



HAL
open science

Mise en correspondance et reconstruction stéréo utilisant une description structurelle des images

Thomas Skordas

► **To cite this version:**

Thomas Skordas. Mise en correspondance et reconstruction stéréo utilisant une description structurelle des images. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG, 1988. Français. NNT: . tel-00329721

HAL Id: tel-00329721

<https://theses.hal.science/tel-00329721v1>

Submitted on 13 Oct 2008

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

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

THESE

présentée par

Thomas Skordas

pour obtenir le titre de **DOCTEUR**

de l'**INSTITUT NATIONAL POLYTECHNIQUE DE
GRENOBLE**

(arrêté ministériel du 5 juillet 1984)

(Spécialité : Informatique)

Mise en correspondance et reconstruction
stéréo utilisant une description structurelle des
images

Date de soutenance : Vendredi 7 Octobre 1988

Composition du jury :

Président : **Philippe JORRAND**
Rapporteurs : **Roger MOHR**
Marc RICHETIN
Examineurs : **Nicholas AYACHE**
Radu HORAUD
Dominique MASSE

Thèse préparée au sein des Laboratoires **D.LETI-CENG / LIFIA-INPG.**

A Marilena

Remerciements

Je tiens à exprimer mes remerciements et ma plus vive gratitude à :

Monsieur *Philippe JORRAND*, Directeur du LIFIA, pour m'avoir fait l'honneur de bien vouloir présider ce jury ;

Messieurs *Roger MOHR* et *Marc RICHTIN*, qui ont accepté d'être rapporteurs de ce mémoire ;

Messieurs *Nicholas AYACHE*, *Radu HORAUD* et *Dominique MASSE*, qui ont bien voulu être examinateurs ;

Monsieur *Jacques MONGE*, de m'avoir accueilli durant trois années dans le Service d'Electronique et d'Informatique pour Systèmes de Traitements (SEIST) qu'il dirige ;

Monsieur *Radu HORAUD*, qui m'a encadré tout au long de ces trois années de travail et qui a joué un rôle tout particulier dans le développement de ce travail de recherche. Je lui suis tout aussi reconnaissant de ses conseils et de l'encouragement qu'il a su me donner et de la confiance qu'il m'a accordée ;

La Fondation Hellénique "*BODOSSAKIS*", qui m'a soutenu financièrement pendant la première année de cette thèse ;

La Fondation Hellénique "*ALEXANDER S. ONASSIS*", qui m'a soutenu financièrement pendant les deux dernières années de cette thèse ;

Messieurs *André JACQUIN* et *Gérard HATCHADOURIAN* pour la gentillesse et la patience dont ils ont fait preuve dans la résolution de mes problèmes "système" ;

Madame *Marie-Claude REOLON* secrétaire du service SEIST pour sa patience et sa volonté de résoudre mes problèmes linguistiques.

Que Messieurs *François CACHARD*, *Jim CROWLEY*, *Dominique DAVID*, *Yves DEMAZEAU*, *François FAVRE-REGUILLON*, *Jean-Luc JACQUOT*, *Augustin LUX*, *Dominique MASSE*, *Alain MERLE* et *Philippe PELTIE* trouvent ici l'expression de ma profonde reconnaissance pour leur disponibilité, leur patience, la qualité de leurs conseils techniques et par-dessus tout leur amitié, qui m'est très chère.

Je tiens à citer la collaboration sympathique et efficace du groupe des chercheurs en Vision par Ordinateur et en général de tous les chercheurs du laboratoire LIFIA.

Je remercie également toutes les personnes que j'ai pu cotoyer, et qui contribuent à cette ambiance chaleureuse et sympathique que j'ai beaucoup appréciée.

Enfin, j'adresse un immense merci à mes *parents* et à ma soeur *Agni*, qui ont sans doute contribué à ma passion pour la recherche, ainsi qu'à *Marilena*, pour sa patience, sa compréhension et tant d'autres choses encore.

Professeurs des Universités

BARIBAUD	Michel	ENSERG	JOUBERT	Jean-Claude	ENSPG
BARRAUD	Alain	ENSIEG	JOURDAIN	Geneviève	ENSIEG
BAUDELET	Bernard	ENSPG	LACOUME	Jean-Louis	ENSIEG
BEAUFILS	Jean-Pierre	ENSEEG	LESIEUR	Marcel	ENSHMG
BLIMAN	Samuel	ENSERG	LESPINARD	Georges	ENSHMG
BLOCH	Daniel	ENSPG	LONGUEUE	Jean-Pierre	ENSPG
BOIS	Philippe	ENSHMG	LOUCHET	François	ENSIEG
BONNETAIN	Lucien	ENSEEG	MASSE	Philippe	ENSIEG
BOUVARD	Maurice	ENSHMG	MASSELOT	Christian	ENSIEG
BRISSONNEAU	Pierre	ENSIEG	MAZARE	Guy	ENSIMAG
BRUNET	Yves	IUFA	MOREAU	René	ENSHMG
CAILLERIE	Denis	ENSHMG	MORET	Roger	ENSIEG
CAVAIGNAC	Jean-François	ENSPG	MOSSIERE	Jacques	ENSIMAG
CHARTIER	Germain	ENSPG	OBLED	Charles	ENSHMG
CHENEVIER	Pierre	ENSERG	OZIL	Patrick	ENSEEG
CHERADAME	Herve	UFR PCP	PARIAUD	Jean-Charles	ENSEEG
CHOVET	Alain	ENSERG	PERRET	René	ENSIEG
COHEN	Joseph	ENSERG	PERRET	Robert	ENSIEG
COUMES	André	ENSERG	PIAU	Jean-Michel	ENSHMG
DARVE	Félix	ENSHMG	POUPOT	Christian	ENSERG
DELLA-DORA	Jean-François	ENSIMAG	RAMEAU	Jean-Jacques	ENSEEG
DEPORTES	Jacques	ENSPG	RENAUD	Maurice	UFR PGP
DESRE	Pierre	ENSEEG	ROBERT	André	UFR PGP
DOLMAZON	Jean-Marc	ENSERG	ROBERT	François	ENSIMAG
DURAND	Francis	ENSEEG	SABONNADIERE	Jean-Claude	ENSIEG
DURAND	Jean-Louis	ENSIEG	SAUCIER	Gabrielle	ENSIMAG
FOGGIA	Albert	ENSIEG	SCHLENKER	Claire	ENSPG
FONLUPT	Jean	ENSIMAG	SCHLENKER	Michel	ENSPG
FOULARD	Claude	ENSIEG	SERMET	Pierre	ENSERG
GANDINI	Alessandro	UFR PGP	SILVY	Jacques	UFR PGP
GAUBERT	Claude	ENSPG	SIRYES	Pierre	ENSHMG
GENTIL	Pierre	ENSERG	SOHM	Jean-Claude	ENSEEG
GREVEN	Hélène	IUFA	SOLER	Jean-Louis	ENSIMAG
QUERIN	Bernard	ENSERG	SOUQUET	Jean-Louis	ENSEEG
GUYOT	Pierre	ENSEEG	TROMPETTE	Philippe	ENSHMG
IVANES	Marcel	ENSIEG	VEILLON	Gérard	ENSIMAG
JAUSSAUD	Pierre	ENSIEG	ZADWORYN	François	ENSERG

Personnes ayant obtenu le diplôme

D'HABILITATION A DIRIGER DES RECHERCHES

BECKER	Monique	DEROO	Daniel	HAMAR	Roger
BINDER	Zdeneck	DIARD	Jean-Paul	LADET	Pierre
CHASSERY	Jean-Marc	DION	Jean-Michel	LATOMBE	Claudine
CHOLLET	Jean-Pierre	DUGARD	Luc	LE GORREC	Bernard
COEY	John	DURAND	Madeleine	MADAR	Roland
COLINET	Catherine	DURAND	Robert	MULLER	Jean
COMMAULT	Christian	GALERIE	Alain	NGUYEN TRONG	Bernadette
CORNUEJOLS	Gérard	CAUTHIER	Jean-Paul	PASTUREL	Alain
COULOMB	Jean-Louis	GENTIL	Sylviane	PLA	Fernand
DALARD	Francis	CHIBAUDO	Gérard	ROUGER	Jean
DANES	Florin	HAMAR	Sylvaine	TCHUENTE	Maurice
				VINCENT	Henri

CHERCHEURS DU C.N.R.S

Directeurs de recherche 1ère Classe

CARRE	René	LANDAU	Ioan
FRUCHART	Robert	VACHAUD	Georges
HOPFINGER	Emile	VERJUS	Jean-Pierre
JORRAND	Philippe		

Directeurs de recherche 2ème Classe

ALEMANY	Antoine	KLEITZ	Michel
ALLIBERT	Colette	KOFMAN	Walter
ALLIBERT	Michel	KAMARINOS	Georges
ANSARA	Ibrahim	LEJEUNE	Gérard
ARMAND	Michel	LE PROVOST	Christian
BERNARD	Claude	MADAR	Roland
BINDER	Gilbert	MERMET	Jean
BONNET	Roland	MICHEL	Jean-Marie
BORNARD	Guy	MUNIER	Jacques
CAILLET	Marcel	PIAU	Monique
CALMET	Jacques	SENATEUR	Jean-Pierre
COURTOIS	Bernard	SIFAKIS	Joseph
DAVID	René	SIMON	Jean-Paul
DRIOLE	Jean	SUERY	Michel
ESCUDIER	Pierre	TEODOSIU	Christian
EUSTATHOPOULOS	Nicolas	VAUCLIN	Michel
GUELIN	Pierre	WACK	Bernard
JOUD	Jean-Charles		

Personnalités agréées à titre permanent à diriger

des travaux de recherche (décision du conseil scientifique)

ENSEEG

CHATILLON	Christian	SARRAZIN	Pierre
HAMMOU	Abdelkader	SIMON	Jean-Paul
MARTIN GARIN	Régina		

ENSERG

BOREL	Joseph		
-------	--------	--	--

ENSIEG

DESCHIZEAUX	Pierre	PERARD	Jacques
GLANGEAUD	François	REINISCH	Raymond

ENSHMG

ROWE	Alain		
------	-------	--	--

ENSIMAG

COURTIN	Jacques		
---------	---------	--	--

EFP

CHARUEL	Robert		
---------	--------	--	--

C.E.N.G

CADET
COEURE
DELHAYE
DUPUY
JOUVE
NICOLAU

Jean
Philippe
Jean-Marc
Michel
Hubert
Yvan

NIFENECKER
PERROUD
PEUZIN
TAIEB
VINCENDON

Hervé
Paul
Jean-Claude
Maurice
Marc

Laboratoires extérieurs :

C.N.E.T

DEVINE
GERBER

Rodericq
Roland

MERCKEL
PAULEAU

Gérard
Yves

UNIVERSITE SCIENTIFIQUE TECHNOLOGIQUE ET MEDICALE
DE GRENOBLE

Président de l'Université :

M. PAYAN Jean Jacques

ANNEE UNIVERSITAIRE 1987 - 1988

MEMBRES DU CORPS ENSEIGNANT DE SCIENCES ET DE GEOGRAPHIE

PROFESSEURS DE 1ERE CLASSE

ARNAUD Paul	Chimie Organique
ARVIEU Robert	Physique Nucléaire I.S.N.
AUBERT Guy	Physique C.N.R.S.
AURIAULT Jean Louis	Mécanique
AYANT Yves	Physique Approfondie
BARBIER Marie Jeanne	Electrochimie
BARJON Robert	Physique Nucléaire I.S.N.
BARNOUD Fernand	Biochimie Macromoléculaire Végétale
BARRA Jean René	Statistiques - Mathématiques Appliquées
BECKER Pierre	Physique
BEGUIN Claude	Chimie Organique
BELORISKY Elie	Physique
BENZAKEN Claude	Mathématiques Pures
BERARD Pierre	Mathématiques Pures
BERNARD Alain	Mathématiques Pures
BERTRANDIAS Françoise	Mathématiques Pures
BERTRANDIAS Jean Paul	Mathématiques Pures
BILLET Jean	Géographie
BOEHLER Jean Paul	Mécanique
BONNIER Jane Marie	Chimie Générale
BOUCHEZ Robert	Physique Nucléaire I.S.N.
BRAVARD Yves	Géographie
CARLIER Georges	Biologie Végétale
CAUQUIS Georges	Chimie Organique
CHARDON Michel	Géographie
CHIBON Pierre	Biologie Animale
COHEN ADDAD Jean Pierre	Physique
COLIN DE VERDIERE Yves	Mathématiques Pures
CYROT Michel	Physique du Solide
DEBELMAS Jacques	Géologie Générale
DEGRANGE Charles	Zoologie
DEMAILLY Jean Pierre	Mathématiques Pures
DENEUVILLE Alain	Physique
DEPORTES Charles	Chimie Minérale
DOLIQUE Jean Michel	Physique des Plasmas
DOUCE Roland	Physiologie Végétale
DUCROS Pierre	Cristallographie
FONTAINE Jean Marc	Mathématiques Pures
GAGNAIRE Didier	Chimie Physique
GERMAIN Jean Pierre	Mécanique
GIRAUD Pierre	Géologie
HICTER Pierre	Chimie
IDELMAN Simon	Physiologie Animale
JANIN Bernard	Géographie
JOLY Jean René	Mathématiques Pure

KAHANE André, détaché	Physique
KAHANE Josette	Physique
KRAKOWIAK Sacha	Mathématiques Appliquées
LAJZEROWICZ Jeanine	Physique
LAJZEROWICZ Joseph	Physique
LAURENT Pierre Jean	Mathématiques Appliquées
LEBRETON Alain	Mathématiques Appliquées
DE LEIRIS Joël	Biologie
LHOMME Jean	Chimie
LLIBOUTRY Louis	Géophysique
LOISEAUX Jean Marie	Sciences Nucléaires I.S.N.
LUNA Domingo	Mathématiques Pures
MACHE Régis	Physiologie Végétale
MASCLE Georges	Géologie
MAYNARD Roger	Physique du Solide
OMONT Alain	Astrophysique
OZENDA Paul	Botanique (Biologie Végétale)
PAYAN Jean Jacques	Mathématiques Pures
PEBAY PEYROULA Jean Claude	Physique
PERRIER Guy	Géophysique
PIERRARD Jean Marie	Mécanique
PIERRE Jean Louis	Chimie Organique
RENARD Michel	Thermodynamique
RINAUDO Marguerite	Chimie C.E.R.M.A.V.
ROSSI André	Biologie
SAXOD Raymond	Biologie Animale
SENGEL Philippe	Biologie Animale
SERGERAERT Francis	Mathématiques Pures
SOUCHIER Bernard	Biologie
SOUTIF Michel	Physique
STUTZ Pierre	Mécanique
TRILLING Laurent	Mathématiques Appliquées
VALENTIN Jacques	Physique Nucléaire I.S.N.
VAN CUTSEM Bernard	Mathématiques Appliquées
VIALON Pierre	Géologie

PROFESSEURS DE 2EME CLASSE

ADIBA Michel	Mathématiques Pures
ANTOINE Pierre	Géologie
ARMAND Gilbert	Géographie
BARET Paul	Chimie
BLANCHI Jean Pierre	S.T.A.P.S.
BLUM Jacques	Mathématiques Appliquées
BOITET Christian	Mathématiques Appliquées
BORNAREL Jean	Physique
BRUANDET Jean François	Physique
BRUGAL Gérard	Biologie
BRUN Gilbert	Biologie
CASTAING Bernard	Physique
CERFF Rudiger	Biologie
CHIARAMELLA Yves	Mathématiques Appliquées
COURT Jean	Chimie
DUFRESNOY Alain	Mathématiques Pures
GASPARD François	Physique
GAUTRON René	Chimie
GENIES Eugène	Chimie
GIDON Maurice	Géologie
GIGNOUX Claude	Sciences Nucléaires
GILLARD Roland	Mathématiques Pures
GIORNI Alain	Sciences Nucléaires
GONZALEZ SPRINBERG Gérardo	Mathématiques Pures
GUIGO Maryse	Géographie
GUMUCHIAN Hervé	Géographie
GUITTON Jacques	Chimie
HACQUES Gérard	Mathématiques Appliquées
HERBIN Jacky	Géographie
HERAULT Jeanny	Physique
JARDON Pierre	Chimie
JOSELEAU Jean Paul	Biochimie
KERCKHOVE Claude	Géologie
LONGEQUEUE Nicole	Sciences Nucléaires I.S.N
LUCAS Robert	Physique
MANDARON Paul	Biologie
MARTINEZ Francis	Mathématiques Appliquées
NEMOZ Alain	Thermodynamique C.N.R.S. C.R.T.B.T.
OUDET Bruno	Mathématiques Appliquées
PECHER Arnaud	Géologie
PELMONT Jean	Biochimie
PERRIN Claude	Sciences Nucléaires I.S.N.
PFISTER Jean Claude	Physique du Solide
PIBOULE Michel	Géologie
RAYNAUD Hervé	Mathématiques Appliquées
RICHARD Jean Marc	Physique
RIEDTMANN Christine	Mathématiques Pures
ROBERT Gilles	Mathématiques Pures
ROBERT Jean Bernard	Chimie Physique
SARROT REYNAULD Jean	Géologie
SAYETAT Françoise	Physique
SERVE Denis	Chimie
STOECKEL Frédéric	Physique
SCHOLL Pierre Claude	Mathématiques Appliquées
SUBRA Robert	Chimie
VALLADE Marcel	Physique
VIDAL Michel	Chimie Organique
VIVIAN Robert	Géographie
VOTTERO Philippe	Chimie

MEMBRES DU CORPS ENSEIGNANT DE L'I.U.T. 1

PROFESSEURS DE 1ERE CLASSE

BUISSON Roger	Physique I.U.T. 1
DODU Jacques	Mécanique Appliquée I.U.T. 1
NEGRE Robert	Génie Civil I.U.T. 1
NOUGARET Marcel	Automatique I.U.T. 1
PERARD Jacques	E.E.A. I.U.T. 1

PROFESSEURS DE 2EME CLASSE

BOUTHINON Michel	E.E.A. I.U.T. 1
CHAMBON René	Génie Mécanique I.U.T. 1.
CHEHIKIAN Alain	E.E.A. I.U.T. 1
CHENAVAS Jean	Physique I.U.T. 1
CHOUTEAU Gérard	Physique I.U.T. 1
CONTE René	Physique I.U.T. 1
GOSSE Jean Pierre	E.E.A. I.U.T. 1
GROS Yves	Physique I.U.T. 1
KUHN Gérard, détaché	Physique I.U.T. 1
MAZUER Jean	Physique I.U.T. 1
MICHOULIER Jean	Physique I.U.T. 1
MONLLOR Christian	E.E.A. I.U.T. 1
PEFFEN René	Métallurgie I.U.T. 1
PERRAUD Robert	Chimie I.U.T. 1
PIERRE Gérard	Chimie I.U.T. 1
TERRIEZ Jean Michel	Génie Mécanique I.U.T. 1
TOUZAIN Philippe	Chimie I.U.T. 1
VINCENDON Marc	Chimie I.U.T. 1

PROFESSEURS DE PHARMACIE

AGNIUS DELORD Claudine	Physique	Faculté 1a Tronche
ALARY Josette	Chimie Analytique	Faculté 1a Tronche
BERIEL Hélène	Physiologie et Pharmacologie	Faculté 1a Tronche
CUSSAC Max	Chimie Therapeutique	Faculté 1a Tronche
DEMENGE Pierre	Pharmacodynamie	Faculté 1a Tronche
FAVIER Alain	Biochimie	C.H.R.G.
JEANNIN Charles	Pharmacie Galenique	Faculté Meylan
LATURAZE Jean	Biochimie	Faculté 1a Tronche
LUU DUC Cuong	Chimie Générale	Faculté 1a Tronche
MARIOTTE Anne-Marie	Pharmacognosie	Faculté 1a Tronche
MARZIN Daniel	Toxicologie	Faculté Meylan
RENAUDET Jacqueline	Bactériologie	Faculté 1a Tronche
ROCHAT Jacques	Hygiène et Hydrologie	Faculté 1a Tronche
SEIGLE MURANDI Françoise	Botanique et Cryptogamie	Faculté Meylan
VERAIN Alice (Chimie thérapeutique)	Pharmacie galénique	Faculté Meylan

MEMBRES DU CORPS ENSEIGNANT DE MEDECINE

PROFESSEURS CLASSE EXCEPTIONNELLE ET IERE CLASSE

AMBLARD Pierre	Dermatologie	C.H.R.G.
AMBROISE-THOMAS Pierre	Parasitologie	C.H.R.G.
BEAUDOING André	Pédiatre - Puériculture	C.H.R.G.
BEZES Henri	Orthopédie - Traumatologie	Hopital Sud
BONNET Jean-Louis	Ophtalmologie	C.H.R.G.
BOUCHET Yves	Anatomie	Faculté 1a Merci
	Chirurgie Générale et Digestive	C.H.R.G.
BUTEL Jean	Orthopédie - Traumatologie	C.H.R.G.
CHAMBAZ Edmond	Biochimie	C.H.R.G.
CHAMPETIER Jean	Anatomie Topographique et Appliquée	C.H.R.G.
CHARACHON Robert	O.R.L.	C.H.R.G.
COLOMB Maurice	Immunologie	Hopital Sud
COUDERC Pierre	Anatomie Pathologique	C.H.R.G.
DELORMAS Pierre	Pneumophtisiologie	C.H.R.G.
DENIS Bernard	Cardiologie	C.H.R.G.
GAVEND Michel	Pharmacologie	Faculté 1a Merci
HOLLARD Daniel	Hématologie	C.H.R.G.
LATREILLE René	Chirurgie Thoracique/Cardiovasculaire	C.H.R.G.
LE NOC Pierre	Bactériologie - Virologie	Faculté 1a Merci
MALINAS Yves	Gynécologie et Obstétrique	C.H.R.G.
MALLION Jean Michel	Médecine du Travail	C.H.R.G.
MICOUD Max	Clinique Médicale/Maladies Infectieuses	C.H.R.G.
MOURIQUAND Claude	Histologie	Faculté 1a Merci
PARAMELLE Bernard	Pneumologie	C.H.R.G.
PERRET Jean	Neurologie	C.H.R.G.
RACHAIL Michel	Hépto-Gastro-Entérologie	C.H.R.G.
DE ROUGEMONT Jacques	Neurochirurgie	C.H.R.G.
SARRAZIN Roger	Clinique Chirurgicale	C.H.R.G.
STIEGLITZ Paul	Anesthésiologie	C.H.R.G.
TANCHE Maurice	Physiologie	Faculté 1a Merci
VIGNAIS Pierre	Biochimie	Faculté 1a Merci

PROFESSEURS 2EME CLASSE

BACHELOT Yvan	Endocrinologie	C.H.R.G.
BARGE Michel	Neurochirurgie	C.H.R.G.
BENABID Alim Louis	Biophysique	Faculté 1a Merci
BENSA Jean Claude	Immunologie	Hopital Sud
BERNARD Pierre	Gynécologie - Obstétrique	C.H.R.G.
BESSARD Germain	Pharmacologie	Abidjan
BOLLA Michel	Radiothérapie	C.H.R.G.
BOST Michel	Pédiatrie	C.H.R.G.
BOUCHARLAT Jacques	Psychiatrie Adultes	Hopital Sud
BRAMBILLA Christian	Pneumologie	C.H.R.G.
CHIROUSSEL Jean Paul	Anatomie - Neurochirurgie	C.H.R.G.
COMET Michel	Biophysique	Faculté 1a Merci
CONTAMIN Charles	Chirurgie Thoracique/Cardiovasculaire	C.H.R.G.
CORDONNIER Daniel	Néphrologie	C.H.R.G.
COULOMB Max	Radiologie	C.H.R.G.
CROUZET Guy	Radiologie	C.H.R.G.
DEBRU Jean Luc	Médecine Interne et Toxicologie	C.H.R.G.
DEMONGEOT Jacques	Biostatistiques/Informatique Médicale	Faculté 1a Merci
DUPRE Alain	Chirurgie Générale	C.H.R.G.
DYON Jean François	Chirurgie Infantile	C.H.R.G.
ETERRADOSSI Jacqueline	Physiologie	Faculté 1a Merci
FAURE Claude	Anatomie et Organogénèse	C.H.R.G.
FAURE Gilbert	Urologie	C.H.R.G.
FOURNET Jacques	Hépto-Gastro-Entérologie	C.H.R.G.
FRANCO Alain	Médecine Interne	C.H.R.G.
GIRARDET Pierre	Anesthésiologie	C.H.R.G.
GUIDICELLI Henri	Chirurgie Générale et Vasculaire	C.H.R.G.
GUIGNIER Michel	Thérapeutique et Réanimation Médicale	C.H.R.G.
HADJIAN Arthur	Biochimie	Faculté 1a Merci
HALIMI Serge	Endocrinologie et Maladies Métaboliques	C.H.R.G.
HOSTEIN Jean	Hépto-Gastro-Entérologie	C.H.R.G.
HUGONOT Robert	Médecine Interne	C.H.R.G.
JALBERT Pierre	Histologie - Cytogénétique	Faculté 1a Merci
JUNIEN-LAVILLAULOY Claude	O.R.L.	C.H.R.G.
KOLODIE Lucien	Hématologie Biologique	C.H.R.G.
LETOUBLON Christian	Chirurgie Générale	C.H.R.G.
MACHECOURT Jacques	Cardiologie et Maladies Vasculaires	C.H.R.G.
MAGNIN Robert	Hygiène	C.H.R.G.
MASSOT Christan	Médecine Interne	C.H.R.G.
MOUILLON Michel	Ophtalmologie	C.H.R.G.
PELLAT Jacques	Neurologie	C.H.R.G.
PHELIP Xavier	Rhumatologie	C.H.R.G.
RACINET Claude	Gynécologie-Obstétrique	Hopital Sud
RAMBAUD Pierre	Pédiatrie	C.H.R.G.
RAPHAEL Bernard	Stomatologie	C.H.R.G.
SCHAERER René	Cancérologie	C.H.R.G.
SEIGNEURIN Jean Marie	Bactériologie - Virologie	Faculté 1a Merci
SELE Bernard	Cytogénétique	Faculté 1a Merci
SOTTO Jean Jacques	Hématologie	C.H.R.G.
STOEBNER Pierre	Anatomie Pathologique	C.H.R.G.
VROUSOS Constantin	Radiothérapie	C.H.R.G.

"Αλογον δὲ ὅλως τὸ ἐξίοντι τινὶ τὴν ὄψιν ὄραν,
 καὶ ἀποτείνεσθαι μέχρι τῶν ἀστρων, ἢ μέχρι τινὸς
 ἐξιοῦσαν συμφύεσθαι, καθάπερ λέγουσιν τινες.
 τούτου μὲν γὰρ βέλτιον τὸ ἐν ἀρχῇ συμφύεσθαι τοῦ
 ὀμματος. ἀλλὰ καὶ τοῦτο εὐήθες· τὸ τε γὰρ συμ-
 30 φύεσθαι τί ἐστὶ φωτὶ πρὸς φῶς; ἢ πῶς οἶόν θ'
 438 b ὑπάρχειν; οὐ γὰρ τῷ τυχόντι συμφύεται τὸ τυχόν.
 τό τ' ἐντὸς τῷ ἐκτὸς πῶς; ἢ γὰρ μῆνυξ μεταξὺ
 ἐστίν. περὶ μὲν οὖν τοῦ ἀνευ φωτὸς μὴ ὄραν
 εἶρηται ἐν ἄλλοις· ἀλλ' εἴτε φῶς εἴτ' ἀήρ ἐστὶ τὸ
 μεταξὺ τοῦ ὀρωμένου καὶ τοῦ ὀμματος, ἢ διὰ
 438 b τούτου κίνησις ἐστὶν ἢ ποιούσα τὸ ὄραν. καὶ
 εὐλόγως τὸ ἐντὸς ἐστὶν ὕδατος· διαφανὲς γὰρ τὸ
 ὕδωρ. ὀράται δὲ ὡσπερ καὶ ἔξω οὐκ ἀνευ φωτὸς,
 οὔτω καὶ ἐντὸς· διαφανὲς ἄρα δεῖ εἶναι. καὶ ἀνάγκη
 ὕδωρ εἶναι, ἐπειδὴ οὐκ ἀήρ. οὐ γὰρ ἐπὶ τοῦ
 ἐσχάτου ὀμματος ἢ ψυχῇ ἢ τῆς ψυχῆς τὸ αἰσθητή-
 10 ριόν ἐστίν, ἀλλὰ δῆλον ὅτι ἐντὸς· διόπερ ἀνάγκη
 διαφανὲς εἶναι καὶ δεκτικὸν φωτὸς τὸ ἐντὸς τοῦ
 ὀμματος.

ARISTOTE

Extrait de "Du sens et des sensibles"

ARISTOTE et la Vision :

Aristote rejette les théories admises jusqu'à son époque, sur la lumière et la vue, selon lesquelles l'œil émet des radiations, comme un corps lumineux, et il admet que celles-ci viennent irriter la rétine et c'est ainsi que l'on a la sensation de la vue. Ensuite cette irritation est transmise au cerveau par des "pores" et c'est de cette façon qu'est produite l'image.

