme justement cette région.

[1] Une région peut posséder plusieurs contours si elle est "trouée", à savoir : son contour externe et ses contours internes.

b. Appariement des points

Les difficultés apparaissent lors de la phase d'appariement des points caractérisant deux régions homologues. Tous les points concernés ont des caractéristiques proches puisqu'ils appartiennent à deux régions homogènes et similaires, dans lesquelles, par définition, il ne se passe rien. La combinatoire du problème est donc excessivement élevée et le résultat n'est pas garanti : n'importe quel point d'une région peut s'associer avec n'importe quel point de la région homologue de l'autre vue, ou presque. Pour ces raisons, cette solution est très rarement retenue.

Remarquons cependant qu'une telle méthode peut être envisagée lorsque les régions sont très texturées, c'est-à-dire formées par des points individuellement dissemblables ; l'ensemble des points formant toutefois une région homogène au sens où nous l'avons défini au chapitre 4.

c. Surfaces tridimensionnelles

Si l'appariement des points est correct entre deux régions homologues, la construction de la surface tridimensionnelle se fait par interpolation de tous les points 3d engendrés. Cette construction est de bonne qualité si ces points sont suffisamment denses (figure 6.4).

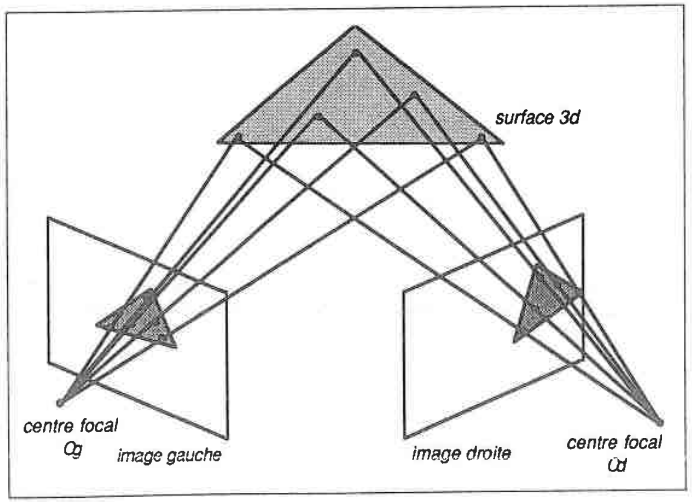

Figure 6.4: *Un appariement correct des points d'un couple de régions homologues permet une bonne construction de la surface 3d dans la scène. Un tel appariement est cependant très difficile, voire impossible dans certains cas, à réaliser.*

L'inconvénient majeur de cette approche est de demander un appariement correct

sur les deux ensembles de points. Une telle mise en correspondance est très difficile, voire impossible à réaliser, avec les images que nous utilisons.

2.2 L'ensemble des points formant le contour de la région

a. Caractérisation des régions

L'idée sous-tendue est de diminuer le nombre de "sous-primitives", c'est-à-dire dans notre cas, de diminuer le nombre de points à apparier. Une solution à ce problème est de travailler avec la primitive duale des régions, c'est-à-dire les lignes de contour. Chaque région peut alors être caractérisée par l'ensemble des points formant son (ses) contour(s).

Dans la suite de ce chapitre, nous parlerons du contour d'une région, même si celle-ci en possède plusieurs.

b. Appariement des points

La combinatoire de l'appariement des points des contours est fortement diminuée par rapport au cas précédent, mais elle reste importante de par le nombre de points, en moyenne, formant les contours des régions. Ce nombre moyen, pour nos exemples, se situe aux environs de 100, comme l'indique le tableau présenté à la fin de ce paragraphe, page 135. Si on ajoute la contrainte épipolaire, la combinatoire du problème devient très faible.

c. Surfaces tridimensionnelles

A partir de chaque couple de points frontières, un point 3d dans l'espace peut être construit. Il s'agit ensuite de construire la surface s'appuyant sur le contour 3d calculé. Une hypothèse simplificatrice peut alors être utilisée : considérer chaque région dans les images comme la projection d'une surface 3d plane, ou quasi-plane.

Cette hypothèse peut paraître très restrictive, mais remarquons que pour certaines images, comme les scènes d'intérieur avec lesquelles nous travaillons, comportent beaucoup d'objets formés de surfaces planes (murs, bureaux, sol, ...). De plus, si nous traitons un objet comportant des surfaces sphériques (lampe, téléphone, ...), nous pouvons attendre de la segmentation d'images, du fait de la différence des intensités lumineuses sur la courbure des surfaces, qu'elle fournisse sur chaque image un ensemble de plusieurs régions représentant la projection de cet objet. L'objet 3d construit est alors un ensemble de facettes planes, que nous pouvons accepter dans un premier temps.

Si l'hypothèse "le monde est localement plan" est formulée, il faut chercher un plan 3d qui passe au mieux par l'ensemble des points 3d que nous venons de calculer (figure 6.5) et représentant son contour.

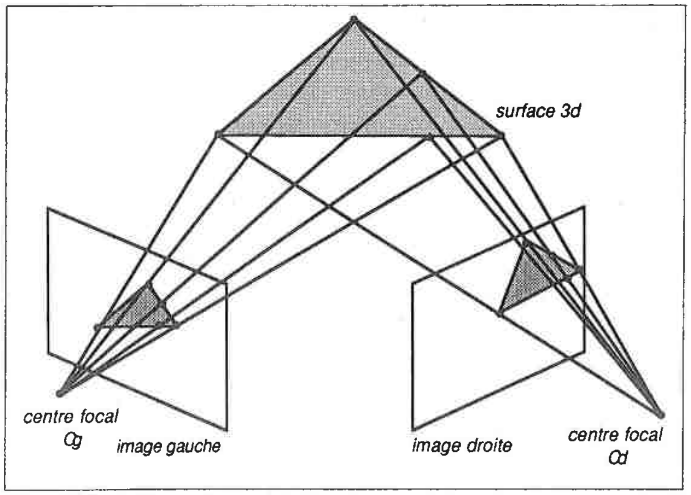

Figure 6.5: *L'appariement des points formant les contours de deux régions homologues permet une construction 3d du contour de la surface dans la scène. Il s'agit ensuite de compléter le "vide", en construisant une surface s'appuyant sur ce contour 3d. Cette surface est généralement approchée par un plan.*

2.3 L'ensemble des points caractéristiques du contour d'une région

a. Caractérisation des régions

Une troisième solution consiste à de caractériser une région par un ensemble de points particuliers situés sur sa frontière. De tels points peuvent être, par exemple, des points d'inflexion dans les chaînes de contours, ou le centre ou les extrémités des segments approchant ces chaînes, etc. Ce dernier choix est souvent effectué : chaque région est alors caractérisée par l'ensemble des segments en approchant le contour.

b. Appariement des points

Si nous choisissons de caractériser les régions par l'approximation polygonale de leur contour, il s'agit de mettre en correspondance les segments homologues avant de passer à un calcul 3d de ce contour.

Pour chaque couple de régions homologues, on cherche à apparier au mieux les segments qui approchent leur frontière. Remarquons que la frontière d'une région est généralement approchée par un très faible nombre de segments. Ceci signifie que la combinatoire du problème est fortement réduite et que toute méthode d'appariement peut être utilisée avec succès. Le tableau présenté page 135 indique le nombre moyen de segments approchant le contour des régions, soit environ 5 segments par région pour les scènes traitées.

c. Surfaces tridimensionnelles

Avant de procéder à la construction des surfaces tridimensionnelles, nous devons calculer, toujours dans le cas de l'utilisation des segments, les segments 3d qui approchent le contour des régions dans la scène.

A partir de chaque paire de segments homologues S_g et S_d, situés sur le contour de deux régions homologues, on construit le plus grand segment 3d S dont les deux projections S'_g et S'_d sur les plans images droit et gauche sont incluses dans les deux primitives images associées S_g et S_d (figure 6.6). Cette construction est détaillée page 146.

Ces constructions fournissent un ensemble de segments 3d dans la scène qui représentent le contour de la surface. L'étape suivante est de déterminer la surface à laquelle tous ces segments appartiennent ; cette étape est facilitée en supposant que toutes les surfaces à construire sont planes (figure 6.7), hypothèse dont nous avons discuté page 131.

Pour information, citons les travaux de S. Kuroe [Kuroe 85] concernant la construction des surfaces tridimensionnelles à partir des couples des régions. Dans cette méthode, chaque région est caractérisée par le polygone convexe qui la circonscrit. Ce choix est guidé d'une part, par le grand nombre de régions dont l'approximation du contour fournit de manière naturelle un polygone convexe, et d'autre part, par l'insensibilité au faible bruit qui peut modifier les contours des régions.

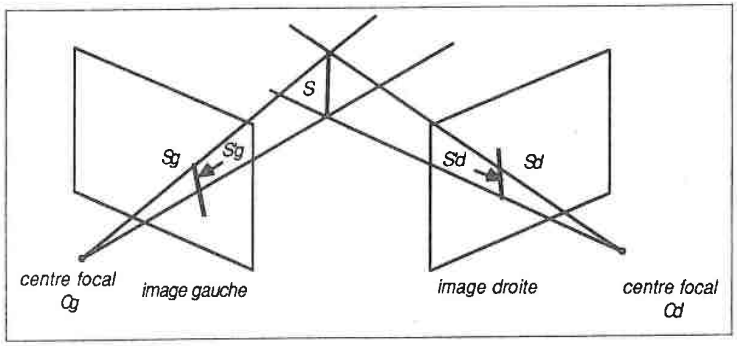

Figure 6.6: *A partir de chaque paire de segments homologues S_g et S_d, on construit le plus grand segment 3d S dont les deux projections S'_g et S'_d sur les plans images gauche et droit sont incluses dans les deux primitives images associées S_g et S_d.*

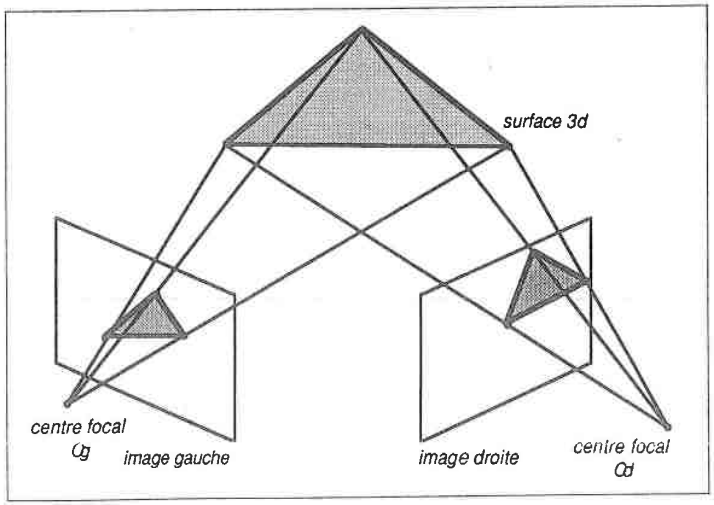

Figure 6.7: *L'appariement des segments formant le contour de deux régions homologues permet une construction 3d du contour de la surface dans la scène. On construit ensuite le plan s'appuyant sur ces éléments de contours.*

L'appariement des polygones convexes homologues fournit, pour chaque région, un ensemble de segments 3d dans la scène. Ces segments sont prolongés pour former un contour 3d fermé qui détermine le plan 3d de la surface cherchée.

Cette méthode testée sur une scène réelle où ne figurent que des objets polyédriques, à facettes planes, sans trous et facilement segmentables, donne de bons résultats.

La construction de surfaces à partir de segments 3d dans la scène peut se faire par une triangulation de Delaunay qui respecte la visibilité des indices [Boissonnat 87]. Les segments sont alors soit des arêtes de Delaunay, soit une réunion d'arêtes de Delaunay. Grâce aux positions des ensembles de segments, il est possible d'utiliser un critère de visibilité et de marquer les tétraèdres vides. La réunion de ces tétraèdres correspond à l'espace vide de la scène, les autres tétraèdres représentent une approximation polyédrale de la surface des obstacles. Cette façon de représenter les données stéréoscopiques est satisfaisante.

Une autre approche en cours de développement par J.D. Boissonnat et O. Monga [Boissonnat 88] concerne la construction des surfaces à partir de l'intersection de faisceaux de plans.

2.4 Tableau des caractéristiques des régions

Nous présentons ci-après le tableau contenant les nombres minimum, maximum et moyens de *points* par régions, de *points de frontière* par régions et de *segments approchant ces frontières* par région. Ces calculs ont été effectués pour quatre paire d'images stéréoscopiques différentes nommées *bureau1*, *bureau2*, *couloir* et *pièce*. Ces images sont présentées dans l'Annexe D consacrée aux résultats.

Dénombrement des éléments caractéristiques des régions

Taille de chacune des images de la scène : 256 × 256 points

Nom de la paire d'images stéréoscopiques	Nombre de *points* par régions			Nombre de *points de frontière* par région			Nombre de *segments de frontière* par région		
	min.	max.	moy.	min.	max.	moy.	min.	max.	moy.
bureau1	2	7543	341.33	6	1597	86.81	1	81	4.72
bureau2	1	8563	320.47	4	1999	84.55	1	96	4.35
couloir	1	10183	355.21	4	892	95.56	1	34	4.08
pièce	1	5167	320.47	4	648	99.25	1	31	4.87

3 Notre approche

Nous venons de présenter une étude concernant les différentes approches permettant de construire des informations 3d à partir des couples de régions. Ceci nous

a conduit à préférer une méthode s'appuyant sur les *segments de contours*. Ce choix repose sur les trois avantages suivants :

- la facilité *d'extraction des segments* de contours des régions à partir de la ligne de leur frontière,
- le *faible nombre* de primitives à apparier,
- l'approximation satisfaisante d'une surface 3d par une *facette plane*.

L'étape de construction de la scène tridimensionnelle se divise alors en quatre phases successives :

1. approximation polygonale des lignes des frontières des régions,
2. mise en correspondance des segments,
3. construction des segments tridimensionnels,
4. construction des surfaces tridimensionnelles.

Nous présentons ci-après les techniques utilisées pour effectuer chacune de ces étapes.

3.1 Approximation polygonale des frontières des régions

Pour chaque région, nous cherchons l'ensemble des segments qui approchent au mieux son contour.

L'approximation polygonale d'une chaîne formée de points est réalisée très simplement par divisions récursives[2] : le principe, illustré figure 6.8, est de créer un segment qui joint les deux extrémités de la chaîne. Si l'écart entre ce segment et la chaîne est jugé trop important, on coupe la chaîne au point où cet écart est maximal. On recommence ensuite l'approximation sur les deux sous-chaînes ainsi créées.

Les deux difficultés générales de tout algorithme d'approximation polygonale qui se posent sont les suivantes :

quand couper : c'est-à-dire déterminer si un segment est une bonne approximation d'une chaîne ou non, donc décider de couper ou non ;

où couper : c'est-à-dire déterminer le point de la chaîne auquel il faut couper.

La technique que nous avons mise en œuvre regroupe ces deux problèmes : le *point d'appréciation* de la qualité de l'approximation de la chaîne par le segment construit et le *point de coupe* de la chaîne sont confondus.

Cette technique peut être appliquée quelle que soit la forme de la courbe à approcher. Si la chaîne est fermée, comme cela est le cas pour le contour des régions, il se pose le problème de choisir les extrémités de la ou des premières sous-chaînes. Une solution coûteuse consiste à couper la chaîne initiale fermée aux deux

[2] Cet algorithme est généralement connu sous le nom d'*algorithme de la corde*.

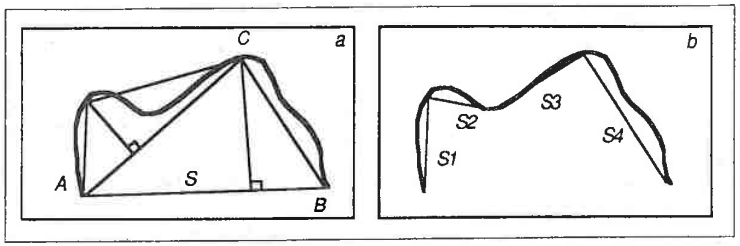

Figure 6.8: **a** : *Le segment S est une mauvaise approximation de la chaîne (A, B). On cherche alors ensuite à approcher les deux chaînes (A, C) et (C, B), où C est le point de la chaîne le plus éloigné du segment S.* **b** : *L'approximation est arrêtée lorsqu'on obtient un ensemble de segments satisfaisant, ici : $\{S_1, S_2, S_3, S_4\}$.*

points géométriquement les plus éloignés et de traiter ensuite les deux sous-chaînes ainsi obtenues. Nous avons implanté une méthode plus simple : deux points voisins sont choisis aléatoirement pour représenter les deux extrémités de la sous-chaîne (le traitement s'effectue alors sur cette seule sous-chaîne).

Nous présentons respectivement, figures 6.9 et 6.10, les partitions en régions et le résultat de l'approximation polygonale des frontières des régions.

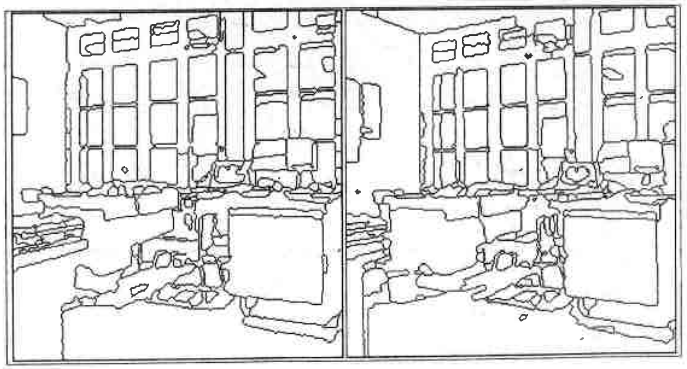

Figure 6.9: *Partitions en régions : image gauche : 190 régions, image droite : 194 régions.*

A la vue de ces résultats, il semble inutile de poursuivre dans cette voie. En effet, on s'aperçoit "à l'oeil nu" sur la figure 6.10 que les segments, avec lesquels on veut travailler, sont mal localisés. Ils sont assez décalés par rapport au contour réel des projections des surfaces de la scène comme en témoigne la figure 6.11 où nous

Figure 6.10: *Approximation polygonale des frontières des régions à apparier ; image gauche : 370 segments, image droite : 360 segments.*

avons superposé l'image originale gauche à ces segments de contours.

Comme nous le montrons au paragraphe 3.4, une incertitude de 1 pixel sur la localisation des primitives images engendre une incertitude importante de la primitive construite dans la scène, cette incertitude provient de la calibration des caméras. Ceci signifie que la précision de localisation des segments 2d à apparier doit être la meilleure possible. Or ce n'est pas le cas.

Face à ce problème, nous pouvons adopter l'une des trois solutions suivantes :

- mieux choisir les deux extrémités de la première sous-chaîne extraite d'une courbe initiale fermée ;
- utiliser un autre traitement d'approximation polygonale, par exemple construire les segments par une méthode d'approximation aux sens des moindres carrés ;
- utiliser d'autres chaînes de points.

Les deux premières approches sont envisageables à condition de disposer de chaînes de contours bien localisées, or ce n'est pas le cas : le processus de segmentation, quoique guidé par une extraction préalable des lignes de contraste, décale légèrement les frontières. Ce léger décalage suffit à perdre une précision considérable lors de la construction des segments 3d.

Nous avons donc retenu l'idée de travailler avec d'autres chaînes de points de contours : nous utilisons les lignes de points de contraste en remplacement des frontières des régions.

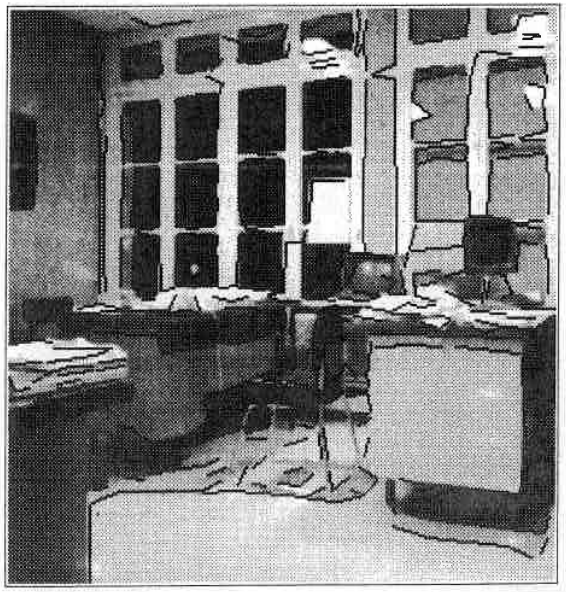

Figure 6.11: *Superposition des segments de contours à l'image originale gauche.*

3.2 Approximation polygonale des lignes de contraste

Nous remplaçons donc les bords des régions par les segments qui approchent les lignes de contraste, ces lignes sont bien placées, elles sont situées avec une incertitude généralement inférieure au pixel, et plus stables que les frontières des régions. Nous avons déjà évoqué cette précision au chapitre 4.

Les chaînes de points de contraste extraites par la méthode développée par R. Deriche [Deriche 87a] sont approchées par des segments en utilisant la technique d'approximation polygonale par divisions récursives que nous venons de présenter.

Nous présentons respectivement figures 6.12 et 6.13 les chaînes de points de contraste extraites des images initiales et le résultat de l'approximation polygonale.

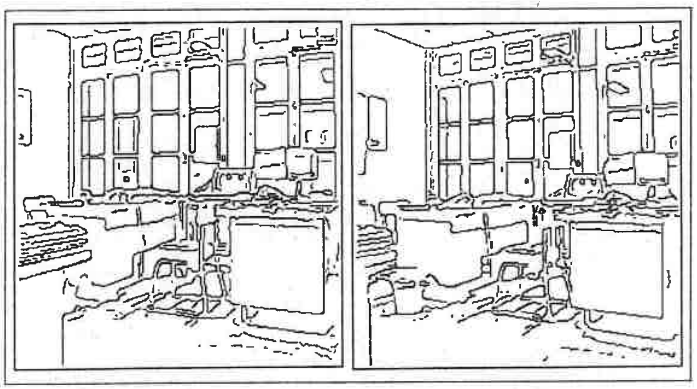

Figure 6.12: *Extraction des chaînes de points de contraste des images stéréoscopiques [Deriche 87a].*

Ces segments sont cette fois-ci beaucoup mieux localisés comme en témoigne la figure 6.14 où nous montrons les segments de contours superposés à l'image initiale gauche.

Les segments trop petits (typiquement d'une longueur inférieure à 5 points pour les résultats présentés) sont éliminés. Il en résulte une perte de la connexité des segments issus de la même chaîne de points. Cette perte n'est cependant qu'apparente ; en effet, on détermine pour chaque segment d'une part les régions dont il est frontière, et d'autre part son rang dans la liste des segments frontières de chacune de ces régions.

Les frontières polygonales des régions restent "visuellement" ouvertes, mais elles délimitent bien des zones fermées possédant des caractéristiques bien définies.

Nous pouvons à présent envisager d'apparier ces segments pour ensuite effectuer le calcul des informations tridimensionnelles.

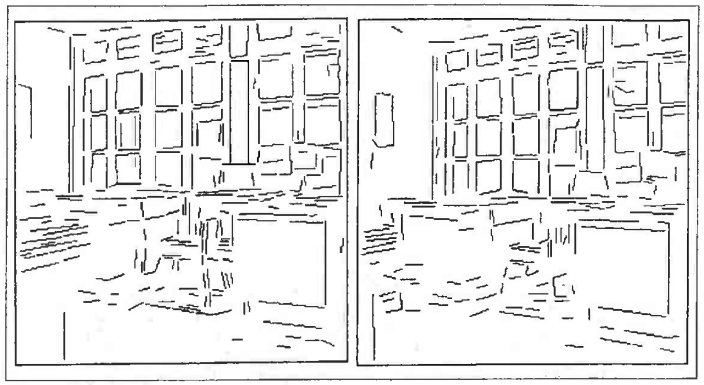

Figure 6.13: *Approximation polygonale des chaînes de points de contraste extraites des images stéréoscopiques ; image gauche : 287 segments, image droite : 261 segments.*

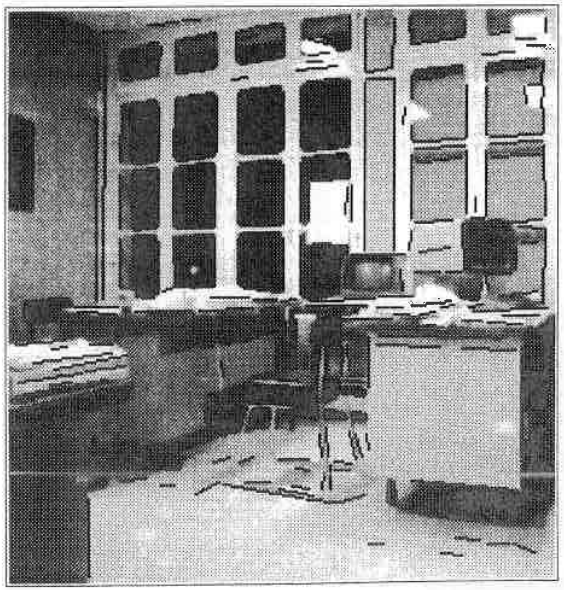

Figure 6.14: *Superposition des segments de contours à l'image originale gauche.*

3.3 Appariement des segments

La mise en correspondance est effectuée en respectant les localisations des segments par rapport aux lignes épipolaires. Les segments intersectant une ligne épipolaire de l'image gauche sont appariés avec les segments intersectant la ligne épipolaire associée dans l'image droite.

L'appariement des segments d'un couple de lignes épipolaires associées est effectué par recherche d'un chemin optimal dans un graphe, selon la technique de programmation dynamique. Cette technique est identique à celle que nous avons présentée au chapitre précédent page 102. Nous en rappelons ci-après seulement brièvement le principe et mettons l'accent sur les différences dues à l'utilisation des segments en remplacement des régions.

La figure 6.15 représente un couple de lignes épipolaires et les segments de l'image (figure 6.13) qui intersectent l'une de ces lignes ; ces segments lignes sont proposés à l'appariement. Le chemin optimal représentant la mise en correspondance de ces segments se trouve figure 6.16 ; les segments de la ligne gauche sont sur l'axe horizontal, ceux de la ligne droite sur l'axe vertical.

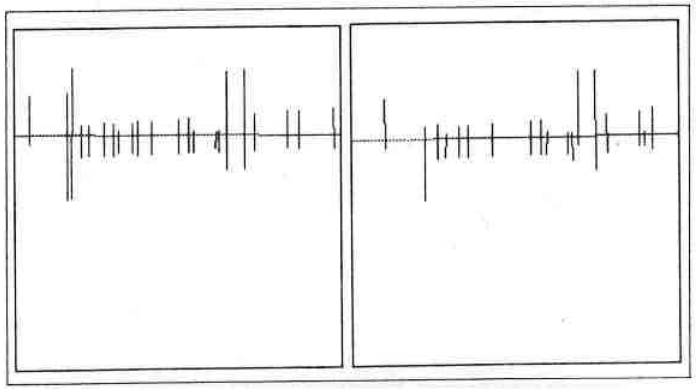

Figure 6.15: *Un couple de lignes épipolaires associées intersectant les segments à apparier.*

Puisque les segments ont une taille supérieure à celle d'un point, un même segment intersecte en fait plusieurs lignes épipolaires. Pour éviter des erreurs ou des conflits d'appariement, nous adoptons la même approche que celle utilisée lors de l'appariement des régions :

- la recherche d'un chemin optimal permet, non pas d'affirmer l'appariement entre segments, mais d'en émettre des *hypothèses* ;
- ces hypothèses sont comptabilisées, ordonnées en fonction de leur nombre d'apparitions, et *validées* sous certaines conditions.

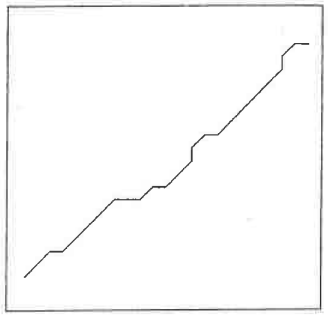

Figure 6.16: *Chemin optimal représentant la mise en correspondance des segments.*

La méthode utilisée se résume de la façon algorithmique suivante :

- *<u>pour</u> chaque couple de lignes épipolaires associées <u>faire</u>*
  - *chercher parmi les segments qui intersectent les lignes épipolaires, ceux qui sont homologues, par construction d'un chemin optimal dans un graphe*
  - *<u>pour</u> chaque couple de segments homologues (S_g, S_d) trouvés <u>faire</u>*
    - *<u>si</u> l'hypothèse d'appariement (S_g, S_d) existe déjà*
      - *<u>alors</u> incrémenter le score de cette hypothèse de 1*
      - *<u>sinon</u> ajouter cette hypothèse dans l'ensemble $\mathcal{H}$ des hypothèses*
- *ordonner les hypothèses de $\mathcal{H}$ en fonction de leur score*
- *<u>tantque</u> l'ensemble $\mathcal{H}$ est non vide <u>faire</u>*
  - *extraire de l'ensemble $\mathcal{H}$ l'hypothèse d'appariement (S_g, S_d) de meilleur score*
  - *calculer la qualité de correspondance Q des deux segments homologues (S_g, S_d)*
  - *<u>si</u> cette qualité est jugée suffisante*
    - *<u>alors</u> valider l'appariement*

La qualité de correspondance de deux segments est fonction de la similitude de leurs attributs et des résultats de l'appariement des régions. L'évaluation de la fonction Q est réalisée de la façon suivante :

$$Q(S_g, S_d) = \sum_{i=1}^{n} C_i \times A_i(S_g, S_d) + C_r \times R(S_g, S_d)$$

avec S_g, S_d : deux segments proposés à l'appariement ;

$A_i(S_g, S_d)$: mesure de la similitude de l'attribut numéro i des segments S_g et S_d ;

C_i : coefficient de pondération de la similitude de l'attribut numéro i

$R(S_g, S_d)$: mesure de l'appariement des régions voisines des segments. R est une fonction qui prend en compte les couples de régions homologues précédemment déterminés, comme le montre la figure 6.17 ;

C_r : coefficient de pondération de la fonction R.

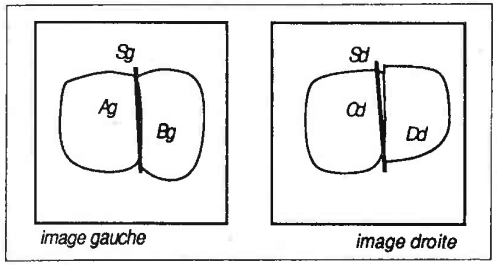

Figure 6.17: *La création des couples de régions homologues (A_g, C_d) et (B_g, D_d) renforce la correspondance du segment S_g avec S_d.*

Les attributs des segments utilisés sont leur longueur et leur direction[3]. On recale les frontières des régions aux segments afin de déterminer leurs régions voisines (figure 6.18).

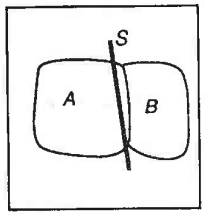

Figure 6.18: *On considère le segment S comme étant sur la frontière entre les régions A et B.*

L'application de cet algorithme sur les données présentées figure 6.13 fournit l'ensemble des appariements visualisés figure 6.19.

[3]La direction d'un segment est évaluée en fonction de sa position par rapport aux lignes épipolaires et non par rapport au "cadre" de l'image

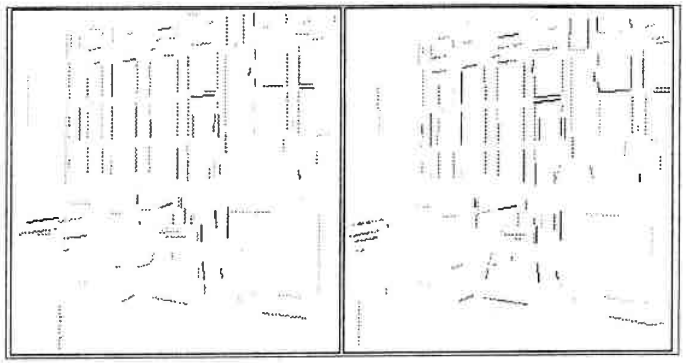

Figure 6.19: *Résultats de l'appariement des segments. Les segments de même niveau de gris sont appariés ; nombre de couples de segments homologues : 105.*

3.4 Méthode de construction des segments tridimensionnels

A partir de chaque paire de segments homologues (S_g, S_d), on construit le segment 3d S correspondant.

Cette construction consiste en fait à réaliser l'intersection de deux faisceaux plans π_g et π_d issus respectivement du centre épipolaire gauche E_g et passant par le segment gauche S_g, et du centre épipolaire droit E_d et passant par le segment droit S_d (figure 6.20). L'intersection de ces deux faisceaux plans forme le segment S 3d cherché.

Nous réalisons ce calcul d'une manière différente mais dont le résultat est absolument similaire. Nous déterminons les zones des segments pouvant s'associer avant de procéder à une construction tridimensionnelle. Pour chaque segment apparié, cette zone est déterminée par son intersection avec la bande épipolaire définie par les extrémités supérieures et inférieures du segment correspondant.

La figure 6.21 représente les deux parties communes S'_g et S'_d des segments S_g et S_d pouvant s'associer. Il suffit ensuite d'apparier respectivement les points supérieurs et inférieurs des deux segments S'_g et S'_d et d'en calculer les deux points 3d correspondants. Ces deux points dans l'espace représentent les extrémités du segment tridimensionnel S cherché (figure 6.22).

Nous présentons figure 6.23 les zones des segments pouvant s'associer. La majorité des segments horizontaux est perdue car ces segments sont portés par les épipolaires ; une grande partie des segments horizontaux qui restent engendre des segments 3d avec une trop grande erreur : ces segments doivent être éliminés pour ne pas perturber la construction des informations tridimensionnelles.

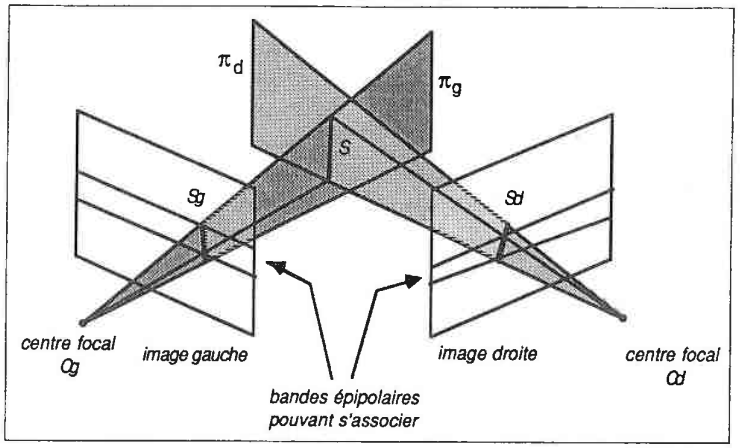

Figure 6.20: *L'intersection des deux faisceaux plans π_g et π_d issus respectivement du centre épipolaire gauche E_g et passant par le segment gauche S_g, et du centre épipolaire droit E_d et passant par le segment droit S_d, forme le segment S 3d cherché.*

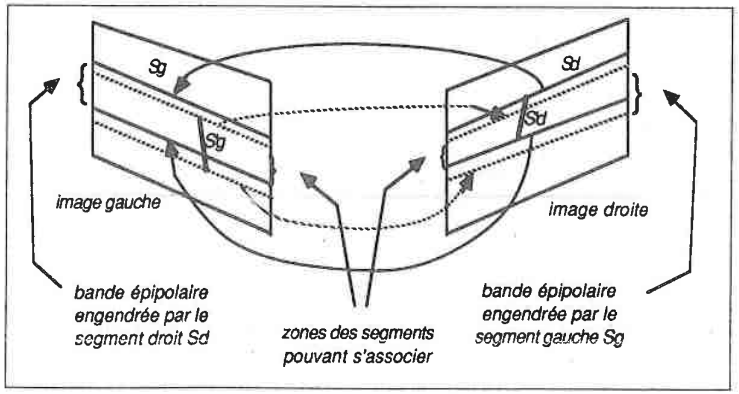

Figure 6.21: *Les zones S'_g et S'_d des deux segments à associer sont définies par l'intersection de chaque segment avec la bande épipolaire issue des extrémités du segment correspondant dans l'autre vue.*

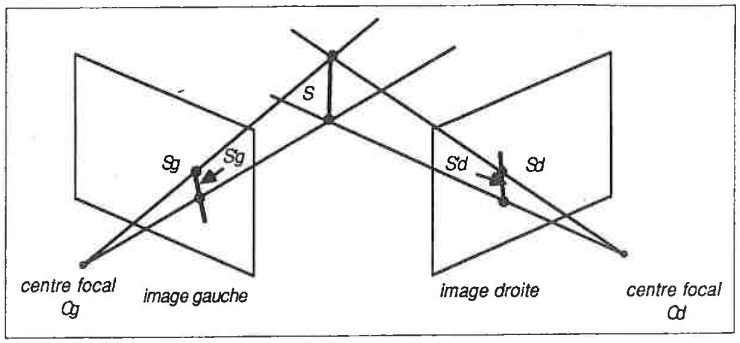

Figure 6.22: *Le segment 3d S est construit en joignant les deux points 3d calculés par l'appariement respectif des extrémités supérieures et inférieures des deux segments S'_g et S'_d.*

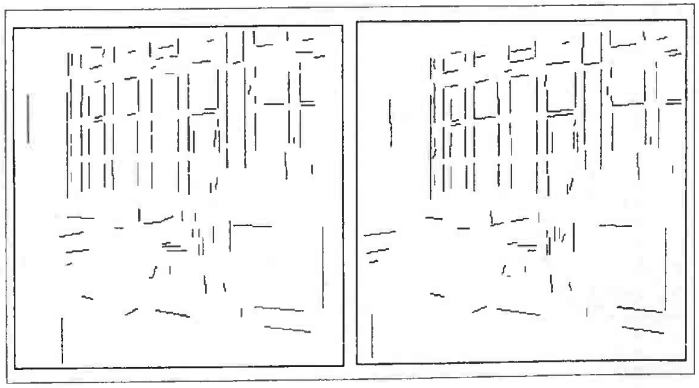

Figure 6.23: *Visualisation des zones des segments pouvant s'associer lors du calcul des segments 3d.*

3.5 Construction des segments tridimensionnels

Résultats :

Nous exposons ici les résultats 3d obtenus pour une scène étudiée. D'autres résultats, plus complets et diversifiés se trouvent en Annexe D.

Il n'est pas facile de visualiser et de représenter les segments obtenus en trois dimensions, aussi nous présentons successivement quatre figures :

- Figure 6.24 : les segments 3d calculés sont projetés sur chacun des plans images et colorés en fonction de leur distance par rapport aux caméras (profondeur de la scène). La mire de coloration utilisée est également visualisée.
- Figure 6.25 : les segments 3d sont projetés sur un plan perpendiculaire aux axes de visée des caméras. Cette vue correspond à la "vue de face" des segments de la scène.
- Figure 6.26 : les segments 3d sont projetés sur un plan parallèle aux axes de visée des caméras. Cette vue correspond à la "vue de côté".
- Figure 6.27 : les segments 3d sont projetés sur le "sol de la scène". Cette vue correspond à la "vue de dessus" et permet de représenter la distance des segments aux caméras, c'est-à-dire la profondeur.

Remarque : Pour les représentations 3d qui suivent, certaines informations sont ajoutées manuellement pour aider à la compréhension :

- l'emplacement du robot est schématisé par la lettre R ;
- les différents segments 3d sont regroupés selon l'objet auquel ils appartiennent. Pour simplifier, nous n'avons considéré que quatre "objets" :
  - le mur du fond : *mur*,
  - le bureau à droite : *bur1*,
  - le bureau à gauche : *bur2*,
  - le bureau au fond à gauche dans la pièce : *bur3*.

Le principal inconvénient de ces résultats est le faible nombre des segments 3d calculés. Ceci correspond, en fait, en majeure partie, à la perte des segments horizontaux du à la position quasi-horizontale des caméras. Nous avons donc un handicap pour la future construction des surfaces 3d car les informations sont incomplètes. Cet inconvénient ne peut être corrigé qu'en plaçant les caméras sur un support oblique.

En revanche, ces segments sont assez bien localisés : la construction des surfaces 3d s'appuyant sur ces segments est prometteuse. A titre d'illustration nous présentons figure 6.28 les résultats obtenus en utilisant les segments qui approchent les contours des régions. Il est aisé de constater que ceux-ci sont très mal localisés.

A titre de complément, nous discutons ci-après des incertitudes obtenues sur les coordonnées des segments 3d.

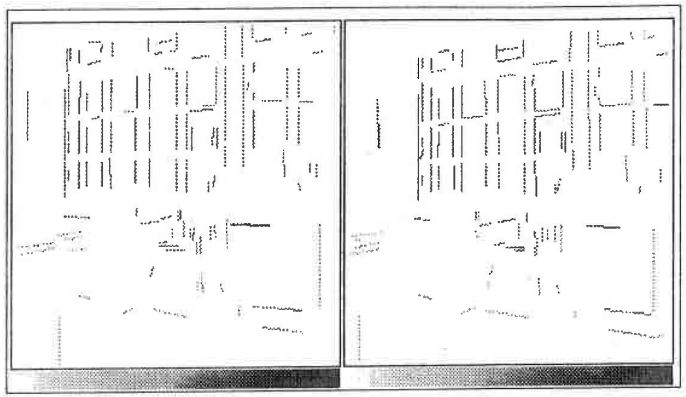

Figure 6.24: *Les segments 3d sont projetés sur les deux plans de vue des caméras et colorés (niveau de gris) en fonction de leur profondeur. En dessous des images se trouve la mire utilisée ; nombre de segments 3d : 105.*

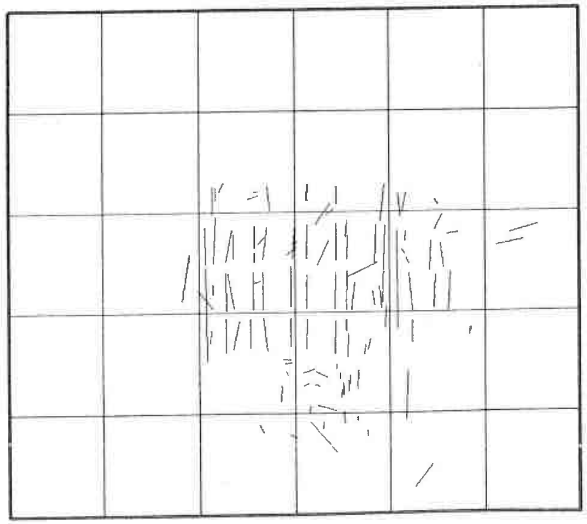

Figure 6.25: *"Vue de face" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 6 mètres, hauteur : 5 mètres ; taille du quadrillage: 1 m × 1 m.*

Figure 6.26: *"Vue de côté" des segments tridimensionnels. Echelle de représentation de la scène : profondeur : 9 mètres, hauteur : 5 mètres ; taille du quadrillage : 1 m × 1 m.*

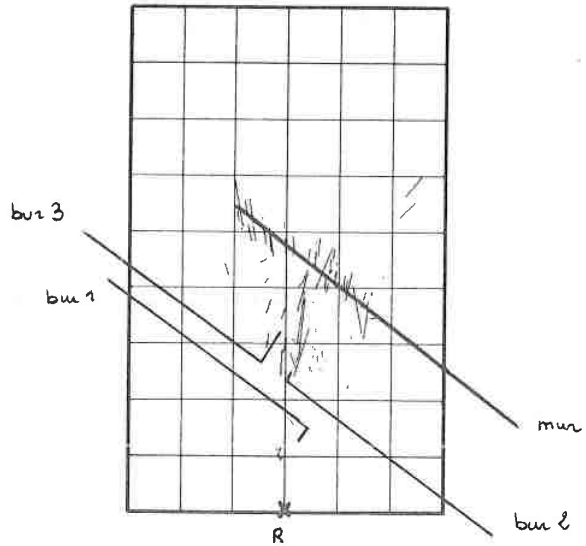

Figure 6.27: *"Vue de dessus" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 6 mètres, profondeur : 9 mètres ; taille du quadrillage : 1 m × 1 m.*

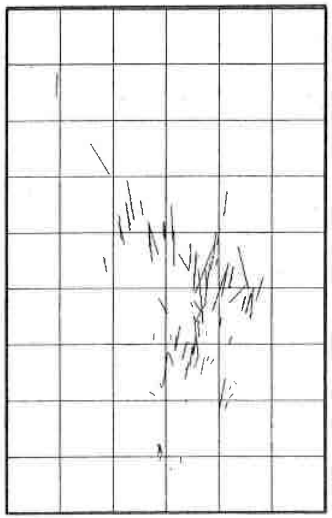

Figure 6.28: *"Vue de dessus" des segments tridimensionnels ; largeur : 6 mètres, profondeur : 9 mètres. (Approximation des frontières des régions).*

Incertitude sur les coordonnées tridimensionnelles :

Les incertitudes de localisation des segments 2d à apparier influent de manière non négligeable sur la position des segments dans l'espace, comme le montre le tableau présenté ci-après. Cette erreur est proportionnelle à la profondeur des segments dans la scène et peut prendre des valeurs très importantes. Le mur des fenêtres est situé à une distance comprise entre 3,50 mètres et 5,50 mètres, et dans ce cas, les erreurs sur les incertitudes des extrémités des segments peuvent atteindre 50 centimètres, ce qui est très important.