L'existence de l'éther est indispensable à la vue. Autrement, s'il y a "vide" il n'y a pas de vue. Comme par exemple, s'il n'y avait pas d'air, il n'y aurait pas d'ouïe.

Il soutient que la lumière se transmet en ligne droite, et se reflète d'après l'angle lui même. Il connaît aussi que pour la transmission de la lumière un certain laps de temps est indispensable. Il décrit les moyens qui permettent d'améliorer la vue. On sait qu'à son époque, on connaissait déjà les lentilles et les miroirs.

Sommaire

1	Introduction	3
1.1	La Vision par Ordinateur et l'extraction d'un modèle de l'espace	3
1.2	Description de la vision stéréoscopique	5
1.3	La mise en correspondance de paires d'images stéréoscopiques	8
1.3.1	Qu'est-ce que la mise en correspondance ? Pourquoi est-elle difficile ?	8
1.3.2	Les approches antérieures	9
1.4	La reconstruction de surfaces	25
1.4.1	Extraction de la profondeur	25
1.4.2	Extraction de surfaces	26
1.5	Notre Approche	31
1.6	Les travaux voisins	33
1.7	Organisation de la thèse	33
2	Description monoculaire d'une image	37
2.1	Organisation de l'information visuelle	37
2.2	Notre approche	38
2.3	Les motivations de ce choix	40
2.4	Le groupement	41
2.4.1	Détection et transformation de jonctions en T	42
2.4.2	Détection de jonctions	46
2.4.3	Détection de bandes associées aux segments	52
2.5	Construction d'une représentation de l'image	54
2.6	Résultats expérimentaux	55
2.7	Conclusions	58

3	La mise en correspondance	63
3.1	Contraintes géométriques pour la mise en correspondance	63
3.1.1	Les contraintes géométriques appliquées aux points image	64
3.1.2	Les contraintes géométriques appliquées aux segments	67
3.1.3	Le rôle des contraintes pour la restriction du nombre d'appariements potentiels	72
3.2	Mesure de similarité associée aux appariements potentiels	73
3.3	Mise en correspondance des structures 2-D	74
3.3.1	Construction des arêtes du graphe de correspondance	75
3.3.2	Exploration du graphe de correspondance	84
3.3.3	Quelques cas difficiles	93
3.3.4	Complexité	95
3.4	Résultats expérimentaux	96
3.4.1	Mise en correspondance locale	96
3.4.2	Mise en correspondance globale	97
3.5	Discussion	100
4	La reconstruction de surfaces	115
4.1	Extraction de la profondeur	115
4.1.1	Extraction de la profondeur pour un couple de points homologues	115
4.1.2	Extraction de la profondeur pour un couple de segments homologues	118
4.1.3	Reconstruction des segments 3-D en pratique	119
4.1.4	Résultats expérimentaux	120
4.2	Reconstruction des surfaces	126
4.2.1	Le rôle du groupement à l'extraction de surfaces planes	126
4.2.2	Caractérisation d'une surface plane	130
4.2.3	Résultats expérimentaux	132
5	Conclusions générales	135
A	Modélisation du capteur stéréo	139
B	L'algorithme des cliques maximales	143
C	Le filtre de Kalman étendu appliqué à notre système	145
C.1	Présentation du filtre de Kalman étendu	145
C.2	Application du filtre de Kalman à notre système	146

Chapitre 1

Introduction

1.1 La Vision par Ordinateur et l'extraction d'un modèle de l'espace

Un être humain n'éprouve aucune difficulté lorsqu'il veut décrire le monde l'entourant, se déplacer dans son environnement ou saisir et manipuler les objets de son voisinage. Une des principales sources d'information que l'homme possède pour interagir avec l'environnement est la vision, un processus qui consiste à découvrir à partir des images ce qui est présent dans le monde qui l'entoure.

La construction des machines de calcul de plus en plus sophistiquées a donné l'occasion à l'homme de franchir une première étape vers son rêve, à savoir la création d'un agent artificiel intelligent à son image et à son service. Etant donné que la vue est d'une importance capitale pour l'intelligence humaine, l'homme s'est alors intéressé d'une part à équiper ses machines de capteurs visuels capables de remplir la fonction de l'oeil humain, et d'autre part à traiter de cette information visuelle pour créer un modèle tridimensionnel de l'espace. C'est ainsi que la Vision par Ordinateur (V.O.) est née.

Voici une définition brève de la V.O. : la V.O. est un processus de traitement de l'information, consistant à prendre en entrée des données sensorielles provenant de capteurs visuels, à traiter ensuite convenablement ces informations visuelles par un certain nombre de méthodes spécifiques et finalement à produire en sortie une *description symbolique* des données de l'entrée en termes d'objets et de relations entre ces objets. La description symbolique de la scène représentée en entrée dans une

séquence d'images est composée d'un certain nombre de connaissances qui peuvent être apportées à tous les niveaux.

Cette thèse concerne la V.O. et plus précisément l'utilisation d'un ensemble de méthodes de la V.O. permettant la restitution de l'information tridimensionnelle à partir de l'information visuelle, c'est-à-dire la récupération de la profondeur et de l'orientation des surfaces et des objets représentés dans une ou plusieurs images de la scène.

L'espace qui nous entoure a une structure géométrique tridimensionnelle. Par conséquent le modèle de l'environnement doit être organisé par la géométrie. L'acquisition d'un modèle géométrique tridimensionnel de la scène est parmi les buts fondamentaux de la V.O. Son importance est essentiellement due aux plusieurs applications qu'elle implique et notamment en robotique, pour la navigation de robots mobiles, la reconnaissance, le positionnement et l'inspection de qualité des objets industriels.

A l'heure actuelle la V.O. n'est pas encore une discipline assez mûre pour pouvoir proposer une solution universelle au problème de l'extraction de la troisième dimension et de reconstruction de scène. Plusieurs méthodes ont été développées recherchant une solution satisfaisante. Ces méthodes se répartissent en trois grandes classes :

- Les méthodes dépendant de l'information monoculaire disponible dans une seule image ;
- Les méthodes traitant directement des informations de distance fournies par un capteur "optiquement actif" ;
- Les méthodes utilisant au moins deux images prises de points de vue différents et éventuellement à des instants différents : *stéréovision généralisée*.

Les méthodes d'*Inférence de formes* monoculaires fournissent des informations 3-D sur les surfaces présentes dans l'image à partir des connaissances géométriques (analyse des contours), géométriques et statistiques (analyse des textures), photométriques (analyse des ombrages) ou colorimétriques (analyse des couleurs). Ces méthodes n'utilisent qu'une seule image pour inférer la profondeur ou l'orientation. Cependant une seule image n'est pas suffisante car les configurations locales de l'intensité image ne contiennent pas suffisamment d'informations pour estimer tous les paramètres caractérisant la forme d'une surface. Ces méthodes se basent alors sur quelques heuristiques supplémentaires reliant les caractéristiques des images et celles de la scène. Par exemple l'une des hypothèses pour l'analyse des éléments de texture est son homogénéité sur la surface, ce qui permet d'affirmer que toute variation observée dans l'image est due au changement soit de forme soit de point de vue. Ces méthodes sont alors limitées à des cas spéciaux, comme à des régions relativement uniformes en reflectance ou en texture. Généralement, elles fournissent une précision sur la profondeur inférieure à celle fournie par les autres méthodes.

Les méthodes actives permettent d'acquérir directement la troisième dimension en faisant intervenir une source supplémentaire d'éclairage, comme la lumière

structurée d'un laser. Le principe de ces méthodes consiste à éclairer ponctuellement la scène par un faisceau laser et à former l'image de l'endroit éclairé sur un capteur non coaxial avec la source d'éclairage. Ces méthodes estiment la profondeur par un calcul de triangulation entre les rayons optiques définis par la source et le(s) capteur(s)-image. Ces méthodes utilisent donc des informations géométriques pour calculer la profondeur et fournissent des résultats assez précis. Néanmoins elles ont des limitations inhérentes : interférence avec le milieu, confusion possible avec les émissions d'autres capteurs actifs voisins, imposition des conditions très contraignantes d'éclairage. Elles ne sont utiles que dans des domaines étroitement contrôlés tels que les applications de l'automatisation industrielle en environnement contraint, et elles sont mal adaptées à des problèmes de vision plus généraux.

Le paradigme classique de la stéréovision généralisée comprend la *vision stéréoscopique* et le *flux optique*, appelé aussi *mouvement*. En flux optique, plusieurs images sont prises d'une façon séquentielle par une seule caméra qui se déplace dans l'espace et sans connaissances à priori quant aux relations géométriques entre les différentes positions spatiales de la caméra. Les différences importantes observées entre les différentes images sont alors dues d'une part aux différentes positions de prise de vue de la scène et d'autre part aux changements éventuels de la scène elle-même, qui est supposée être une scène dynamique, c'est-à-dire un ou plusieurs objets de la scène traitée peuvent être en mouvement.

Déterminer le mouvement est un objectif ambitieux et difficile à résoudre car il faut trouver plusieurs inconnues à la fois : le mouvement de la caméra, le mouvement des objets qui se déplacent par rapport à la caméra ou les deux simultanément. Ensuite, à partir de ces mesures, le mouvement infère la structure 3-D de la scène.

Le cas de la vision stéréoscopique peut être considéré comme un cas spécial du mouvement et le même formalisme mathématique peut s'appliquer dans tous les cas. Par rapport au mouvement, en vision stéréoscopique on suppose que la scène traitée reste invariante et statique d'une image à l'autre, par conséquent les deux images sont différentes uniquement à cause de leur prise sous des angles de vue différents. Qui plus est, en vision stéréoscopique la relation géométrique entre les positions spatiales des caméras est connue à l'avance par une phase préalable de calibrage.

Dans le cadre de cette thèse nous nous proposons d'étudier le problème de la vision stéréoscopique. La vision stéréoscopique est une source d'informations très séduisante pour la perception de l'environnement. Comparée à des techniques monoculaires, elle conduit à des mesures de distance plus directes et moins ambiguës et elle est généralement dépourvue d'hypothèses statistiques ou photométriques sur la scène à traiter. De plus, c'est une méthode passive, donc elle est facilement adaptable à plusieurs problèmes généraux de la vision.

1.2 Description de la vision stéréoscopique

La vision stéréoscopique aussi appelée *vision binoculaire* est une des méthodes les plus populaires et les plus étudiées pour reconstituer la forme et la profondeur des surfaces des objets. La raison principale est que la vision stéréoscopique humaine

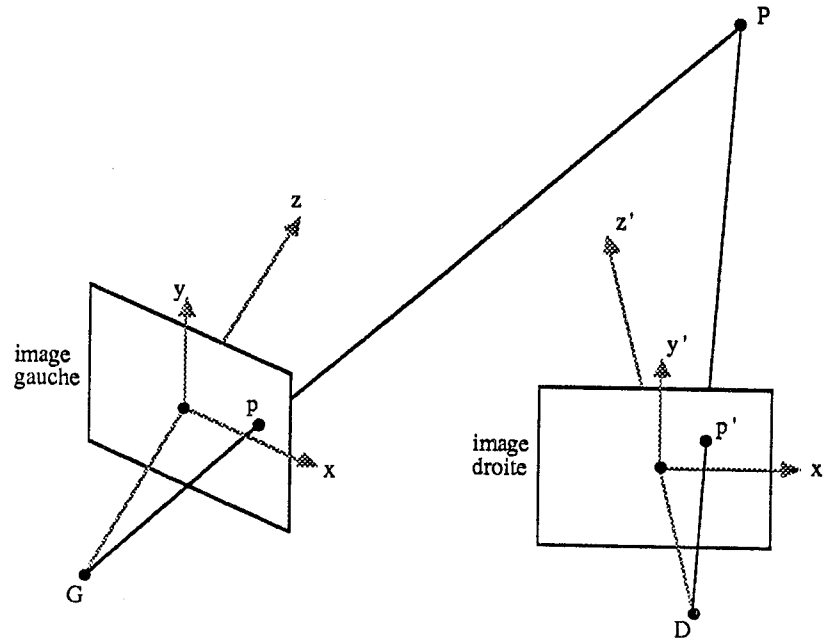


Figure 1.1: La géométrie d'un système stéréoscopique.

constitue une des sources primaires fondamentales que possède l'être humain pour percevoir son environnement.

La Figure 1.1 illustre le principe bien connu de la vision stéréoscopique. Deux images sont acquises simultanément par deux caméras légèrement éloignées l'une de l'autre. Le modèle du système optique généralement utilisé en V.O. est le modèle de sténopé, où une caméra est modélisée par son axe optique perpendiculaire au plan image et par son centre optique situé sur l'axe optique à une distance f (distance focale) du plan image. Selon ce modèle, l'image d'un point physique est la projection perspective de ce point sur le plan image. La relation géométrique entre les deux positions spatiales des caméras est parfaitement connue à l'avance. A cause de la position différente des deux caméras, leurs vues de la scène 3-D sont légèrement différentes. Le but de la vision stéréoscopique est de mettre en correspondance les informations bidimensionnelles homologues des deux images et de reconstituer ensuite l'information 3-D de la scène.

Décrivons succinctement le modèle géométrique présenté sur la Figure 1.1. Un point P dans l'espace se projette dans l'image gauche au point p et dans l'image droite au point p' . Du fait que l'on observe la scène de deux positions différentes, les coordonnées (x, y) et (x', y') dans chaque plan image des deux projections p et p' sont différentes.

Définissons à l'occasion la notion de *disparité*. La disparité binoculaire associée au point physique P est définie comme la différence des positions de ses deux projections sur chaque plan image. Pour le système stéréo de la Figure 1.1, la disparité du point P est donnée par la distance :

$$D = \sqrt{(x - x')^2 + (y - y')^2} \quad (1.1)$$

La disparité de P dépend de sa position dans la scène et de la relation géométrique des caméras du système stéréo.

En stéréoscopie, on dispose alors d'un ensemble de projections p dans l'image gauche et d'un autre ensemble de projections p' dans l'image droite. Le but est de retrouver la position des points P dans l'espace. Le problème se décompose ainsi en deux étapes élémentaires :

1. Dans les deux images, établir la correspondance entre les points homologues provenant du même point physique de la scène ; ceci est équivalent à déterminer en chaque point image la fonction de disparité.
2. Connaissant la géométrie des caméras, reconstruire par triangulation le point P dans l'espace. P se localisera à l'intersection des deux lignes droites pG et $p'D$, où G et D sont les centres optiques des caméras gauche et droite, respectivement.

Cette approche logique du problème stéréo implique dans sa conception un certain nombre de problèmes associés non négligeables : quels sont les indices visuels qu'il faut extraire des images, qui soient les mieux adaptés et les plus facilement reconnaissables pour les mettre en correspondance ? Comment faut-il les apparier ? Comment modéliser correctement la géométrie du système stéréo afin de retrouver avec précision la troisième dimension ? Comment extraire la profondeur tout en sachant que l'intervention de phénomènes de bruit est telle que, les deux lignes de vue pG et $p'D$ n'ont pas d'intersection pratiquement jamais dans l'espace ? Comment retrouver les surfaces, auxquelles les différents indices visuels apparierés appartiennent ? Une description détaillée de ces problèmes se trouve en [10] et en [30].

Tenant compte de tous ces problèmes, un système de vision stéréoscopique peut se décomposer en trois étapes principales :

1. La modélisation du système (la modélisation des caractéristiques internes et externes des deux caméras composant le système stéréoscopique).
2. La mise en correspondance de la paire d'images stéréoscopique.
3. La reconstruction des surfaces de la scène.

En ce qui concerne la première étape, nous renvoyons le lecteur intéressé aux nombreux travaux bibliographiques existants, [31], [45], [55], [83]. Nous allons donc nous pencher sur les deux étapes suivantes, en commençant par celle considérée la plus difficile, à savoir la mise en correspondance des images.

1.3 La mise en correspondance de paires d'images stéréoscopiques

1.3.1 Qu'est-ce que la mise en correspondance ? Pourquoi est-elle difficile ?

Il est généralement admis que l'étape clé de la vision stéréoscopique est le processus de mise en correspondance des images. Ce processus consiste à identifier entre les deux images des points homologues, c'est-à-dire des points provenant d'un même point physique de l'espace 3-D.

Le problème de la mise en correspondance est difficile à résoudre, essentiellement à cause de son aspect combinatoire. D. Marr [53] discute brièvement ce problème par rapport au problème de fausse cible. Lorsque l'on réalise l'enregistrement d'une scène 3-D par une caméra, il existe une relation du type *projection plane* entre les points 3-D du monde réel et leurs correspondants dans le plan image. Cette projection perspective déforme les informations 3-D représentant la scène quand celle-ci est projetée sur le récepteur. Les distorsions géométriques sont telles que les propriétés physiques et la géométrie des caractéristiques de scène ne sont en général plus conservées au plan image. Cette déformation des informations 3-D est d'ailleurs différente d'une image à l'autre à cause de la non-stabilité de la projection perspective d'un endroit à l'autre ; ceci implique que les caractéristiques 3-D et leurs relations n'apparaissent pas de la même manière dans les deux images et par conséquent la géométrie locale d'une image est assez différente de celle de l'autre. Il en résulte une première difficulté quant à l'identification de caractéristiques homologues dans les deux images. De plus, des points visibles dans une image sont absents dans l'autre à cause du changement de point de vue.

En outre, ce dont on dispose, ce sont des données photométriques d'une scène physique projetée dans deux images différentes. Etant donné un point dans une image, il peut être apparié à un grand nombre de points ayant les mêmes propriétés photométriques dans l'autre image. Par ailleurs, l'appariement des deux images selon ce principe devient encore plus difficile du fait de points homologues ne présentant pas de propriétés photométriques stables d'une image à l'autre. En effet leur luminance peut être assez différente à cause de conditions d'éclairage différentes des deux prises de vue et du bruit intrinsèque à tout système de formation d'images.

La présence de structures périodiques dans la scène peut aussi troubler la mise en correspondance, lorsque les caractéristiques-image générées par ces structures sont assez proches comparées à leur disparité. Il est probable que dans ce cas le processus de mise en correspondance confonde une caractéristique d'une structure dans une image avec des caractéristiques des parties voisines de la structure dans l'autre image.

Finalement, un autre problème qui perturbe souvent la mise en correspondance est le problème des reflets, provoqués par la présence dans la scène de surfaces polies, agissant comme un miroir. En réalité, les deux caméras ne voient pas la surface polie, mais les images réfléchies par cette surface compte tenu de leur direction de

vue. Sur les endroits des images occupés par ce type de surfaces, quasiment aucune correspondance n'est possible.

1.3.2 Les approches antérieures

Le problème de mise en correspondance des images stéréo a suscité depuis quelques années de nombreux travaux de recherche et il continue d'être à l'heure actuelle un champ de recherche très actif. De nouveaux algorithmes sont étudiés soit pour réduire la complexité combinatoire soit pour augmenter la probabilité d'appariements corrects.

Les algorithmes divers de mise en correspondance d'une paire d'images stéréo diffèrent sur trois points fondamentaux :

- Le choix des caractéristiques-image à mettre en correspondance ;
- les contraintes mises en oeuvre pour réduire la complexité de la mise en correspondance ;
- les stratégies utilisées pour appairer les caractéristiques choisies.

Nous allons discuter brièvement dans la suite chacun de ces trois aspects, illustrés par un certain nombre de travaux proposés dans la littérature.

1.3.2.1 Le choix des caractéristiques

Un problème préliminaire du problème de mise en correspondance des images est celui du choix des caractéristiques à appairer dans les deux images.

Idéalement, la mise en correspondance devrait s'établir sur chaque pixel individuel et ceci pour toutes les deux images. Cela engendrerait une carte très dense lors de la reconstruction de la scène. Néanmoins, il est évident que l'information contenue dans la valeur de l'intensité lumineuse d'un seul pixel est très pauvre pour résoudre le problème combinatoire. Pour limiter le nombre de candidats potentiels et diminuer la complexité algorithmique, il est nécessaire d'extraire au préalable des deux images un ensemble de caractéristiques visuelles qui contienne des informations pertinentes dans les images et qui le distingue clairement de son voisinage. De cette manière, trois grandes classes d'approches ont été développées :

- Des techniques basées sur l'appariement des régions.
- Des techniques basées sur la mise en correspondance de points d'intérêt.
- Des techniques qui mettent en correspondance des caractéristiques géométriques extraites séparément de deux images.

Appariement des régions

Cette classe d'approches peut se scinder en deux catégories.

La première met en correspondance simplement des ensembles de pixels voisins (fenêtres) dans les deux images, sans aucun souci concernant le contenu de ces fenêtres. Ces méthodes utilisent des opérateurs bidimensionnels qui mesurent le degré de ressemblance des intensités locales dans des fenêtres des deux images [38]. Pour réduire la complexité, le choix de fenêtres est souvent réalisé en sélectionnant de fenêtres autour d'un point image présentant une variance d'intensité élevée par rapport à ses pixels voisins [32].

La deuxième catégorie met en correspondance des régions, c'est-à-dire des *zones* des images provenant d'une même entité physique de la scène [22], [62], [87]. Ces zones sont constituées cette fois-ci de collections de points voisins possédant un certain nombre de propriétés communes (intensité lumineuse, texture, couleur, etc.). Ce genre d'approches propose alors de découvrir d'abord dans chaque image séparément des ensembles de pixels *cohérents* (segmentation des images en régions) et ensuite de mettre en correspondance les régions des deux images ayant été caractérisées par des propriétés physiques quasi-identiques. Il s'agit d'une approche à priori séduisante car les régions portent une information intrinsèque très riche (paramètres de surface, de forme, etc.). Malheureusement, il est très difficile d'extraire correctement des régions dans de scènes naturelles et les différentes techniques utilisées dans la littérature ne fournissent pas de résultats satisfaisants. Une présentation détaillée du problème de l'extraction des régions se trouve en [63].

Ces deux catégories d'appariement de régions ne tiennent pas compte de la géométrie de la scène : elles n'utilisent que ses propriétés photométriques. Par conséquent, ces techniques sont sensibles aux changements de l'intensité absolue, du contraste et de l'illumination. Dans la pratique, elles s'avèrent extrêmement instables.

Appariement des points d'intérêt

Ces techniques n'utilisent que quelques points dans les deux images qui se distinguent clairement de leurs voisins. Dans la plupart des cas il s'agit de points extraits par l'opérateur de Moravec [64] ; cet opérateur sélectionne les points images qui présentent une grande variance de luminance par rapport à ses pixels voisins dans les quatre directions. Si le même opérateur s'applique aux deux images stéréo, alors les points choisis ne sont pas forcément homologues. L'approche consiste donc à localiser dans une image avec l'opérateur d'intérêt des endroits (zones de pixels) qui peuvent être appariés avec une certaine confiance à des points correspondants dans la seconde image [37].

Les reproches que nous faisons à ces techniques sont les mêmes que ceux que nous faisons aux techniques d'appariement des régions. De plus, la reconstruction de la scène produite à partir de cette approche est une représentation tridimensionnelle très pauvre : elle fournit une carte de profondeur très creuse.

Mise en correspondance de caractéristiques-image

La troisième catégorie d'approches met en correspondance de caractéristiques extraites de chaque image. Il s'agit de caractéristiques provenant de changements de l'intensité lumineuse plutôt que de valeurs individuelles de l'intensité. Ce genre d'approches essaie de découvrir des propriétés géométriques de la scène en transformant les données de la matrice d'intensité lumineuse en une forme plus abstraite. Ces approches extraient des paramètres physiques importantes sur quelques points des images et elles comparent ensuite ces paramètres physiques pour établir la mise en correspondance.

Les caractéristiques extraites des images proviennent des changements brusques des valeurs voisines de la matrice de l'intensité lumineuse. Le principe sous-jacent est que les discontinuités en intensité représentent des discontinuités sur les surfaces physiques de la scène ; ces discontinuités sont provoquées par des discontinuités de profondeur, de changement de l'orientation, de réflectance ou d'illumination des surfaces physiques. Mettant en correspondance les caractéristiques-image dérivées de ces discontinuités, on met en correspondance des points physiques sur les surfaces des objets.

L'inconvénient principal de ce genre d'approches est que, comme elles établissent la correspondance sur quelques points des images seulement, elles engendrent des cartes de profondeurs éparses, qui nécessitent des traitements supplémentaires considérables pour reconstruire une description symbolique de scène utile et riche. Néanmoins, ces techniques possèdent deux avantages considérables vis-à-vis des techniques des deux premières classes :

- elles sont moins sensibles à des variations photométriques, puisqu'elles exploitent des propriétés géométriques de la scène ;
- en général, le résultat de leur mise en correspondance est plus précis.

Les caractéristiques-image les plus souvent utilisées dans la littérature pour la mise en correspondance sont les points de contour. Probablement l'algorithme le plus connu en vision stéréo mettant en correspondance de points de contour est celui de Marr et Poggio [54], implanté par Grimson [36]. Bien que cet algorithme paraisse très élégant dans sa conception et satisfaisant intuitivement, il est très onéreux en calcul. Sa contribution la plus importante est indubitablement son rôle comme un modèle informatique des premiers étages de la stéréovision humaine.

Un point de contour n'est pas une primitive très discriminante parce qu'elle ne contient qu'une quantité assez faible d'information : intensité lumineuse, contraste et voisinage locaux et le problème combinatoire reste assez élevé.

Afin de franchir cet inconvénient, à partir de points de contour on peut envisager d'aller plus loin et d'obtenir des représentations dérivées plus synthétiques et plus riches : grouper les points de contour en quelques primitives géométriques ; approcher un ensemble de points de contour par des segments de droite [5], [59], [75], ou bien détecter des points communs à plusieurs segments de droite (détection de jonctions), [26], [41], [49].

L'avantage de ces dernières approches est qu'elles utilisent un nombre de primitives géométriques relativement limité dans les images. Par conséquent, il y a une réduction considérable du problème combinatoire associé à la mise en correspondance. De plus, ces primitives contiennent des paramètres intrinsèques beaucoup plus riches et donc facilitent leur appariement. Enfin, il s'agit de primitives invariantes en projection perspective, ce qui signifie que leurs propriétés restent stables d'une image à l'autre.

1.3.2.2 Les contraintes mises en oeuvre

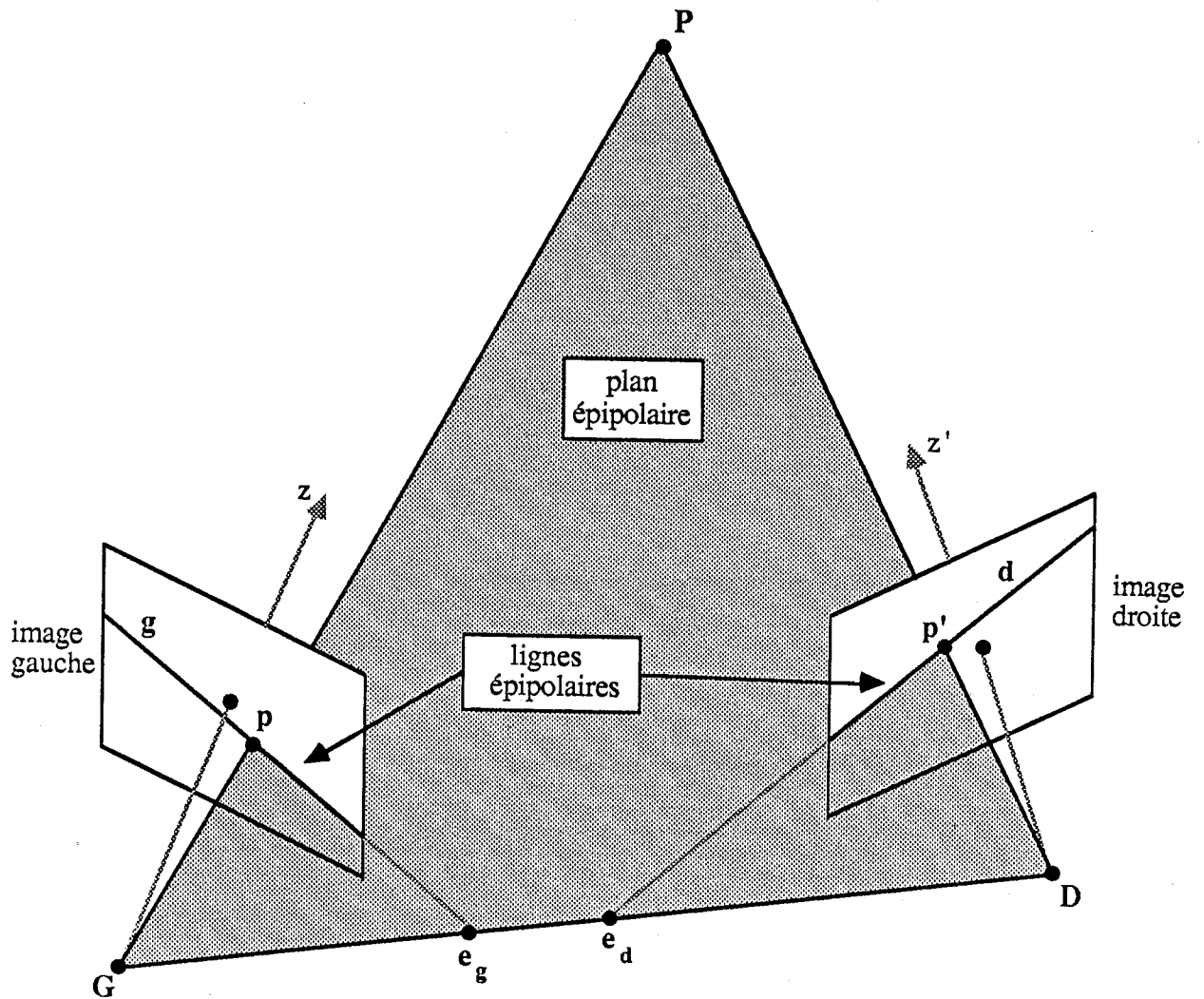
Même l'utilisation des primitives assez riches ne supprime pas complètement la complexité du problème. La résolution des appariements ambigus est possible en considérant quelques contraintes dictées par la scène et par la géométrie du capteur stéréoscopique. Parmi ces contraintes, certaines sont très simples et directes et d'autres plus subtiles. Les contraintes les plus importantes sont les suivantes :

La contrainte d'épipolarité

Un point de l'espace ainsi que les deux centres optiques des deux caméras du système stéréo définissent un plan, appelé *plan épipolaire*. L'intersection d'un plan épipolaire avec chacun de deux plans images du système stéréo définit une ligne droite, appelée *ligne épipolaire*. De cette manière, chaque plan épipolaire différent détermine dans les deux plans images deux lignes épipolaires correspondantes. Une ligne épipolaire représente en fait la projection de la ligne de vue d'un point physique par la première caméra sur le plan image de la seconde caméra. La contrainte d'épipolarité est alors exprimée par la règle : *tout point d'une ligne épipolaire de la première image doit correspondre à un point se trouvant sur la ligne épipolaire correspondante de la seconde image*.