Il faut cependant remarquer que les images originales sont numérisées à une taille de 512 × 512 points. Ces images sont ensuite réduites par sous-échantillonnage à une taille de 256 × 256 afin de diminuer le volume des données et de certains traitements. Le fait de travailler avec des images réduites nous fait perdre, ici, de la précision.

Incertitude sur la profondeur

Calculs effectués sur l'image de la scène de bureau dont nous venons de présenter les résultats. Les mesures sont présentées en centimètres.

Erreur maximale dans la profondeur d'un segment provoquée par une erreur de 1 pixel dans la position de l'un des deux segments 2d	Profondeur du segment
1.846	50
3.148	100
6.497	150
8.824	200
12.146	250
16.054	300
22.116	350
27.378	400
57.513	450
40.478	500
44.416	550
62.358	600
79.004	750

Les incertitudes croissent avec la profondeur mais également avec la distance au centre de la scène (largeur et hauteur), de façon toutefois moins importante, comme nous le constatons sur les valeurs du tableau suivant ainsi que sur la figure 6.30.

Incertitude sur les coordonnées des segments 3d

Mesures présentées en centimètres.

Erreur maximale			Erreur moyenne			Profondeur
en x (largeur)	en y (hauteur)	en z (profondeur)	en x (largeur)	en y (hauteur)	en z (profondeur)	
0.185	0.183	1.846	0.005	0.076	0.464	50
0.948	1.622	3.148	0.350	0.447	1.264	100
0.636	0.631	6.497	0.090	0.234	1.637	150
3.291	2.806	8.823	1.507	1.076	4.405	200
0.760	2.103	10.378	0.114	0.597	2.608	250
4.565	5.543	16.054	0.674	1.571	6.773	300
7.517	7.205	22.116	2.347	2.371	10.243	350
9.385	9.192	27.377	2.599	2.441	12.611	400
13.063	16.499	57.512	1.294	2.833	15.782	450
9.142	11.476	40.478	2.575	2.593	18.574	500
9.520	12.613	44.414	3.824	3.634	20.030	550
14.402	20.355	62.358	3.792	5.236	15.360	600
22.601	2.441	79.004	6.219	0.826	21.643	750

Les erreurs présentées dans ces deux tableaux sont calculées "sur le terrain". Nous n'avons pas tenter de déterminer l'équation qui évalue ces erreurs en fonction des matrices de transformation perspective liées à chaque caméra. Les erreurs sont calculées de la façon suivante :

- Pour chaque couple de segments, nous déterminons le segment 3d appelé "segment de référence" qui en résulte. Nous calculons ensuite tous les segments 3d engendrés par les deux segments du couple lors d'une variation de 1 pixel de leurs extrémités (figure 6.29). Les écarts maximum et moyens sont ensuite calculés par différence des segments 3d avec le "segment de référence".

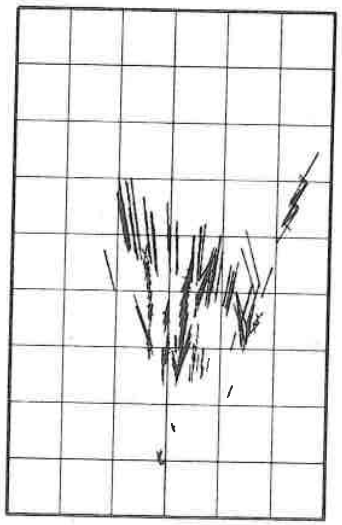

Figure 6.29: *"Vue de dessus" des segments 3d avec leur zone d'incertitude. L'incertitude est proportionnelle à la distance des segments aux caméras ; largeur : 6 mètres, profondeur : 9 mètres.*

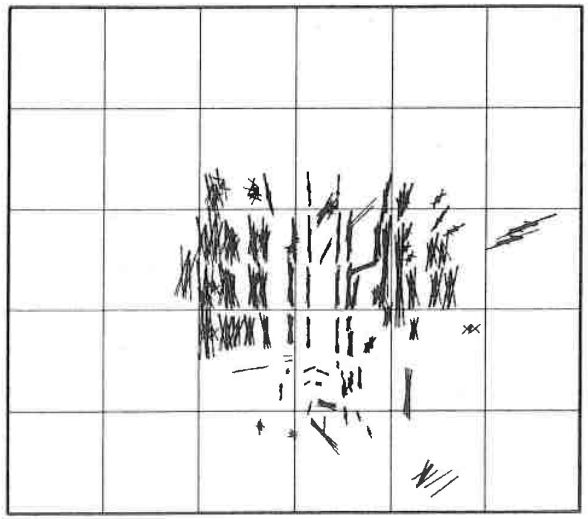

Figure 6.30: *"Vue de face" des segments 3d avec leur zone d'incertitude. L'incertitude est proportionnelle à la distance des segments au centre de la scène ; largeur : 6 mètres, hauteur : 5 mètres.*

4 Construction des surfaces tridimensionnelles

4.1 Position du problème

Les données dont nous disposons sont les suivantes : un ensemble de segments 3d dans la scène. Ces segments sont situés sur la frontière des surfaces que nous avons à construire. Pour chaque région, nous connaissons l'ensemble des segments 3d qui forme son contour dans l'espace. L'idée la plus simple est de construire les plans qui passent au mieux par ces ensembles de segments [WrobelDautcourt 87].

C'est l'approche suivie par S. Kuroe [Kuroe 85], méthode que nous avons déjà brièvement évoquée. Mais les images stéréoscopiques avec lesquelles nous travaillons sont issues de scènes naturelles beaucoup plus complexes, lesquelles posent essentiellement deux problèmes :

- qualité imparfaite des partitions d'images en régions,
- non appartenance de certains segments 3d au contour réel de la région.

Ce dernier point prend en compte la possibilité d'observer des surfaces cachantes : lorsqu'une surface recouvre partiellement l'une de ses voisines et se trouve donc en avant de la scène, les segments qui approchent la frontière située entre ces deux régions appartiennent seulement à la surface cachante. Ce cas, très fréquent dans les images de scènes d'intérieur, est illustré figure 6.31.

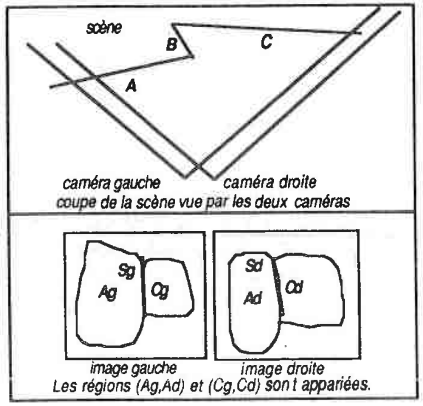

Figure 6.31: *Le segment 3d S calculé grâce à l'appariement des segments S_g et S_d représente une approximation de la frontière de la surface A de la scène, mais non de la surface C. Le segment construit dans la scène appartient à la surface cachante.*

La méthode proposée par S. Kuroe ne prend pas en compte de genre de situation :

aucune surface observée n'en cache une autre. Le mur du fond de la scène et le sol de la scène qui eux sont partiellement cachés, ne sont pas construit.

4.2 Méthode de construction

Nous nous proposons de résoudre le problème des surfaces cachantes en commençant la construction des surfaces par celle se trouvant la plus proche des caméras, soit donc la plus en avant de la scène.

Dans la majorité des cas, la surface "en avant" est entièrement vue, et tous les segments qui en représentent la frontière lui appartiennent. Sa construction doit se faire facilement. Ensuite, il s'agit de traiter les surfaces selon l'ordre croissant de la profondeur de leur contour. Nous devons respecter le fait que certains segments représentant la frontière des surfaces en cours de création, peuvent ne pas leur appartenir. En règle générale, si la construction d'un plan s'avère particulièrement difficile; il faut éliminer en premier lieu les segments sur lesquels il s'appuie et qui appartiennent déjà au contour d'une surface précédemment construite.

Algorithme :

Ce traitement s'écrit sous la forme algorithmique suivante :

- *construire $\mathcal{S}$, l'ensemble des surfaces à déterminer ; à chaque surface est attachée la liste des segments 3d qui en approchent le contour 3d dans la scène*
- *caractériser chaque surface par la profondeur de leur segment de frontière 3d le plus éloigné des caméras*
- *<u>tantque</u> $\mathcal{S}$ est non vide <u>faire</u>*
  - *extraire de $\mathcal{S}$ la surface dont le segment caractéristique est le plus proche des caméras*
  - *construire $\mathcal{F}$, l'ensemble des segments de frontière 3d de cette surface*
  - *essayer de construire la surface plane s'appuyant sur l'ensemble des segments de $\mathcal{F}$*
  - *<u>tantque</u> la surface construite n'est pas satisfaisante <u>faire</u>*
    - *modifier l'ensemble $\mathcal{F}$*
    - *essayer de construire la surface plane s'appuyant sur le nouvel ensemble des segments $\mathcal{F}$*

Construction d'une surface plane s'appuyant sur un ensemble de segments 3d :

Nous construisons le plan qui passe au mieux par l'ensemble des segments 3d par un méthode d'approximation au sens des moindres carrés. Cette approximation est effectuée sur les extrémités des segments, et non, par exemple, sur tous les points des segments. On calcule le plan qui passe au mieux par le nuage de points ainsi défini.

La surface plane recherchée appartient à ce plan et est bordée par les projections des segments sur ce plan (figure 6.32).

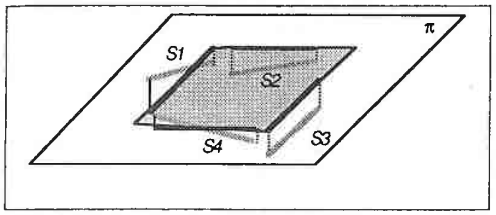

Figure 6.32: *Le plan π qui s'appuie sur les segments S_1, S_2, S_3 et S_4 est construit par une méthode d'approximation au sens des moindres carrés. La surface plane recherchée appartient au plan π et est délimitée par les projections des segments sur π.*

La méthode de construction d'une surface 3d ne tient pas compte de la qualité des mesures des coordonnées tridimensionnelles, elles ont toutes le même coefficient de pondération. Un segment mal placé perturbe alors fortement la position de la surface 3d calculée.

Modification de l'ensemble des segments :

Une surface 3d construite trop "éloignée" d'un ou plusieurs segments peut être considérée comme non satisfaisante. Une telle situation peut survenir pour l'une des trois causes suivantes :

(a) Un ou plusieurs segments 3d sont trop bruités et faussent ou empêchent la construction correcte de la surface.

(b) Un ou plusieurs segments 3d de la frontière de la surface ne lui appartiennent pas.

(c) La surface à construire n'est pas plane. Ceci se produit essentiellement quand le processus de segmentation est mis en échec : une frontière n'a pas été détectée, une seule région représente en fait la projection d'une ou plusieurs surfaces 3d planes.

Ces cas d'échecs entraînent généralement la modification de l'ensemble $\mathcal{F}$ des segments 3d de chaque région. Une stratégie différente doit être adoptée pour chacun des trois cas **(a)**, **(b)** et **(c)** :

(a) Eliminer de l'ensemble $\mathcal{F}$ les segments mal placés. L'étude et la modélisation des incertitudes sur les segments indiquent les segments qui ont la probabilité la plus forte d'être mal placés ; ceci peut être généralisé par l'élimination en premier lieu des segments les plus éloignés des caméras. Cette méthode est malheureusement assez faible.

(b) Eliminer de l'ensemble $\mathcal{F}$ les segments qui n'appartiennent pas à la surface, c'est-à-dire qui se trouvent à la limite de sa (ses) partie(s) cachée(s).

(c) Découper l'ensemble $\mathcal{F}$ des segments 3d en autant d'ensemble $\mathcal{F}_i$ que de surfaces planes possibles, tel que :

$$\mathcal{F} = \bigcup_{i=1}^{n} \mathcal{F}_i$$

pour ensuite traiter séparément chacune des nouvelles régions ainsi créées.

Tout le problème est de trouver dans lequel de ces trois cas l'on se trouve lors de l'échec de la construction d'une surface plane.

De plus, la modification de l'ensemble $\mathcal{F}$ des segments 3d pour le cas **(c)** requiert une interaction avec un processus de segmentation : il faut segmenter à nouveau une zone de l'image en ayant une connaissance supplémentaire. Ce cas n'est donc actuellement pas traité car il demande un retour arrière sur les images originales et l'intégration des processus de segmentation et d'appariement. Ceci fait partie des extensions possibles de la méthode.

Nous avons donc choisi de modifier l'ensemble $\mathcal{F}$ des segments 3d d'une surface à construire selon les cas **(b)** et **(a)** dans cet ordre.

Une étude préliminaire des images stéréoscopiques segmentées en régions permet cependant de construire les ensembles $\mathcal{F}$ minimum : on élimine de chaque ensemble $\mathcal{F}$ d'une surface, les segments dont on est sûr qu'ils n'appartiennent pas cette surface. Ces ensembles $\mathcal{F}$ ne contiennent que les segments dont on ne peut rien dire, c'est-à-dire qui vont servir de support aux essais de construction des surfaces.

Les résultats d'une telle étude sont présentés ci-après.

Règles d'appartenance des segments aux surfaces qu'ils délimitent :

Les contraintes géométriques qui lient les deux images d'un couple stéréoscopique nous ont permis de construire les trois règles suivantes :

Règle 1 :

Soit S_g un segment dans l'image gauche situé entre deux régions A_g et B_g. Soit S_d un segment dans l'image droite situé entre deux régions A_d et B_d.

Si l'appariement des régions fournit les deux couples de régions homologues suivants :

$$(A_g, A_d) \text{ et } (B_g, B_d)$$

alors le segment tridimensionnel S construit dans la scène a les trois possibilités :

- S appartient aux deux surfaces A et B, cas (a) figure 6.33 ;

- S appartient à la surface A seulement, cas (b) figure 6.33 ;
- S appartient à la surface B seulement, cas (c) figure 6.33.

Dans ce cas, rien ne nous permet pas de conclure sur l'appartenance de S aux régions qu'il délimite.

Règle 2 :

Soit S_g un segment dans l'image gauche situé entre deux régions A_g et B. Soit S_d un segment dans l'image droite situé entre deux régions A_d et C.

Si l'appariement des régions fournit le seul couple de régions :

$$(A_g, A_d)$$

alors le segment tridimensionnel S construit dans la scène a les deux possibilités :

- S appartient aux deux surfaces A et C, cas (a) figure 6.34 ;
- S appartient à la surface A seulement, cas (b) figure 6.34.

Ceci nous permet alors d'affirmer que le segment S appartient à la région A. Cette surface est en avant de la scène, elle cache la surface B et éventuellement la surface C.

Règle 2' : Le symétrique de la règle précédente.

Règle 3 :

Soit S_g un segment dans l'image gauche situé entre deux régions A et B. Soit S_d un segment dans l'image droite situé entre deux régions C et D.

Si aucun homologue n'a été trouvé pour ces quatre régions, alors le segment tridimensionnel S construit dans la scène a une seule possibilité :

- S appartient aux deux surfaces A et D, cas (a) figure 6.35.

On s'aperçoit que ce sont les régions restées sans homologue qui nous permettent de tirer des conclusions quant à l'appartenance des segments 3d aux surfaces qu'ils délimitent. Ces régions, vues par une seule caméra, permettent de déterminer les surfaces cachantes.

L'inconvénient est que peu de régions entrent dans cette catégorie avec les scènes que nous traitons. Dans 80 % des cas environ, nous ne pouvons affirmer l'appartenance ou la non-appartenance des segments aux surfaces.

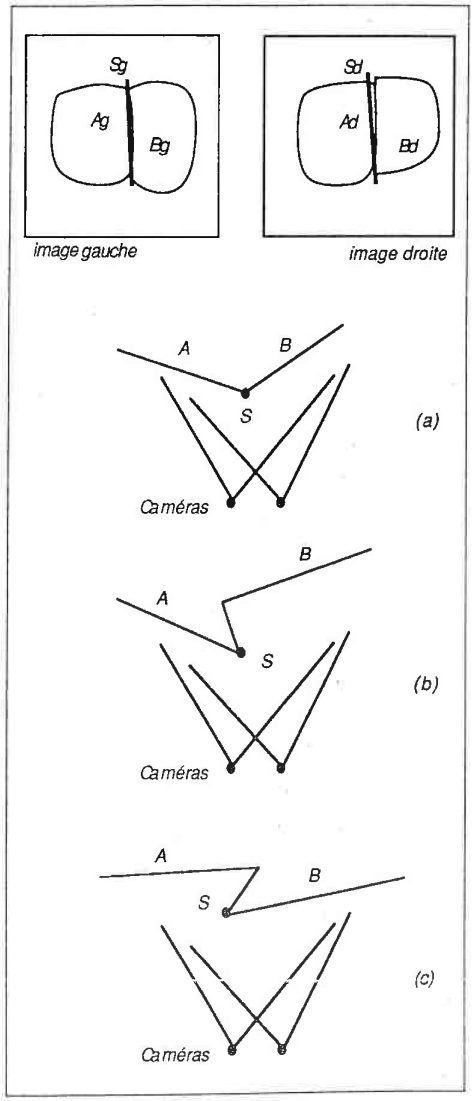

Figure 6.33: *Deux couples de régions homologues : (A_g, A_d) et (B_g, B_d).* **Cas (a)** *: $S \in A, S \in B$.* **Cas (b)** *: $S \in A$ seulement.* **Cas (c)** *: $S \in B$ seulement.*

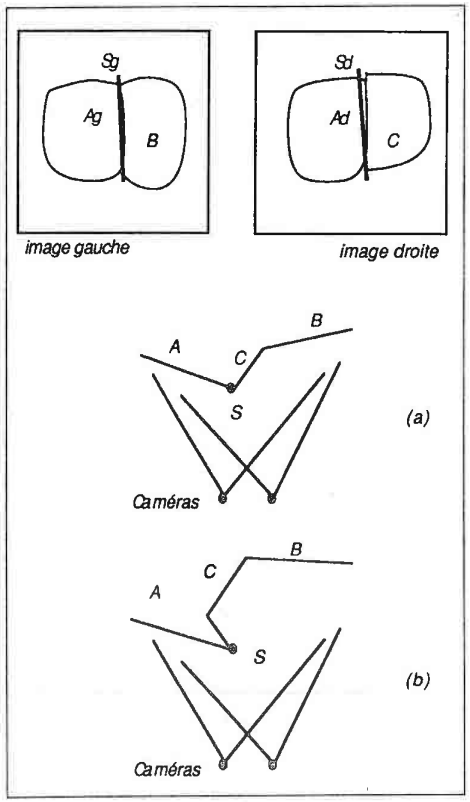

Figure 6.34: *Un seul couple de régions homologues :* (A_g, A_d). **Cas (a)** : $S \in A, S \in C$. **Cas (b)** : $S \in A$ *seulement.*

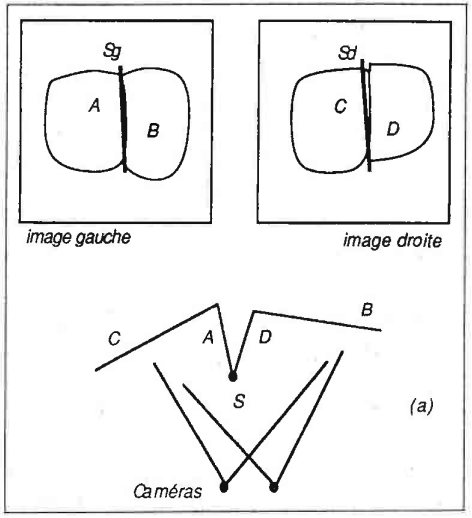

Figure 6.35: *Aucun couple de régions homologues.* **Cas (a)** : $S \in A, S \in D$.

Méthode :

Nous pouvons à présent détailler l'algorithme décrit précédemment :

- *construire $\mathcal{S}$ l'ensemble des surfaces à déterminer. A chaque surface est attachée la liste des segments 3d qui approche son contour dans la scène*
- *<u>pour</u> chaque surface de l'ensemble $\mathcal{S}$ <u>faire</u>*
  - *utiliser les quatre règles pré-citées pour :*
    - *enlever les segments qui n'appartiennent pas à cette surface*
    - *marquer les segments qui doivent appartenir à cette surface*
  - *caractériser la surface par la profondeur de son segment frontière 3d le plus éloigné des caméras*
- *<u>tantque</u> l'ensemble $\mathcal{S}$ est non vide <u>faire</u>*
  - *extraire de $\mathcal{S}$ la surface dont le segment caractéristique est le plus proche des caméras*
  - *construire $\mathcal{F}$, l'ensemble des segments frontières tridimensionnels de cette surface*

- *essayer de construire la surface plane s'appuyant sur l'ensemble des segments de $\mathcal{F}$*
- *tantque la surface construite n'est pas satisfaisante faire*
  - *choisir dans $\mathcal{F}$ un segment à retirer :*
    - *soit en accord avec les règles pré-citées*
    - *soit celui le plus éloigné des caméras, car généralement le plus incertain*
  - *essayer de construire la surface plane s'appuyant sur le nouvel ensemble des segments $\mathcal{F}$*

Ces règles d'appartenance ne nous permettent pas de construire correctement les surfaces visualisées par "un trou" d'une autre surface (figure 6.36).

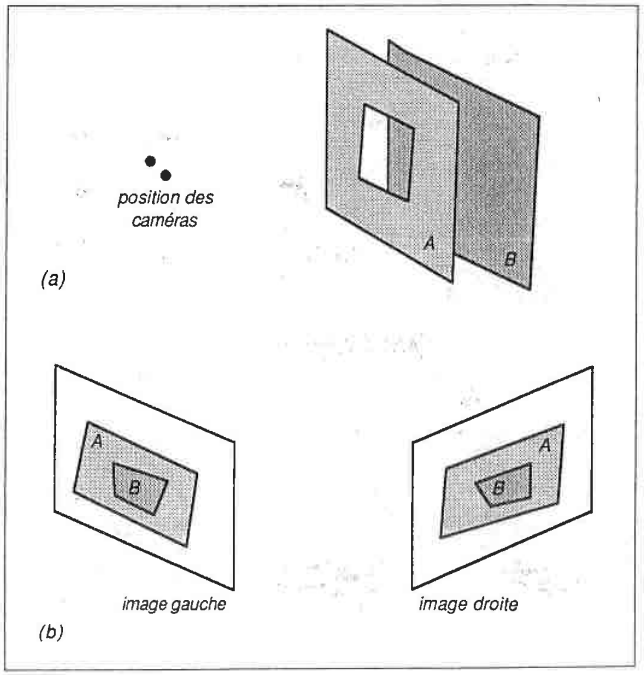

Figure 6.36: **(a)** : *La surface B est visualisée par un "trou" dans la surface A. Les deux surfaces sont construites sur le même plan 3d, aucune indication ne permet de construire la surface B en arrière plan de la surface A.* **(b)** : *Image gauche et image droite de la scène.*

5 Résultats et conclusion

La construction des surfaces tridimensionnelles est une phase délicate et n'est que partiellement résolue. Nous avons seulement mis en œuvre la construction des surfaces sans tenir compte ni des surfaces cachantes, ni des incertitudes de localisation des segments, mais les résultats sont néanmoins encourageants.

Le schéma récapitulatif des traitements effectués pour cette troisième partie de notre module de stéréovision est présenté figure 6.37.

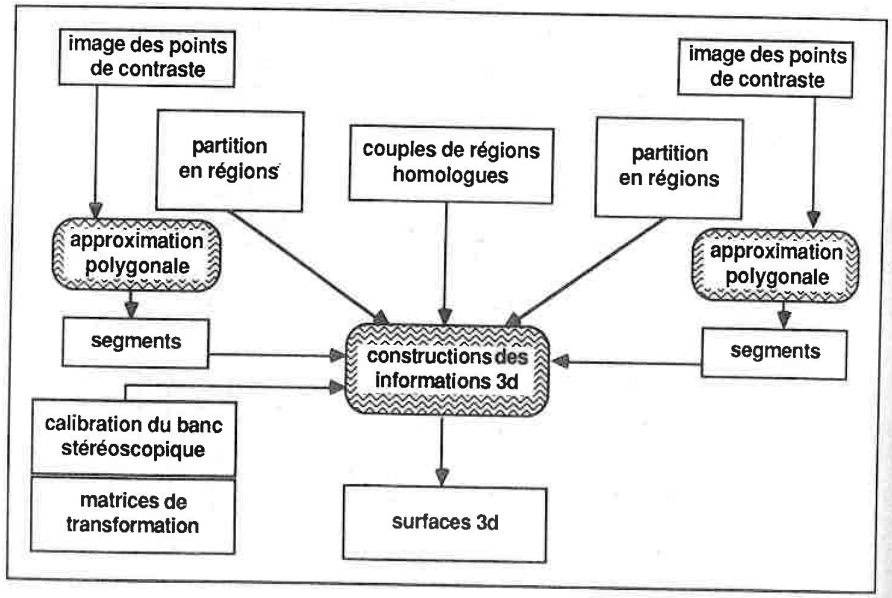

Figure 6.37: *Schéma récapitulatif des traitements effectués pour la construction des surfaces 3d à partir des couples de régions homologues.*

Cette phase de construction des surfaces à partir des ensembles de segments 3d a été développée par Y. Belaïd dans notre laboratoire et nous présentons ci-après les résultats obtenus sur la scène *bureau1*.

La figure 6.38 représente les différentes surfaces 3d construites. La scène est visualisée de face et les différentes régions sont colorées en fonction de leur distance au robot. L'erreur autorisée par la méthode d'approximation au sens des moindres carrés est faible, aussi, seules les régions bordées par un faible nombre de segments sont construites ; pour la scène traitée, ce sont essentiellement les carreaux des fenêtres.

Si on augmente l'erreur autorisée pour la construction, on obtient les régions

bordées par un grand nombre de segments comme les montants des fenêtres (figure 6.39). Les montants des fenêtres et les carreaux sont des surfaces appartenant à un même plan, donc de même niveau de gris et visuellement confondus sur l'image présentée.

La figure 6.40 représente la "projection sur le sol" des plans 3d ainsi construits : les segments 3d sont modifiés de façon à appartenir au plan qu'ils ont servis à construire, et nous présentons une vue aérienne de ces "nouveaux" segments.

D'autres résultats sont présentés en Annexe D.

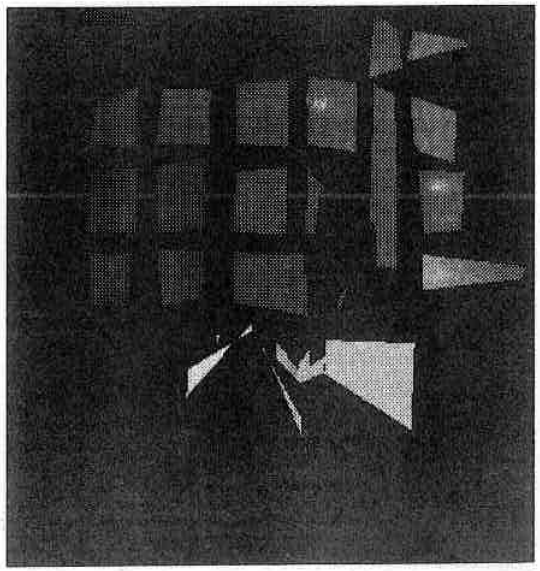

Figure 6.38: *Construction des surfaces tridimensionnelles de la scène avec une faible erreur pour la méthode d'approximation au sens des moindres carrés. Les surfaces 3d sont colorées en fonction de la distance au robot. L'intensité lumineuse croît (du blanc au noir) avec la distance de la fenêtre au robot.*

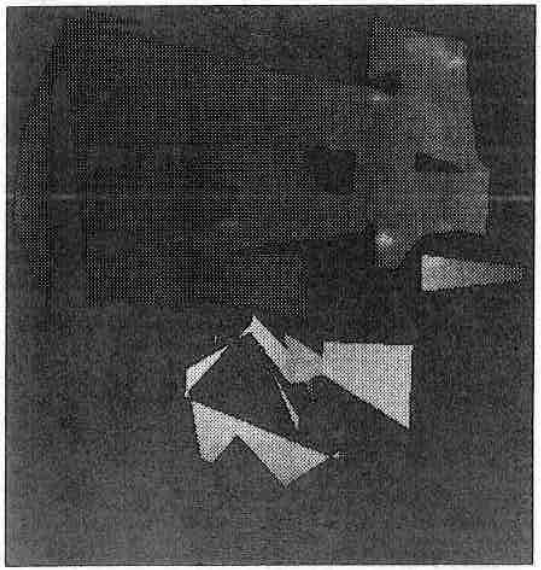

Figure 6.39: *Construction de toutes les surfaces tridimensionnelles de la scène. L'erreur autorisée pour la méthode d'approximation au sens des moindres carrés est plus élevée que pour la construction obtenue pour la figure précédente. Les surfaces 3d sont colorées en fonction de la distance au robot, les montants des fenêtres et les carreaux, appartenant au même plan, sont donc colorés de manière identique.*

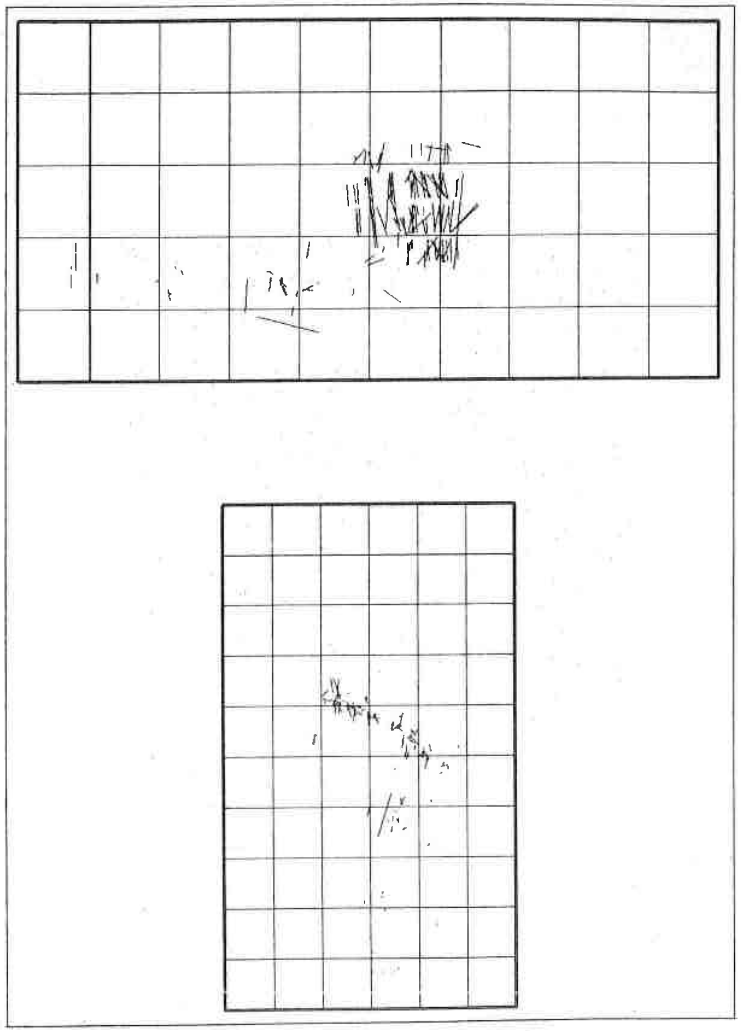

Figure 6.40: *Vue des segments tridimensionnels "corrigés".* **haut :** *Vue de "côté".* **bas :** *Vue de "dessus", projection des segments sur le sol de la pièce. Echelle de représentation de la scène : largeur 6 mètres, hauteur : 5 mètres, profondeur : 10 mètres ; taille du quadrillage : 1 m × 1 m.*

Lors de l'observation de ces résultats, on constate que les surfaces sont construites avec une assez bonne précision. Cependant certaines régions très importantes sont absentes, comme les bureaux, le mur de gauche, et ceci pour essentiellement deux raisons :

- perte de certains segments,
- problème de segmentation en régions.

Perte de certains segments :

Certaines régions ne sont pas bordées par suffisamment de segments pour être construites :

- les régions "coupées" par les bords de l'image (comme le mur de gauche de la scène traitée) ont une frontière incomplète ;
- la perte des segments très bruités comme les segments parallèles aux épipolaires (segments horizontaux pour la scène traitée) réduit de moitié environ le nombre de segments par régions.

Dans ce dernier cas, la stratégie pour augmenter le nombre de segments consiste à placer les caméras sur un support oblique. Le banc stéréoscopique utilisé pour capter la paire nommée *pièce*, et présentée en Annexe D, a cette caractéristique : les segments verticaux et horizontaux sont en majeure partie conservés.

L'incertitude de localisation des segments tridimensionnels est également suffisamment importante pour perburber la qualité des résultats. Cependant, l'étape de calibration des caméras permet de connaître la matrice de covariance sur les coefficients des matrices de transformation perspective ; il est alors possible de prédire l'erreur de construction 3d en utilisant le filtre de Kalman [Toscani 87a]. La modélisation de cette incertitude peut alors être intégrée à la construction des surfaces 3d.

Une difficulté apparaît également pour retrouver les segments qui se situent sur la frontière des régions : le problème majeur est de diviser les segments qui bordent plusieurs régions ; ce problème, assez complexe, n'est que partiellement résolu.

Mauvaise segmentation en régions :

L'absence des deux bureaux de gauche de la scène traitée illustre le second cas : l'origine de la non-construction tridimensionnelle de ces régions provient de leur non-appariement, lui-même du à une différence de segmentation des images. Cette différence n'a pas été réduite par fusions de régions. La nécessité de mettre en œuvre un processus plus complet et performant apparaît alors.

Un axe de recherche est de concevoir une coopération entre le processus de construction et ceux qui le précédent, à savoir : la segmentation et l'appariement des régions puis des segments. Ceci peut être implanté, par exemple, si la scène 3d construite est trop incomplète, ou si on refuse de conserver une surface construite avec

un trop grand écart entre le plan qui la porte et ses segments de contours[4]. Une solution consiste à demander à nouveau la segmentation sur une portion d'image pour laquelle nous avons à présent une plus grande connaissance : nombre approximatif de régions à trouver (en fonction du nombre de plans issus des segments 3d), leur localisation dans la zone de l'image, ... Il suffit ensuite d'enchaîner à nouveau les traitements d'appariement.

Cette coopération est exposée de façon simpliste, son implantation n'est pas immédiate. Ce point de coopération est abordé dans la conclusion générale.

[4] Ce cas se produit quand la surface 3d construite s'appuie sur un ensemble de segments non planaires.

Chapitre 7

CONCLUSION

1 Stéréovision par appariement de régions

Des deux grandes approches utilisées pour résoudre le problème de la stéréovision :

- appariement de points, calcul des points 3d, puis recherche dans cette scène de plans, surfaces courbes, ... ;
- appariement de structures (segments, régions, ...) puis calcul de ces primitives en trois dimensions ;

nous avons choisi de mettre en œuvre une méthode du second type ; cette méthode est fondée sur l'appariement des régions puisque ces primitives contiennent une information de surface nécessaire à une construction la plus dense possible de la scène.

Cependant, ce type d'approche comporte généralement les trois limitations suivantes : extraction difficile des régions, mauvaise localisation des bords des régions et temps de calcul important. Nous discutons de ces trois points ci-après, en fonction de l'application réalisée.

Extraction des régions :

La région n'est pas une entité stable et il est plus facile de détecter des discontinuités (points de contraste) que des zones homogènes dans une image.

Nous obtenons cependant des partitions correctes d'une part en guidant la segmentation en régions par une carte de contours préalablement détectée et d'autre part en ménageant une étroite **interaction** entre les deux processus d'extraction et d'appariement. L'idée de cette interaction est d'utiliser une redondance d'information : deux images de la même scène pour poursuivre le processus de segmentation. Les résultats sont d'une part un appariement correct des régions de chacune des images et d'autre part deux partitions améliorées des images initiales.

Localisation des régions :

Il est exact que les points de contraste sont bien localisés sur les images, avec un précision théorique inférieure au pixel, alors que les bords des régions sont légèrement décalés du fait des différents traitements appliqués aux images.

L'idée retenue est d'utiliser des **segments bien placés** pour effectuer la construction des informations tridimensionnelles : les segments choisis sont ceux qui approchent les lignes de points de contraste en remplacement des segments qui approchent le bord des régions. La précision des résultats est nettement améliorée.

Temps de calcul :

Les temps de calculs sont actuellement importants : 90 minutes environ de traitement sur SM90 pour obtenir les segments tridimensionnels à partir des deux vues initiales. La méthode actuellement implantée est longue car composée de nombreux traitements.

Il convient cependant d'évaluer la complexité des algorithmes mis en œuvre, plutôt que de chronométrer les temps d'exécution qui dépendent pour une grande part du matériel utilisé. Cette complexité est en :

Segmentation :

$$\mathcal{O}(F \times M \times log(M \times N))$$

avec F : nombre de fusions réalisées ;
M : nombre de régions voisines d'une région susceptibles de lui être fusionnées ;
N : nombre total de régions dans la partition.

Appariement des régions :

- Phase de génération des hypothèses (recherche d'un chemin optimal dans un graphe, programmation dynamique) :

$$\mathcal{O}(nbe \times n_e^2)$$

avec nbe : nombre de couples de lignes épipolaires associées (environ 256 pour les couples d'images où les épipolaires sont quasi-horizontales) ;
n_e : nombre moyen de points frontières sur une ligne épipolaire.

- Phase de validation des hypothèses :

$$\mathcal{O}(N \times nbh)$$

avec N : nombre de régions dans une image ;
nbh : nombre moyen d'hypothèses d'appariement différentes concernant une région.

Appariement des segments :

$$\mathcal{O}(nbe * n_e)$$

avec nbe : nombre de couples de lignes épipolaires associées ;
n_e : nombre moyen de segments intersectant une ligne épipolaire, ces segments approchent les lignes de contraste, n_e représente également le nombre moyen de points de contraste situés sur une ligne épipolaire.

Construction des surfaces tridimensionnelles :

Ce traitement comporte deux étapes de complexités différentes :

- la construction des facettes dans la scène, chaque facette est représentée par les coordonnées tridimensionnelles de ces sommets :

$$\mathcal{O}(N)$$

avec N : nombre de régions.

- la création des points tridimensionnels à partir de ces facettes :

$$\mathcal{O}(np)$$

avec np : nombre de points.

La complexité du traitement général est donc assez difficile à évaluer car définie pour chaque processus en fonction d'objets différents : régions, points contours, segments, ...

Nous pouvons cependant répartir les différents traitements en deux catégories en fonction de leur complexité :

Traitements de forte complexité : La programmation dynamique, employée à deux étapes différentes de la chaîne de traitements est très coûteuse. Rappelons que la programmation dynamique est utilisée pour l'appariement des points frontières des régions pour engendrer les hypothèses d'appariement des régions, et pour l'appariement des segments avant leur construction tridimensionnelle dans la scène.

Il faut cependant remarquer que la programmation dynamique est immédiatement **parallélisable** : l'appariement de chaque couple de lignes épipolaires est entièrement indépendant des appariements des autres couples de lignes.

Traitements de faible complexité : Les autres traitements mis en œuvre sont de complexité "raisonnable", au plus en $\mathcal{O}(n\,log(n))$ où n représente le nombre de régions. Ces traitements sont cependant entièrement **séquentiels**.

L'utilisation d'une machine à architecture parallèle s'avère dans ce cas quasi-inutile.

On peut cependant remarquer que les traitements indépendants effectués sur chaque image du couple stéréoscopique peuvent s'exécuter de façon parallèle ; citons dans ce cas :

. *extraction des points de contraste*
. *segmentation en régions*
. *approximation polygonale des lignes de contraste*

De plus, la détermination de la partition initiale par divisions récursives, utilisée lors de la segmentation d'images, peut être typiquement implantée sur une machine multi-processeurs avec une architecture pyramidale [?].

Il est également possible, pour le problème du temps d'exécution, de considérer les deux cas de traitement suivants :

- si le robot est immobile et est attelé par exemple à une tâche de localisation et/ou de reconnaissance, la rapidité de l'acquisition de l'information 3d est importante mais non fondamentale ;
- en revanche, si le robot se déplace, nous pouvons envisager d'utiliser les segmentations et les appariements de la paire d'images stéréoscopiques précédente pour guider le traitement de la paire d'images courante et ainsi réduire le temps de calcul.

Pour conclure, nous pouvons cependant imaginer de câbler la version définitive et figée de nos algorithmes afin de répondre aux exigences demandées par le temps réel. Mais, dans sa version actuelle, notre méthode doit plutôt être considérée comme :

- une initialisation de la base de données sur laquelle travaillent d'autres processus, correspondant au traitement de la première paire stéréoscopique lors du déplacement du robot ;
- ou en complément d'autres traitements quand ceux-ci rencontrent des problèmes, par exemple pour aider un processus travaillant exclusivement à l'appariement des segments et confronté à un problème d'ambiguïté.

Programmes :

Le schéma figure 7.1 représente tous les traitements participants au module de stéréovision réalisé.

A titre d'information, mentionnons que tous les programmes sont écrits en langage C ; un tel langage est bien adapté aux traitements de relativement bas-niveau effectués.

Les différentes caractéristiques du module de stéréovision sont réparties comme suit, les temps sont donnés pour les deux ordinateurs utilisés : Vax[1] et Sm90[2].

Caractéristiques moyennes des différents traitements intervenants dans le module de stéréovision

	nombre de lignes de programme en langage C	taille du module exécutable (octets)	temps moyen de calcul en secondes
Segmentation en régions : 2 images	2 000	86 000	405[1]
Appariement :			
Génération des hypothèses	4 000	1 788 500	1 556[2]
Validation des hypothèses	2 300	76 100	938.5[2]
Construction 3d :			
Segments 3d	2 900	186 700	1 985.5[2]
Surfaces 3d			
. construction des facettes 3d	700	40 000	0.5[1]
. construction des points 3d	2 200	35 000	540.5[1]

Remarquons pour terminer que les images traitées sont relativement simples : elles comportent beaucoup de structures linéaires, dont l'extraction est plus stable, et peu de surfaces fortement texturées. Le traitement appliqué directement sur des scènes d'extérieur comportant des arbres, par exemple, poserait quelques problèmes, si les arbres sont assez proches pour que l'on perçoive leur texture.