Décrivons la géométrie épipolaire, telle que est illustrée sur la Figure 1.2. Soit P un point de l'espace, G et D les centres optiques des caméras gauche et droite respectivement et p et p' les deux projections du point P dans les deux plans images. La ligne droite GD qui relie les deux centres optiques est appelée *base* du système stéréo. Le plan épipolaire associé au point P est le plan PGD ; de manière évidente, les deux projections p et p' sont contenues dans ce plan. L'intersection du plan épipolaire PGD et des plans images des caméras gauche et droite détermine les deux lignes épipolaires correspondantes, qui sont les lignes droites g et d sur le plan image gauche et droit respectivement. Notons ici que les deux projections p et p' se trouvent forcément sur ces deux lignes épipolaires.

Ce dont on dispose en stéréo ce sont les différentes projections p et p' des points physiques de la scène sur les deux plans images et on cherche à établir les couples de points homologues (p, p') . Recherchons la correspondance du point p dans le plan image droit. Le plan épipolaire peut être déterminé comme le plan passant par le point p et par la base GD du système stéréo. Le point correspondant au point p dans l'image droite doit appartenir aussi au plan épipolaire pGD ; par conséquent,



GD : base du système stéréoscopique

$PpGDp'$: plan épipolaire

g : ligne épipolaire gauche

d : ligne épipolaire droite

e_g : épipole gauche

e_d : épipole droit

Figure 1.2: La géométrie épipolaire d'un système stéréo.

il doit se situer le long de l'intersection du plan pGD et du plan image droit, c'est-à-dire le long de la ligne épipolaire droite d . De cette manière, la recherche du point correspondant au point p au lieu de s'effectuer partout dans le plan image droit, se limitera à la ligne épipolaire droite d . L'espace de recherche des points correspondants se restreint ainsi de deux dimensions à une seule. Il en découle une réduction énorme de la complexité combinatoire de l'appariement des deux images stéréo.

Explorons encore plus la géométrie épipolaire du système de la Figure 1.2. Tous les plans épipolaires contiennent en réalité la base du système stéréo. Soit e_g le point de l'intersection du plan image gauche et de la base GD . e_g est en fait la projection du centre optique de la caméra droite sur le plan image gauche. Toutes les lignes épipolaires de l'image gauche passent par ce point, qui est appelé *épipole gauche*. Ceci est expliqué par le fait que le point e_g appartient en même temps au plan image gauche et au plan épipolaire, car e_g se trouve sur la base du système stéréo, par conséquent il se trouve aussi sur leur intersection, qui est la ligne épipolaire gauche. De façon similaire, il existe pour le plan image droit l'épipole droit e_d d'où passent toutes les lignes épipolaires droites.

La contrainte d'épipolarité est la seule contrainte qui est dérivée purement de la géométrie du capteur stéréoscopique ; elle ne s'applique que lorsque le banc stéréoscopique est proprement calibré [31], [83], c'est-à-dire quand les caractéristiques internes et la relation spatiale des deux caméras sont parfaitement connues à l'avance.

Quoique cette contrainte apporte une réduction considérable au problème combinatoire de la mise en correspondance, seule son utilisation s'avère insuffisante pour résoudre complètement les ambiguïtés des correspondances multiples. Les autres contraintes utilisées sont dérivées d'une combinaison de quelques propriétés géométriques du capteur stéréo et surtout de quelques hypothèses et heuristiques concernant les propriétés physiques d'une scène.

La contrainte d'unicité

Elle est représentée par la règle : *chaque caractéristique descriptive d'une image peut être appariée au plus à une caractéristique descriptive de l'autre image* [54].

Cette contrainte reflète la contrainte physique qu'un point d'une surface a une position unique dans l'espace à un moment donné. Sur les images, elle établit que chaque point imagé provenant d'un seul point physique de l'espace ne peut correspondre qu'à un seul point dans l'autre image.

Les quelques exceptions rares de cette règle se produisent lorsque deux points de l'espace se trouvent le long de la même ligne de vue par la première caméra mais ils sont séparés par la deuxième caméra (voir Figure 1.3). Cette règle exclut donc la possibilité d'un alignement accidentel et évidemment elle n'est plus valable lorsque l'on traite des objets transparents.

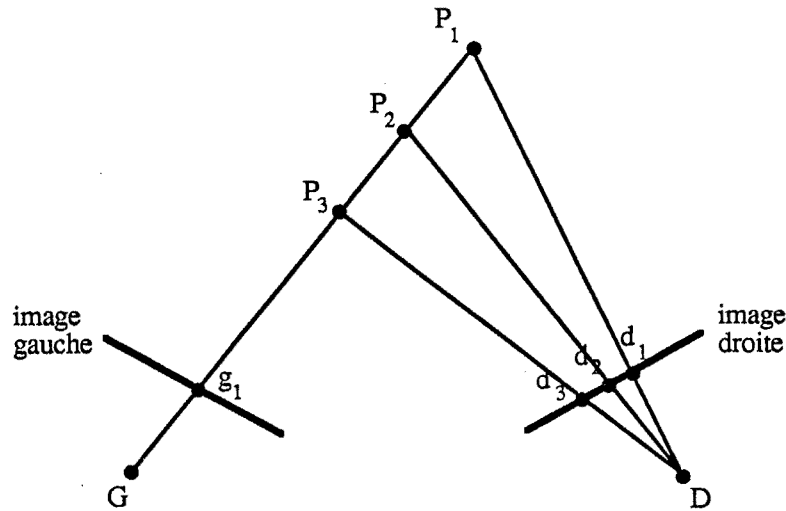


Figure 1.3: Violation de la règle d'unicité : le point g_1 de l'image gauche correspond aux points d_1 , d_2 et d_3 de l'image droite.

La contrainte due à la continuité de la disparité

Elle s'appuie sur l'hypothèse suivante : *les surfaces physiques sont localement continues et la variation de leur profondeur par rapport à l'observateur est généralement faible*. Implicitement elle suppose que les changements discontinus sont très rares, puisqu'ils n'apparaissent qu'aux frontières physiques des surfaces [54]. Ceci implique dans les images qu'il n'y a pas de variation brutale de la disparité, donc le champ de disparité varie **presque partout** d'une façon continue.

Cette règle a été implantée de la façon suivante : un appariement initial entre deux caractéristiques est confirmé ou infirmé par la similitude ou la dissimilitude de la disparité d'autres caractéristiques respectivement voisines aux deux premières. Si un nombre suffisamment grand de caractéristiques de la région voisine porte de disparités similaires à la disparité du couple initial, alors ce couple est confirmé, sinon il est rejeté.

La quasi-totalité des techniques existantes utilise cette hypothèse de la continuité de la disparité d'une manière ou d'une autre. Quelques systèmes l'exploitent à un niveau local [36], où des couples des appariements potentiels initiaux reçoivent leur confirmation par le vote (similitude de la disparité) des points voisins. Néanmoins, des interprétations erronées peuvent souvent se produire lorsque l'information locale est assez pauvre (région voisine contenant peu de caractéristiques) ou assez ambiguë (présence de structures périodiques voisines portant des valeurs de disparité voisines). Pour lever l'ambiguïté au niveau local, d'autres systèmes propagent le critère de la continuité de la disparité à des régions plus grandes afin d'aboutir à une interprétation globalement consistante [5], [11], [59].

La contrainte due au gradient de disparité

Quelques systèmes stéréo ont implanté l'hypothèse de continuité des surfaces physiques comme une contrainte du gradient de disparité [5], [70], [88].

Considérons le système stéréo montré sur la Figure 1.4 et où les deux plans images sont parallèles. Soient A et B deux points de l'espace et (A_g, B_g) et (A_d, B_d) ses projections dans l'image gauche et droite, respectivement. Soient (A_c, B_c) les projections des points A et B sur une troisième caméra, dont le centre optique se situe au milieu de la base GD du système stéréo. L'image de cette caméra est appelée *image cyclopéenne*. Le gradient de disparité entre les points A et B est défini comme la différence de leur disparité divisée par leur séparation cyclopéenne, qui est donnée par la distance Euclidienne entre les points A_c et B_c dans l'image cyclopéenne. Pour le système stéréo de la Figure 1.4, le gradient de disparité entre les points A et B est donné par :

$$gd_{AB} = \frac{|x' - x|}{\sqrt{y^2 + \frac{(x+x')^2}{4}}} \quad (1.2)$$

La contrainte due au gradient de disparité a été utilisée pour éliminer des couples d'appariements impossibles en imposant que : *le gradient de disparité entre des appariements corrects est petit presque partout dans les images, (il ne doit pas dépasser la valeur critique de 2, [20], [70]), tandis que ceci ne sera pas respecté pour des appariements incorrects.*

La contrainte d'ordre

L'opacité des surfaces de la scène et la continuité du monde visuel imposent une contrainte d'ordre le long des lignes épipolaires correspondantes : *si la caractéristique q se situe à gauche de la caractéristique p dans l'image gauche, alors cette même relation mutuelle existe généralement aussi entre leurs caractéristiques homologues (q' et p') dans l'image droite.* La Figure 1.5 illustre la contrainte d'ordre.

Soit P un point de l'espace observé par les deux caméras. En supposant que la scène est composée d'objets opaques, on peut associer au point P une *zone interdite*, définie par le cône passant par P et par la base stéréo. La Figure 1.6 illustre la zone interdite associée au point P . Les points appartenant à l'intérieur de cette zone ne peuvent pas être observés par les deux caméras à la fois.

Supposons maintenant que par P passe une surface transparente et soit Q un point se trouvant à l'intérieur de la zone interdite associée à P . Les projections de Q dans les deux images violent la contrainte d'ordre par rapport aux deux projections de P , (cf. Figure 1.6). Ainsi, la zone interdite associée à un point physique peut être considérée comme l'ensemble de points dans l'espace qui enfreignent la contrainte d'ordre.

Cette contrainte peut ne pas être respectée dans le cas de surfaces transparentes et d'objets isolés de faibles dimensions (par exemple les points P et Q de la Fi-

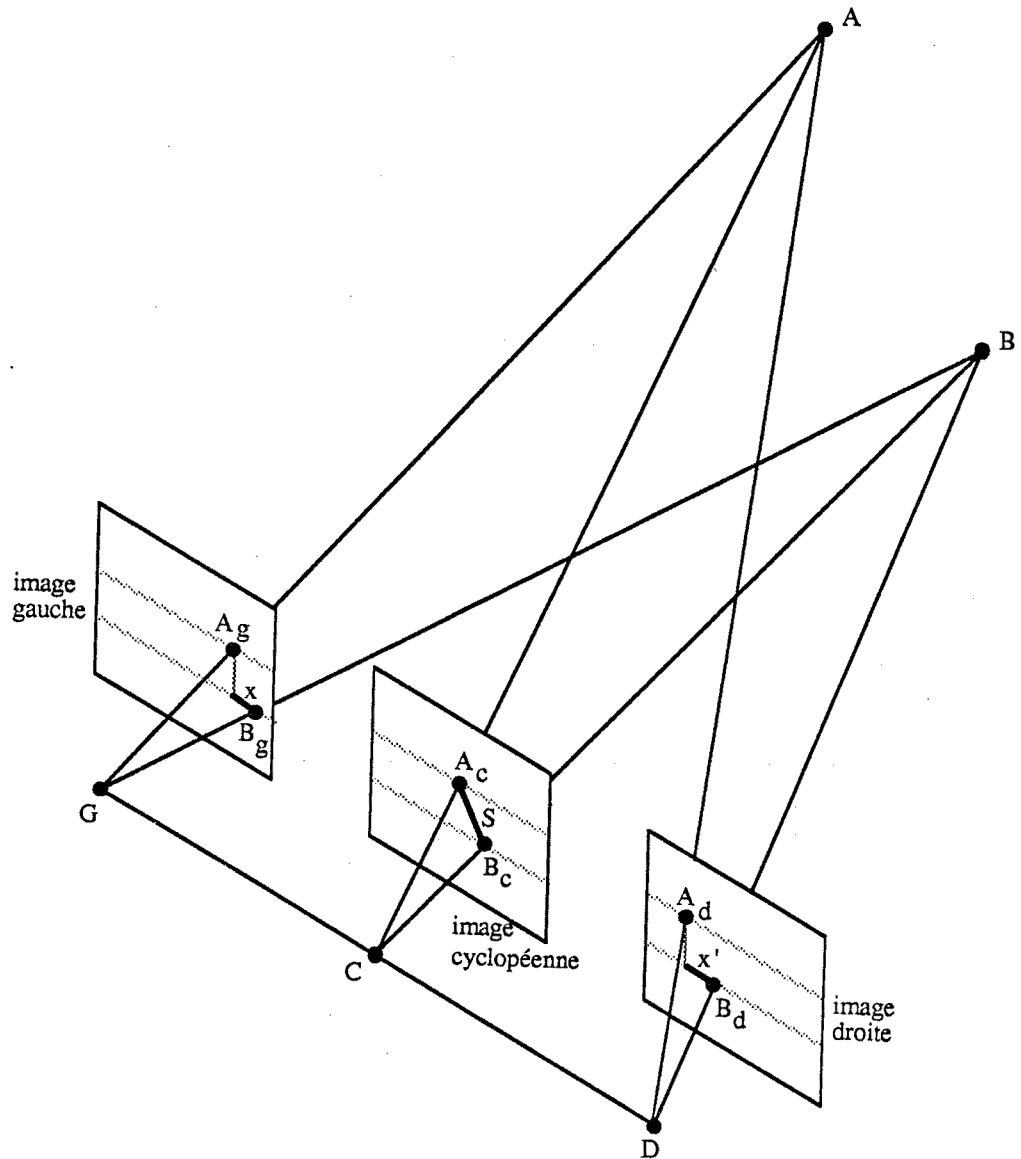


Figure 1.4: Le système stéréo de cette figure a les deux plans images parallèles. L'image de la camera C , située au milieu de la base GD , est l'image cyclopéenne. Le gradient de disparité entre les points A et B est défini par la différence de leur disparité $|x' - x|$ divisée par la distance S entre les points A_c et B_c , (cf. équation (1.2)).

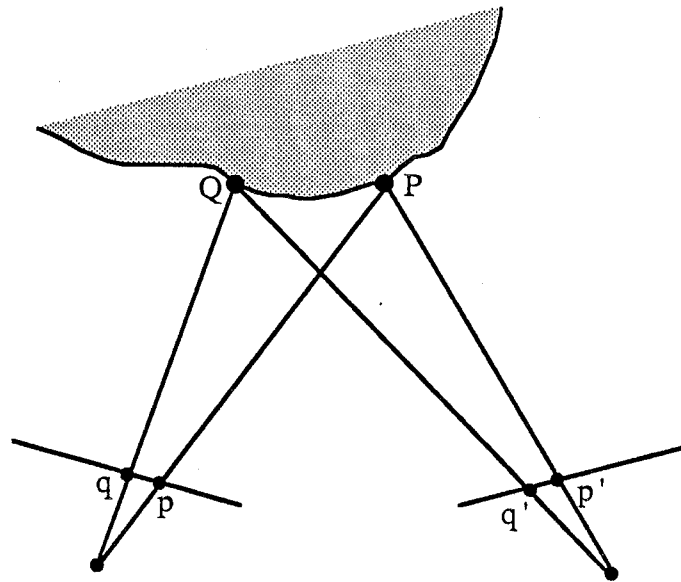


Figure 1.5: L'ordre des caractéristiques p , q dans l'image gauche est le même avec celui de leurs caractéristiques homologues p' , q' dans l'image droite.

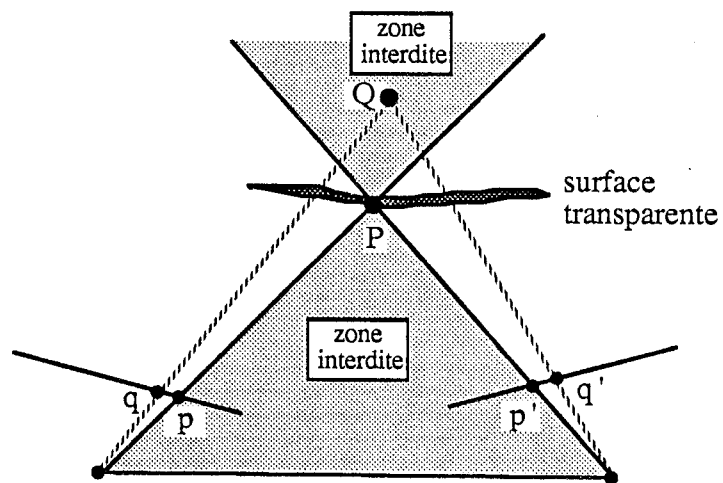


Figure 1.6: La zone interdite associée à un point P de l'espace.

gure 1.6). En général elle peut être mise en défaut dès lors que cohabitent dans l'image des champs de disparité très différents.

En réalité, sous l'hypothèse que nous traitons de surfaces opaques, la contrainte d'ordre n'est violée que très localement dans les images. Dans la pratique, elle est utilisée pour éliminer des paires d'appariements potentiels, dont la position sur les deux lignes épipolaires correspondantes viole cette contrainte par rapport à un appariement déjà validé se situant sur ces lignes épipolaires. Notons ici que le système développé par Ayache et Faverjon [5] réussit à s'affranchir de cette contrainte.

Une étude approfondie de la zone interdite, de son utilisation par rapport à la contrainte d'ordre et de sa formulation de manière indépendante de la géométrie épipolaire a été effectuée par Yuille et Poggio [88].

La contrainte de continuité de formes

La continuité d'une courbe physique est une propriété qui se conserve par projection perspective : une courbe physique de l'espace se projette en des courbes continues dans les deux images. Ceci implique que la disparité varie le long d'un contour de façon continue, c'est-à-dire il ne doit pas y avoir de discontinuité de disparité. *Une séquence de points de contour connectés dans une image devrait apparaître comme une séquence connectée de points de contour dans l'autre image* [57].

Cette information est exploitée convenablement en travers des lignes épipolaires : soit durant le processus de mise en correspondance pour assurer la continuité entre deux points de contour voisins [35], [65] ; soit après la mise en correspondance pour détecter les appariements incorrects et les rejeter [7], ou les corriger [60] ; soit implicitement, en utilisant des caractéristiques-image incluant dans leur nature cette continuité, comme par exemple les segments de droite.

1.3.2.3 Les stratégies de mise en correspondance

Dans cette partie nous décrivons succinctement les différentes méthodes algorithmiques développées qui, en conjonction avec les différentes contraintes résolvent le problème de la mise en correspondance stéréo. Une bonne revue des systèmes stéréo existants se trouve en [69] et aussi en [10] concernant les systèmes développés avant 1982.

La corrélation

Le principe de cette méthode est le suivant : la meilleure correspondance d'une primitive de l'image gauche avec une primitive de l'image droite est établie en recherchant la meilleure ressemblance possible de mesures de corrélations entre des fenêtres centrées aux deux primitives respectivement.

La corrélation est une technique utilisée notamment pour la mise en correspondance des régions-fenêtres composées d'ensembles de pixels voisins [37], [21], où l'intercorrélation est utilisée pour déterminer des appariements entre de fenêtres

d'une image et de fenêtres de l'autre. L'intercorrélation normalisée permet de compenser les différences du contraste et de la luminosité entre les images.

Dans la plupart des systèmes l'utilisant, l'intercorrélation n'est pas appliquée à tous les pixels, car cela coûte cher en temps de calcul, mais de manière sélective seulement aux pixels, dont la variance locale est assez élevée. Selon cette approche, la mesure de variance est utilisée comme un filtre pour limiter le nombre de correspondances possibles ; la corrélation est ensuite utilisée pour sélectionner parmi les candidats potentiels le meilleur possible [39], [64].

Les systèmes basés sur la corrélation souffrent de quelques limitations inhérentes à la corrélation. Ils sont sensibles à l'absence de texture, à la présence de structures répétitives, aux distorsions de perspective, aux occlusions et aux discontinuités des surfaces. Plus précisément : pour des régions non texturées, il y a une infinité de correspondances ; des correspondances multiples sont très probables sous la présence de structures répétitives ; aux discontinuités brusques de la profondeur, aucune correspondance n'est possible entre les régions croisant ces discontinuités. A cause de toutes ces raisons, la technique de corrélation a connu un certain succès uniquement dans le domaine d'interprétation des images stéréoscopiques aériennes, où les terrains sont relativement lisses et continus.

La stratégie hiérarchique des niveaux de résolution (*coarse-to-fine*)

Cette stratégie traite de la mise en correspondance utilisant une pyramide d'images de niveaux de résolution différents. Le principe de cette approche consiste à descendre à un niveau de résolution des images assez bas, où une première mise en correspondance est établie. La mise en correspondance à un niveau de résolution grossier est facile car il y a une diminution considérable du nombre de caractéristiques-image à apparier. Qui plus est, cette mise en correspondance préliminaire apporte une première idée sur la variation du champ de disparité entre les deux images. Ensuite, cette mise en correspondance est propagée graduellement aux niveaux de résolution supérieurs, jusqu'à atteindre le niveau le plus fin. La détermination du champ de disparité effectuée à un niveau de résolution précédent guide la recherche des points correspondants au niveau de résolution supérieur. Ainsi cette recherche se restreint sur un petit voisinage autour des points appariés au niveau de résolution inférieur.

Ce genre de stratégie a été mis en oeuvre par Hannah [39], qui met en correspondance les points de Moravec et par Grimson [35], qui met en correspondance les points de contour. P. Stelmaszyk et al. [78], plutôt que d'utiliser une approche du type *coarse-to-fine*, ont développé une technique de mise en correspondance qui intègre l'information contenue dans les différents niveaux de résolution au moyen de la transformée de DOLP [24].

La programmation dynamique

La programmation dynamique est utile pour mettre en correspondance deux séquences. Elle résout un processus de décision de N-étapes comme N processus

individuels d'une seule étape. Ceci réduit la complexité informatique par le logarithme de la complexité initiale. Le processus de décision doit être Markovien : à tout moment, l'état du processus dépend uniquement de l'état courant et pas de son passé.

La programmation dynamique a été notamment utilisée par Baker et Binford [7] et par Ohta et Kanade [65]. Dans les deux systèmes développés par ces auteurs, les lignes épipolaires correspondantes sont les mêmes lignes de balayage horizontal et les primitives appariées sont les points de contour.

Un couple des points correspondants doit être recherché à l'intérieur des mêmes lignes de balayage (contrainte d'épipolarité). Cette recherche se ramène à la recherche d'un chemin optimal de mise en correspondance dans un plan bidimensionnel, dont les axes vertical et horizontal sont les lignes de balayage des images droite et gauche, respectivement. Baker et Binford ont utilisé un algorithme de Viterbi (une version de la programmation dynamique) pour que cette recherche 2-D soit effectuée de façon efficace. Le chemin optimal dans le plan 2-D est obtenu en itérant la sélection d'un chemin optimal à tous les noeuds 2-D (un noeud étant l'intersection d'un contour de l'image gauche avec un contour de l'image droite). Après un premier passage où les correspondances sont déterminées sur chaque ligne de balayage séparément, la contrainte de continuité de formes est utilisée pour rejeter les correspondances qui ne respectent pas cette contrainte en travers des lignes épipolaires.

Mohr et Wrobel [61] ont proposé une approche voisine ; la différence essentielle est que l'information de continuité des chaînes de contour est dans ce cas exploitée par un algorithme de relaxation.

La méthode de Baker et Binford n'utilise pas la contrainte de continuité de formes pendant le processus de recherche des correspondances, par conséquent son résultat n'est guère garanti d'être optimal. Ohta et Kanade ont étendu la technique de programmation dynamique pour inclure cette recherche pendant le processus de mise en correspondance. Le problème se ramène ainsi à une recherche d'un ensemble de chemins optimaux dans un espace tridimensionnel ; cet espace est en fait l'empilement des plans 2-D construits à partir de la recherche des correspondances dans chaque ligne de balayage. Comme dans le cas 2-D, l'ensemble de chemins optimaux dans l'espace 3-D est obtenu en itérant la sélection d'un ensemble optimal de chemins à chaque noeud 3-D de l'espace 3-D (un noeud 3-D étant un ensemble de noeuds 2-D qui appartiennent au même contour de l'image gauche et de l'image droite). Cette recherche nécessite cependant un nombre d'opérations immense.

Afin de réduire la complexité algorithmique, Raju et al. [73] proposent de mettre d'abord en correspondance les chaînes verticales de contour en travers des lignes épipolaires. Un algorithme de Viterbi modifié est utilisé, qui recherche le chemin optimal dans un plan 2-D constitué de chaînes de contour de l'image gauche et de l'image droite ; le chemin optimal est obtenu en maximisant une fonction de ressemblance des points de contour composant les différentes chaînes de contour. Les points de contour appariés peuvent ensuite être utilisés comme des contraintes fortes pour mettre en correspondance les intervalles entre ces points de contour le long des lignes épipolaires correspondantes. Cette mise en correspondance est

INSTITUT IMAG

Informatique, Mathématiques Appliquées de Grenoble

CNRS - INPG - IMAG

1000 LA TRONCHE

38000

38000 Grenoble Cedex 9

effectuée par une recherche 2-D similaire à celle de la mise en correspondance de chaînes de contour. Malgré la réduction considérable des calculs que cette nouvelle approche apporte, les résultats obtenus sont relativement pauvres (25% des points de contour initiaux sont appariés en moyenne).

La relaxation

Le problème de mise en correspondance stéréo est formulé comme un problème d'optimisation d'étiquetage et la relaxation est ensuite utilisée comme moyen de résolution du problème. Le problème d'optimisation d'étiquetage prend la forme suivante dans le cas de la stéréo : chaque primitive de la première image est considérée comme un objet ; on lui associe des étiquettes, qui sont tous les candidats potentiels possibles de cet objet dans l'autre image. Ainsi, chaque étiquette représente une disparité possible. Parmi les étiquettes existe aussi l'étiquette NUL, qui correspond à la possibilité que cet objet n'ait pas de correspondant. A chaque étiquette de disparité est associée une mesure de vraisemblance initiale ; cette mesure est établie à partir de quelques critères locaux de ressemblance entre l'objet et son étiquette. A un problème donné donc, il existe un grand nombre de combinaisons potentielles objets-étiquettes. A ce moment on fait intervenir les différentes contraintes stéréo et notamment la contrainte due à la continuité de la disparité pour tester la compatibilité des différents objets étiquetés. Les estimées initiales de vraisemblance sont améliorées itérativement par un algorithme de relaxation, où la validité de l'association d'une étiquette à un objet est une fonction des objets étiquetés voisins :

- Si plusieurs voisins ont des étiquettes similaires, alors l'étiquette est confirmée et sa probabilité de vraisemblance augmente ;
- si aucune étiquette ne reçoit de soutien fort, la mesure de vraisemblance de l'étiquette NUL augmente.