[1] Vax 785 sous UNIX 4.3 BSD
[2] Sm90 sous SMX V.1, 3 Mo de mémoire centrale, sans processeur arithmétique

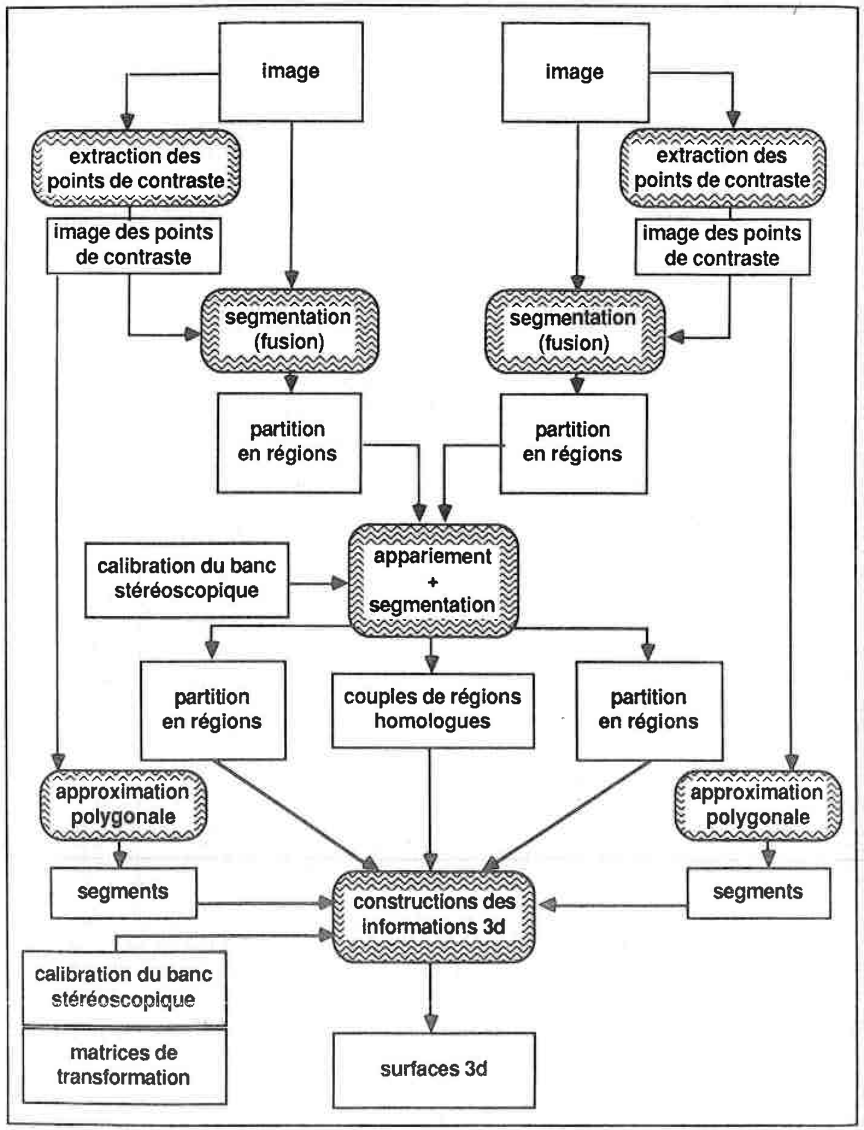

Figure 7.1: *Schéma de tous les traitements nécessaires au module de stéréovision.*

2 Points importants

L'application de stéréovision à laquelle nous nous sommes intéressés a pour données des images qualifiées généralement de difficiles et a un but bien ambitieux : trouver l'information 3d la plus complète et précise possible. Le développement de ce mémoire présentant la méthode implantée pour tendre vers ce but met en évidence deux idées maîtresses : l'utilisation de plusieurs sources d'informations et l'interaction entre différents processus.

Utilisation de plusieurs sources d'informations :

Notre méthode de stéréovision par appariement de régions utilise les contours, information duale des régions, en trois endroits :

- collaboration à la segmentation en régions,
- génération des hypothèses d'appariement des régions,
- point de passage indispensable pour calculer les surfaces 3d.

Plus généralement, différentes sources d'information sont utilisées lors des divers traitements :

Segmentation d'images :

- Nous utilisons conjointement deux informations duales issues des caractéristiques de l'image :

  . discontinuités locales : points de contraste,
  . homogénéité des groupes de points adjacents : régions.

- La connaissance de deux partitions issues indépendamment des images d'un couple stéréoscopique permet de les améliorer. Rappelons que chaque processus de segmentation est arrêté précocement avant de faire des fusions intempestives de régions par manque d'informations pertinentes.

Appariement des régions : tous les attributs permettant de caractériser les régions sont évalués, qu'ils soient :

. photométriques,
. géométriques,
. morphologiques,
. topographiques.

Construction des informations 3d : les segments de contours approchant les lignes de points de contraste s'avèrent être un point de passage obligé pour la construction des surfaces 3d. Une étude plus pertinente des caractéristiques photométriques et morphologiques des régions devrait permettre une meilleure construction.

Interaction entre les traitements :

Lors du développement de cette application de stéréovision, le schéma classique : extraction des primitives, appariement, puis construction 3d, s'est avéré insuffisant. Notre solution a été de permettre une interaction entre les processus :

Interaction (extraction - appariement des régions) : Cette interaction est actuellement limitée à la poursuite de la segmentation par le processus d'appariement, mais c'est un premier pas vers une plus étroite collaboration.

Interaction (segmentation - calcul 3d) et/ou interaction (appariement - calcul 3d) : Les fondements d'une telle interaction se dégagent lors de la mise en œuvre du processus de construction des informations tridimensionnelles, mais nous ne l'avons pas implantée.

Notre application de stéréovision mêle intimement les techniques de Reconnaissance des Formes et d'Intelligence Artificielle et prouvent, une fois de plus, si cela était nécessaire, les liens étroits entre ces deux disciplines dès que nous travaillons sur un système de vision "élaboré".

3 Perspectives

Nous pouvons dégager les perspectives à deux niveaux, l'un concernant l'aspect "stéréovision", le second, plus général, pour l'aspect "système de vision".

Stéréovision : L'approche par régions est bonne à condition de trouver des caractéristiques pertinentes pour ces primitives. Il apparaît intéressant de poursuivre des efforts pour trouver d'autres attributs utilisés par les différents traitements. Par exemple, l'étude des dégradés de l'intensité lumineuse sur les régions ("shape from shading") et l'étude de la forme des contours ("shape from contours") peuvent fournir de précieuses informations quant à la courbure des surfaces.

Système de vision : L'interaction entre les différents processus peut être beaucoup plus étroite que celle mise en œuvre et concerner tous les traitements intervenants dans un système de vision comme celui étudié pour le projet ORASIS. Si une telle démarche est adoptée, il convient alors d'organiser les communications entre processus de façon totalement différente : autoriser les échanges dynamiques et éventuellement avoir un processus de contrôle, une base de connaissance et un moteur d'inférence, ... Les différents modules concernés peuvent être les suivants :

. *extraction des points de contraste*
. *segmentation en régions*
. *appariement des régions*
. *calcul des informations 3d*
. *recalage scène - modèle (pour l'identification)*
. *localisation*
. *navigation*
. *apprentissage*

Chaque module doit pouvoir travailler :

- indépendamment, sans connaissance a priori, sur les données nécessaires à son fonctionnement,
- sur la demande d'un autre processus qui lui fournit d'autres informations.

N'oublions pas un point essentiel concernant le matériel utilisé : le robot est mobile. Ceci signifie qu'un déplacement peut être également demandé pour parfaire l'acquisition d'une partie de la scène. La connaissance des zones incertaines ou inconnues permet de définir une stratégie sensorielle : comment augmenter la précision d'un objet, quel déplacement du robot envisager ?

La vision par ordinateur, discipline dans laquelle s'insère notre étude, est un sujet vaste et passionnant et notre contribution à la réalisation d'un système de stéréovision reste modeste. Résoudre un problème intuitivement simple comme l'acquisition d'une carte de profondeur à partir de deux vues s'avère en fait assez complexe.

La perception de l'environnement par un robot mobile est un problème riche et étendu, nous en proposons une approche, mais de nombreuses autres voies restent ouvertes.

Annexe A

Partition initiale par "quad-tree"

Nous décrivons ici la création de la partition initiale utilisée lors par le processus du segmentation d'image en régions. Cette partition est obtenue par divisions récursives de l'image originale en quatre parties égales jusqu'à ce que chaque zone vérifie un prédicat d'homogénéité P ; l'image est alors décrite sous forme d'un "quad-tree" [Pavlidis 77].

Le niveau (n) d'un quad-tree définit une image de dimension $2^n \times 2^n$ dont chaque élément correspond à quatre éléments du niveau inférieur (n+1). Le niveau (N), tel que $2^N \times 2^N$ soit la dimension de l'image, correspond à tous les points de l'image. La figure A.1 représente une partie de quad-tree associé à une image.

La partition initiale des régions est obtenue en deux étapes :

- Construction d'un quad-tree associé à l'image originale en attachant à chaque élément d'un niveau (n) quelconque les attributs des 2^{N-n} points de l'image auxquels il correspond. Ces attributs sont nécessaires pour l'évaluation du prédicat P. Ce quad-tree est construit par fusions, c'est-à-dire en partant du niveau (N), puis en calculant le niveau (n) en fonction du niveau (n+1).

- Détermination des éléments du quad-tree tels que les points de la portion d'image originale associée vérifient le prédicat P. Ceci est effectué en parcourant le quad-tree du niveau (0) vers le niveau (N) et en mettant à jour deux listes contenant respectivement, lors du passage du niveau (n) au niveau (n+1), les éléments validés des niveaux précédents (c'est-à-dire inférieurs au niveau (n+1)) et les éléments du niveau (n) dont il faut examiner les fils au niveau (n+1).

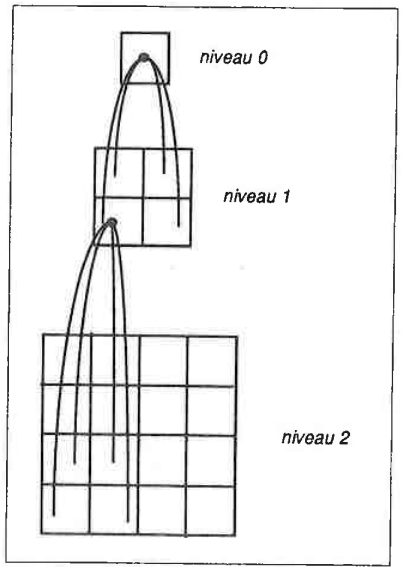

Figure A.1: *Partie d'un "quad-tree" associé à une image. Le niveau (0) se divise en quatre zones pour donner le niveau (1) qui se divise également en quatre pour donner le niveau (2), etc. Le niveau (N) représente tous les points de l'image.*

Annexe B

Exemple de déroulement de la segmentation

Nous présentons dans cette annexe un exemple de déroulement du processus de segmentation en régions, le but est d'illustrer par différents schémas l'évolution des structures de données adaptées à une partition en régions modifiée par des fusions successives.

Soit une partition en régions schématisée figure B.1. Le graphe d'adjacence et l'arbre binaire de recherche de cette partition sont représentés figures B.2 et B.3.

Le prédicat d'homogénéité choisi pour dérouler cet exemple est le suivant :

$$P(R_i \cup R_j) = [\mid Moy_i - Moy_j \mid < Seuil]$$

avec R_i et R_j : les deux régions proposées au fusionnement si le prédicat est vérifié ;
Moy_i et Moy_j : moyennes respectives des niveaux de gris des points des régions R_i et R_j ;
$Seuil$: seuil de fusionnement.

Pour notre exemple, le seuil de fusionnement choisi est 5.

Avec la valeur du seuil choisie, le processus de segmentation modifie successivement la partition et l'arbre binaire de recherche (figures B.4 à B.11) et fournit la partition figure B.10.

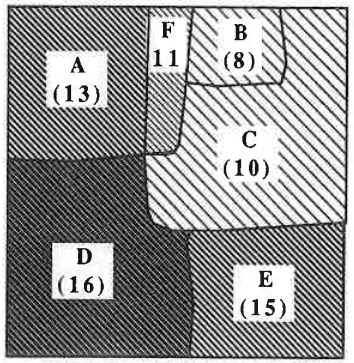

Figure B.1: *La partition à traiter est ici composée de 6 régions : A, B, C, D, E et F. Pour chaque région, nous avons indiqué son niveau de gris moyen.*

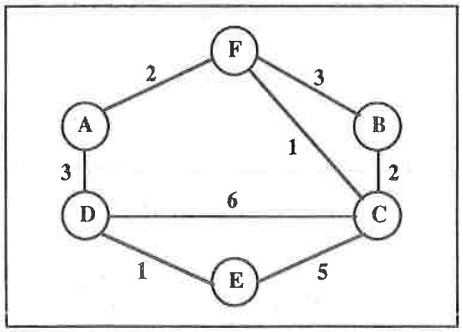

Figure B.2: *Graphe d'adjacence valué associé à la partition de la figure précédente. Les valeurs des arcs représentent les différences des moyennes des niveaux de gris des régions.*

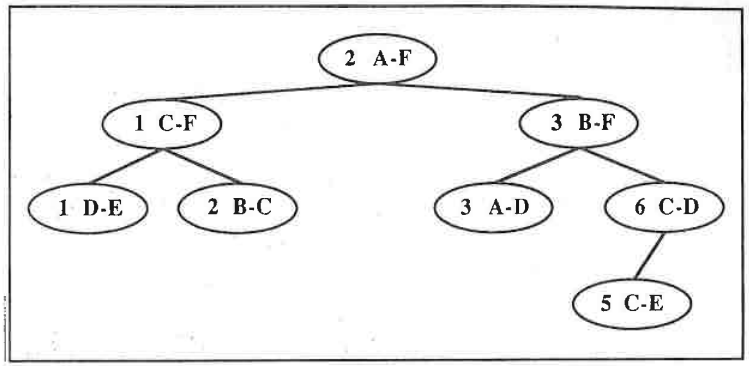

Figure B.3: *Arbre binaire de recherche associé à la partition de la figure précédente. Les valeurs des nœuds représentent les différences des moyennes des niveaux de gris des régions.*

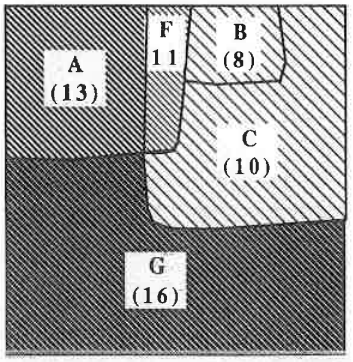

Figure B.4: *Le nœud de plus petite valeur $(1DE)$ extrait de l'arbre binaire de recherche indique la fusion entre les régions D et E pour donner la nouvelle région G.*

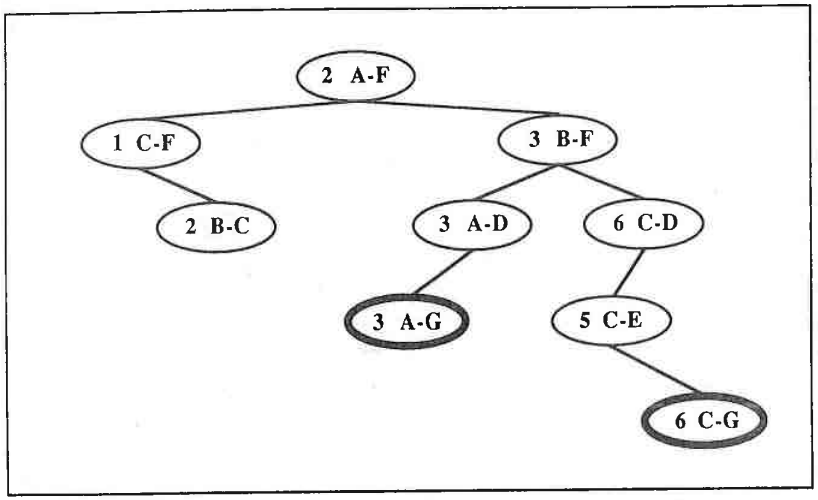

Figure B.5: *Arbre binaire de recherche associé à la partition de la figure précédente. On a ajouté les nœuds concernant la nouvelle région G. Les régions D et E n'existent plus, mais les nœuds les concernant restent dans l'arbre binaire de recherche.*

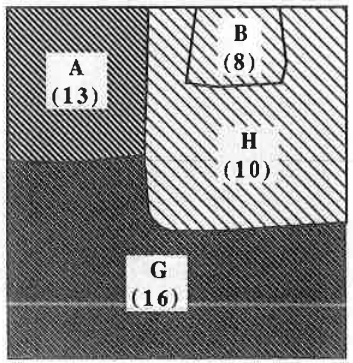

Figure B.6: *Le nœud de plus petite valeur ($1CF$) extrait de l'arbre binaire de recherche indique la fusion entre les régions C et F pour donner la nouvelle région H.*

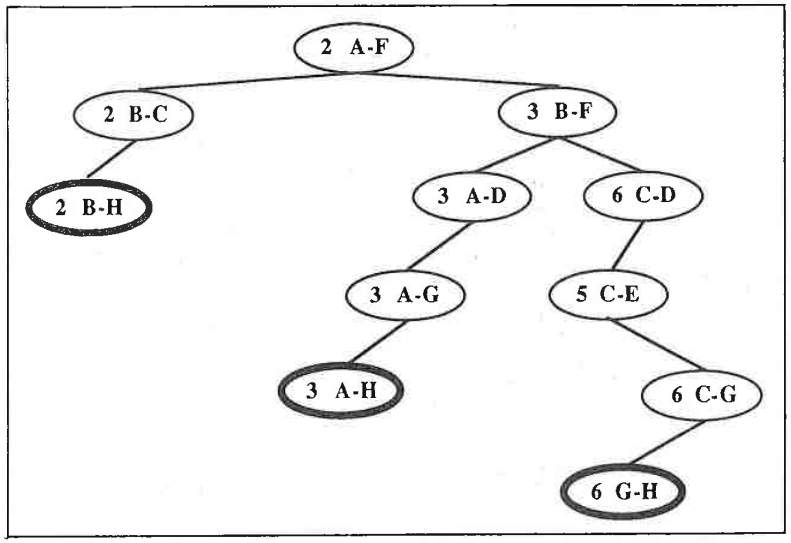

Figure B.7: *Arbre binaire de recherche associé à la partition de la figure précédente. On a ajouté les nœuds concernant la nouvelle région H. Les régions D, E, C et F n'existent plus, mais les nœuds les concernant restent dans l'arbre binaire de recherche.*

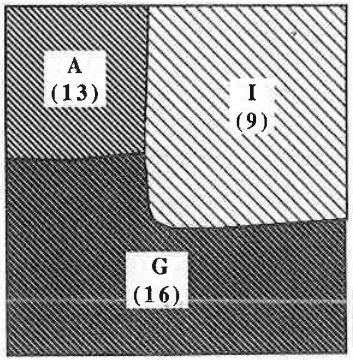

Figure B.8: *Le nœud de plus petite valeur ($1BH$) extrait de l'arbre binaire de recherche indique la fusion entre les régions B et H pour donner la nouvelle région I.*

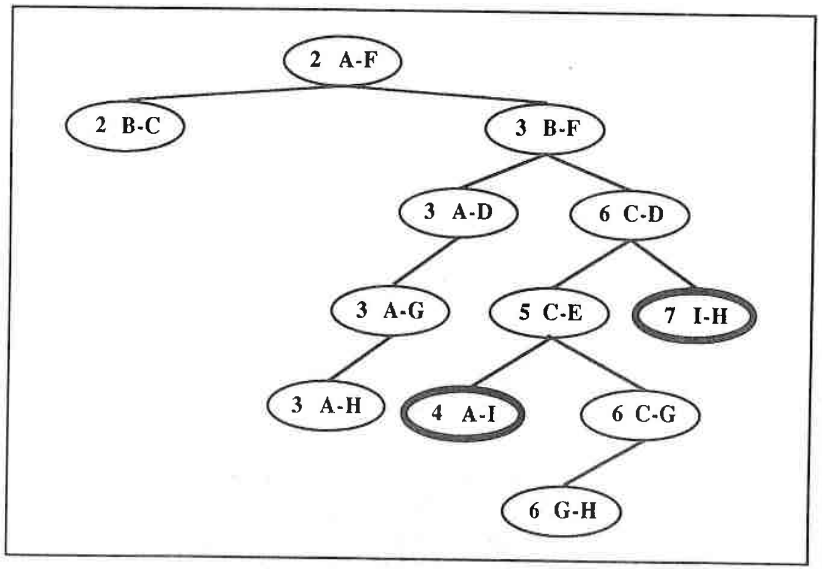

Figure B.9: *Arbre binaire de recherche associé à la partition de la figure précédente. On a ajouté les nœuds concernant la nouvelle région I. Les régions D, E, C, F, B et H n'existent plus, mais les nœuds les concernant restent dans l'arbre binaire de recherche.*

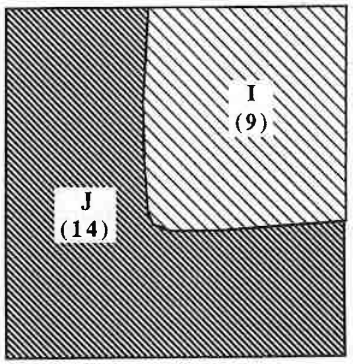

Figure B.10: *Le nœud de plus petite valeur (3AG) extrait de l'arbre binaire de recherche indique la fusion entre les régions A et G pour donner la nouvelle région J.*

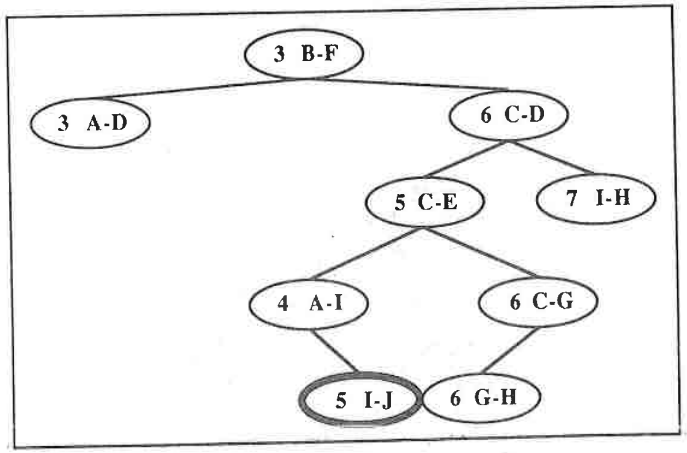

Figure B.11: *Arbre binaire de recherche associé à la partition de la figure précédente. On a ajouté les nœuds concernant la nouvelle région J. Les régions D, E, C, F, B, H, A et G n'existent plus, mais les nœuds les concernant restent dans l'arbre binaire de recherche.*

Annexe C

Calibration du banc stéréoscopique

Nous présentons, dans cette annexe, les différentes matrices de calibration associées aux couples stéréoscopiques d'images des scènes traitées.

1 Matrices de transformation perspective

Chaque caméra est calibrée de façon à pouvoir, connaissant les coordonnées images d'un point, déterminer la demi-droite issue du centre focal de la caméra et passant par ce point image.

1.1 Paires stéréoscopiques *bureau1*, *bureau2* et *couloir*

Caméra gauche :

$$T_g = \begin{pmatrix} 20.88264465 & 0.1500499249 & 6.426446915 & -109.2985382 \\ 0.5257893801 & 20.04711914 & 6.900261879 & -368.3737793 \\ 0.002918525599 & -0.0001615366815 & 0.02940095589 & 1 \end{pmatrix}$$

Caméra droite :

$$T_d = \begin{pmatrix} 18.76075745 & 0.2198942900 & 7.608737946 & -498.8466797 \\ -0.1805693507 & 18.36303711 & 6.694689751 & -307.0510254 \\ 0.00006556042488 & -0.0002896967344 & 0.02740492672 & 1 \end{pmatrix}$$

1.2 Paire stéréoscopique *pièce*

Caméra gauche :

$$T_g = \begin{pmatrix} 2.154522389 & 0.040438956 & 0.934180256 & 113.276973881 \\ -0.023074243 & 3.000277342 & 0.625608288 & 105.550642197 \\ -0.000228769 & 0.000229301 & 0.002865840 & 1 \end{pmatrix}$$

Caméra droite :

$$T_d = \begin{pmatrix} 2.406741147 & 0.109191320 & 0.653840707 & 96.593928594 \\ -0.039225538 & 3.153902807 & 0.798611890 & 112.172803841 \\ 0.000209578 & 0.000139188 & 0.003026424 & 1 \end{pmatrix}$$

2 Relation géométrique entre les caméras

2.1 Paires stéréoscopiques *bureau1*, *bureau2* et *couloir*

Coordonnées des centres épipolaires :

$$M_g = \begin{pmatrix} 10706.21875 & 232.81560 \end{pmatrix} \; et \; M_d = \begin{pmatrix} -21609.62109 & 307.15503 \end{pmatrix}$$

Lignes épipolaires : Soit $P_g(u_g, v_g)$ un point de l'image gauche, son homologue se trouve sur la droite passant par le centre épipolaire droit M_d et de vecteur directeur $\vec{A_d}$ (figure C.1) :

$$\vec{A_d} = \begin{pmatrix} x_d \\ y_d \end{pmatrix} = \begin{pmatrix} \alpha_{d11} + u_g \times \alpha_{d12} + v_g \times \alpha_{d13} \\ \alpha_{d21} + u_g \times \alpha_{d22} + v_g \times \alpha_{d23} \end{pmatrix}$$

avec :

$$\begin{pmatrix} \alpha_{d11} & \alpha_{d12} & \alpha_{d13} \\ \alpha_{d21} & \alpha_{d22} & \alpha_{d23} \end{pmatrix} = \begin{pmatrix} 249330.38 & -23.258224 & -1.4192057 \\ -3291.9807 & 0.068780303 & 10.977427 \end{pmatrix}$$

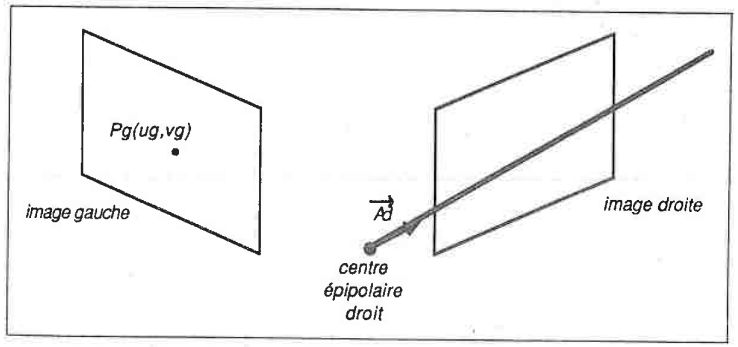

Figure C.1: *Centre épipolaire droit et ligne épipolaire droite associée à un point P_g de l'image gauche.*

Soit $P_d(u_d, v_d)$ un point de l'image droite, son homologue se trouve sur la droite passant par le centre épipolaire gauche M_g et de vecteur directeur $\vec{A_g}$ (figure C.2) :

$$\vec{A_g} = \begin{pmatrix} x_g \\ y_g \end{pmatrix} = \begin{pmatrix} \alpha_{g11} + u_d \times \alpha_{g12} + v_d \times \alpha_{g13} \\ \alpha_{g21} + u_d \times \alpha_{g22} + v_d \times \alpha_{g23} \end{pmatrix}$$

avec :

$$\begin{pmatrix} \alpha_{g11} & \alpha_{g12} & \alpha_{g13} \\ \alpha_{g21} & \alpha_{g22} & \alpha_{g23} \end{pmatrix} = \begin{pmatrix} -105044.88 & -4.8703070 & -0.62960821 \\ -2509.8982 & 0.030500591 & 10.317852 \end{pmatrix}$$

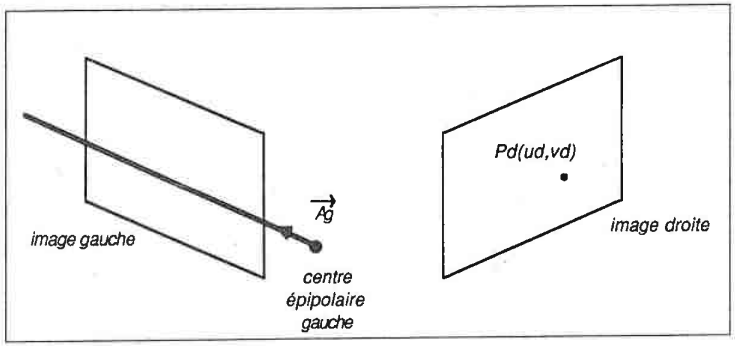

Figure C.2: *Centre épipolaire gauche et ligne épipolaire gauche associée à un point P_d de l'image droite.*

Sur la figure C.3 sont représentés quelques couples de lignes épipolaires calculés à l'aide de ces matrices.

2.2 Paire stéréoscopique *pièce*

Coordonnées des centres épipolaires :

$$M_g = \begin{pmatrix} -2518.40454 & 203.87158 \end{pmatrix} \; et \; M_d = \begin{pmatrix} -6433.31348 & 2941.91455 \end{pmatrix}$$

Lignes épipolaires :

$$\begin{pmatrix} \alpha_{d11} & \alpha_{d12} & \alpha_{d13} \\ \alpha_{d21} & \alpha_{d22} & \alpha_{d23} \end{pmatrix} = \begin{pmatrix} 118.29686 & 0.045375752 & -0.0033412864 \\ -54.335675 & -0.011275503 & 0.021546694 \end{pmatrix}$$

$$\begin{pmatrix} \alpha_{g11} & \alpha_{g12} & \alpha_{g13} \\ \alpha_{g21} & \alpha_{g22} & \alpha_{g23} \end{pmatrix} = \begin{pmatrix} 57.589318 & 0.0096349649 & 0.0014941158 \\ -27.256094 & 0.0050420400 & 0.020290548 \end{pmatrix}$$

Sur la figure C.4 sont représentés quelques couples de lignes épipolaires calculés à l'aide de ces matrices.

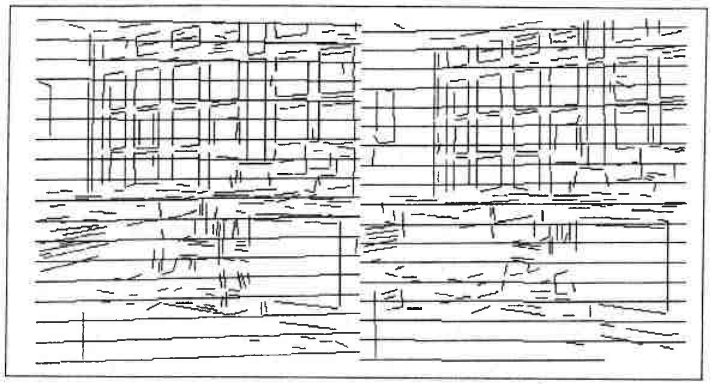

Figure C.3: *Couples de lignes épipolaires associées.*

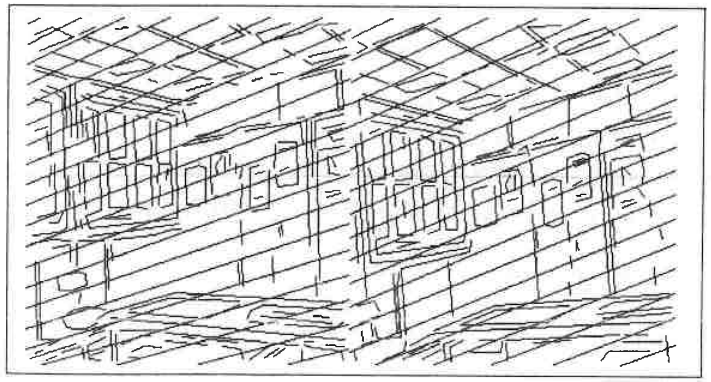

Figure C.4: *Couples de lignes épipolaires associées.*

Annexe D

Résultats

Nous présentons les résultats de la méthode de stéréovision que nous avons développée sur quatre paires d'images stéréoscopiques appelées : *bureau1*, *bureau2*, *couloir* et *pièce*. Ces couples d'images, ainsi que la calibration des caméras et du banc stéréoscopique nous ont été fournies par l'Inria Rocquencourt. Les images sont acquises par des caméras CCD, numérisées à 512 × 512 points et réduites par sous-échantillonnage à une taille de 256 × 256 points.

Pour chaque couple d'images, nous présentons :

- les images originales,
- la segmentation initiale des régions,
- l'appariement des régions,
- la segmentation en régions améliorée par le processus d'appariement,
- la "vue de face" des segments 3d construits dans la scène,
- la "vue de dessus" des segments 3d construits dans la scène,
- la construction des surfaces 3d visualisées en fonction de la distance.

La scène *bureau2* est identique à la scène *bureau1* mais elle est captée après un déplacement du robot vers les fenêtres.

Remarquons que, lors de l'acquisition des images stéréoscopiques de la scène *pièce*, les caméras ont été placées sur un support légèrement oblique. Les segments horizontaux, perdus lors du traitement des autres scènes, sont presque tous conservés.

1 Scene *bureau1*

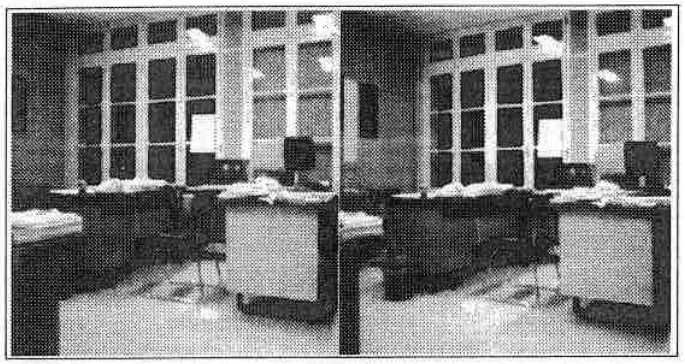

Figure D.1: *Couple stéréoscopique. Images initiales.*

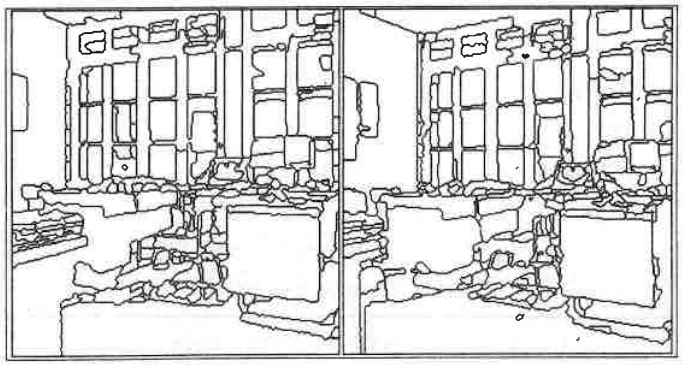

Figure D.2: *Partitions initiales en régions (image gauche : 247 régions, image droite : 254 régions).*

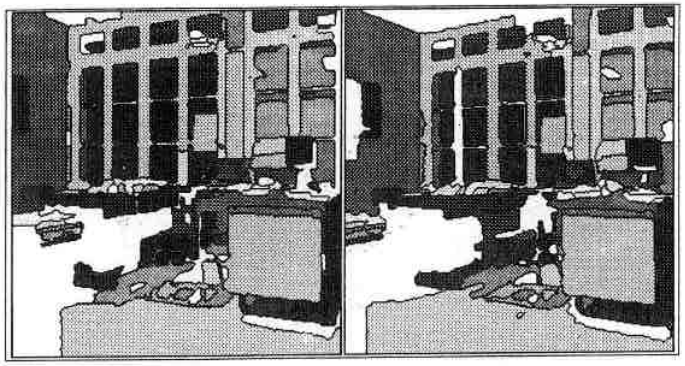

Figure D.3: *Mise en correspondance des régions (113 régions appariées). Les régions appariées sont visualisées avec le même niveau de gris, les régions laissées en blanc n'ont pas d'homologue.*

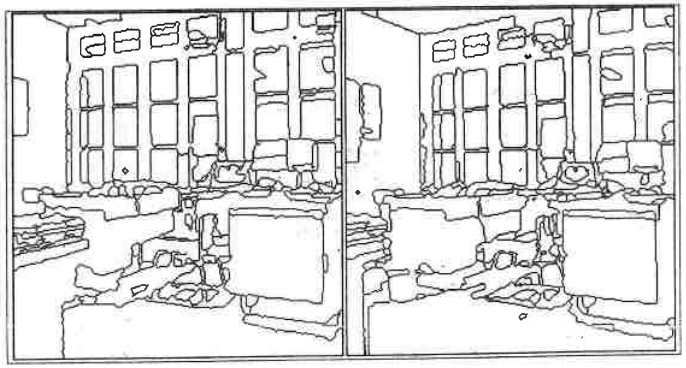

Figure D.4: *Partitions finales en régions. Le processus a réalisé 117 fusions (image gauche : 190 régions, image droite : 194 régions).*

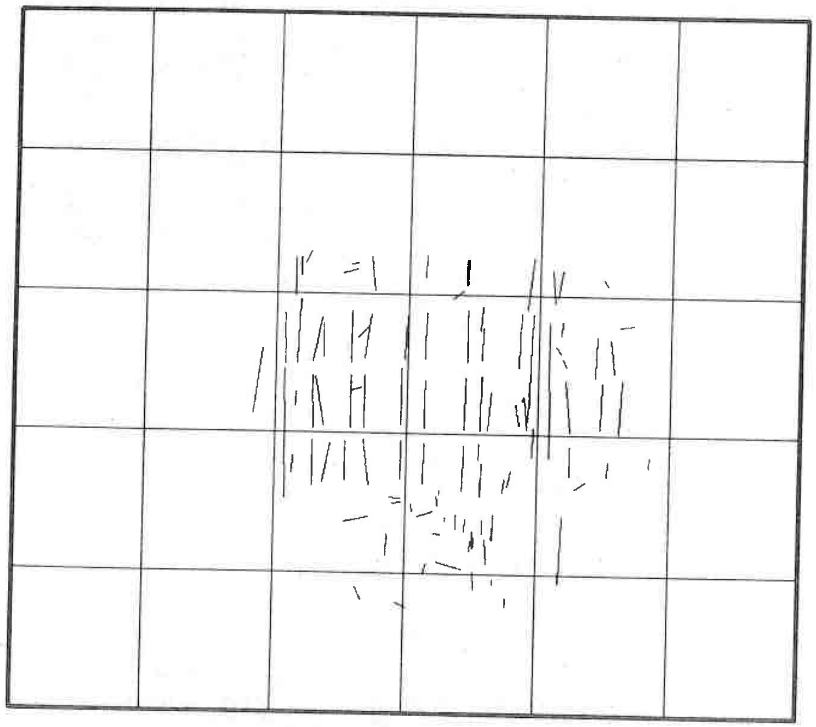

Figure D.5: *"Vue de face" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 5 mètres, hauteur : 4 mètres ; taille du quadrillage : 1 m × 1 m (100 segments 3d).*

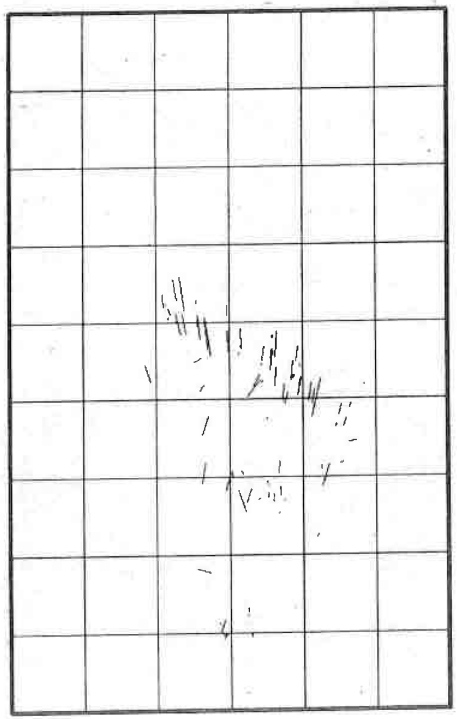

Figure D.6: *"Vue de dessus" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 6 mètres, profondeur : 9 mètres ; taille du quadrillage : 1 m × 1 m (100 segments 3d).*

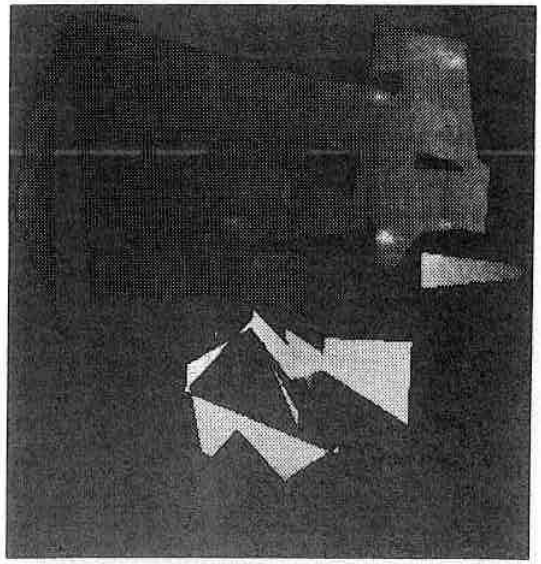

Figure D.7: *Surfaces tridimensionnelles de la scène - Premiers résultats.*

2 Scène *bureau2*

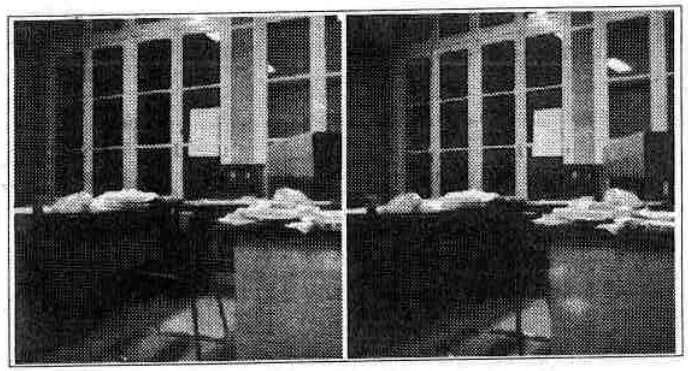

Figure D.8: *Couple stéréoscopique. Images initiales.*

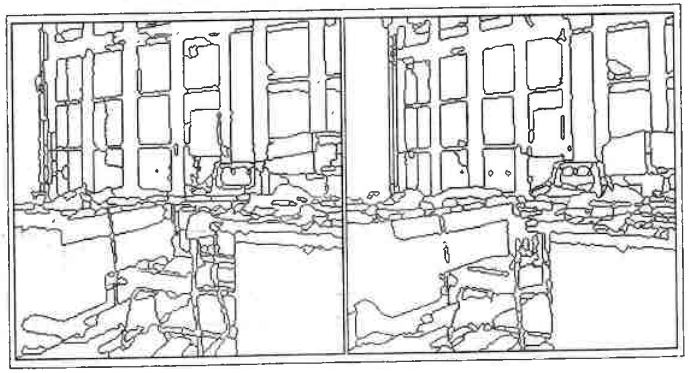

Figure D.9: *Partitions initiales en régions (image gauche : 254 régions, image droite : 251 régions.*

Figure D.10: *Mise en correspondance des régions (113 régions appariées). Les régions appariées sont visualisées avec le même niveau de gris, les régions laissées en blanc n'ont pas d'homologue.*

Figure D.11: *Partitions finales en régions. Le processus a réalisé 94 fusions (image gauche : 206 régions, image droite : 205 régions).*

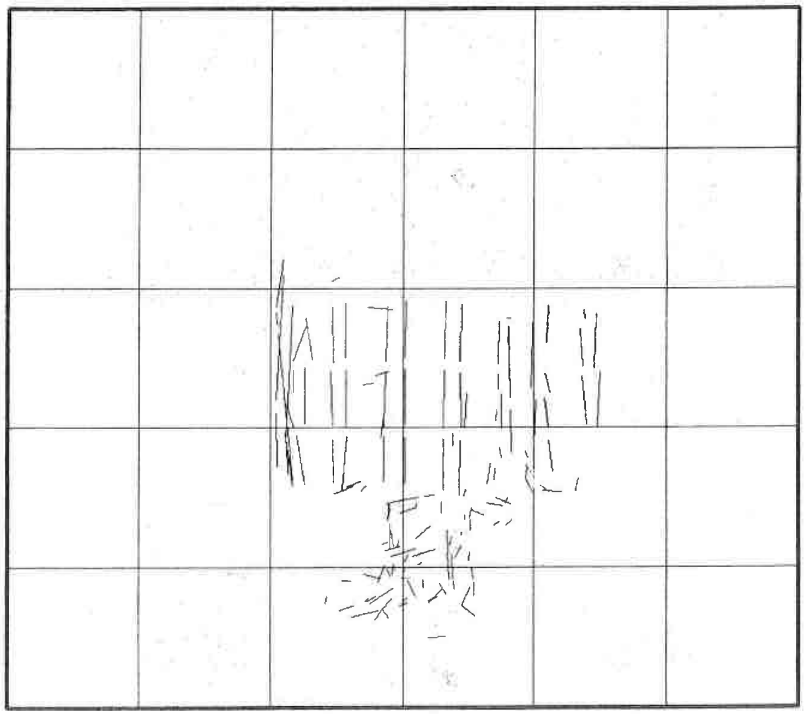

Figure D.12: *"Vue de face" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 5 mètres, hauteur : 4 mètres ; taille du quadrillage : 1 m × 1 m (133 segments 3d).*