Après un nombre d'itérations donné, les estimées de vraisemblance convergent vers des classifications de disparité uniques pour chaque objet.

Plusieurs systèmes ont utilisé différentes versions de cette technique pour mettre en correspondance des points images [70], [71], des points de Moravec [11], des segments de droite [50], [59], ou des descriptions structurelles simples des images [19].

La recherche dans un graphe

Le principe de ces méthodes est le suivant : toutes les correspondances possibles des primitives de l'image gauche sont recherchées dans l'image droite utilisant certains critères de ressemblance locaux en conjonction avec la contrainte d'épipolarité. Les appariements potentiels ainsi établis sont représentés par de noeuds dans un graphe. Le but est d'explorer ce graphe pour découvrir le meilleur ensemble de noeuds qui sont tous mutuellement compatibles. La compatibilité entre les différents noeuds est établie en utilisant les différentes contraintes dérivées d'hypothèses sur la scène.

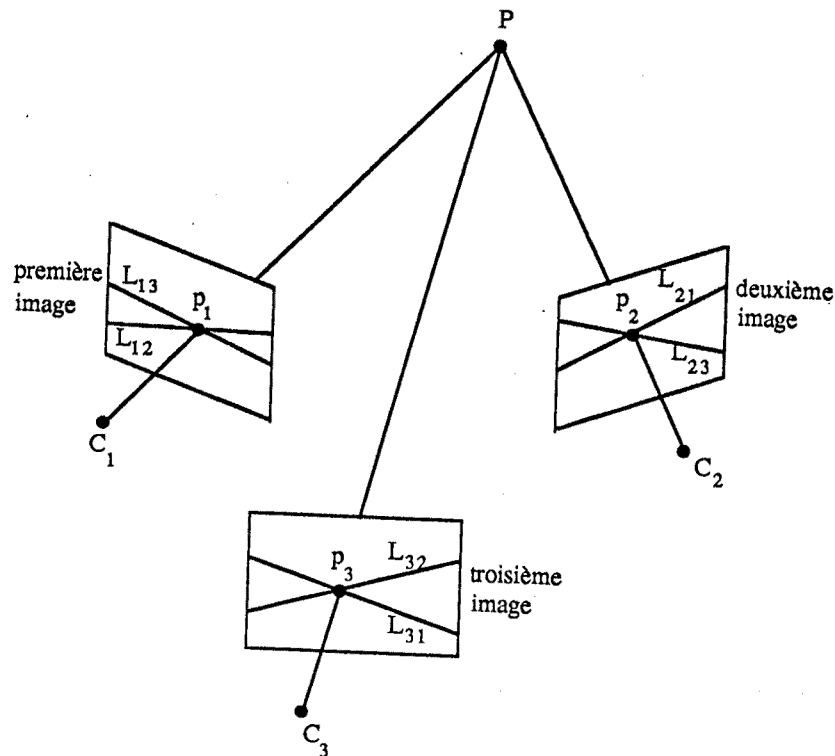


Figure 1.7: La géométrie d'un système stéréo trinoculaire.

Pour Ayache et Faverjon [5] et Robert de Saint Vincent [75], les noeuds du graphe sont composés d'appariements potentiels de segments de droite. Le critère de compatibilité utilisé est la continuité de la disparité entre des segments voisins. Chaque image est décrite par un graphe d'adjacence de segments de droite ; la correspondance stéréo est réalisée en explorant le graphe des appariements potentiels par une technique de prédiction et vérification récursive d'hypothèses.

Pour Herman et Kanade [41] les noeuds du graphe représentent d'appariements potentiels de jonctions de segments de droite. A chaque noeud est associé un coût local selon un critère de ressemblance locale (similitude de l'intensité lumineuse des régions correspondantes se trouvant autour de jonctions homologues). La cohérence entre deux appariements de jonctions voisines est représentée par un coût global. La meilleure correspondance est réalisée en recherchant dans le graphe le chemin optimal minimisant une fonction de coût, qui est la somme des coûts locaux et globaux des noeuds participant à cette correspondance.

Les stratégies basées sur la redondance d'informations

L'idée est d'utiliser une troisième caméra pour obtenir une redondance d'informations permettant de réduire la complexité combinatoire. La Figure 1.7 présente la géométrie d'un tel système stéréo trinoculaire.

Le principe est le suivant : soit P un point physique se projetant sur les trois caméras aux points p_1 , p_2 et p_3 respectivement. La géométrie d'un tel système trino-

culaire impose que, n'importe quel point du triplet (p_1, p_2, p_3) se situe à l'intersection des deux lignes épipolaires définies par les deux autres points. Ainsi la recherche des points images homologues se réduit à une simple vérification dans la troisième image à un endroit précis. Par exemple la vérification que le couple (p_1, p_2) est homologue consiste à vérifier dans l'image de la troisième caméra la présence du point p_3 à l'intersection des lignes épipolaires L_{31} et L_{32} , (cf. Fig. 1.7).

A l'heure actuelle, plusieurs systèmes stéréo ont été développés se basant sur ce principe : [6], [44], [52], [67]. La difficulté principale de ce genre d'approches est de traiter du problème des occlusions dont la probabilité d'apparition est beaucoup plus élevée par rapport à un système stéréo classique. Dans l'exemple précédent, si le point p_3 est absent dans la troisième image, alors deux hypothèses sont équiprobables ; soit le couple (p_1, p_2) n'est pas homologue, soit P est invisible pour la troisième caméra. Une solution possible serait de tenir compte seulement de triplets de points validés, mais ceci n'aboutirait qu'à une reconstruction de scène très pauvre.

D'autres approches

S. Barnard [9] ramène le problème de mise en correspondance stéréo en un problème d'optimisation stochastique. Il s'agit d'une approche inspirée par les concepts de la mécanique statistique. La correspondance stéréo est présentée comme une recherche d'une carte de disparité satisfaisant deux contraintes concurrentes : la contrainte que les points correspondants devraient avoir de valeurs voisines d'intensité lumineuse et la contrainte que la carte de disparité devrait être continue. Ces deux contraintes sont exprimées par une fonction d'énergie, qui peut être évaluée localement et ensuite un algorithme de "réduit simulé" est utilisé pour la minimiser. Il en résulte une carte de disparité dense rendant inutile l'interpolation.

M. Kass dans [46] présente le problème de la stéréocorrespondance comme un calcul de mesures de similitude entre les sorties de filtres linéaires appliqués sur les deux images de la paire stéréo. Son algorithme calcule la correspondance en se basant sur de combinaisons d'indications provenant de l'application sur chaque point image d'un ensemble de filtres linéaires presque indépendants. La décision si un point de l'image gauche correspond à un point de l'image droite est prise en comparant les vecteurs de sortie des filtres linéaires aux deux points. Si la sortie de n'importe quel filtre linéaire diffère dans les deux images de plus d'un seuil, adapté à cet opérateur, alors la correspondance est rejetée. La famille des filtres linéaires utilisée est un ensemble de filtres dont les réponses impulsionnelles sont des Gaussiennes.

Dans une approche ultérieure [47], Kass propose d'utiliser une estimée de la distorsion géométrique due à la perspective pour détecter dans les images les points, où la distorsion géométrique est considérable et risque de poser de problèmes. Pour cela, il propose d'ajuster dynamiquement les seuils associés à chaque opérateur linéaire précédent, de façon que la pondération relative des différentes filtres dépende de leur degré d'invariance par rapport aux changements de la géométrie.

1.3.2.4 Quelques critiques

La majeure partie des techniques existantes suppose que les caractéristiques images à apparier sont des marques sur des surfaces lisses telles que les points de texture. Par conséquent, leur stratégie de mise en correspondance utilise la contrainte due à la continuité de la disparité. Les algorithmes proposés fonctionnent bien quand ils sont appliqués à des domaines assez spécifiques tels que l'interprétation de photographies aériennes et la mise en correspondance de surfaces artificielles, où les surfaces traitées sont relativement continues. Par contre, ces algorithmes rencontrent des problèmes considérables lorsqu'ils traitent de scènes où les discontinuités sont prépondérantes.

Quoiqu'il apparaisse correct d'utiliser la continuité de la disparité pour mettre en correspondance des marques sur des surfaces lisses, il n'est pas correct d'appliquer cette contrainte aux points de discontinuité, c'est-à-dire aux points de contour qui résultent de discontinuités d'orientation des surfaces ou de discontinuités de profondeur. Effectivement, à ces endroits la surface physique n'est pas continue ; dans les images on obtient une cohabitation locale de deux champs de disparité très différents, d'où la non-validité de l'hypothèse de la continuité de la disparité.

1.4 La reconstruction de surfaces

Un système de vision stéréo n'est utile que s'il arrive à récupérer les caractéristiques 3-D de la scène. Son but final doit être non seulement de reconstruire de caractéristiques individuelles 3-D, mais de reconstituer la structure géométrique de la scène en la décrivant en termes de primitives surfaciques et d'objets contenant les caractéristiques 3-D construites.

1.4.1 Extraction de la profondeur

En supposant que l'on dispose d'une liste de caractéristiques appariées des images gauche et droite, le processus de reconstruction de surfaces commence par l'extraction de la profondeur à l'aide d'une méthode géométrique appelée *triangulation* (cf. Figure 1.1) : connaissant les paramètres optiques des deux caméras, leur relation géométrique spatiale et le fait que les points p et p' sont homologues, le point physique P se localisera à l'intersection des deux demi-droites Gp et Dp' .

Bien que cette étape paraisse simple à première vue, elle présente un certain nombre de difficultés importantes. En effet, à cause de l'échantillonnage et de la quantification des données de départ et du bruit inhérent à tout système de formation d'images, le système stéréo va localiser les points p et p' avec une certaine erreur, qui à son tour va engendrer une erreur sur la position estimée de P .

La relation entre la géométrie du capteur stéréo et la précision des informations 3-D obtenues a attiré l'attention d'un certain nombre de chercheurs. Ainsi McVey et Lee [58] ont analysé l'erreur maximale sur la résolution que le plan image doit avoir pour arriver à une mesure de profondeur d'une précision donnée. Blostein et Huang [13] ont modélisé l'erreur sur la profondeur comme une quantification spa-

tiale du plan image et ils ont présenté une méthode qui calcule la probabilité pour que les estimations de la profondeur résultant de la triangulation ne dépassent pas une tolérance d'erreur spécifiée d'avance. Crowley et Ramparany [25], Kriegman, Triendl et Binford [48] et Matthies et Shafer [56] ont proposé de modéliser l'erreur de triangulation en utilisant la théorie des variables aléatoires multidimensionnelles de distribution Gaussienne. L'approche consiste à supposer une erreur Gaussienne sur les coordonnées-image mesurées et à dériver ensuite l'incertitude des coordonnées 3-D résultant de la triangulation. Ayache et Faugeras [3], [4] proposent de ne pas effectuer la triangulation de façon explicite, mais une relation linéaire entre les coordonnées 2-D et 3-D est exprimée via une matrice de projection 3×4 (matrice de calibrage) ; l'information sur la géométrie de l'environnement et sur l'incertitude des paramètres définissant cette géométrie est modélisée à l'aide des filtres de Kalman étendus.

Dans la plupart des systèmes existants, les caractéristiques à apparier sont localisées avec une précision d'un pixel, imposée par les limitations inhérentes à la résolution spatiale des capteurs optiques. De manière évidente, la précision des mesures 3-D peut considérablement être améliorée si les caractéristiques appariées sont localisées avec une précision au dessous du pixel. De telles caractéristiques pourraient être par exemple des intersections de courbes. Il en découle donc que la précision des données 3-D dépend de manière critique du choix des caractéristiques images à apparier.

De manière générale, l'erreur des mesures de distance est inversement proportionnelle à la longueur de la base du système stéréo. Néanmoins allonger la longueur de la base complique le processus de mise en correspondance puisqu'on augmente d'une part le champ de disparité (la région qui doit être recherchée) et d'autre part la différence en apparence des caractéristiques à apparier.

Enfin, la précision des mesures 3-D peut être améliorée en combinant d'informations provenant de plusieurs points de vue. Un appariement issu d'une seule paire des vues représente une estimation de profondeur possédant une incertitude assez grande. Des appariements d'autres paires des vues pourraient être moyennés afin de réduire cette incertitude. Les variables aléatoires multidimensionnelles de distribution Gaussienne et les filtres de Kalman étendus ont été utilisés dans la bibliographie spécialisée comme les deux outils mathématiques principaux permettant de représenter, combiner et propager les incertitudes des données visuelles provenant de plusieurs points de vue et/ou de capteurs différents [3], [25], [28], [56], [77].

1.4.2 Extraction de surfaces

Si la mise en correspondance entre les deux images s'établissait sur chaque pixel, ceci engendrerait une carte très dense de la profondeur lors de la reconstruction de la scène. Néanmoins, nous avons vu que pour de raisons de complexité énorme, la mise en correspondance n'est appliquée que sur quelques caractéristiques, qui sont distribuées irrégulièrement et de manière éparse dans les deux images. Qui plus est, pour des raisons diverses, le processus de mise en correspondance n'apparie finale-

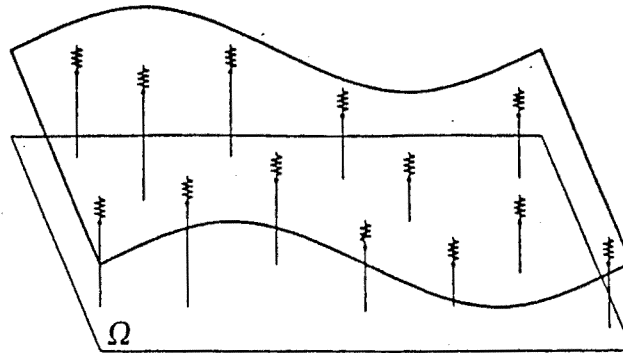


Figure 1.8: Le modèle physique du plateau fin et flexible pour la reconstruction de surfaces.

ment qu'une partie de ces caractéristiques. Par conséquent, lorsque l'on reconstruit les caractéristiques appariées, on aboutit à une carte creuse et bruitée de la profondeur, qui est en fait une représentation très pauvre de l'environnement vis-à-vis de sa richesse.

Le problème qui se pose donc est comment découvrir dans cette carte creuse des données 3-D les surfaces de la scène qui contiennent les caractéristiques reconstruites, tout en sachant qu'il est pratiquement impossible de récupérer la "physionomie" exacte de la scène.

Plusieurs méthodes approximatives ont été proposées dans la bibliographie spécialisée. Toutes se basent sur l'hypothèse que les caractéristiques 3-D construites sont suffisantes pour pouvoir saisir les changements importants de la profondeur. Les méthodes existantes peuvent se répartir en trois grandes catégories :

1. Les méthodes qui interpolent la carte creuse de la profondeur à l'aide d'un modèle physique qui assure la continuité des surfaces.
2. Les méthodes intégrant plusieurs processus afin d'extraire les surfaces de scènes.
3. Les méthodes qui approximent directement les mesures 3-D par de surfaces polyédriques.

Nous allons présenter maintenant brièvement les travaux les plus importants qui sont répertoriés dans chacune de ces catégories.

Reconstruction de surfaces à l'aide d'un modèle physique

Les mesures 3-D sont interpolées par de surfaces lisses et continues dont le modèle physique est donné par un plateau fin (thin plate) qui peut fléchir mais pas se plisser. Ce modèle est constitué d'une surface élastique idéale qui est contrainte par de ressorts idéaux dont la tension est proportionnelle à la profondeur des mesures 3-D, (cf. Figure 1.8). Ainsi, le problème d'interpolation est formulé comme un problème de minimisation de l'énergie possédée par cette surface.

Les algorithmes basés sur ce modèle sont ceux de Grimson [36], Terzopoulos [81], [80] et Blake et Zisserman [12].

Grimson [36] a noté que l'absence de caractéristiques à apparier implique une limite quant à la variabilité de la surface à interpoler et il a proposé un algorithme de relaxation qui interpole progressivement l'ensemble de données 3-D par une surface unique selon le modèle du plateau fin et flexible.

Terzopoulos [81] a amélioré la méthode de Grimson de plusieurs façons, la plus impressionnante étant en efficacité, atteinte via une technique de multirésolution. Le problème de reconstruction de surfaces est formulé mathématiquement comme un problème de principe variationnel décrivant l'équilibre du plateau fin et flexible. Des approximations discrètes optimales sont obtenues via la méthode d'élément fini, qui utilise une représentation locale, polynomiale par morceaux des surfaces. Finalement, un algorithme multirésolution est proposé pour résoudre rapidement la hiérarchie des problèmes discrets résultant.

Le modèle de reconstruction proposé par Grimson et Terzopoulos rencontre des problèmes considérables lorsque dans la scène il y a des surfaces ayant des discontinuités en profondeur ou en orientation. En effet, le processus d'interpolation basé sur le modèle du plateau fin et flexible va tenter de reconstruire une surface unique et lisse sur l'image entière et la surface reconstruite donnera ainsi l'impression d'une espèce de nappe jetée sur le modèle 3-D de la scène. Ce modèle lisse donc toute discontinuité en profondeur présente dans la scène ; qui plus est, il n'est pas capable de reconstruire une surface discontinue en orientation, car sa propriété physique de lissage qui en fait un bon modèle d'interpolation, l'empêche justement de se plisser aux endroits où il y a une discontinuité en orientation.

Afin d'améliorer la précision de la forme de la surface reconstruite, Grimson [34] propose d'examiner d'autres indices visuels pouvant fournir d'information quant à la forme des surfaces ; plus précisément, il propose d'utiliser l'information d'ombrage disponible avec les images stéréo pour déterminer l'orientation des surfaces aux points 3-D et ceci à l'aide d'une méthode similaire à la technique de stéréo photométrique. Une carte de reflectance est construite sur chaque contour apparié, sous l'hypothèse que l'on traite de surfaces composées d'une classe particulière de matériaux. Enfin, l'orientation de surface sur chaque contour est résolue, le problème majeur étant la stabilité numérique des calculs et des solutions proposées.

Terzopoulos, dans une approche ultérieure [80], propose un modèle de reconstruction, où il n'y a pas une seule surface physique, mais un ensemble de surfaces lisses et continues par morceaux selon le modèle physique suivant : le modèle du plateau fin est cassé en morceaux le long des contours détectés ; ces morceaux sont complètement disjoints le long d'un contour provenant d'une discontinuité en profondeur et ils se joignent par de bandes minces de membranes le long des contours issus d'une discontinuité en orientation.

Terzopoulos donc suppose qu'avant la phase de reconstruction, il existe d'une manière ou d'une autre une classification des contours détectés en contours provenant d'une discontinuité en profondeur ou d'une discontinuité en orientation. Or, la difficulté principale est justement cette détection explicite des contours.

Blake et Zisserman [12] proposent de résoudre le problème de reconstruction en introduisant la notion de contrainte de continuité faible. Il s'agit en fait d'un mécanisme pour exprimer l'espérance que la profondeur des surfaces varie de façon continue presque partout, à l'exception de quelques endroits où des changements brusques sont permis. La formalisation mathématique du problème est en réalité une généralisation de la formalisation proposée par Terzopoulos.

Intégration de plusieurs processus pour l'extraction de surfaces

Eastman et Waxman [29] et Hoff et Ahuja [42] suggèrent d'inclure les processus de mise en correspondance et de reconstruction en une seule étape utilisant une représentation explicite des surfaces.

En ce qui concerne Eastman et Waxman, ils proposent de représenter localement dans les images la structure des surfaces de scène par un modèle polynomial (de deuxième ordre) du champ de disparité, appelé *fonctionnel de disparité* (disparity functional). Des estimations locales de ce modèle sont fournies par une réaction réciproque d'appariements potentiels des points de contour. Leur algorithme intègre alors le fonctionnel de disparité avec une mise en correspondance en résolutions multiples des contours pour extraire la profondeur de différents éléments de surfaces, qui sont ainsi obtenues comme un sous-produit de mise en correspondance.

Quant à Hoff et Ahuja [42], ils intègrent trois processus simultanément, à savoir celui d'extraction et de détection explicite des contours et les deux processus de mise en correspondance et d'interpolation. Le but de cette intégration est de renforcer l'hypothèse que les surfaces à reconstruire sont lisses et continues localement. L'intégration des trois processus est effectuée à l'aide d'un algorithme hiérarchique des niveaux de résolution : à un niveau de résolution donné, les contours sont extraits et appariés de manière que la surface, résultant de l'extraction de la profondeur des contours appariés et de son interpolation, soit lisse et continue localement ; ensuite la surface estimée est utilisée pour prédire les emplacements de futurs contours à appairer au niveau de résolution plus fin.

Extraction de surfaces à partir de données brutes 3-D

Les travaux les plus importants sont ceux de Herman et Kanade [41], Thonnat [82], Robert de Saint Vincent [75], Wrobel [87] et Boissonnat et al. [14].

Herman et Kanade [41] ont développé un système qui acquiert de proche en proche un modèle 3-D de scènes urbaines et complexes à partir de plusieurs paires d'images stéréo. La première paire d'images fournit un premier modèle 3-D assez incomplet. Ce modèle est composé de surfaces polyédriques et de leurs composantes (faces-contours-sommets) ; il est construit à partir d'un modèle intermédiaire du type fil de fer résultant de la reconstruction des caractéristiques appariées (jonctions de segments de droite). Avec chaque nouvelle vue, le modèle 3-D de la scène est mis à jour et ainsi il devient graduellement plus précis et plus complet. Une limitation de cette méthode est le fait que la construction du modèle fil de fer et du modèle

de scène et leur mise à jour sont effectuées à l'aide de quelques heuristiques et connaissances sémantiques sur la scène traitée.

Une autre approche au problème de reconstruction est celle proposée par M. Thonnat [82], qui se caractérise aussi par l'incorporation d'information sémantique dans le processus de modélisation de scènes : les données 3-D (segments 3-D) sont d'abord triées selon leur longueur, orientation et profondeur et ensuite elles sont groupées progressivement en surfaces planes à l'aide de quelques règles et d'un moteur d'inférence afin de pouvoir retrouver les structures principales de la scène (murs, sol, plafond, etc.).

Robert de Saint Vincent et Wrobel ont proposé d'approximer les données éparses 3-D (segments 3-D) par un ensemble de surfaces planes passant au mieux par ces données.

En ce qui concerne Robert de Saint Vincent [75], les segments de droite appariés sont projetés dans l'espace 3-D et sont ensuite utilisés pour modéliser la scène. Etant donné que les segments 3-D obtenus sont assez bruités, l'approche consiste à extraire seulement de plans verticaux. Les données initiales 3-D sont partitionnées en un ensemble de voisinages et une méthode de propagation/vérification d'hypothèses est utilisée pour rassembler à l'intérieur de chaque voisinage (à l'aide d'un critère de moindres carrés) un certain nombre de segments en un plan vertical et pour intégrer ensuite à ce plan d'autres segments 3-D.

Quant à B. Wrobel [87], après avoir apparié de régions homologues dans les deux images, elle propose d'approximer chaque région 3-D résultant par une surface plane. Pour cela, elle construit d'abord les segments 3-D représentant les frontières des régions appariées (abandonnant ainsi la richesse des régions) et ensuite elle construit les surfaces planes qui passent au mieux par ces ensembles de segments 3-D.

Une dernière méthode est celle proposée par Boissonnat, Faugeras et Le Bras Mehlman [14]. Cette méthode est basée sur la "Triangulation de Delaunay Contrainte" [15]. Les données 3-D (segments 3-D) sont interpolées à l'aide d'une surface polyédrique (composée d'un ensemble de tétraèdres), qui est obtenue en utilisant un critère de visibilité pour marquer les tétraèdres vides. La réunion de ces tétraèdres représente l'espace vide d'obstacles et sa frontière est une représentation polyédrique de la surface des objets. Cette méthode a deux avantages principaux : premièrement, elle résout explicitement le problème des discontinuités en profondeur, puisque la triangulation de Delaunay préserve ces discontinuités et deuxièmement, elle met à jour rapidement le modèle 3-D de la scène en intégrant facilement de données nouvelles acquises d'autres points de vue et/ou de capteurs différents.

Quelques critiques

Toutes les méthodes des deux premières catégories fournissent de bons résultats, lorsque la scène traitée est abondante de surfaces texturées, c'est-à-dire quand il y a suffisamment d'information locale pour pouvoir interpoler et reconstruire les surfaces. Par contre, dans les environnements créés par l'homme, la grande majorité des

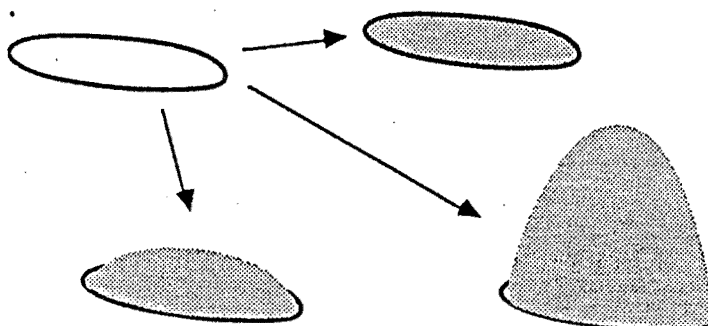


Figure 1.9: Ambiguïté quant à la reconstruction d'une surface.

surfaces est dépourvue de texture et la reconstruction de ce genre de surfaces pose de graves problèmes d'ambiguïté : par exemple, un cercle donné dans l'espace pourrait être la frontière d'une infinité d'éléments lisses (cf. Figure 1.9). Par conséquent, ces méthodes nécessitent des données initiales relativement denses et apparemment elles sont mal adaptées pour être appliquées aux scènes où la plupart des surfaces sont non-texturées.

Un autre reproche que nous faisons aux systèmes de reconstruction stéréo existants est qu'ils s'appuient en général seulement sur l'information tridimensionnelle portée par les caractéristiques reconstruites et n'utilisent guère l'information géométrique et les relations structurelles entre les caractéristiques images dont les caractéristiques 3-D sont issues. Les méthodes proposées souffrent par le manque de généralisation puisqu'elles marchent bien uniquement lorsque les hypothèses et heuristiques utilisées sont présentes dans la scène.

1.5 Notre Approche

Dans cette thèse, nous proposons une méthode qui met en correspondance et reconstruit des contours sans utiliser la contrainte de continuité du monde physique. Pour cela, nous suggérons de tenir compte de contraintes disponibles avec des propriétés de scène qui restent quasi-invariantes en projection perspective. Les contraintes de scène traduisent la structure qui sous-tend le monde physique, structure qui est indépendante de la position de l'observateur. Nous proposons d'utiliser ces contraintes pour découvrir dans chaque image séparément des structures appropriées reflétant directement les propriétés sous-jacentes de la scène. Ces contraintes guident alors une analyse monoculaire effectuée sur chaque image. Le but de cette analyse est d'établir une description monoculaire qui tienne explicitement compte de ces contraintes.

Par rapport à l'approche classique du problème stéréo présentée à la Section 1.2 (page 7), nous allons donc introduire un nouveau concept, à savoir l'analyse et la description monoculaire de chaque image. Cette description va incorporer à la fois des entités géométriques extraites de l'image ainsi que des relations topologiques entre ces entités. Nous suggérons qu'une telle description peut être acquise en

extrayant des entités-image et en les groupant convenablement. L'entité-image de base que nous allons utiliser est le segment de droite.

Plus précisément pour arriver à une telle description les étapes suivantes sont effectuées : des points de contour sont détectés dans les images, puis les chaînes de contour seront approchées par des segments de droite. Ces segments sont groupés convenablement afin de former des structures locales 2-D composées de segments colinéaires, de segments connexes et de segments délimitant une même région. De cette manière, chaque image est décrite par des segments de droite avec leurs attributs géométriques et par des relations entre les segments voisins. Sur la base de ces groupements, nous décrivons ensuite chaque image séparément par un graphe relationnel (que nous appelons *description monoculaire*) dont les noeuds représentent les segments avec leurs propriétés géométriques et les arcs représentent les relations structurelles entre ces segments.

Nous aboutissons ainsi à deux descriptions monoculaires représentant la paire stéréo initiale et le problème de mise en correspondance se ramène à une recherche d'isomorphisme entre les deux descriptions monoculaires. La mise en correspondance est obtenue en trouvant des structures similaires entre les deux descriptions monoculaires. Pour ceci, utilisant des contraintes dictées par la géométrie du capteur, nous établissons d'abord pour chaque segment d'une image une liste de candidats potentiels composée de segments de l'autre image. Ensuite, nous utilisons les propriétés des segments voisins dans les deux graphes relationnels afin d'établir la compatibilité entre les différents appariements potentiels. Les appariements potentiels sont représentés par des noeuds dans un graphe (que nous appelons *graphe de correspondance*) et leurs compatibilités sont représentées par des arêtes dans ce graphe. Des correspondances entre les deux images sont équivalentes à des ensembles d'appariements mutuellement compatibles. Ces ensembles sont fournis par les cliques maximales de ce graphe. La meilleure correspondance est choisie parmi les cliques maximales trouvées selon un critère d'évaluation photométrique, géométrique et topologique. L'algorithme que nous proposons tient compte de dissimilarités éventuelles entre les deux images dues à des occlusions, des alignements accidentels, des inversions des caractéristiques (violation locale de la contrainte d'ordre) et/ou à des défaillances des processus d'extraction et de groupement des caractéristiques.