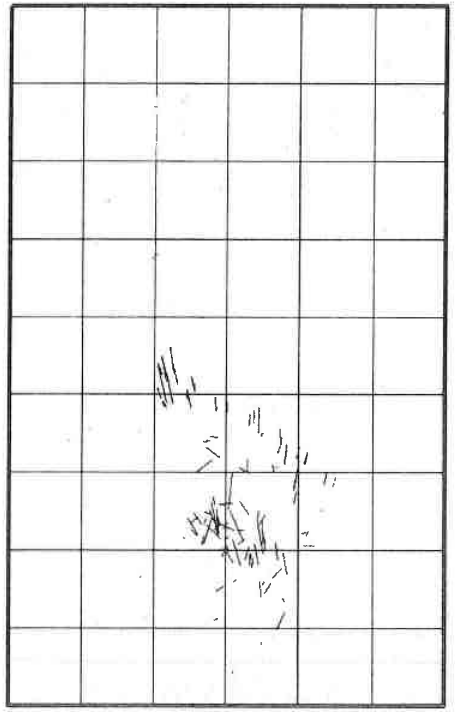

Figure D.13: *"Vue de dessus" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 6 mètres, profondeur : 9 mètres ; taille du quadrillage : 1 m × 1 m (133 segments 3d).*

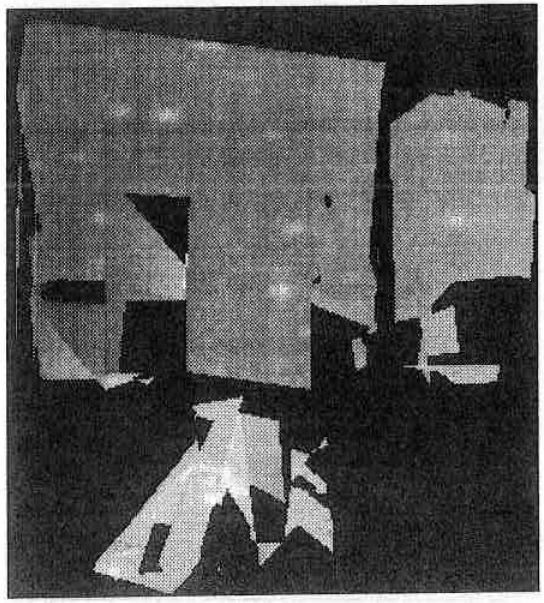

Figure D.14: *Surfaces tridimensionnelles de la scène - Premiers résultats.*

3 Scène *couloir*

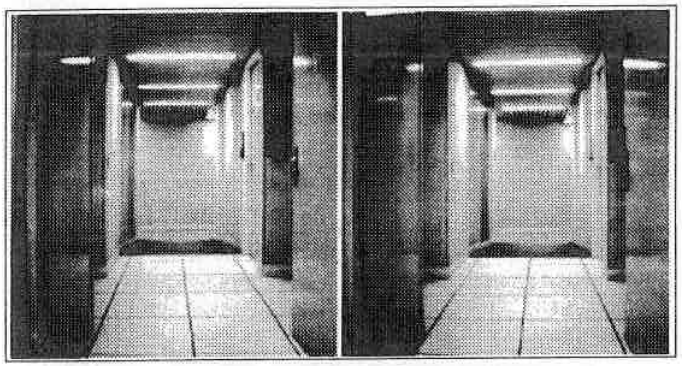

Figure D.15: *Couple stéréoscopique. Images initiales.*

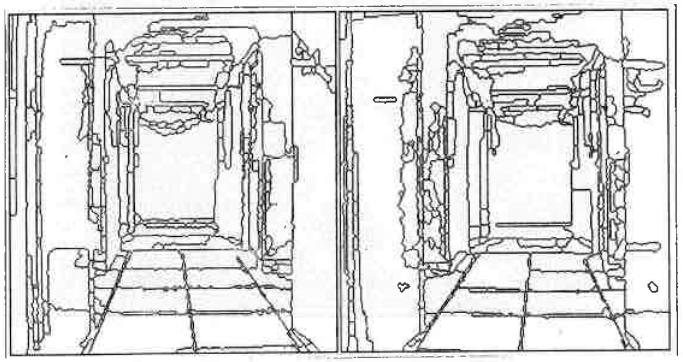

Figure D.16: *Partitions initiales en régions (image gauche : 249 régions, image droite : 254 régions).*

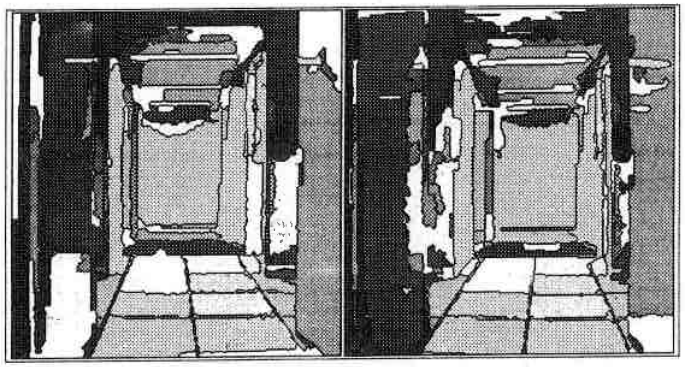

Figure D.17: *Mise en correspondance des régions (113 régions appariées). Les régions appariées sont visualisées avec le même niveau de gris, les régions laissées en blanc n'ont pas d'homologue.*

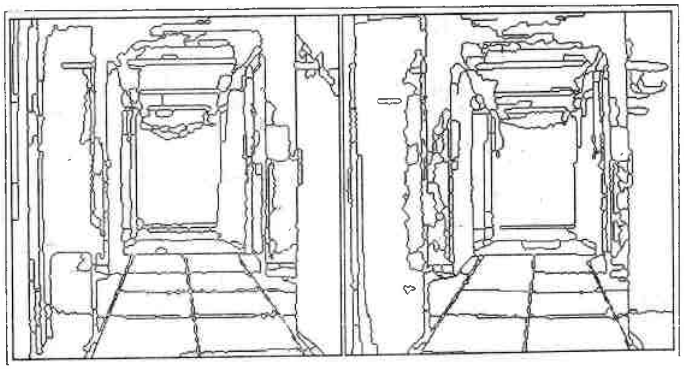

Figure D.18: *Partitions finales en régions. Le processus a réalisé 134 fusions (image gauche : 181 régions, image droite : 188 régions).*

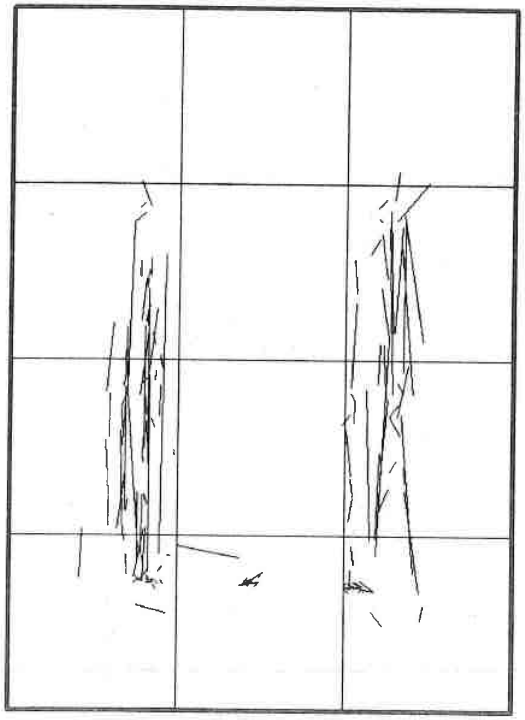

Figure D.19: *"Vue de face" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 3 mètres, hauteur : 4 mètres ; taille du quadrillage : 1 m × 1 m (115 segments 3d).*

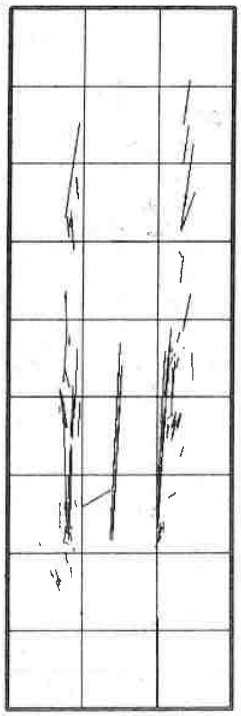

Figure D.20: *"Vue de dessus" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 3 mètres, profondeur : 8 mètres ; taille du quadrillage : 1 m × 1 m (115 segments 3d).*

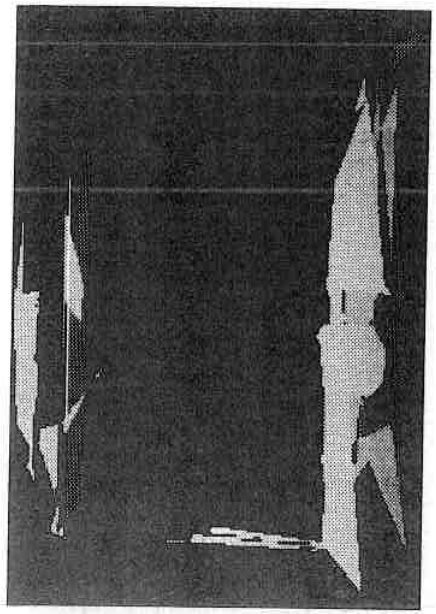

Figure D.21: *Surfaces tridimensionnelles de la scène - Premiers résultats.*

4 Scène *pièce*

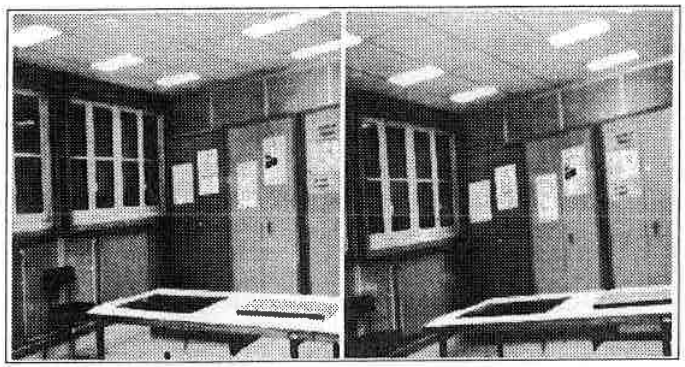

Figure D.22: *Couple stéréoscopique. Images initiales.*

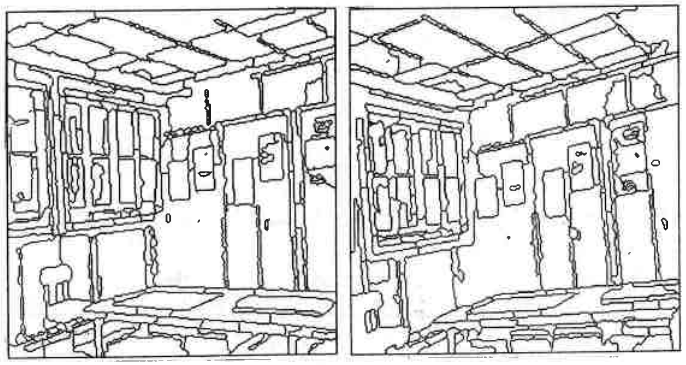

Figure D.23: *Partitions initiales en régions (image gauche : 254 régions, image droite : 254 régions.)*

Figure D.24: *Mise en correspondance des régions (117 régions appariées). Les régions appariées sont visualisées avec le même niveau de gris, les régions laissées en blanc n'ont pas d'homologue.*

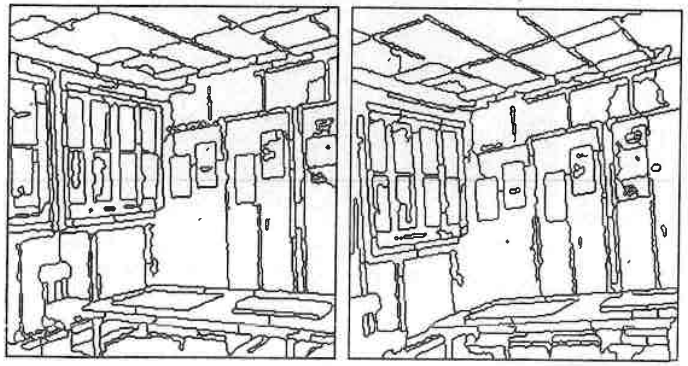

Figure D.25: *Partitions finales en régions. Le processus a réalisé 85 fusions (image gauche : 217 régions, image droite : 206 régions).*

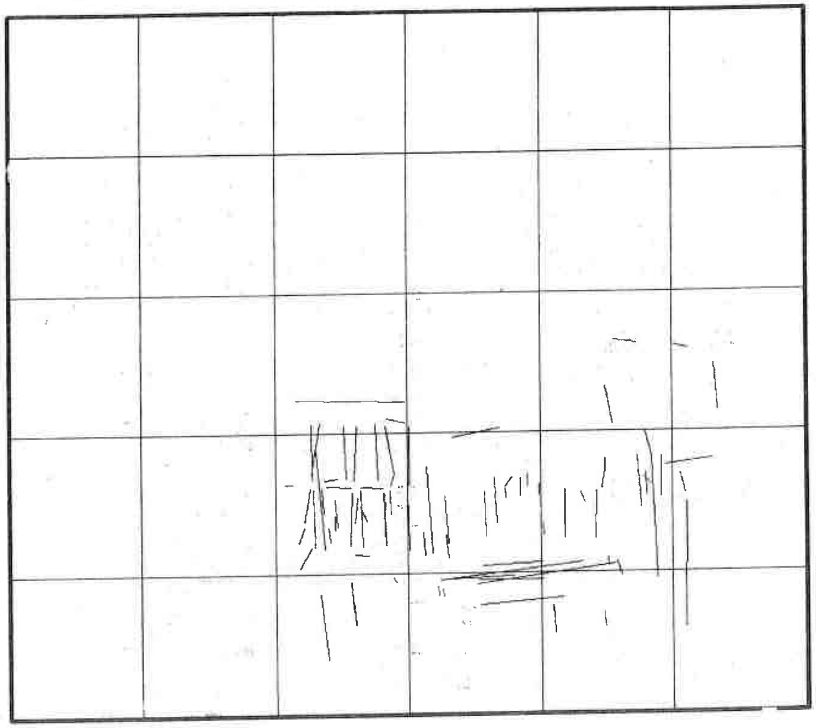

Figure D.26: *"Vue de face" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 5 mètres, hauteur : 4 mètres ; taille du quadrillage : 1 m × 1 m (79 segments 3d).*

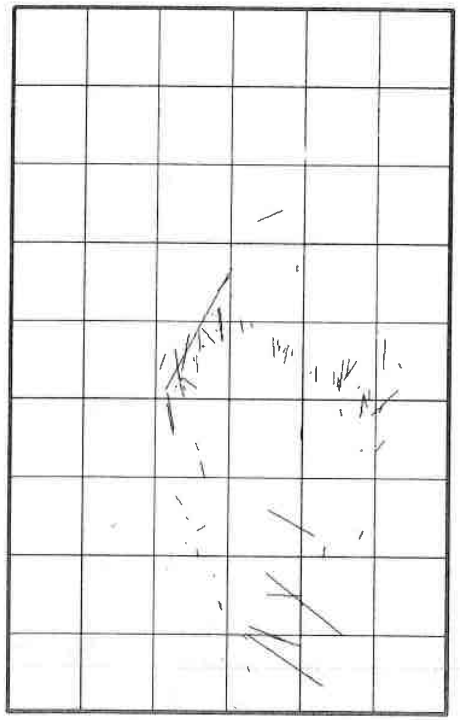

Figure D.27: *"Vue de dessus" des segments tridimensionnels. Echelle de représentation de la scène : largeur : 6 mètres, profondeur : 9 mètres : taille du quadrillage : 1 m × 1 m (79 segments 3d).*

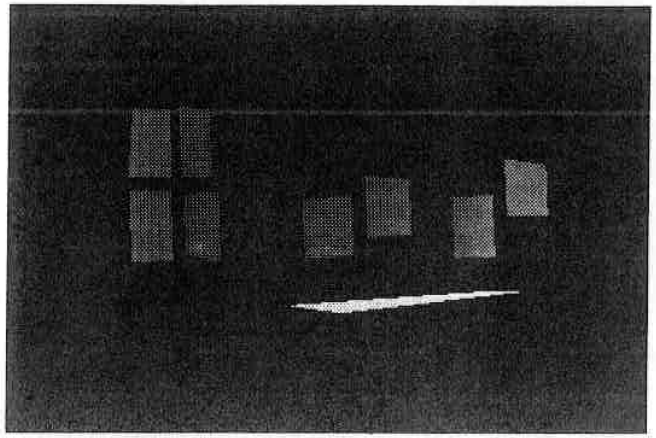

Figure D.28: *Surfaces tridimensionnelles de la scène - Reconstruction des surfaces correspondant aux papiers affichés sur les armoires, à quelques carreaux des fenêtres et au plateau de la table.*

Bibliographie

[Aho 87] A.H. Aho, J.E. Hopcroft, et J.D. Ullman. *The Design and Analysis of Computer Algorithms.* Publishing Company, Reading, Mass., 1974. *Structures de données et algorithmes*, Inter Editions, une traduction d'une édition récente de l'ouvrage précédent, 1987.

[Alami 87] M. Alami, A. Osorio, et L. Peralta. K2D avec des caractéristiques de niveau moyen. Dans *Actes du Congrès* MARI-*87*, pages 189–196, vol. 2, Paris, 1987.

[Ayache 85] N. Ayache et B. Faverjon. Un algorithme de stéréoscopie passive utilisant la prédiction-vérification récursive d'hypothèses. Dans *Actes du 5ème Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 1217–1232, Grenoble, 1985.

[Ayache 86] N. Ayache et O.D. Faugeras. HYPER: a New Approach for the Recognition and Positioning of 2d Objects. IEEE *Transactions on Pattern Analysis and Machine Intelligence (*PAMI*)*, 8(1):44–54, 1986.

[Ayache 87] N. Ayache et F. Lustman. Fast and Reliable Passive Trinocular Stereovision. Dans *Proceedings of the First International Conference on Computer Vision (*ICCV*)*, pages 422–427, Londres, 1987.

[Backer 81] H.H. Backer et T.O. Binford. Depth from Edge and Intensity Based Stereo. Dans *Proceedings of the 7th International Joint Conference on Artificial Intelligence (*IJCAI*)*, pages 631–636, Vancouver B.C Canada, 1981.

[Bajcsy 76] R. Bajcsy et L. Lieberman. Textured Gradient as a Depth Cue. *Computer Vision and Image Processing*, 5:52–67, 1976.

[Bajcsy 80] R. Bajcsy. Three-dimensional Scene Analysis. Dans *Proceedings of the 5th International Conference on Pattern Recognition (*ICPR*)*, pages 1064–1074, Miami Fl. (USA), 1980.

[Bauzil 81] G. Bauzil, M. Briot, et P. Ribes. A Navigation Sub-System Using Ultrasonics Sensors for the Mobile Robot HILARE. Dans *First International Joint Conference on Robot Vision and Sensor Control*, pages 47–48, 1981.

[Belaid 85] A. Belaid et J.P. Haton. La reconnaissance de l'écriture. *La Recherche*, 170:1188–1197, 1985.

[Belaid 86] Y. Belaid, M. Camonin, G. Masini, R. Mohr, et E. Thirion. *Le système de vision* TRIDENT. Rapport final du contrat ADI,CRIN–86–R–012, Centre de Recherche en Informatique de Nancy, 1986.

[Bellman 57] J. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[Benard 83] M. Benard. *Restitution automatique en stéréophotogrammétrie*. Rapport ENST–E–83012, Ecole Nationale Supérieure des Télécommunications, Paris, 1983.

[Berthod 84] M. Berthod et P. Long. Stéréoscopie par mise en correspondance de graphe. Dans *Premier Colloque Image*, pages 1011–1016, Biarritz, 1984.

[Berthod 87] M. Berthod, F. Bretaudeau, et C. Proy. Extraction automatique d'informations structurelles à partir de données SPOT et premiers résultats. Dans *Actes du Congrès* MARI–*87*, pages 243–248, vol. 2, Paris, 1987.

[Binford 81] T.O. Binford. Inferring Surfaces from Images. *Artificial Intelligence*, 17:205–244, 1981.

[Boissonnat 87] J.D. Boissonnat, O.D. Faugeras, et E. Lebras-Melhman. Reconstruction de surfaces à partir de données stéréo par la triangulation de Delaunay. Dans *Actes du* 6ème *Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 687–701, Antibes, 1987.

[Boissonnat 88] J.D. Boissonnat et O. Monga. Surface Reconstruction from Stereo Data by Planar Cross Sections. Dans *Proceedings of the* IEEE *Robotics and Automation (article soumis)*, Philadelphia (USA), 1988.

[Bonton 87] P. Bonton, S. Gidel, L. Grouche, G. Rongier, A. Geneix, P. Malet, et B. Perissel. Tri chromosomique par traitement d'images. Dans *Actes du Congrès* MARI–*87*, pages 326–333, vol. 2, Paris, 1987.

[Borianne 84] P.L. Borianne. *Contributions à la vision par ordinateur tridimensionnelle*. Thèse de 3ème Cycle, Institut National Polytechnique de Grenoble, 1984.

[Bouthemy 87] P. Bouthemy. Estimation et structuration d'indices spatio-temporels pour l'analyse du mouvement dans une séquence d'images. *Revue Traitement du Signal* (TS), *Numéro spécial : Vision par ordinateur*, 4:239–257, 1987.

[Boyer 87] A. Boyer. *Application des techniques de programmation dynamique et de quantification vectorielle à la reconnaissance des mots isolés et des mots enchaînés*. Thèse de Doctorat, Université de Nancy 1, 1987.

[Brooks 78] R.A. Brooks et T.O. Binford. The Acronym Model-based Vision System. *Proceedings* ARPA *Image Understanding Workshop*, pages 145–151, Pittsburg Penn. (USA), 1978.

[Brooks 81] R.A. Brooks. Symbolic Reasoning among 3d Models and 2d Images. *Artificial Intelligence*, 17:285–348, 1981.

[Camillerapp 87] J. Camillerapp, I. Leplumey, et A. Walter. Acquisition d'un modèle tridimensionnel du réseau vasculaire cérébral à partir d'un couple stéréoscopique. *Actes du* 6ème *Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 341–349, Antibes, 1987.

[Camonin 87] M. Camonin. *Le système* MEPHISTO *: Validation de modèles tridimensionnels*. Thèse de 3ème Cycle, Université de Nancy 1, 1987.

[Castan 84a] S. Castan et J. Shen. A Stereo Vision Algorithm taking into account the Perspective Distorsions. *Proceedings of the* 7th *International Conference on Pattern Recognition* (ICPR), pages 444–446, Montréal Qué, Canada, 1984.

[Castan 84b] S. Castan et J. Shen. Une méthode photométrique pour déterminer l'orientation des surfaces possédant certaines propriétés de réflectance. *Actes du* 4ème *Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 81–92, Paris, 1984.

[Clowes 71] M.B. Clowes. On Seeings Things. *Artificial Intelligence*, 2, 1971.

[Cocquerez 87] J.P. Cocquerez et O. Monga. Matching Regions in Stereovision. *Proceedings of the* 5th *Scandinavian Conference on Image Analysis*, pages 295–308, Stockholm, 1987.

[Coleman 82] E.N. Coleman et J. Ramesh. Obtaining Three-dimensionnal Shape of Textured and Specular Surfaces Using Four-Source Photometry. *Computer Vision and Image Processing*, 18:309–328, 1982.

[Coulon 85] P.Y. Coulon et J.P. Charras. Trajectométrie automatique par caméra. *Actes du 5ème Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 849–856, Grenoble, 1985.

[Crowley 85] J. Crowley. Navigation for an Intelligent Mobile Robot. IEEE *Journal of Robotics and Automation*, 1(1), 1985.

[Demazeau 85] Y. Demazeau et J.L. Gordillo. A color Stereo Vision System Applied to Wire Identification and Localization. *Proceedings of the 4th Canadian cad/cam and Robotics*, Toronto, 1985.

[Demazeau 86] Y. Demazeau. *Niveaux de représentation pour la vision par ordinateur, indices d'image et indices de scène.* Thèse de Doctorat de l'Institut National Polytechnique de Grenoble, 1986.

[Deriche 87a] R. Deriche. Optimal Edge Detection Using Recursive Filtering. *Proceedings of the First International Conference on Computer Vision (*ICCV*)*, pages 501–505, Londres, 1987.

[Deriche 87b] R. Deriche et J.P. Cocquerez. Extraction de composantes connexes basée sur une détection optimale des contours. *Actes du Congrès* MARI-*87*, pages 1–9, vol. 2, Paris, 1987.

[Dyer 82] C.R. Dyer. *Pyramids Algorithms and Machines.* Multicomputer and Multiprogramming, Preston K. Jr, Uhr L., pages 409–412, 1982.

[Faugeras 86a] O.D. Faugeras et M. Hebert. The Representation, Recognition, and Locating of 3d Objects. *The International Journal of Robotics Research*, 5(3):27–52, 1986.

[Faugeras 86b] O.D. Faugeras et G. Toscani. The Calibration Problem for Stereo. *Proceedings of International Conference on Computer Vision and Pattern Recognition (*CVPR*)*, pages 15–20, Miami Fl. (USA), 1986.

[Faugeras 87] O.D. Faugeras et G. Toscani. Camera Calibration for 3d Machine Vision. *Proceedings of the International Workshop on Machine Vision and Machine Intelligence*, pages 240–247, Tokyo, 1987.

[Forney 73] G.D. Forney. The Viterbi Algorithm. *Proceedings of* IEEE, *vol. 61,n. 3*, 1973.

[Gagalowicz 85] A. Gagalowicz et O. Monga. Un algorithme de segmentation hiérarchique. *Actes du 5ème Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 163–177, Grenoble, 1985.

[Gambotto 85] J.P. Gambotto et O. Monga. A Parallel and Hierarchical Algorithm for Region Growing. *Proceedings of International Conference on Computer Vision and Pattern Recognition (*CVPR*)*, San-Francisco Ca. (USA), 1985.

[Gandelin 87] M.H. Gandelin et M. Thonnat. Etude pour l'automatisation de la reconnaissance d'organismes zooplanctoniques à l'aide d'un système expert. *Actes du 6ème Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 491–496, Antibes, 1987.

[Geographic 87] National Geographic. Medecine's New Vision. *National Geographic*, 171(1):241, 1987.

[Gerhard 86] A. Gerhard, H. Platzer, J. Steurer, et R. Lenz. Depth Extraction by Stereo Triples and a Fast Correspondance Estimation Algorithm. *Proceedings of the 8th International Conference on Pattern Recognition (*ICPR*)*, pages 512–515, Paris, 1986.

[Giraudon 87] G. Giraudon. An Efficient Edge Following Algorithm. *Proceedings of the 5th Scandinavian Conference on Image Analysis*, pages 547–554, Stockholm, 1987.

[Grimson 81] W.E.L. Grimson. *From Images to Surfaces. A Computational Study of the Human Early Visual System.* The Massachussets Institute of Technology Press, 1981.

[Guilloux 84] Y. Le Guilloux. *Détermination automatique du mouvement d'une séquence d'images ; intérêt pour l'interprétation.* Thèse de Docteur Ingénieur, ENST-84-E012, Ecole Nationale Supérieure des Télécommunications, Paris, 1984.

[Hannah 84] M.J. Hannah. *Description of* SRI*'s Baseline Stereo System.* Technical Note 342, SRI International Artificial Intelligence Center, 1984.

[Haralick 85] R.M. Haralick et L. Shapiro. "Survey": Image Segmentation Techniques. *Computer Vision, Graphics and Image Processing*, 29:100–132, 1985.

[Henderson 78] R.L. Henderson, W.J. Miller, et C.B. Grosh. A Flexible Approach to Digital Stereo Matching. *Photogrammetry Engineering and Remote Sensing*, 1499–1512, 1978.

[Hoff 87] W. Hoff et N. Ahuja. Extracting Surfaces from Stereo Images: an Integrated Approach. *Proceedings of the First International Conference on Computer Vision (ICCV)*, pages 284–294, Londres, 1987.

[Horaud 87a] R. Horaud. La vision artificielle - Editorial. *Revue Traitement du Signal (TS), Numéro spécial : Vision par ordinateur*, 4:163–167, 1987.

[Horaud 87b] R. Horaud. New Methods for Matching 3d Objects with Single Perspective views. IEEE *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 9(3):401–412, 1987.

[Horaud 88] R. Horaud, T. Skordas, et Y. Demazeau. Mise en correspondance de structures 2d dans une paire stéréo - article soumis. *Techniques et Sciences Informatiques*, 1988.

[Horn 70] B. Horn et R. Stoberg. *Shape from Shading, A Method for Obtaining the Shape of a Smooth Opaque Object from one View.* Technical Report 79, Massachussets Institute of Technology Laboratory for Computer Science, 1970.

[Horowitz 78] E. Horowitz et S. Sahni. *Fondamentals of Computer Algorithms.* Computer Science Press, 1978.

[Huffman 71] D.A. Huffman. Impossible Objects as Nonsense Sentences. *Machine Intelligence, B. Meltzer and D. Michie Ed., American Elsevier*, 6, 1971.

[Hwang 82] J.J. Hwang et E.L. Hall. Matching of Featured Objects Using Relational Tables from Stereo Images. *Computer Vision and Image Processing*, 20:22–42, 1982.

[Jarvis 83] R.A. Jarvis. A Laser Time of Flight Range Scanner for Robotic Vision. IEEE *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 5(5):505–511, 1983.

[Julesz 71] B. Julesz. *Fondations of Cyclopean Perception.* University of Chicago Press, 1971.

[Kent 86] E.W. Kent et M.O. Shneier. Eyes for Automatons. *Spectrum*, 23(3):37–45, 1986.

[Keskes 82] N. Keskes, A. Boulanouar, et O.D. Faugeras. Application of Image Analysis Techniques to Seismic Data. *International Conference on Acoustic Signal and Speech Processing (ICASSP)*, pages 855–857, Paris, 1982.

[Kuroe 85] S. Kuroe, K. Nakajima, I. Nose, et K. Tanoshima. Segmented-Region Based Stereo Reconstruction. *Proceedings of the 15th International Symposium on Industrial Robots (ISIR)*, pages 375–382, Tokyo, 1985.

[Lelandais 84] S. Lelandais, A. Clainchard, J. Lopez-Krahe, H. Maître, et F. Schmitt. Numérisation de formes tridimensionnelles : acquisition, traitement. *Actes du 4ème Congrès AFCET Reconnaissance des Formes et Intelligence Artificielle*, pages 339–356, Paris, 1984.

[LimozinLong 86] P. Limozin-Long. *Vision stéréoscopique appliquée à la robotique.* Thèse de Doctorat, Université de Nice, 1986.

[Lustman 87] F. Lustman. *Vision stéréoscopique et peception du mouvement en vision artificielle.* Thèse de doctorat, Université de Paris-Sud, Orsay, 1987.

[Lux 84] A. Lux et V. Souvignier. PVV : un système de vision appliquant une stratégie de prédiction-vérification. *Actes du 4ème Congrès AFCET Reconnaissance des Formes et Intelligence Artificielle*, pages 223–224, Paris, 1984.

[Marr 76] D. Marr et T. Poggio. Cooperative Computation of Stereo Disparity. *Science*, 194:283–287, 1976.

[Marr 77] D. Marr et T. Poggio. A Theory of Human Stereo Vision. *Artificial Intelligence*, 451, 1977.

[Marr 80] D. Marr et E.C. Hildreth. Theory of Edge Detection. *Proceedings of the Royal Society of London B207*, 187–217, 1980.

[Medioni 85] G. Medioni et R. Nevatia. Segment Based Stereo Matching. *Computer Vision, Graphics and Image Processing*, 31:2–18, 1985.

[Mohr 84] R. Mohr et B. Wrobel. La correspondance stéréoscopique vue comme une recherche de chemin optimal dans un graphe. *Actes du 4ème Congrès AFCET Reconnaissance des Formes et Intelligence Artificielle*, pages 71–79, Paris, 1984.

[Monga 86] O. Monga et N. Keskes. A Hierarchical Algorithm for the Segmentation of 3d Images. *Proceedings of the 8th International Conference on Pattern Recognition (ICPR)*, pages 735–738, Paris, 1986.

[Monga 87] O. Monga et B. Wrobel. Segmentation d'images : vers une méthodologie. *Revue Traitement du Signal (TS)*,

Numéro spécial : Vision par ordinateur, 4:169–193, 1987.

[Montanvert 87] A. Montanvert, J.M. Chassery, et C. Charles. Décomposition de formes discrètes complexes appliquées à un Problème de trajectographie sous-marine. *Actes du Congrès* MARI-*87*, pages 411–418, vol. 2, Paris, 1987.

[Moravec 77] H.P. Moravec. Towards Automatic Visual Obstacle Avoidance. *Proceedings of the 5th International Joint Conference on Artificial Intelligence (*IJCAI*)*, Cambridge Ma. (USA), 1977.

[Moravec 80] H.P. Moravec. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover.* Ph. D. Thesis, Stanford Artificial Intelligence Laboratory, 1980.

[Moravec 85] H.P. Moravec. High-Resolution Maps fron Wide-Angle Sonar. *Proceedings of the IEEE Computer Society Conference on Robotics and Automation*, St-Louis Missouri, (USA), 1985.

[Morris 86] O.J. Morris, J. Lee, et A.G. Constantinides. Graph Theory for Image Analysis: an Approach based on the Shortest Spanning Tree. *Proceedings of the* IEEE, 133(2):146–151, 1986.

[MullerBelaid 84] Y. Muller-Belaid. *Détection de surfaces par transformée de Hough.* Thèse de 3ème Cycle, Université de Nancy 1, 1984.

[Nilsonn 69] Nilsonn. A Mobile Automaton: an Application of Artificial Intelligence Techniques. *Proceedings of the First International Joint Conference on Artificial Intelligence (*IJCAI*)*, Washington D.C. (USA), 1969.

[Nishihara 84] H.K. Nishihara. PRISM, a Practical Real-Time Imaging Stereo Matcher. Artificial Intelligence Memo 780, MIT, 1984.

[Ohta 83] Y. Ohta et T. Kanade. Stereo by Intra- and Interscanline Search Using Dynamic Programming. Computer Science Department, Carnegie-Mellon University, Pittsburg PA, 1983.

[Ohta 86a] Y. Ohta, K. Takano, et K. Ikeda. A Highspeed Stereo Matching System Based on Dynamic Programming. *Proceedings of the First International Conference on Computer Vision (*ICCV*)*, pages 335–342, Londres, 1986.

[Ohta 86b] Y. Ohta, M. Watanebe, et K. Ikeda. Improving Depth Map by Right-Angled Trinocular Stereo. *Proceedings of the 8th International Conference on Pattern Recognition (*ICPR*)*, pages 519–521, Paris, 1986.

[Pavlidis 77] T. Pavlidis. *Structural Pattern Recognition.* Springer Verlag, 1977.

[Pavlidis 86] T. Pavlidis. A Critical Survey of Image Analysis Methods. *Proceedings of the 8th International Conference on Pattern Recognition (*ICPR*)*, pages 502–511, Paris, 1986.

[Pettengill 81] G. Pettengill et al. Pioneer Venus Radar Results: Altimetry and Surfaces Proprieties. *Journal of Geophysic Research*, 85:8261, 1981.

[Pietikainen 86] M. Pietikainen et D. Harwwod. Multiple-Camera Contour Stereo. *Proceedings of the 8th International Conference on Pattern Recognition (*ICPR*)*, pages 1269–1272, Paris, 1986.

[Remion 87] Y. Remion, H. Maître, et J.M. Lopez. Recalage par zones de deux vues stéréoscopiques d'un univers composé d'objets plans. *Actes du Congrès* MARI-*87*, pages 174–179, vol. 2, Paris, 1987.

[RobertdeSaintVincent 86] A. Robert-de-Saint-Vincent. *Perception et modélisation de l'environnement d'un robot mobile : une approche par stéréovision.* Thèse de Doctorat, Université Paul Sabatier, Toulouse, 1986.

[Rogala 87] J.P. Rogala et S. Simon. Extraction auto-adaptative de primitives pour la localisation de composants électroniques. *Actes du Congrès* MARI-*87*, pages 16–22, vol. 2, Paris, 1987.

[Rosenberg 78] D. Rosenberg, M.D. Levine, et S.W. Zucker. Computing Relative Depth Relationships from Occlusion Cues. *Proceedings of the 4th International Conference on Pattern Recognition (*ICPR*)*, pages 765–769, Kyoto, 1978.

[Rosenfeld 82] A. Rosenfeld et A.C. Kak. *Digital Picture Processing.* Volume 2.8, Academic-Press, New-York, 1982.

[Serra 83] J. Serra. Images et morphologie mathématique. *La Recherche*, 14(144):723–732, 1983.

[Shen 87] J. Shen et S. Castan. Fast Binary Image Correlation Algorithms for Stereo Vision. *Proceedings of the 5th Scandinavian Conference on Image Analysis*, pages 317–323, Stockholm, 1987.

[Shihari 81] S.N. Shihari. Representation of Three-Dimensional Digital Image. *Computer Surveys*, 13(4):399–424, 1981.

[Shihari 86] S.N. Shihari et G.W. Zack. Document Image Analysis. *Proceedings of the 8th International Conference on Pattern Recognition (ICPR)*, pages 434–436, Paris, 1986.

[Shirai 79] Y. Shirai. *Three-Dimensionnal Computer Vision, Computer Vision and Sensor-Based Robots.* PLENUM Press, 1979.

[Stansfield 86] S.A. Stansfield. Primitives, Features, and Exploratory Features: Building a Robot Tactile Perception System. *Proceedings of the IEEE Computer Society Conference on Robotics and Automation*, 1986.

[Stelmaszyk 85] P. Stelmaszyk, P. Bonnet, et J.G. Postaire. Analyse de scènes dynamiques par recherche des contours en mouvement. *Actes du 5ème Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 1181–1193, Grenoble, 1985.

[Thonnat 86] M. Thonnat. Semantic Interpretation of 3d Stereo Data: Finding the Main Structures. *Proceedings of the 8th International Conference on Pattern Recognition (ICPR)*, pages 1051–1054, Paris, 1986.

[Thorpe 83] C. Thorpe et S. Shafer. Correspondance in Line Drawings of Multiple Views of Object. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 959–965, Karlsruhe, 1983.

[Thorpe 85] C. Thorpe. Experiments and Thoughts on Visual Navigation. *Proceedings of the IEEE Computer Society Conference on Robotics and Automation*, St-Louis Missouri, (USA), 1985.

[Tombre 85] K. Tombre et B. Wrobel. *Extraction de points contours dans une image.* Rapport CRIN 85–R–96, Centre de Recherche en Informatique de Nancy, 1985.

[Tombre 87] K. Tombre. *La saisie automatisée de documents composites : reconnaissance, codage et interprétation des parties graphiques.* Thèse de Doctorat, Institut National Polytechnique de Lorraine, 1987.

[Toscani 87a] G. Toscani. *Systèmes de calibration optique et perception du mouvement en vision.* Thèse de Doctorat, Université de Paris-Sud, Orsay, 1987.

[Toscani 87b] G. Toscani et O.D. Faugeras. Mouvement par reconstruction et reprojection. *Actes du 11ème Colloque sur le Traitement du Signal et des Images (GRETSI)*, pages 535–538, Nice, 1987.

[Tsuji 85] S. Tsuji. Dynamic Scene Analysis for a Mobile Robot in a Man-made Environment. *Proceedings of the IEEE Computer Society Conference on Robotics and Automation*, St-Louis Missouri, (USA), 1985.

[Vintsyuk 68] T.K. Vintsyuk. *Reconnaissance de mots par programmation dynamique.* Kibernetica, 1968.

[Wallace 85] R. Wallace. First results in Robot Road Following. *Proceedings of the 9th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1089–1095, Los Angeles, Ca. (USA), 1985.

[Waxman 86] A.M. Waxman. A Visual Navigation System. *Proceedings of the IEEE Computer Society Conference on Robotics and Automation*, San-Francisco Ca., (USA), 1986.

[Witkin 81] A.P. Witkin. Recovering Surface Shape and Orientation from Texture. *Artificial Intelligence*, 17:17–46, 1981.

[Wrobel 85] B. Wrobel. *Segmentation d'images.* Rapport CRIN 85–R–103, Centre de Recherche en Informatique de Nancy, 1985.

[Wrobel 87a] B. Wrobel. Stéréovision : coopération entre l'extraction et la mise en correspondance symbolique de régions. *Actes du Congrès* MARI–*87*, pages 181–188, vol. 2, Paris, 1987.

[Wrobel 87b] B. Wrobel et O. Monga. Segmentation d'images naturelles : coopération entre un détecteur-contour et un détecteur-région. *Actes du 11ème Colloque sur le Traitement du Signal et des Images (GRETSI)*, pages 539–542, Nice, 1987.

[WrobelDautcourt 87] B. Wrobel-Dautcourt. Surfaces tridimensionnelles obtenues par appariement stéréoscopique de régions. *Actes du 6ème Congrès* AFCET *Reconnaissance des Formes et Intelligence Artificielle*, pages 299–313, Antibes, 1987.

[Yakimovsky 78] Y. Yakimovsky et R. Cunningham. A System for Extraction Three-Dimensionnal Measurements from a Stereo Pair of TV Cameras. *Computer Vision, Graphics and Image Processing*, 7(4):195–210, 1978.

[Zaroli 83] F. Zaroli. *Réalisation d'un système d'analyse et d'interprétation d'images tridimensionnelles.* Thèse de 3ème Cycle, Université de Nancy 1, 1983.

[Zhao 84] C. Zhao. *Localisation absolue d'un robot mobile par télémètre laser.* Thèse de Docteur Ingénieur, INSA, Rennes, 1984.

[Zucker 76] S. Zucker. "Survey": Region Growing: Childhood and Adolescence. *Computer Vision and Image Processing*, 5:382–399, 1976.

institut
national
polytechnique
de lorraine

Le Président,

AUTORISATION SOUTENANCE DE THESE
DU DOCTORAT DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

VU LES RAPPORTS ETABLIS PAR :
Monsieur LESCANNE Pierre, Directeur de Recherche CRIN
Monsieur MAITRE Henri, Professeur ENST Paris.

Le Président de l'Institut National Polytechnique de Lorraine,
autorise :
Madame WROBEL-DAUTCOURT Brigitte

à soutenir devant l'I.N.P.L. une thèse de Doctorat,
intitulée :
"Perception de la distance par mise en correspondance de régions entre des images stéréoscopiques"

en vue de l'obtention du titre de :

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

Spécialité : "Informatique"

Fait à Vandoeuvre, le 23 Février 1988

Le Président de l'I.N.P.L.,

M. CANTOIS

2, avenue de la Forêt-de-Haye - Vandœuvre
B.P. 3 - 54501 Vandœuvre Cedex - Téléphone 83.57.48.48 (lignes groupées)

Résumé

Notre travail s'inscrit dans le cadre du projet ORASIS dont l'objectif est le contrôle de la navigation d'un robot mobile autonome se déplaçant dans un environnement de bureau ou d'atelier sous des conditions naturelles d'éclairage. Ce robot doit être capable de voir, reconnaître et apprendre l'univers dans lequel il évolue.

Cette thèse présente le résultat de notre travail sur la vision du robot, plus précisément la *perception de la distance par stéréovision.* Le principe de base de la stéréovision est d'apparier des indices visuels des deux images traitées puis de déterminer par un calcul géométrique leur position dans la scène tridimensionnelle. Les indices images considérés sont les régions, puisque celles-ci contiennent une information de surface nécessaire à la construction d'une scène tridimensionnelle la plus complète possible.

Notre travail se scinde en trois étapes : extraction des régions, appariement puis construction des informations tridimensionnelles. Les difficultés de cette approche sont essentiellement dans la mise en correspondance puisque les régions à apparier présentent de grandes différences aussi bien photométriques que morphologiques. De plus, les deux surfaces visualisées d'un même objet sont partitionnées en un nombre inégal de régions. La solution apportée s'appuie sur l'utilisation de plusieurs indices visuels (région, contours, ...) afin de bénéficier d'informations plus riches d'une part, et sur le renforcement de la segmentation à chaque niveau du système d'autre part.

Notre travail montre également les limites d'un traitement de "bas-niveau" sur les images et la nécessité d'inclure la compétence de traitements de plus "haut-niveau" ; la méthode de stéréovision que nous avons développée en est un exemple.

Mots clés :

- robot mobile - stéréovision - extraction de régions -
- appariement - segments tridimensionnels - surfaces tridimensionnelles -
- interaction entre processus -