Après avoir apparié les deux images, nous procédons à la construction d'un modèle géométrique de la scène. Pour ceci, à l'aide de la technique de triangulation et de la théorie des filtres de Kalman, nous déterminons d'abord la position tridimensionnelle des segments appariés et nous leur associons une région ellipsoïdale représentant leurs incertitudes en position. Nous créons ainsi un premier modèle de la scène du type *fil de fer*. Enfin, nous présentons une technique pour redécouvrir des éléments de surfaces de scène qui contiennent les segments 3-D que nous avons construits. Pour diminuer la complexité du problème, nous proposons de décrire la scène en l'approximant en termes d'un ensemble de facettes planes et afin d'arriver à une telle description, nous suggérons de combiner simultanément l'information tridimensionnelle portée par les segments 3-D et l'information monoculaire contenue dans les deux images stéréo.

La Figure 1.10 récapitule l'architecture globale de notre méthode.

1.6 Les travaux voisins

Ayache et Faverjon [5] et Robert de Saint Vincent [75] ont mis en oeuvre une technique de recherche dans un graphe qui met en correspondance des segments de contours : cette mise en correspondance est réalisée par une technique récursive de prédiction/vérification d'hypothèses. Medioni et Nevatia [59] ont aussi apparié ce genre de caractéristiques en utilisant une méthode de relaxation. L'avantage de ces approches est qu'elles utilisent des primitives géométriques riches : les segments de contours restent relativement stables d'une image à l'autre, leur nombre est relativement limité dans les images et par nature ils incluent la contrainte de continuité de formes. Par conséquent leur mise en correspondance devient séduisante. Néanmoins, pour lever l'ambiguïté entre des appariements locaux et pour propager la mise en correspondance, ces approches s'appuient sur l'hypothèse de la continuité de la disparité. Afin de s'affranchir de cette contrainte, par rapport à ces approches nous avons introduit le processus de groupement des caractéristiques-image. Ce processus exploite des contraintes associées à des propriétés de scène qui restent invariantes en projection perspective ; par conséquent, l'utilisation du groupement d'indices semble être la solution appropriée pour lever les ambiguïtés au niveau local et pour aboutir à une correspondance globale compatible avec des propriétés locales.

Le groupement a été utilisée par Herman et Kanade [41] et par Lim et Binford [49] : ils mettent en correspondance des jonctions de segments de droite. Herman et Kanade tiennent aussi compte de quelques connaissances sémantiques disponibles dans les scènes à traiter. Bien que nous ayons aussi implanté une technique robuste d'extraction de jonctions, nous pensons qu'une mise en correspondance basée uniquement sur des jonctions fournit des résultats relativement pauvres ; d'une part parce qu'il y a relativement peu de jonctions dans une image et d'autre part parce que les techniques d'extraction de jonctions sont considérablement influencées par le bruit. Nous proposons de mettre en correspondance des structures 2-D composées de segments de droite et de relations entre eux, telles que : la colinéarité, la connexité (jonctions) et la position relative entre des segments voisins.

1.7 Organisation de la thèse

La suite de cette thèse est organisée de la manière suivante :

- Le Chapitre 2 présente la description monoculaire utilisée par notre système est la façon dont une telle description est construite à partir des données initiales.
- Le Chapitre 3 décrit notre algorithme de mise en correspondance des structures 2-D extraites des deux images. Il illustre également les résultats de l'application de cet algorithme sur quelques paires d'images stéréoscopiques.

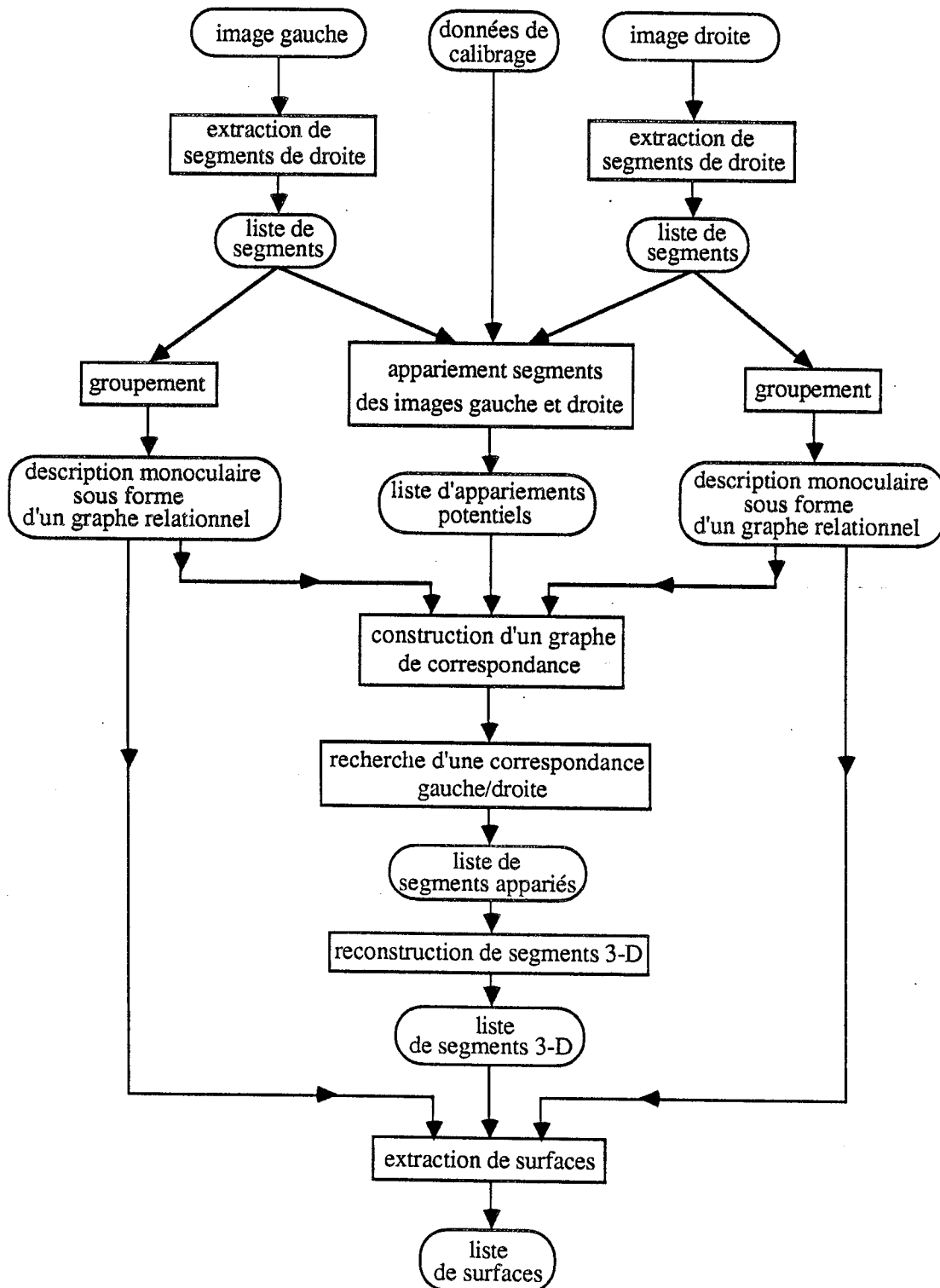


Figure 1.10: L'architecture globale de notre système stéréo.

- Le Chapitre 4 présente l'extraction de la profondeur des structures 2-D appariées et la modélisation des incertitudes qui leurs sont associées. Il aborde ensuite le problème de reconstruction des surfaces de scène qui contiennent les structures 3-D construites ; il présente aussi quelques résultats expérimentaux illustrant notre méthode d'interpolation des surfaces.
- Enfin, le Chapitre 5 résume sommairement notre approche, discute ses avantages et ses inconvénients et indique les améliorations et extensions envisagées pour l'avenir.



Chapitre 2

Description monoculaire d'une image

2.1 Organisation de l'information visuelle

Le défi principal de la Vision par Ordinateur est d'inférer des éléments concernant le monde à partir de l'information contenue dans une ou plusieurs images. La représentation de l'information visuelle et sa structuration sont parmi les problèmes centraux de la Vision par Ordinateur. En réalité, tout système d'interprétation d'une image doit faire face à deux problèmes essentiels :

1. Comment définir les sémantiques de la scène en termes de processus de description de l'image ?
2. Comment réduire l'énorme espace de recherche qui contient la projection des événements intéressants de la scène à une taille maniable et comment l'exprimer sous une forme utile et efficace ?

Pour résoudre de manière efficace ces problèmes, il a été largement soutenu qu'un système de vision devrait posséder une succession de niveaux de représentation [53]. Les premiers niveaux dépendent de ce qui est possible d'extraire directement de l'image et sont relativement indépendants de ce qui est finalement souhaitable. Les représentations supérieures, quant à elles, sont plutôt liées aux informations qui définissent les buts (tels que la reconnaissance d'objets ou la navigation de robots

mobiles). De manière évidente, la construction de l'interprétation de l'image dépend de façon critique du choix des niveaux de représentation, car un choix impropre peut conduire à des calculs lourds et inefficaces.

Une matrice d'intensité lumineuse contient deux types d'information en chaque point : la valeur de l'intensité lumineuse et sa position à partir desquels on peut déterminer les changements de l'intensité lumineuse et la géométrie locale de l'image. Cependant, le "niveau pixel" est trop bas pour être exploité convenablement. L'interprétation d'une image implique donc dans son processus de nombreux autres sous-processus de description et d'analyse intermédiaires, qui transforment la matrice de l'intensité lumineuse en une forme appropriée aux buts du système d'interprétation. Il s'agit en fait de sous-processus qui, commençant par le niveau pixel, génèrent au fur et à mesure de primitives de plus en plus abstraites, jusqu'à atteindre la génération de primitives sémantiques prêtes à servir à l'interprétation. De cette manière, d'une part on *condense* le volume des données à traiter en diminuant ainsi l'espace de recherche et d'autre part on *augmente* la quantité de l'information portée par chaque primitive, fournissant ainsi les primitives sémantiques nécessaires au processus de l'interprétation.

L'organisation de l'information visuelle, sa structuration en niveaux d'abstraction ainsi que sa relation avec les événements correspondants de la scène ont attiré l'attention de plusieurs chercheurs et des nombreux travaux existent dans la littérature spécialisée [26], [51], [76], [85], [86]. Une discussion excellente quant au traitement et à la représentation de l'information visuelle se trouve en [53].

2.2 Notre approche

Une description monoculaire riche doit incorporer des représentations explicites de l'information visuelle contenue dans une matrice d'intensité lumineuse. Elle doit contenir à la fois des entités géométriques extraites de l'image, leurs propriétés intrinsèques, ainsi que des relations topologiques entre ces entités. Une façon de décrire en termes symboliques les propriétés géométriques et topologiques d'une image est de représenter les entités-image et leurs relations sous forme d'un graphe relationnel, où les noeuds représentent les entités et leurs propriétés géométriques et les arcs représentent les relations topologiques entre ces entités.

Le schéma que nous avons adopté pour le processus de l'organisation de l'information visuelle est le suivant : l'entité-image de base sera un segment de droite détecté dans l'image. Nous désirons lui associer deux ensembles de segments qui lui sont connectés à ses extrémités, et deux régions adjacentes, une à sa gauche et une autre à sa droite (nous appelons ces régions *bandes associées aux segments*). Pour cela, nous effectuons un groupement des segments en jonctions et une détection de bandes de part et d'autre de chaque segment. Finalement, nous construisons un graphe relationnel pour représenter les segments et leurs relations. La Figure 2.1 illustre l'architecture globale de notre système de description monoculaire.

Pour mieux illustrer cette représentation, considérons l'exemple de la Figure 2.22 (page 56). Un ensemble de six segments de droite a été extrait d'une image (l_1 à l_6).

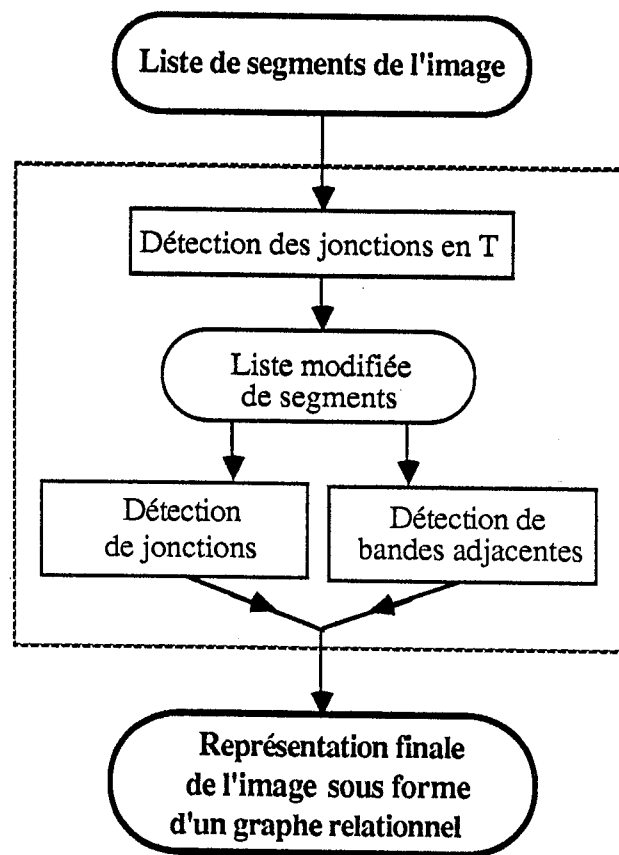


Figure 2.1: L'architecture globale du système de description monoculaire d'une image.

La description que l'on cherche à atteindre est illustrée sur la Figure 2.23. Sur cette figure, la description monoculaire de l'image est représentée par un graphe relationnel, dont les noeuds sont les segments de droite avec leurs attributs géométriques ; les arcs reliant deux noeuds de ce graphe représentent une relation entre deux segments voisins. Nous avons choisi quatre types de relations : 1-même jonction ; 2-colinéarité ; 3-à gauche ; 4-à droite.

2.3 Les motivations de ce choix

Lorsque l'on observe une scène de différents points de vue, l'apparence des objets 3-D composant la scène peut changer complètement d'un point de vue à l'autre. Néanmoins, il est également vrai que certains aspects de la projection d'un objet restent invariants, comme par exemple la connexité, la colinéarité, le parallélisme, les textures répétitives et certains types de symétrie.

Le rôle d'une analyse détaillée de l'image est de détecter ce genre de structures, qui ont peu de probabilité d'apparaître par accident de position ou de prise de vue. Une fois détectées, ces structures et leurs relations pourraient contribuer considérablement à l'interprétation finale de l'image : les relations topologiques locales entre des caractéristiques extraites de l'image sont si improbables d'apparaître par hasard, que, lorsque ces relations sont détectées d'une façon systématique et décrites d'une manière cohérente, elles pourraient simplifier considérablement la complexité de la recherche d'une correspondance image/scène. Par conséquent, la détection et l'utilisation des attributs-image proéminents ou de structures-image de haut niveau faciliteraient beaucoup la tâche de compréhension d'images : elles impliqueraient la sélection et la génération d'hypothèses plus fiables concernant l'identité d'objets présents dans la scène. Le processus d'extraction des configurations locales de l'image est connu sous le nom de *groupement* ou aussi d'*organisation perceptuelle*.

Notre motivation était donc de construire une description de l'image en termes de quelques structures géométriques reflétant directement la richesse sémantique de la scène.

En général, les entités qui sont organisées pour décrire l'image sont de type point, courbe ou région. Les entités de type point ne portent pas d'information sémantique assez riche. Quant à la description par des primitives de type région, bien que les régions représentent de manière explicite quelques zones des images très importantes de point de vue sémantique, cependant beaucoup de structures géométriques importantes ne sont représentées qu'implicitement aux frontières de ces régions. En ce qui nous concerne, nous avons opté pour le segment de droite. Les segments de droite peuvent décrire de façon satisfaisante un grand nombre d'objets et de scènes et en même temps portent explicitement une structure géométrique importante. Néanmoins, de structures géométriques plus complexes, telles que les régions fermées, les collections de lignes colinéaires ou parallèles ne sont qu'implicitement aux segments. Afin de les rendre explicites, nous groupons les segments en structures plus grandes contenant une information de régions, de connexité et de colinéarité.

2.4 Le groupement

Dans cette section nous allons présenter la description monoculaire utilisée par notre système, et la façon dont une telle description est construite à partir des données initiales.

L'analyse de l'image commence avec un processus classique de détection de segments de droite (segmentation) : détection et chaînage des points de contour et approximation polygonale de chaînes de contour. Cette première analyse aboutit à une liste de segments de droite avec leurs propriétés géométriques intrinsèques, telles que la position de leurs deux extrémités, la longueur et l'orientation (exprimée pour l'instant modulo π).

Le processus de groupement prend en entrée la liste de segments extraits et effectue une analyse systématique pour détecter les segments concourants en un point commun (jonctions) et pour extraire deux bandes adjacentes aux segments. L'analyse que nous envisageons met l'accent sur le rôle de tous les segments présents dans l'image. Un segment, même s'il est très court, constitue la frontière physique de deux régions de l'image adjacentes et assure en même temps la connexité entre deux segments ou entre deux jonctions voisines¹ ; cette connexité est réalisée via les deux extrémités du segment. Les principes suivants sont donc nécessaires pour notre analyse :

Principe 1

Un segment-image ne peut avoir que deux régions adjacentes, une à sa gauche et une autre à sa droite.

Principe 2

Un segment-image participe tout au plus à deux jonctions, une à chaque extrémité.

Principe 3

Un segment-image (même s'il est très court) ne peut pas avoir ses deux extrémités incluses dans la même jonction.

La Figure 2.2 illustre la structure locale que nous voulons associer à un segment-image selon les principes ci-dessus.

Toutefois, ce modèle de structure locale d'un segment n'est pas respecté lorsque la configuration de la Figure 2.3 est rencontrée dans l'image. Regardons en détail la situation montrée sur cette figure. Le segment l_2 touche le segment l_1 en un point différent des extrémités de ce dernier ; plus précisément, leur point d'intersection tombe à l'intérieur du segment l_1 , ce qui viole le deuxième principe de notre approche. De plus, si l'on voulait attacher des régions adjacentes au segment l_1 , alors dans le demi-plan (q) situé à gauche du segment l_1 il y aurait une région ($reg1$), tan-

¹La plupart des approches antérieures considèrent ces segments comme du bruit et par conséquent les éliminent de la description finale.

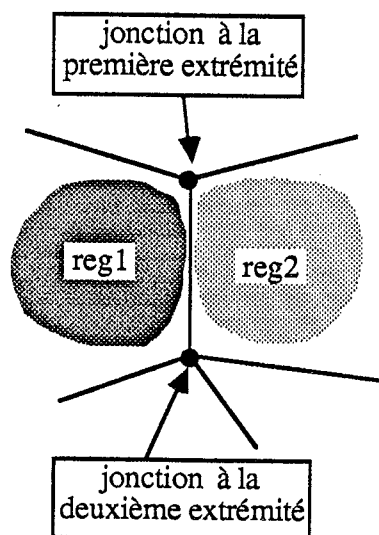


Figure 2.2: La structure locale que nous voulons associer à un segment-image.

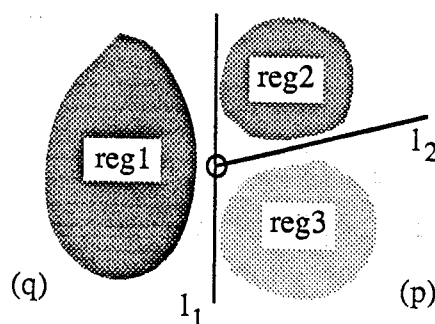


Figure 2.3: La configuration locale d'une jonction en T.

dis que dans le demi-plan (p) situé à droite du segment l_1 , il y aurait deux régions différentes ($reg2$ et $reg3$), séparées par le segment l_2 , ce qui viole le premier des principes ci-dessus.

La configuration de la Figure 2.3 est le cas d'une jonction en "T". Afin de nous affranchir des inconvénients qu'elle soulève, notre analyse nécessite alors une étape préalable, celle de la détection et de la transformation des jonctions en T. Ainsi, l'architecture de notre système de groupement présentée sur la Figure 2.1 est justifiée.

2.4.1 Détection et transformation de jonctions en T

Deux segments d'une image forment une jonction en T, lorsqu'ils se croisent en un point se trouvant à l'intérieur d'un des deux segments. La Figure 2.4-a illustre le cas d'une jonction en T *parfaite*, tandis que les Figures 2.4-b,-c,-d montrent quelques exemples de situations difficiles à traiter.

Pour détecter un T, l'algorithme procède à une analyse locale autour de chaque extrémité d'un segment ; il ouvre une fenêtre de taille $w \times w$ pixels et il teste

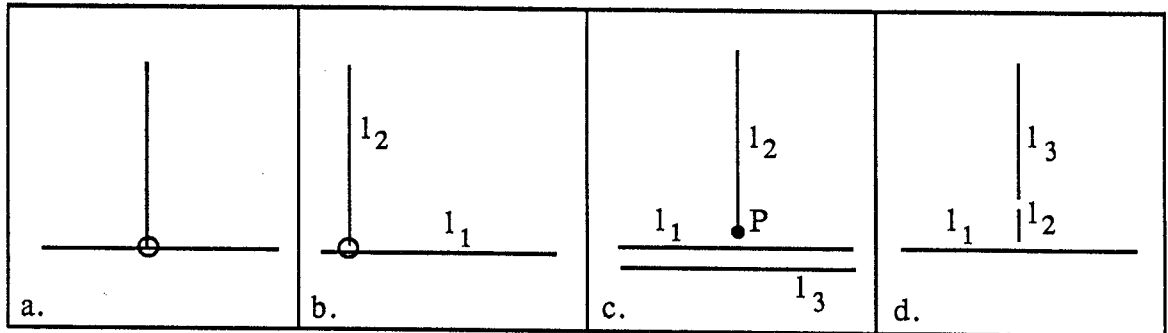


Figure 2.4: Jonction en T *parfaite* (a), et quelques structures image où la détection d'un T devient difficile (b, c, d).

l'intersection du segment courant avec tous les segments qui croisent cette fenêtre. Considérons un tel segment l_1 se trouvant à l'intérieur de la fenêtre ouverte autour de l'extrémité B_1 d'un autre segment l_2 (voir Figure 2.5). Soit $\Delta\theta$ l'angle relatif entre les deux segments, Q leur point d'intersection, d la distance QB_1 et d_1 , d_2 les distances QA_1 et QA_2 respectivement. Pour qu'un T soit détecté, les conditions suivantes doivent être satisfaites :

1. $\epsilon_1 \leq \Delta\theta \leq \pi - \epsilon_1$
2. $d \leq w/2$
3. $0 < \alpha < 1$
4. $d_1 \geq \epsilon_2$
5. $d_2 \geq \epsilon_2$
6. il ne doit pas y avoir de segment-image qui coupe le segment QB_1 ;
7. il ne doit pas y avoir de segment-image contenu dans une fenêtre ouverte autour du segment QB_1 et dont l'orientation relative par rapport à celle du segment QB_1 soit inférieure à ϵ_1 .

(1.) ϵ_1 est un seuil sur l'orientation relative des deux segments et il permet d'éviter de traiter les segments qui sont presque parallèles.

(2.) Le deuxième test permet d'éliminer rapidement les segments dont l'intersection avec le segment l_2 se situe loin de l'extrémité B_1 (le seuil étant fixé à la moitié de la taille w de la fenêtre de travail).

(3.) α est un coefficient exprimant la position de la jonction en T sur le segment l_1 ; soient (x_Q, y_Q) , (x_{A_1}, y_{A_1}) et (x_{A_2}, y_{A_2}) les coordonnées-image des points Q , A_1 et A_2 , respectivement. α est donné par la formule :

$$\alpha = \frac{y_Q - y_{A_1}}{y_{A_2} - y_{A_1}} \quad \text{si } |y_{A_2} - y_{A_1}| \geq |x_{A_2} - x_{A_1}|$$

$$\alpha = \frac{x_Q - x_{A_1}}{x_{A_2} - x_{A_1}} \quad \text{autrement.}$$

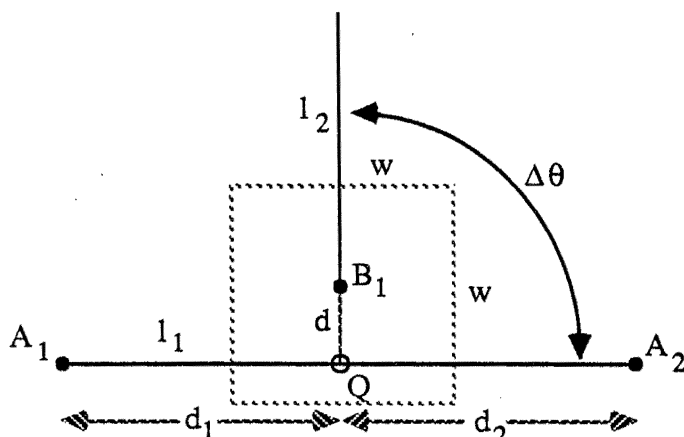


Figure 2.5: Critères géométriques pour la détection d'une jonction en T.

Des valeurs de α inférieures à 0 ou supérieures à 1 signifient que le point d'intersection se situe à l'extérieur du segment l_1 , tandis que des valeurs de α entre 0 et 1 signifient que le point d'intersection se situe à l'intérieur du segment l_1 —cas d'une jonction en T. Des valeurs de α autour de 0.5 signifient que le T est formé à peu près au milieu du segment l_1 .

(4. et 5.) Le quatrième et le cinquième tests ont été ajoutés pour élucider le cas ambigu de la Figure 2.4-b. Dans ce cas, le T est formé près d'une des deux extrémités. Faut-il le considérer ? Les deux tests permettent justement une telle décision : un T est détecté uniquement quand la distance du point d'intersection à l'extrémité la plus proche du segment l_1 est supérieure au seuil ϵ_2 .

(6. et 7.) Finalement, les deux derniers critères permettent de traiter les cas difficiles des Figures 2.4-c et 2.4-d. Ainsi, le sixième test interdit la formation d'un T entre les segments l_2 et l_3 , parce qu'il existe un segment (l_1) entre l_3 et l'extrémité P de l_2 . Dans ce cas, il n'y aura donc qu'un seul T, formé par les segments l_1 et l_2 . Quant à la Figure 2.4-d, la présence du segment l_2 à l'intérieur de la fenêtre entre l'extrémité de l_3 et le point d'intersection des l_1 et l_3 interdit la formation d'un T sur le segment l_1 (dernier critère). Le T sera formé par les segments l_1 et l_2 .

Les Figures 2.6-a,-b,-c récapitulent les solutions proposées aux situations difficiles des Figures 2.4-b,-c,-d, respectivement.

Lors de la détection d'un T, on coupe le segment formant la barre supérieure du T en deux, créant ainsi deux nouveaux segments. Ainsi le segment l_3 de la Figure 2.7-a est coupé en trois segments : l_3 , l_4 et l_5 (Figure 2.7-b). Il en résulte une augmentation du nombre total des segments de l'image. (Lorsqu'un segment est découpé en plusieurs petits morceaux, l'information de connexion et de colinéarité entre les différents morceaux est conservée).

Cette fragmentation des segments lors de la détection d'un T permet de résoudre de manière élégante les problèmes soulevés par les T ; notre analyse peut maintenant se baser sur les trois principes que nous avons présentés à la Section 2.4 pour détecter les jonctions associées aux extrémités des segments et pour extraire les deux bandes adjacentes aux segments.

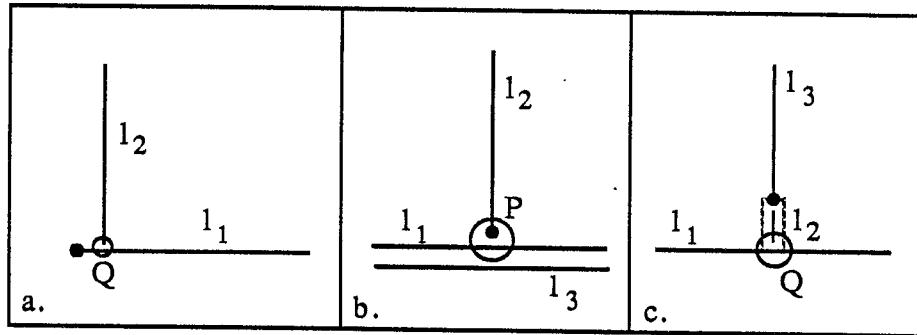


Figure 2.6: Le résultat de l'algorithme de détection des jonctions en T appliqué aux cas *difficiles* des Figures 2.4-b,-c,-d. La présence d'un cercle indique la détection d'un T.

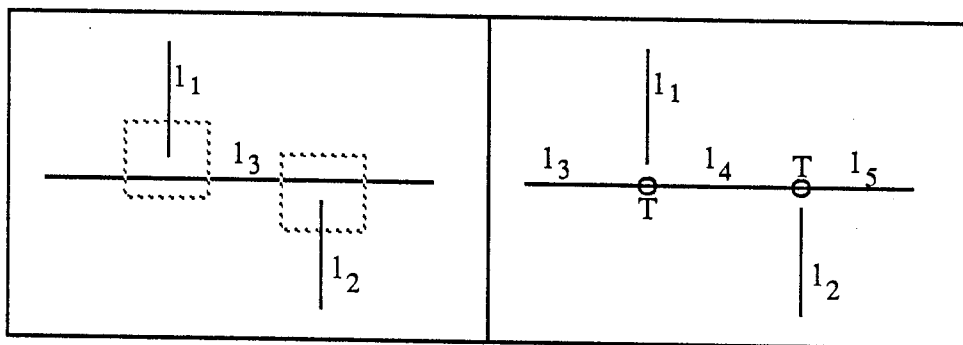


Figure 2.7: Détection des jonctions en T. l_3 devient l_3 , l_4 et l_5 .

2.4.2 Détection de jonctions

2.4.2.1 Les difficultés de la détection

Une jonction est un groupe de deux segments au minimum, qui se rencontrent en un point commun.

Un algorithme générique de détection de jonctions nécessiterait de considérer tout d'abord tous les segments de l'image, deux par deux, de calculer ensuite leur point d'intersection et de chercher finalement s'il y a d'autres segments qui passent par ce même point. La complexité algorithmique d'une telle approche est de l'ordre de $O(N^2)$, où N est le nombre de segments-image.

Afin de réduire la combinatoire associée au processus de groupement, une idée, proposée dans [41] et [84], serait d'analyser localement le voisinage de chaque segment² : on ouvre une petite fenêtre à chaque extrémité d'un segment et on analyse seulement les segments qui traversent cette fenêtre. Toutefois, cette approche ne garantit guère l'obtention de bons résultats.

En effet, la principale difficulté de l'extraction de jonctions est due aux erreurs inhérentes à la segmentation de l'image. La présence de bruit, d'ombres, de reflets, d'objets qui se touchent ou qui se recouvrent influence les résultats et la segmentation est entachée d'erreurs.

Voici énumérés quelques aspects dont il faut tenir compte lors de la détection d'une jonction :

- un segment-image peut être brisé en plusieurs petits morceaux ;
- une partie du segment peut ne pas être détectée ;
- un segment-image peut être légèrement incliné par rapport à son orientation théorique.

La Figure 2.8-a illustre le cas d'une jonction *parfaite*, tandis que les Figures 2.8-b,-c,-d montrent le même type de jonctions *déformées* par les processus de formation et de segmentation de l'image.

Un autre phénomène qui perturbe souvent le processus de groupement est dû aux structures images dont la présence crée des ambiguïtés quant à la décision à prendre. Quelques exemples de situations difficiles sont montrés sur la Figure 2.9 : faut-il grouper tous les segments en une seule jonction, ou bien faut-il les considérer seulement deux par deux ?

En conclusion, nous pensons qu'une analyse explicite de la géométrie et de la topographie locales de l'image est indispensable si l'on veut obtenir de bons résultats

²Cette idée est matérialisée de la manière suivante : une image artificielle est créée à partir des segments de droite, où à chaque pixel on associe l'information suivante :

- -1, s'il n'y a pas de segment passant par ce pixel ;
- un indice correspondant au segment qui passe par ce pixel ; cet indice est en fait l'indice que possède ce segment dans la liste des segments.

Ainsi, il est possible de savoir quel est le segment de droite qui passe par n'importe quel pixel de l'image.

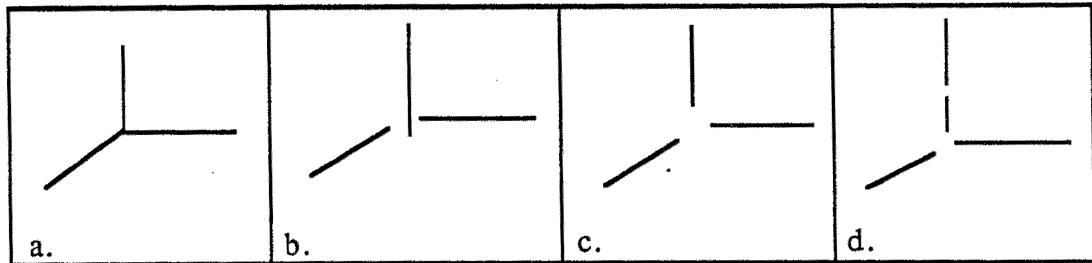


Figure 2.8: Jonction *parfaite* (a), et jonctions *déformées* par le processus de segmentation (b, c, d).

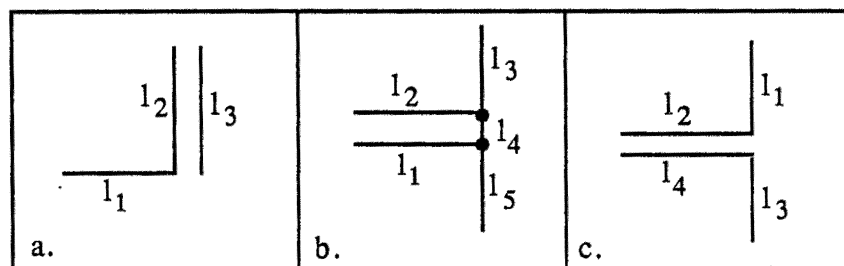


Figure 2.9: Quelques structures-image où le groupement devient difficile car il n'y a pas de point d'intersection commun pour tous les segments.

de groupement.

2.4.2.2 L'algorithme de détection des jonctions

Nous voulons associer à un segment-image deux ensembles de segments (deux jonctions) chacun se situant à chaque extrémité du segment. Pour ceci, nous avons introduit les Principes 2 et 3 et nous avons commencé notre analyse par la détection de jonctions en T. Afin de faciliter encore plus cette analyse, nous avons décomposé l'algorithme de détection de jonctions en deux étapes successives que nous justifierons par la suite :

- détection des angles ;
- rassemblement des angles en jonctions.

Détection des angles

Soient les exemples de la Figure 2.9. Intuitivement on voudrait trouver une jonction (Figure 2.9-a), deux jonctions reliées par un segment (Figure 2.9-b) et une jonction (Figure 2.9-c), respectivement. Cependant, dans les cas 2.9-a et 2.9-c il n'y a pas d'intersection unique, commune à tous les segments. Nous allons donc tout d'abord considérer les segments deux par deux et déterminer s'ils forment un angle.

Pour détecter un angle entre deux segments-image, l'algorithme procède à une analyse locale autour de chaque extrémité d'un segment : il ouvre une fenêtre de

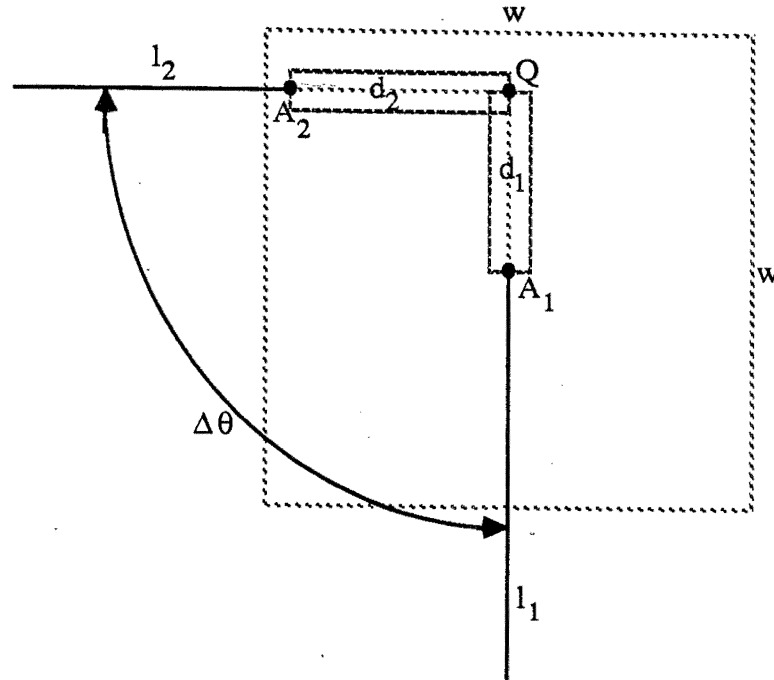


Figure 2.10: Critères géométriques pour la détection d'un angle.

taille $w \times w$ pixels et il élimine tous les segments de l'image dont au moins une extrémité ne passe pas par cette fenêtre. Soit l_2 un segment-image se trouvant à l'intérieur de la fenêtre ouverte autour de l'extrémité A_1 d'un autre segment-image l_1 . (Voir Figure 2.10). Trois mesures contribuent à déterminer si les segments l_1 et l_2 forment un angle : $\Delta\theta$, d_1 et d_2 . $\Delta\theta$ est tout simplement l'angle entre les segments l_1 et l_2 . Soit Q le point d'intersection de l_1 et l_2 et soient A_1 et A_2 les extrémités des segments l_1 et l_2 les plus proches du point Q . d_1 est la distance entre Q et A_1 et d_2 la distance entre Q et A_2 . Pour que les segments l_1 et l_2 forment un angle, les conditions suivantes doivent être satisfaites :

- $\epsilon_1 \leq \Delta\theta \leq \pi - \epsilon_1$
- $d_1 \leq w/2$
- $d_2 \leq w/2$
- il ne doit pas y avoir de segment-image contenu dans une fenêtre ouverte autour du segment QA_1 et dont l'orientation relative par rapport à celle du segment QA_1 soit inférieure à ϵ_1 ;
- le critère précédent, appliqué cette fois-ci à une fenêtre ouverte autour du segment QA_2 .

Montrons l'utilisation de ces critères sur l'exemple illustré sur la Figure 2.11. Soit l_1 un segment, et soit A_1 une extrémité de ce segment. On va essayer de détecter tous les angles possibles formés par le segment l_1 et par les autres segments se trouvant

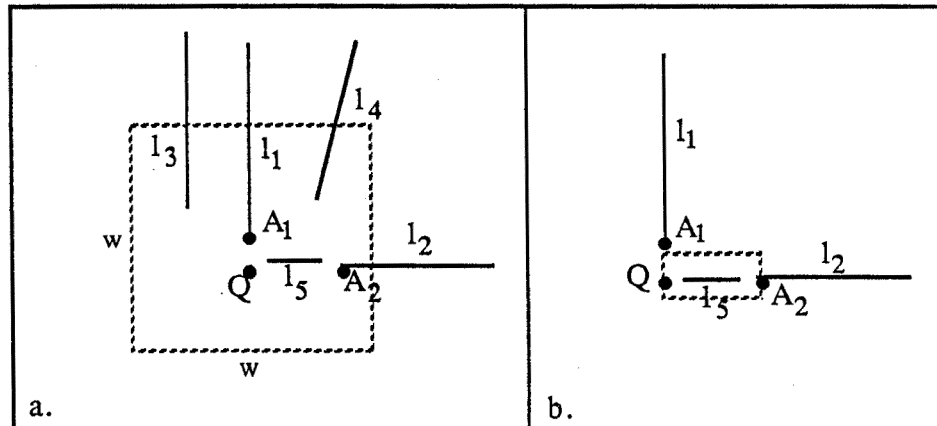


Figure 2.11: Un exemple de détection des angles. On cherche les angles pouvant être formés par le segment l_1 et les segments-image se trouvant autour de l'extrémité A_1 du segment l_1 . Un seul angle est finalement formé, l'angle $(l_1 l_5)$.

autour du point A_1 . Pour cela, on ouvre une fenêtre de taille $w \times w$ pixels autour de A_1 . On élimine tous les segments de l'image qui ne passent pas par cette fenêtre. On va donc considérer uniquement les paires de segments $(l_1 l_2)$, $(l_1 l_3)$, $(l_1 l_4)$ et $(l_1 l_5)$.

Le premier critère géométrique permet d'éviter de calculer l'intersection entre deux segments-image qui sont presque parallèles, ce parallélisme étant déterminé par le seuil ϵ_1 . Ainsi ce critère élimine d'emblée le couple $(l_1 l_3)$. Pour les trois paires de segments qui restent, on calcule leur point d'intersection. Par exemple, Q est le point d'intersection de l_1 avec l_2 . On sélectionne ensuite, pour chaque segment, l'extrémité la plus proche du point d'intersection et on calcule la distance entre cette extrémité et le point d'intersection. Le deuxième et le troisième critères sont ensuite utilisés pour éliminer tous les segments pour lesquels cette distance est supérieure à un seuil $(w/2)$. Par exemple, la paire $(l_1 l_4)$ ne peut pas former un angle car cette distance est trop grande.

Finalement on ouvre une fenêtre entre le point d'intersection et chaque extrémité (Figure 2.11-b). Pour accepter la paire $(l_1 l_i)$ comme angle, on vérifie que les fenêtres QA_1 et QA_i ne contiennent aucun autre segment. Ce test permet d'éliminer la paire $(l_1 l_2)$ car la fenêtre QA_2 contient le segment l_5 . Dans cet exemple seule la paire $(l_1 l_5)$ est acceptée comme angle.

Dans le cas de la Figure 2.9-a cet algorithme détecte les angles $(l_1 l_2)$ et $(l_1 l_3)$.

Rassemblement des angles en jonctions

Ayant extrait les angles, le passage aux jonctions est relativement direct. En effet, un raisonnement purement logique suffit pour grouper en une jonction les angles qui partagent un segment. On considère d'abord un angle dans la liste des angles ; on parcourt ensuite la liste des angles afin de trouver ceux qui partagent un segment avec l'angle considéré initialement.

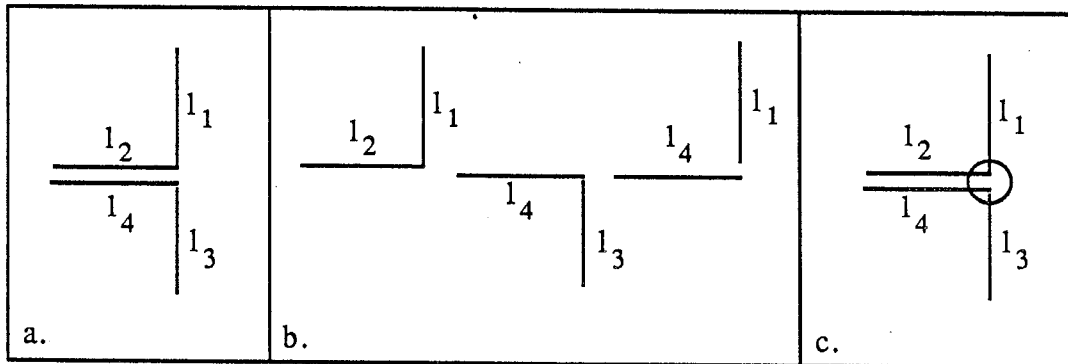


Figure 2.12: Formation d'une jonction à partir des angles. (a) Situation initiale. (b) Angles détectés. (c) La jonction formée.

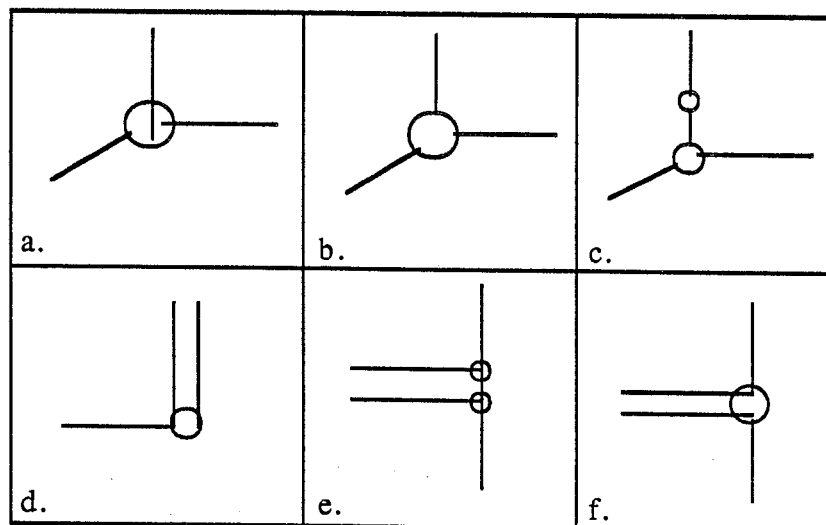


Figure 2.13: Le résultat de l'algorithme de groupement appliqué aux cas *difficiles* des Figures 2.8 et 2.9.

Illustrons cette méthode sur l'exemple de la Figure 2.9-c reproduit sur la Figure 2.12-a. Supposons que nous ayons détecté les angles $(l_1 l_2)$, $(l_1 l_4)$ et $(l_3 l_4)$ (Figure 2.12-b). D'une part, $(l_1 l_2)$ et $(l_1 l_4)$ ont l_1 en commun. D'autre part, $(l_1 l_4)$ et $(l_3 l_4)$ ont l_4 en commun. Finalement on obtient la jonction $(l_1 l_2 l_3 l_4)$ (Figure 2.12-c).

La Figure 2.13 montre le résultat de la détection de jonctions appliquée aux cas *difficiles* des Figures 2.8 et 2.9.

Finalement, la Figure 2.14 illustre le comportement de notre algorithme en présence d'une structure répétitive : supposons que parmi les segments parallèles l_2 , l_3 , l_4 et l_5 , seuls les segments l_2 et l_3 forment un angle au sens de notre analyse avec le segment l_1 . Dans ce cas, on va obtenir la jonction $(l_1 l_2 l_3)$ (Figure 2.14-b).

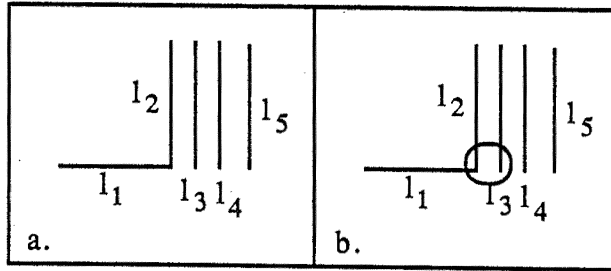


Figure 2.14: Le résultat de l'algorithme de groupement appliqué à une structure répétitive : supposons que l'on ait détecté les angles $(l_1 l_2)$ et $(l_1 l_3)$; alors la jonction formée sera la jonction $(l_1 l_2 l_3)$.

Détection de segments colinéaires

L'algorithme de détection de jonctions que nous venons de présenter ne calcule pas explicitement une relation très intéressante entre deux segments, à savoir la colinéarité. La colinéarité est un cas spécial du parallélisme et pour l'inclure, nous avons ajouté à notre algorithme une analyse supplémentaire qui la détecte de manière explicite.

La Figure 2.15 illustre les critères géométriques qui doivent être satisfaits pour que deux segments $(l_1$ et $l_2)$ soient colinéaires. Autour de l'extrémité A_1 du segment l_1 , on ouvre une fenêtre "étroite" de taille $w \times v$ pixels (dans la pratique, $v = w/3$) et on considère tous les segments qui passent par cette fenêtre. Soit l_2 un tel segment, soit A_2 son extrémité la plus proche à A_1 et soit $\Delta\theta$ l'angle qu'il forme avec l_1 . Pour que l_1 et l_2 soient colinéaires, ils doivent satisfaire les conditions suivantes :

- $\Delta\theta \leq \epsilon_1$ ou $\pi - \epsilon_1 \leq \Delta\theta \leq \pi$
- $recouvrement(l_1, l_2) < max_recouvrement$
- il ne doit pas y avoir de segment-image contenu dans une fenêtre ouverte autour du segment A_1A_2 et dont l'orientation relative par rapport à celle du segment l_1 soit inférieure à ϵ_1 .

La fonction $recouvrement()$, retourne la zone de recouvrement entre deux segments. Pour les deux segments l_1 et l_2 de la Figure 2.15, leur zone de recouvrement est déterminée par le segment commun entre l_1 et la projection de l_2 sur l_1 . Pour ces segments donc, la fonction $recouvrement(l_1, l_2)$ va retourner la longueur du segment $A_1A'_2$. Le deuxième critère s'énonce alors : la longueur de la zone de recouvrement doit être inférieure à un seuil ($max_recouvrement$), pour que l_2 soit colinéaire avec l_1 . Dans toutes nos expérimentations, ce seuil était égal à $L/10$, où L est la plus courte longueur entre l_1 et l_2 .

Illustrons les critères ci-dessus sur l'exemple de la Figure 2.16. Dans la fenêtre ouverte autour de l'extrémité A_1 du segment l_1 , il y a quatre segments (l_2 à l_5). Le segment l_5 est d'emblée éliminé car l_5 n'est pas parallèle à l_1 . Le critère de recouvrement élimine ensuite le segment l_4 . Finalement, le troisième critère élimine

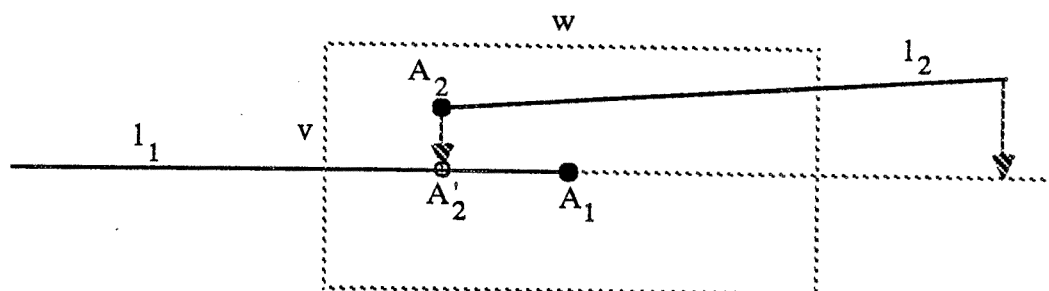


Figure 2.15: Critères géométriques pour la détection de deux segments colinéaires.

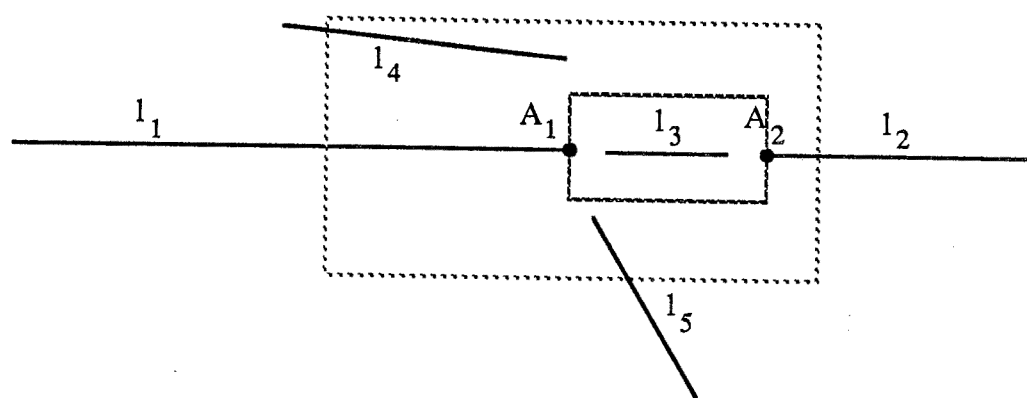


Figure 2.16: Un exemple de détection des segments colinéaires avec le segment l_1 à l'extrémité A_1 . Seul le segment l_3 est colinéaire avec l_1 .

le segment l_2 car à l'intérieur de la fenêtre A_1A_2 il y a le segment l_3 . Seul le segment l_3 est colinéaire avec le segment l_1 . Par ailleurs on trouve que l_2 et l_3 sont colinéaires.

2.4.3 Détection de bandes associées aux segments

Nous voulons associer à chaque segment deux régions adjacentes, une à sa gauche et une autre à sa droite que nous appelons *bandes associées à un segment*.

La détection de ces bandes est réalisée de la façon suivante : considérons un point du segment (Figure 2.17). A partir de ce point, on se déplace pixel par pixel dans une direction perpendiculaire au segment. Cette opération est effectuée à gauche et à droite du segment jusqu'à la rencontre d'un point de contour-image. On répète ce processus pour tous les points du segment. Ainsi, chaque bande (gauche et droite) est bornée d'un côté par le segment lui-même et de l'autre côté, par un ensemble de points de contour ou par l'ensemble de segments passant par ces points de contour (Figure 2.18). De plus, une telle bande est caractérisée par la valeur moyenne des niveaux de gris de ses pixels. Il est possible maintenant d'**orienter** le segment relativement aux bandes qui lui sont associées, de façon que la bande la plus sombre se trouve à gauche du segment et la bande la plus claire à sa droite. La Figure 2.17, illustre cette description.

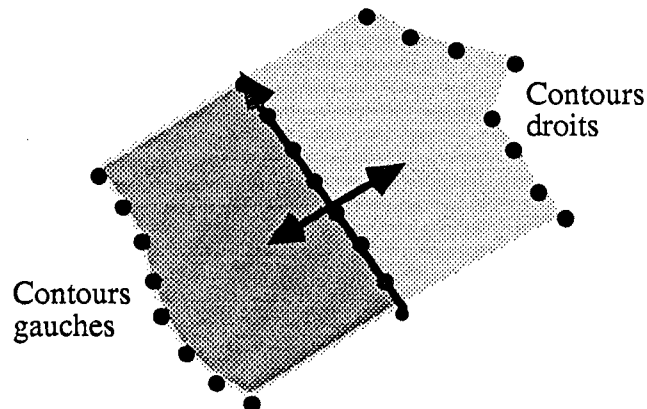


Figure 2.17: Un segment-image et la détection des deux bandes adjacentes qui lui sont associées.

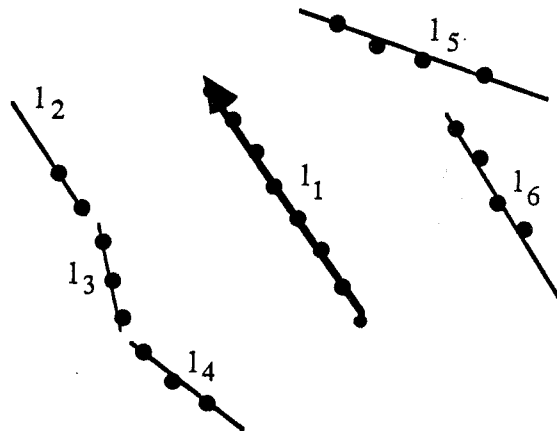


Figure 2.18: Les segments-image bornant les deux bandes associées au segment de la figure précédente.

Selon leur appartenance à la bande gauche ou à la bande droite du segment l_1 , il est possible de caractériser maintenant les segments bornant ces deux bandes par une relation de type *à gauche* ou *à droite*, par rapport au segment l_1 . Ainsi, les segments l_2 , l_3 et l_4 se trouvent à gauche du segment l_1 , tandis que les segments l_5 et l_6 se trouvent à droite du segment l_1 (Figure 2.18).

La Figure 2.19 illustre une difficulté quant à la détermination d'une relation entre les segments l_1 et l_3 : est-ce qu'il faut considérer que le segment l_3 se trouve à droite du segment l_1 ? En fait, le segment l_3 ne participe qu'avec un seul point à la frontière de la bande droite du segment l_1 . Pour accepter une relation de type *à gauche* ou *à droite*, un segment doit border une bande avec au moins un certain nombre de points, suffisants à rendre explicitement claire une telle relation³. Ainsi, sur la Figure 2.19, seuls les segments l_2 et l_4 se trouvent à droite du segment l_1 .

³Dans nos expérimentations, nous avons fixé ce seuil à 3 points.

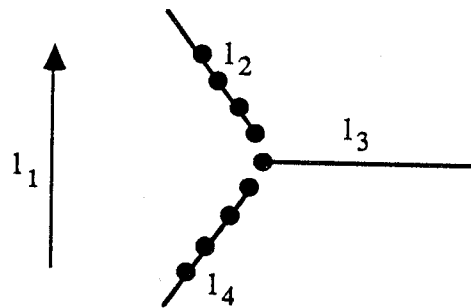


Figure 2.19: Un exemple où la relation entre les segments l_1 et l_3 est ambiguë : doit-on considérer que l_3 appartient à la frontière de la bande droite associée à l_1 ?

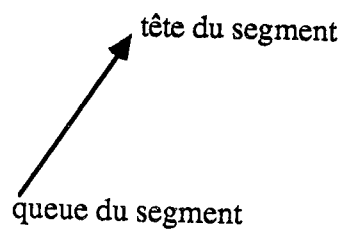


Figure 2.20: La signification de tête et de queue que nous associons aux extrémités d'un segment par rapport à son orientation.

Le but de cette analyse d'extraction des bandes de part et d'autre des segments n'est pas de former de régions dans l'image, mais :

- d'établir l'orientation du segment modulo 2π ;
- d'associer un contraste à chaque segment ; le contraste est défini par la différence des niveaux de gris moyens associés à chacune de deux bandes associées au segment ;
- d'établir une liste de segments se trouvant à gauche du segment courant et une autre liste de segments se trouvant à droite du segment courant.

Puisque nous avons établi l'orientation d'un segment modulo 2π , à partir de maintenant, nous pouvons distinguer les deux extrémités du segment en *tête* et *queue*. Ceci est illustré sur la Figure 2.20.

2.5 Construction d'une représentation de l'image

Selon l'analyse décrite jusqu'à maintenant, nous associons à un segment-image une structure qui comprend :

- les attributs locaux du segment (position, orientation, longueur, contraste) ;
- une liste de segments constituant la frontière de la bande gauche qui lui est associée ;

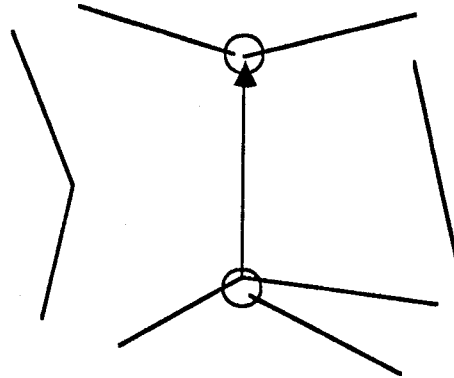


Figure 2.21: La structure locale associée à un segment-image.

- une liste de segments constituant la frontière de la bande droite qui lui est associée ;
- une liste de segments connectés à sa tête ;
- une liste de segments connectés à sa queue.

Cette structure locale est illustrée sur la Figure 2.21. Il résulte une représentation des propriétés géométriques et topologiques locales de l'image sous la forme d'un graphe relationnel. Chaque noeud dans ce graphe représente un segment de l'image avec ses attributs géométriques ; chaque arc reliant deux noeuds représente une propriété reliant deux segments voisins. Il y a quatre types de propriétés :

1. même jonction
2. colinéarité.
3. à gauche ;
4. à droite ;

Nous appelons ces propriétés *relations topographiques*.

Présentons sous forme d'un exemple une telle description. La Figure 2.22 montre un ensemble initial de six segments d'une image (les flèches indiquant l'orientation) et la Figure 2.23 montre la représentation sous forme de graphe relationnel de leur configuration. Nous appelons un tel graphe *description monoculaire*. Cette description reflète localement la structure des contours polygonaux d'une image.

2.6 Résultats expérimentaux

Nous allons illustrer le processus de groupement de caractéristiques sur l'image de la Figure 2.25. La Figure 2.26 montre le résultat d'extraction de segments de droite utilisant l'algorithme de détection de contour de Deriche [27] suivi par un algorithme d'approximation polygonale par morceaux dû à Pavlidis et Horowitz [66].

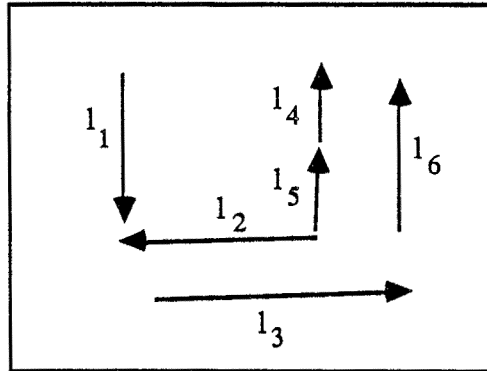


Figure 2.22: Un ensemble de six segments d'une image avec leurs orientations.

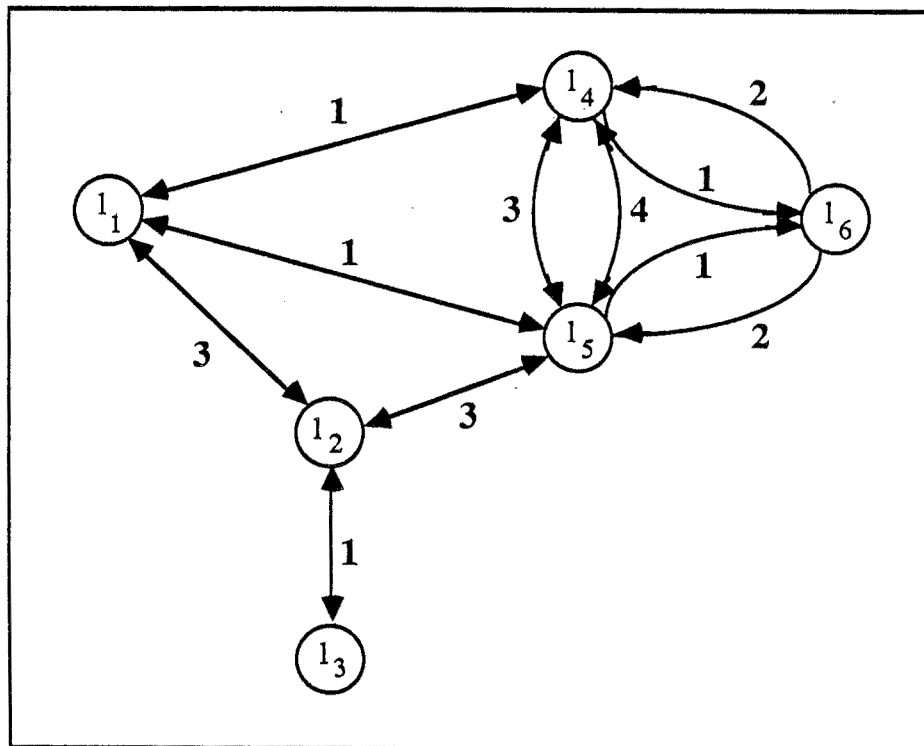


Figure 2.23: Le graphe relationnel qui décrit la configuration des six segments de la Figure 2.22. Les numéros aux arcs de ce graphe représentent les quatre relations topographiques suivantes : 1-à gauche ; 2-à droite ; 3-même jonction et 4-colinéarité.

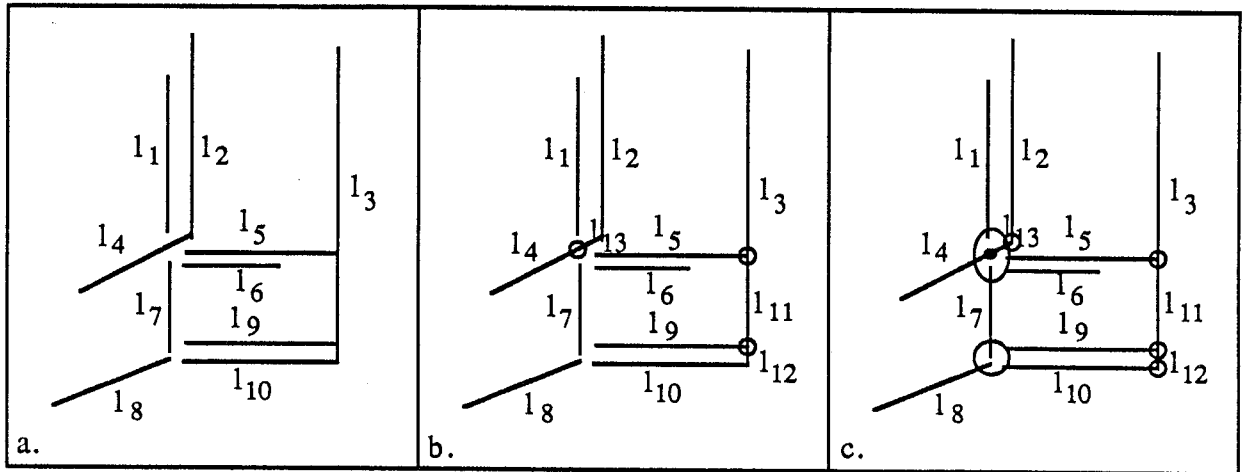


Figure 2.24: Le résultat du groupement à l'intérieur de la fenêtre montrée sur la Figure 2.26 et agrandie ici. (a) Le contenu de cette fenêtre. (b) Trois jonctions en T ont été détectées. (c) six jonctions ont été finalement formées (voir texte).

La Figure 2.27 montre le résultat de la détection des T ; de 345 segments initialement extraits, il en résulte 415 segments (70 jonctions en T ayant été détectées). L'algorithme de détection des angles détecte 462 angles et l'algorithme de groupement logique de ces angles fournit 236 jonctions, dont 110 de deux segments (Figure 2.28), 79 de trois segments (Figure 2.29) et 47 de quatre segments ou plus (Figure 2.30). Finalement, l'algorithme de détection de segments colinéaires détecte 103 paires de segments colinéaires, montrées sur la Figure 2.31. Pour cette image, le temps de groupement est de 11 secondes CPU sur un VAX 11/780 (détection des T, des angles, des jonctions et des segments colinéaires) et la fenêtre locale choisie était de 4×4 pixels.

Détaillons le résultat du groupement à l'intérieur d'une petite fenêtre montrée sur la Figure 2.26 et reproduite en agrandissement sur la Figure 2.24-a. A première vue, il n'y a pas de structure évidente en cet endroit de l'image. Présentons ce que notre algorithme en extrait. Il y a dix segments dans cette fenêtre (l_1 à l_{10} , sur la Figure 2.24-a). Trois jonctions en T ont été détectées (Figure 2.24-b) ; les deux sur le segment l_3 , formées par l_5 et l_9 respectivement et la troisième sur le segment l_4 formée par l_1 (en fait l_5 et l_7 forment aussi la même configuration). Fragmentant le segment l_3 en trois et le segment l_4 en deux, le nombre de segments-image dans cette fenêtre devient 13 (l_1 à l_{13} , sur la Figure 2.24-b). Le groupement de ces segments est présenté sur la Figure 2.24-c. Six jonctions ont été détectées dont : deux à deux segments, (l_2 l_{13}) et (l_{10} l_{12}) ; deux à trois segments, (l_3 l_5 l_{11}) et (l_9 l_{11} l_{12}) ; une à quatre segments, (l_7 l_8 l_9 l_{10}) et une à six segments, (l_1 l_4 l_5 l_6 l_7 l_{13}). Finalement, trois paires de segments colinéaires ont été détectées : (l_3 l_{11}), (l_4 l_{13}) et (l_{11} l_{12}). (Cette information était déjà connue, puisque ces segments proviennent de la détection de jonctions en T).

Le Tableau 2.1 récapitule tous les résultats expérimentaux présentés ci-dessus.

RESULTATS DU GROUPEMENT	Image	Fenêtre
No. segments initiaux	345	10
No. jonctions en T	70	3
No. segments après détection des T	415	13
No. jonctions de deux segments	110	2
No. jonctions de trois segments	79	2
No. jonctions de quatre segments ou plus	47	2
No. paires de segments colinéaires	103	3
Temps CPU (secondes)	11	0.4

Table 2.1: Résultats du groupement.

2.7 Conclusions

Nous avons présenté une méthode de description d'une image en termes de propriétés géométriques et de relations topologiques locales. Cette représentation en contours et en régions est valable à un niveau de résolution fixé et il s'avère qu'on peut aboutir à une représentation de même nature pour chaque niveau de résolution.

Les horizons d'application d'une telle représentation sont assez ouverts. Voici une liste (non exhaustive) d'améliorations et d'extensions possibles :

1. enrichir les propriétés des bandes adjacentes associées aux segments par une analyse statistique explicite ou en ajoutant l'information de couleur et de texture ;
2. élargir notre notion de bande en l'étendant à des régions homogènes complètes (régions complètement bornées par un ensemble de segments) ;
3. enrichir les relations entre des segments voisins en caractérisant précisément leur relation géométrique (par exemple : valeur de leur angle, distance et longueur de leur zone de recouvrement en cas de segments parallèles, etc.) ;
4. extraire des structures images plus globales, telles que des collections de lignes colinéaires, parallèles ou orthogonales. Un tel groupement a été effectué par Reynolds et Beveridge [74] qui ont utilisé une représentation de l'image très similaire à la nôtre.
5. extraire des polygones fermés et les caractériser par leur forme (carrés, parallélogrammes, etc.), concave ou convexe, etc.

Le description que nous avons proposée reste quand même à un niveau d'abstraction assez bas, puisqu'elle n'exploite que l'information provenant directement de l'image. Néanmoins, selon le but souhaité, la représentation choisie pourrait facilement s'étendre pour inclure des informations de plus haut niveau en introduisant des connaissances supplémentaires, telles que la forme ou le modèle des objets à reconnaître.

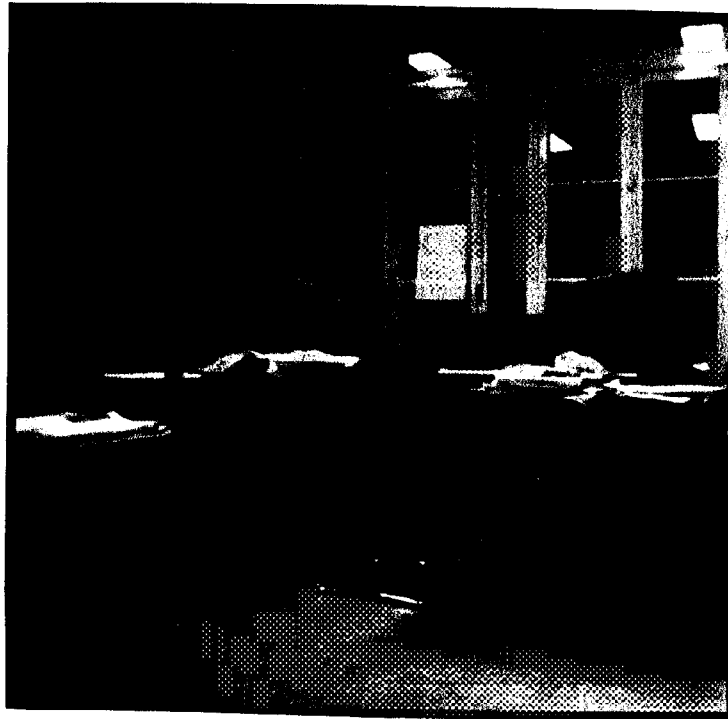


Figure 2.25: L'image d'un bureau.

Notre description monoculaire saisit partiellement la structure locale de l'image. En fait, pour nous cette description ne constitue qu'un premier pas vers la tâche complexe de la compréhension du contenu des images. Elle est donc considérée comme une étape préliminaire, mais indispensable, de tout système de vision qui vise ce but. Dans ce contexte, cette description pourrait être vue comme une partie importante d'un système d'apprentissage de formes bidimensionnelles à partir d'images, comme celui proposé par Connell et Brady [23].



Figure 2.26: Les segments de droite extraits de l'image précédente. La fenêtre sur cette figure, est la fenêtre que nous avons agrandie sur la Figure 2.24 et à l'intérieur de laquelle nous avons détaillé le résultat du groupement.

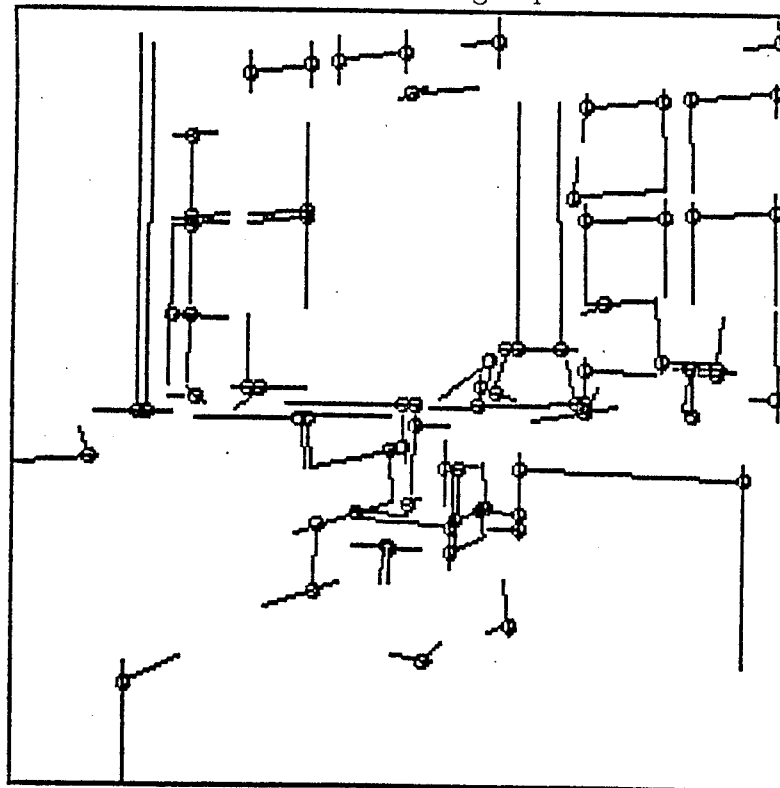


Figure 2.27: Les jonctions en T.

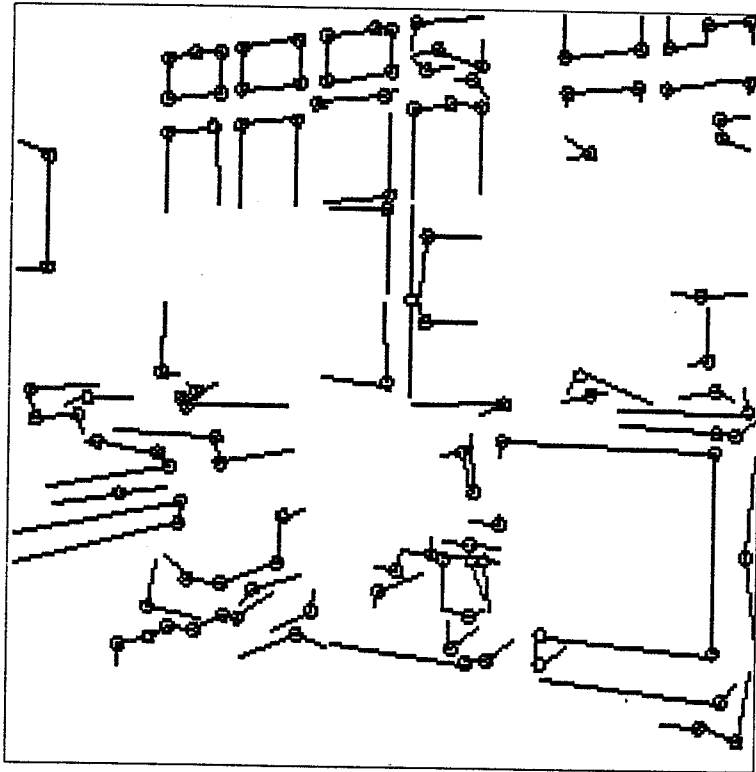


Figure 2.28: Les jonctions à deux segments.

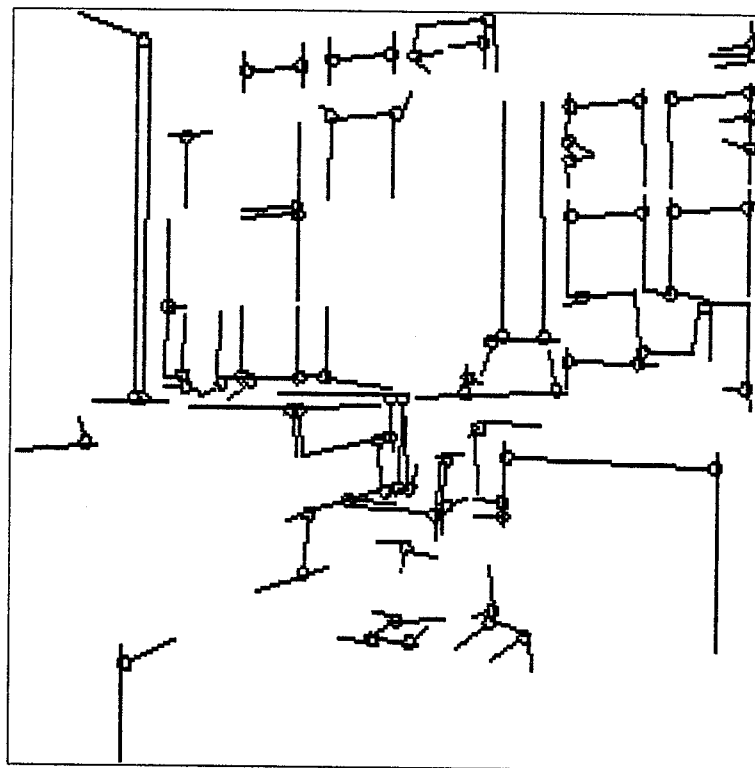


Figure 2.29: Les jonctions à trois segments.

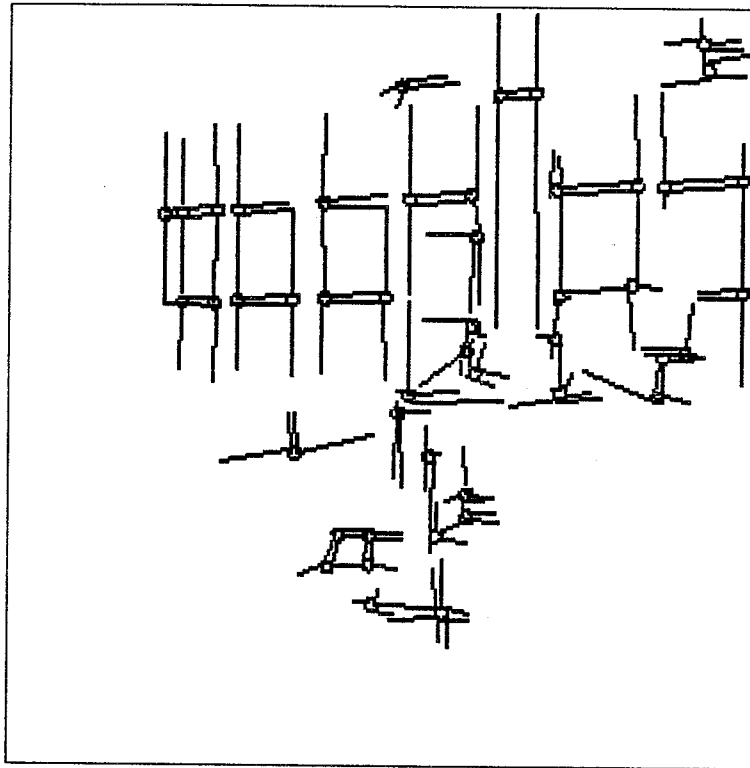


Figure 2.30: Les jonctions à quatre segments ou plus.

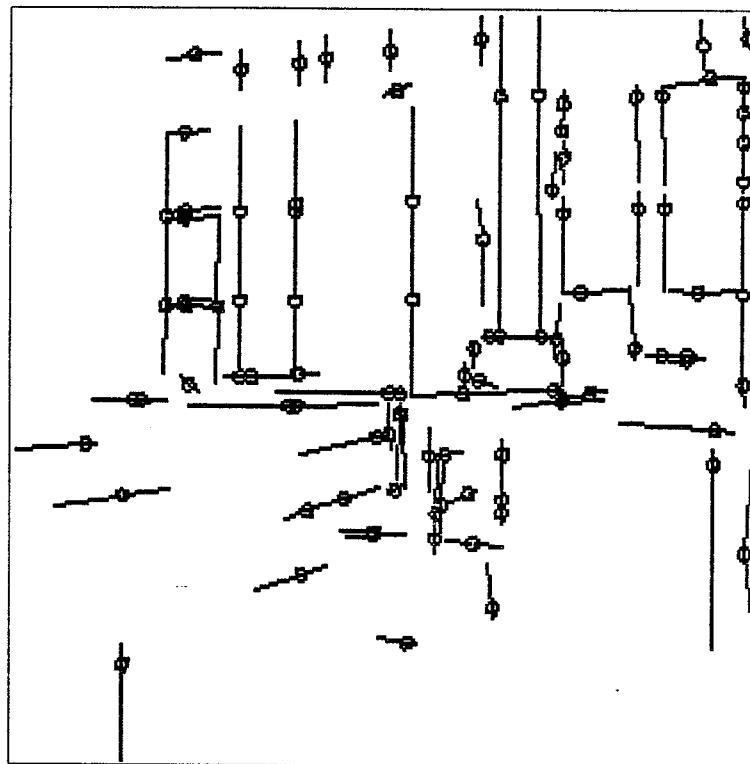


Figure 2.31: Les segments colinéaires.

Chapitre 3

La mise en correspondance

Nous avons présenté au chapitre précédent, une méthode de description monoculaire de chaque image en termes d'un graphe relationnel. Nous disposons alors de deux graphes relationnels, qui décrivent l'image gauche et l'image droite séparément. En théorie, ces deux descriptions devraient se ressembler considérablement. Malheureusement, dans la pratique ceci n'est pas vrai. En effet, les deux descriptions monoculaires peuvent être différentes à cause des prises de vue différentes (occlusions, variations photométriques) ou à cause des défaillances des processus d'extraction et de groupement des segments. Nous désirons alors nous munir d'un mécanisme robuste de mise en correspondance de l'image gauche avec l'image droite permettant de tenir compte des dissimilarités éventuelles entre les images gauche et droite.

3.1 Contraintes géométriques pour la mise en correspondance

Nous allons considérer maintenant une caractéristique dans une image et nous allons essayer de restreindre le nombre de candidates potentielles qui lui sont associées dans l'autre image. Plus précisément, étant donné un segment de droite ayant une position et une orientation connues dans l'image gauche, nous allons rechercher dans l'image droite le lieu des positions et orientations possibles des segments qui pourraient lui correspondre. Les différentes contraintes sont obtenues par la géométrie

du capteur stéréo.

Considérons la géométrie du capteur présenté sur la Figure 3.1¹. A chaque caméra on associe un repère local. Dans chaque repère local l'axe des z coïncide avec l'axe optique de la caméra associée et les deux plans images sont parallèles aux plans xy à une distance $f = 1$ de chaque origine, respectivement. Le repère de coordonnées de la caméra gauche est obtenu en faisant subir une rotation (d'un angle ϕ autour de l'axe des y) et une translation (d'un vecteur $O\vec{O}$ – la base du capteur) au système des coordonnées de la caméra droite. Cette transformation est représentée par une matrice A quatre par quatre (coordonnées homogènes) :

$$A = \begin{pmatrix} \cos \phi & 0 & \sin \phi & b_x \\ 0 & 1 & 0 & 0 \\ -\sin \phi & 0 & \cos \phi & b_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.3)$$

Dans l'Annexe A nous montrons comment cette modélisation peut être déduite à partir des données de calibrage fournies pour chaque caméra.

3.1.1 Les contraintes géométriques appliquées aux points-image

Soit P un point de la scène et soient (X, Y, Z) ses coordonnées par rapport au repère de la caméra gauche. Exprimons les coordonnées du point P dans le repère de la caméra droite (X', Y', Z') :

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = A \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.4)$$

Soient p et p' les deux projections du point P sur les plans images gauche et droite, respectivement (cf. Figure 3.1). Les coordonnées de p et p' sont :

$$\text{pour } p : x = X/Z \text{ et } y = Y/Z \quad (3.5)$$

$$\text{pour } p' : x' = X'/Z' \text{ et } y' = Y'/Z' \quad (3.6)$$

En remplaçant X' , Y' et Z' par leurs expressions données par l'équation (3.4) et en observant que $X = xZ$ et $Y = yZ$, nous pouvons exprimer les coordonnées du point p' en fonction des coordonnées x et y du point p , de la profondeur Z et des paramètres ϕ , b_x et b_z du capteur :

$$x' = \frac{Z(x \cos \phi + \sin \phi) + b_x}{Z(\cos \phi - x \sin \phi) + b_z} \quad (3.7)$$

$$y' = \frac{Z y}{Z(\cos \phi - x \sin \phi) + b_z} \quad (3.8)$$

¹Nous avons choisi ce cas simplifié de la réalité pour deux raisons. Premièrement, parce qu'il correspondait aux données que nous disposions et deuxièmement car les calculs présentés à la suite de ce mémoire deviennent un peu plus courts. Le cas complet se trouve en [31].

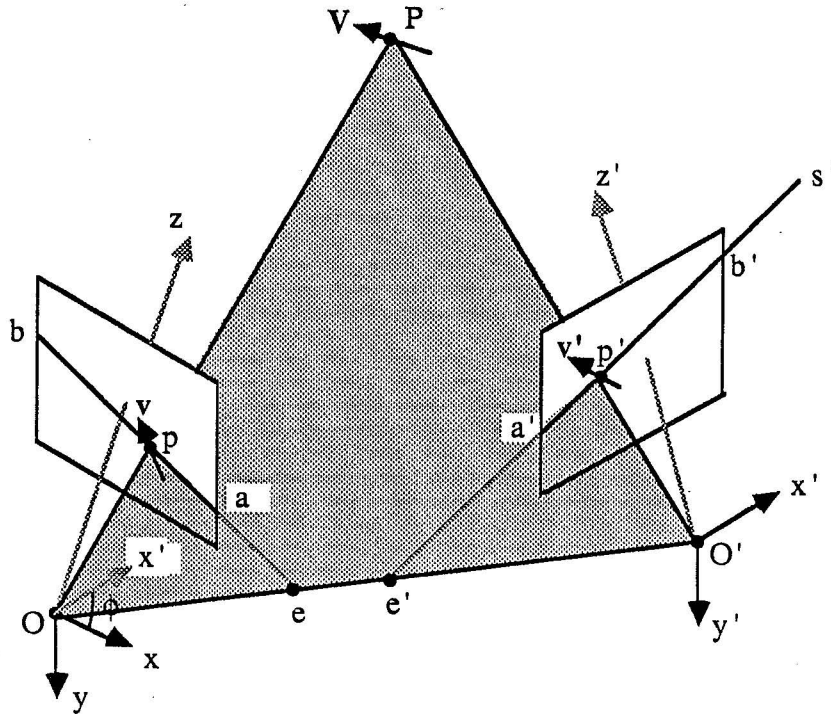


Figure 3.1: La géométrie du capteur stéréo.

La contrainte d'épipolarité

Éliminons Z entre les équations (3.7) et (3.8). Nous pouvons ainsi obtenir une équation linéaire entre les coordonnées x' et y' du point p' en fonction des coordonnées x et y du point p et des paramètres connus ϕ , b_x et b_z du capteur :

$$(yb_z) x' + (b_x(\cos \phi - x \sin \phi) - b_z(x \cos \phi + \sin \phi)) y' - yb_x = 0 \quad (3.9)$$

Étant donné donc un point p de l'image gauche, l'équation (3.9) exprime le lieu géométrique des points p' de l'image droite correspondant au point p . Par conséquent l'équation (3.9) est la ligne épipolaire droite associée au point p . Notons ici que toutes les lignes épipolaires droites passent par l'épipôle droit e' (Figure 3.1) (la démonstration se trouve dans l'Annexe A).

Notons aussi qu'en réarrangeant les termes de l'équation (3.9), nous pouvons obtenir une équation linéaire entre les coordonnées x et y du point p en fonction des coordonnées du point p' et des paramètres du capteur :

$$(-y'(b_x \sin \phi + b_z \cos \phi)) x + (b_z x' - b_x) y + y'(b_x \cos \phi - b_z \sin \phi) = 0 \quad (3.10)$$

Cette équation est en fait l'équation de la ligne épipolaire gauche associée au point p' .

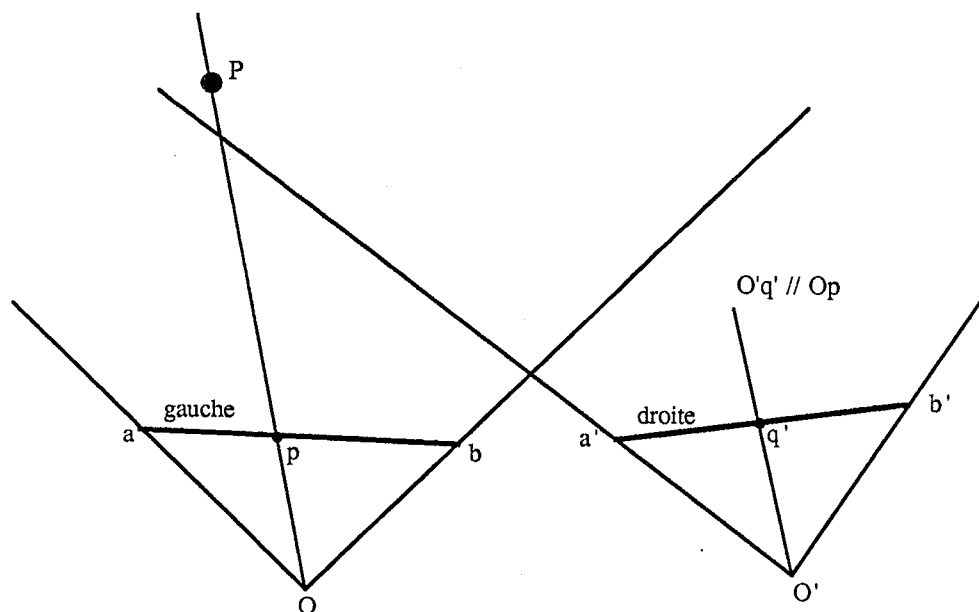


Figure 3.2: Une vue de dessus du plan épipolaire défini par p , O et O' . La contrainte de position impose que les points de l'image droite, correspondant au point p de l'image gauche, doivent se situer le long du segment $a'q'$ de la ligne épipolaire $a'b'$. $O'q'$ est parallèle à Op .

La contrainte de position

Nous avons vu que le point correspondant à un point p de l'image gauche doit être trouvé le long de la ligne épipolaire droite $e's'$ (Fig. 3.1) et plus précisément le long du segment $a'b'$ de cette ligne (condition pour que le point physique P soit visible par les deux caméras). Examinons maintenant si l'on peut restreindre encore plus le lieu des positions possibles des candidats potentiels p' du point p .

Pour mieux illustrer une telle possibilité, considérons la Figure 3.2, qui est une vue de dessus du plan épipolaire passant par les points P , p , O et O' . P doit se situer le long de la ligne droite Op . Lorsque P se déplace le long de cette droite, le point p' homologue du point p doit trouver quelque part entre a' et q' ; a' est le bord gauche de l'image droite et $O'q'$ est parallèle à Op . La position de q' peut être obtenue de façon analytique : ses coordonnées sont égales aux limites des équations (3.7) et (3.8) quand Z tend vers l'infini. Le segment $a'q'$ est le lieu des positions possibles du point p' . Quand p se rapproche de a , le segment $a'q'$ devient plus court et quand p se rapproche de b , $a'q'$ devient plus long, q' tendant vers b' . Ceci suggère qu'il y a moins de candidats potentiels pour les points de l'image gauche plus proches du bord gauche que pour ceux plus proches du bord droit ! Ceci est également vrai pour les points de l'image droite qui sont plus proches du bord droit que du bord gauche.

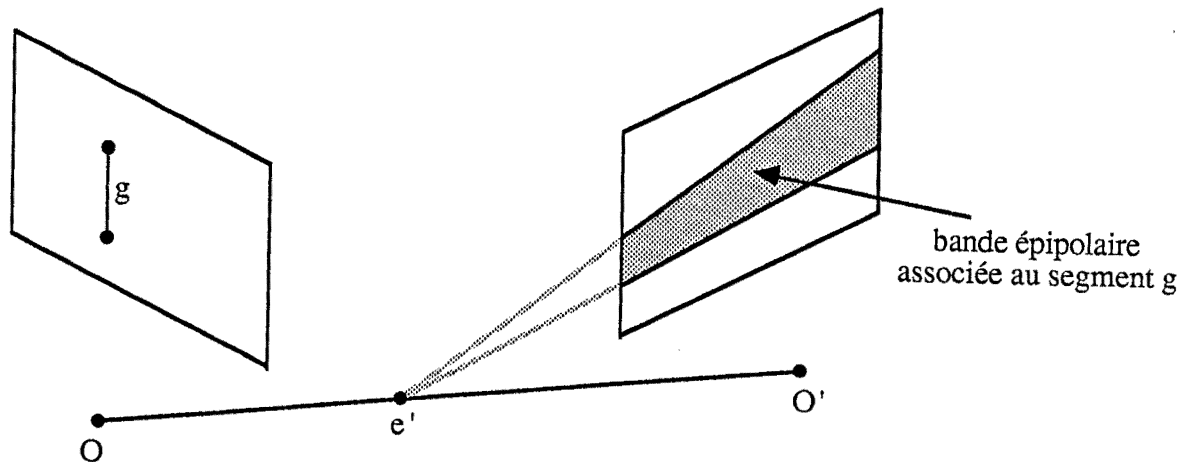


Figure 3.3: Un segment de l'image gauche et la bande épipolaire qui lui est associée dans l'image droite. La bande épipolaire droite est la région délimitée par les deux lignes épipolaires droites associées aux deux extrémités du segment.

3.1.2 Les contraintes géométriques appliquées aux segments

Nous désirons adapter les contraintes d'épipolarité et de position que nous avons présentées au cas des points, pour les appliquer au cas des segments de droite.

La contrainte d'épipolarité

Considérons un segment g de l'image gauche. Chaque extrémité de g détermine sur l'image droite une ligne épipolaire correspondante. Ainsi, les deux lignes épipolaires droites associées aux deux extrémités du segment g définissent dans l'image droite une région, que nous appelons *bande épipolaire droite*. La Figure 3.3 illustre la bande épipolaire droite associée au segment g de l'image gauche. La contrainte d'épipolarité impose maintenant que le segment correspondant à g doit être contenu dans la bande épipolaire droite associée à g . Ceci suppose donc que les extrémités de g et celles de son segment homologue se correspondent parfaitement. Or, en réalité à cause de défaillances du processus de segmentation et/ou d'occlusions possibles d'une image à l'autre, les extrémités d'un couple de segments homologues ne se correspondent pratiquement jamais. Qui plus est, à cause de mêmes raisons, un segment long d'une image peut apparaître comme un ensemble de segments courts et colinéaires dans l'autre image. Afin de pouvoir traiter tous ces cas, nous acceptons comme candidats potentiels de g tous les segments qui traversent la bande épipolaire qui lui est associée. (Voir Figure. 3.5-a).

Toutefois, le problème suivant se pose : si le segment considéré a la même direction que les lignes épipolaires gauches, (cas qui se produit lorsque le segment provient d'un segment dans l'espace qui est coplanaire avec la base OO' du capteur), alors les deux extrémités du segment se trouveront le long d'une seule ligne épipolaire gauche (cf. Figure 3.4). Dans ce cas, la bande épipolaire droite, que les

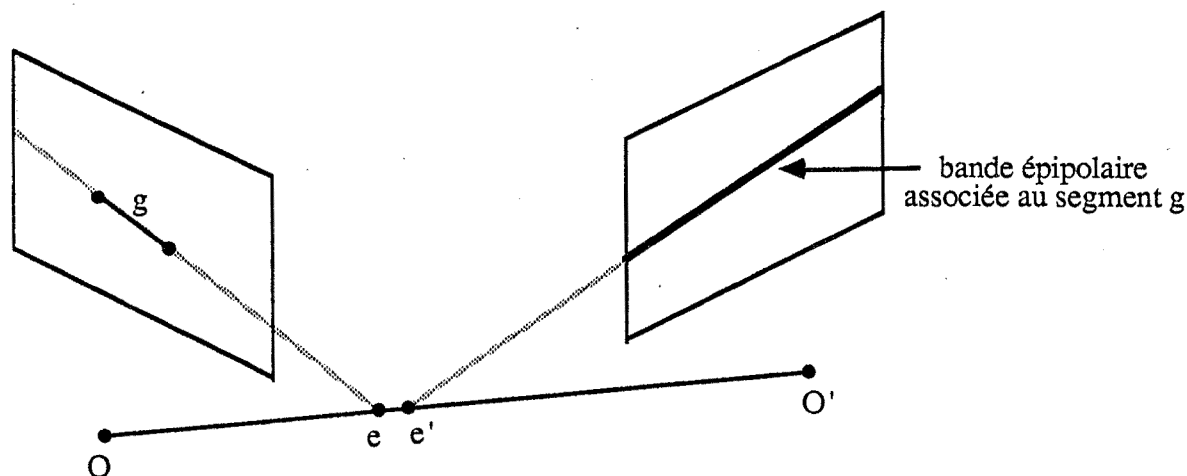


Figure 3.4: La bande épipolaire droite associée à un segment de l'image gauche, lorsque celui-ci a une direction parallèle à la direction des lignes épipolaires gauches. Dans ce cas, la bande épipolaire est réduite en une ligne droite.

deux extrémités du segment vont définir dans l'image droite, sera réduite en une ligne droite ; cette ligne est en réalité la ligne épipolaire droite correspondant à la ligne épipolaire gauche sur laquelle les deux extrémités se trouvent.

Afin de résoudre le problème posé par les segments dont l'orientation est presque la même que celle des lignes épipolaires², leur bande épipolaire n'est plus définie par leurs deux extrémités, mais elle est élargie de la manière suivante : on considère deux points fictifs dans une direction perpendiculaire au segment, chacun se situant à une distance égale à la mi-longueur du segment et on associe au segment la bande épipolaire définie par les lignes épipolaires correspondant aux deux points fictifs précédents. Etant donné donc que l'on a élargi la vraie bande épipolaire correspondant au segment, on accepte cette fois-ci comme candidats potentiels de ce segment, tous les segments se trouvant complètement à l'intérieur de la bande épipolaire que nous lui avons associée. (Voir Figure 3.5-b). Néanmoins, selon cette approche, si un segment est très long, la bande épipolaire, qui lui sera associée, sera très large et par conséquent on associera à ce segment un très grand nombre de candidats potentiels. Afin d'éviter ce problème, nous avons décidé que la largeur des bandes épipolaires ne devrait pas dépasser une largeur maximale³.

La contrainte de position

La contrainte de position présentée pour les points n'est pas directement applicable aux segments parce que, comme nous l'avons déjà mentionné, pour un couple de segments homologues, leurs extrémités ne se correspondent pas forcément. Nous

² Les segments considérés comme parallèles aux lignes épipolaires sont ceux dont la direction diffère $\pm 30^\circ$ par rapport à la direction des lignes épipolaires.

³ Dans toutes nos expérimentations, la largeur maximale d'une bande épipolaire était égale à $1/8$ de la taille horizontale des images traitées.

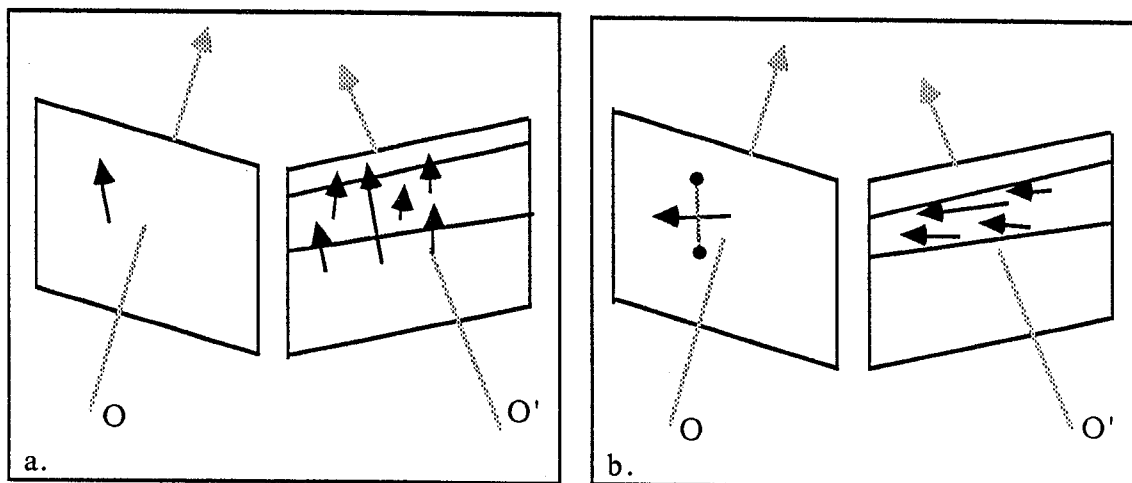


Figure 3.5: Recherche des candidats potentiels dans l'image droite d'un segment de l'image gauche. (a) Les candidats potentiels doivent avoir au moins quelques points à l'intérieur de la bande épipolaire associée au segment et définie par les deux extrémités du segment. (b) Pour un segment dont la direction est parallèle à celle des lignes épipolaires, la bande épipolaire qui lui est associée est élargie dans une direction perpendiculaire au segment ; ses candidats potentiels doivent être contenus complètement dans la bande épipolaire qui lui est associée.

avons donc implanté la contrainte de position de la façon suivante : soit g un segment de l'image gauche et soit d un de ses candidats potentiels dans l'image droite ; on trouve sur les deux segments un couple de points se correspondant, et on applique la contrainte de position sur ce couple de points.

Pour les segments dont la direction est parallèle à celle des lignes épipolaires, il n'est pas possible de déterminer de points potentiellement homologues entre ces segments et leurs candidats potentiels. Dans ce cas, nous avons considéré comme points homologues, les milieux des segments et de leurs candidats et nous avons appliqué la contrainte de position sur ces points.

Outre ces deux contraintes, dans le cas des segments, il est possible d'utiliser deux contraintes supplémentaires afin de restreindre davantage le nombre de leurs candidats potentiels. Ces contraintes sont la contrainte d'orientation et la contrainte de contraste que nous présentons ci-dessous.

La contrainte d'orientation

Considérons un segment de droite de la scène, passant par un point P , dont la direction est donnée par le vecteur unitaire \vec{V} (cf. Fig 3.1). Les équations (3.4) à (3.8) expriment les coordonnées de P et de ses projections (p et p') dans les deux images par rapport aux repères associés aux deux caméras. Soient maintenant (V_x, V_y, V_z) et (V'_x, V'_y, V'_z) les coordonnées du vecteur \vec{V} dans les repères de la caméra gauche et de la caméra droite respectivement et soient \vec{v} et \vec{v}' les projections gauche et droite de \vec{V} . La projection gauche est donnée par la formule : (voir Figure 3.6)

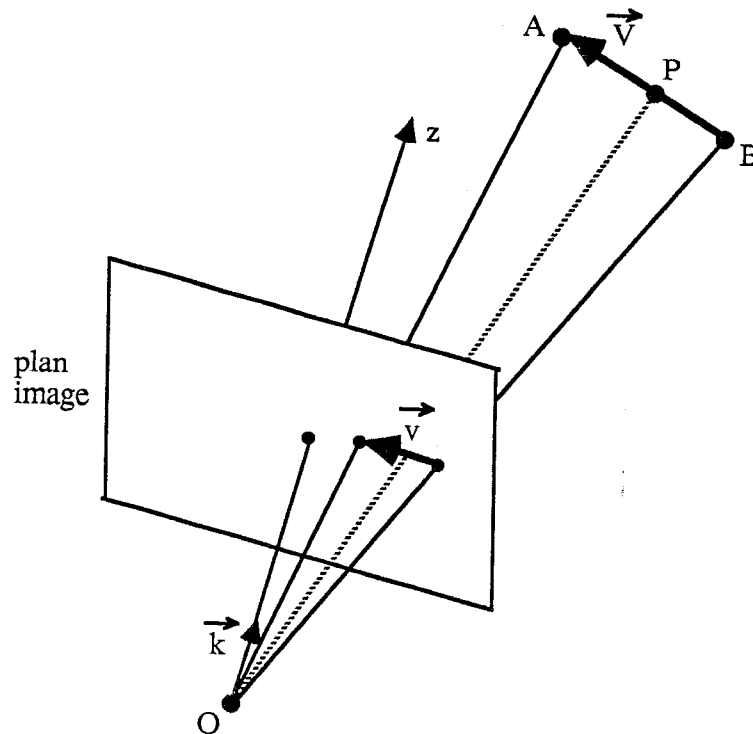


Figure 3.6: La projection \vec{v} du vecteur \vec{V} appartient à la fois au plan image représenté par son vecteur unitaire \vec{k} et au plan OAB (appelé plan d'interprétation) et représenté par le vecteur $\vec{l} = \vec{V} \times \vec{OP}$; par conséquent \vec{v} est donné par $\vec{v} = \vec{k} \times \vec{l} = \vec{k} \times (\vec{V} \times \vec{OP})$.

$$\vec{v} = \vec{k} \times (\vec{V} \times \vec{OP}) \quad (3.11)$$

où \vec{k} est le vecteur unitaire associé à l'axe des z. Les coordonnées-image de \vec{v} sont :

$$v_x = ZV_x - XV_z \quad (3.12)$$

$$v_y = ZV_y - YV_z \quad (3.13)$$

Utilisant les coordonnées (x, y) du point p données par l'équation (3.5), v_x et v_y peuvent s'écrire :

$$v_x = Z(V_x - xV_z) \quad (3.14)$$

$$v_y = Z(V_y - yV_z) \quad (3.15)$$

La pente du vecteur \vec{v} est :

$$\tan \theta = \frac{v_y}{v_x} = \frac{V_y - yV_z}{V_x - xV_z} \quad (3.16)$$

où θ est l'angle formé par \vec{v} et l'axe des x. De façon similaire, il est possible d'exprimer la pente ($\tan \theta'$) de \vec{v}' , en fonction des coordonnées (x', y') du point p' et de V'_x , V'_y et V'_z . Utilisant la matrice A (donnée par l'équation (3.3)), V'_x , V'_y , et V'_z sont exprimés en fonction de V_x , V_y et V_z . En remplaçant ensuite x' et y' par leurs expressions données par les équations (3.7) et (3.8) nous obtenons finalement pour la pente de \vec{v}' :

$$\tan \theta' = \frac{\tan \theta (V_x - xV_z) \cos \phi + (yV_x - xV_y) \sin \phi + \frac{b_x}{Z} V_y}{V_x - xV_z + V_x \left(\frac{b_x}{Z} \cos \phi + \frac{b_z}{Z} \sin \phi \right) + V_z \left(\frac{b_x}{Z} \sin \phi - \frac{b_z}{Z} \cos \phi \right)} \quad (3.17)$$

Cette équation montre que la relation entre θ' et θ est une fonction de la position de la ligne dans l'image gauche, x , y , de l'orientation de la ligne dans la scène V_x , V_y , et V_z , de la profondeur Z et des paramètres b_x , b_z et ϕ du capteur. La disparité en orientation $|\theta' - \theta|$ est la plus petite possible pour $x \rightarrow 0$, $y \rightarrow 0$ et $Z \rightarrow \infty$:

$$\tan \theta' = \tan \theta \cos \phi \quad (3.18)$$

Cette équation détermine une limite inférieure pour la disparité en orientation. Essayons de déterminer une limite supérieure pour cette disparité ; la limite à calculer correspondra au cas le plus défavorable - la ligne est le plus près possible de capteur. Sans perte de généralité, nous considérons le cas $x = 0$, $y = 0$, c'est-à-dire que la position de la ligne dans l'image gauche est autour de l'axe optique. Notre capteur a les caractéristiques suivantes : $\phi = 5.5^\circ$, $b_x = 25\text{cm}$, $b_z = 1.6\text{cm}$.

Nous pouvons maintenant définir une fonction continue qui fasse correspondre des points appartenant au domaine $V_x \times V_y \times V_z$ à des paires d'angles $\theta' \times \theta$. Etant donné cette représentation, il est possible de transformer des distributions de probabilité d'un domaine à l'autre. En supposant que les orientations des lignes de l'espace aient une distribution uniforme (toutes les orientations sont équiprobables) nous voulons établir si quelques combinaisons entre θ et θ' sont plus probables que d'autres. Arnold et Binford [1] ont déterminé cette fonction de distribution de probabilité. Leurs expérimentations ont donné les résultats suivants :

- Pour $\phi = 0$, $b_z = 0$ et $\frac{b_x}{Z} = 0.7$ ils ont obtenu une fonction de distribution de probabilité ayant une selle étroite le long la ligne $\theta' = \theta$ et la demi-largeur à demi-hauteur (DLDH) au centre était de 30° .
- Pour $\phi = 0$, $b_z = 0$ et $\frac{b_x}{Z} = 0.07$ la selle était encore plus étroite et DLDH au centre était de 3° .
- $\theta' = \theta$ est la combinaison la plus probable.

Dans notre cas $\frac{b_x}{Z} = 0.7$ correspond à $Z = 35\text{cm}$ qui est très proche au capteur. Etant donné les caractéristiques de notre capteur, nous pouvons extrapoler les résultats d'Arnold et Binford. Ainsi nous avons décidé de limiter la disparité en orientation entre 0° et 30° .

La contrainte de contraste

Les contraintes que nous avons présentées jusqu'à maintenant sont des contraintes purement géométriques. De plus nous avons ajouté un critère photométrique pour éliminer quelques appariements relativement improbables : soit c_g le contraste d'un segment g (la notion de contraste a été définie au § 2.4.3, page 54) et soit c_d le contraste d'un segment d se trouvant dans la liste des candidats potentiels de g . La disparité de contraste entre g et d est donnée par $|c_g - c_d|$. On élimine d de la liste des candidats potentiels de g si la disparité de contraste dépasse un seuil⁴.

Les quatre contraintes ci-dessus sont appliquées à tous les segments de l'image gauche et ensuite à tous les segments de l'image droite. Finalement, un appariement potentiel n'est (pour l'instant) retenu que si le segment de l'image gauche se trouve parmi les candidats potentiels associés au segment de l'image droite, et le segment de l'image droite se trouve parmi les candidats potentiels associés au segment de l'image gauche.

3.1.3 Le rôle des contraintes pour la restriction du nombre d'appariements potentiels

Dans ce paragraphe nous allons présenter le rôle des quatre contraintes utilisées quant à la restriction du nombre d'appariements potentiels que leur utilisation provoque.

Supposons qu'un segment d'une image contient dans sa liste des candidats potentiels tous les segments de l'autre image. Nos expérimentations ont montré que, parmi les quatre tests, trois (orientation, position et contraste) sont de tests très rapides quant à la mise en oeuvre et très discriminants, qui éliminent d'emblée un très grand nombre de faux candidats potentiels. Plus précisément, effectuant une étude statistique sur plusieurs paires d'images stéréoscopiques quant au pourcentage des candidats éliminés par chaque test indépendamment, nous avons constaté que :

- le test sur l'orientation des segments élimine environ 80% des candidats potentiels ;

⁴Le seuil de disparité de contraste était de 25% de la dynamique des images traitées.

- le test sur la position élimine environ 50% des candidats potentiels ;
- le test sur le contraste rejette environ 40% des candidats potentiels.

Un enchaînement séquentiel de ces trois tests élimine en moyenne 93% des candidats potentiels, tandis que le pourcentage de candidats corrects éliminés par ces tests n'était que de l'ordre de 0.4%.

En ajoutant la contrainte d'épipolarité et en faisant ensuite l'intersection des deux listes de candidats potentiels des segments des images gauche et droite, on restreint le nombre moyen de candidats potentiels par segment à cinq ou six.

3.2 Mesure de similarité associée aux appariements potentiels

Jusqu'à maintenant, nous avons établi, pour chaque segment d'une image, une liste de candidats potentiels composée de segments de l'autre image. Intuitivement, parmi ces candidats potentiels, certains sont beaucoup plus probables d'être les candidats corrects que d'autres, parce que, en quelque sorte, ils "ressemblent" beaucoup plus à un segment de l'autre image plutôt qu'à un autre. La question qui se pose est donc la suivante : comment mesurer la qualité d'un appariement ?

Un appariement est composé d'un segment de l'image gauche mis en correspondance avec un segment de l'image droite. Lors de la description monoculaire des deux images, nous avons associé aux segments un certain nombre de propriétés géométriques et topologiques. Les propriétés disponibles sont les attributs géométriques des segments et leurs relations avec les segments voisins et elles sont fournies par les deux graphes relationnels représentant les deux images. Or, on pourrait se servir justement de ces propriétés et les comparer pour mesurer le degré de ressemblance entre deux segments composant un appariement. Voici énumérées quelques unes de ces propriétés, dont on pourrait se servir pour établir la qualité d'un appariement potentiel :

- orientation ;
- longueur ;
- contraste ;
- valeur moyenne de niveau de gris de la bande gauche ;
- valeur moyenne de niveau de gris de la bande droite ;
- nombre de segments voisins.

Parmi ces propriétés, seule l'orientation est assez robuste et varie peu avec des points de vues proches, [1]. La longueur est une propriété sensible au bruit, aux erreurs inhérentes au processus de segmentation et aux occlusions dues aux changements de point de vue. Quant au contraste et (surtout) aux valeurs moyennes

de niveau de gris, elles sont très sensibles aux variations photométriques dues aux changements de point de vue. Finalement, le nombre de segments voisins peut aussi changer considérablement d'un point de vue à l'autre.

Tenant compte de ces observations, nous avons associé à chaque appariement potentiel une mesure de similarité V ; V exprime la similitude entre les différentes propriétés présentées ci-dessus associées aux segments composant un appariement. Le degré de similitude des différentes propriétés homologues des deux segments est pondéré par un coefficient reflétant leur importance et leur invariance aux changements de point de vue. La mesure de similarité combine donc la différence en orientation (O), en longueur (L) et en contraste (C) entre les deux segments, la différence entre les valeurs moyennes de niveau de gris des bandes à gauche (BG) et à droite (BD), ainsi que la différence entre le nombre de leurs segments voisins (N). V est donné par la formule :

$$V = 1/8 \left(3 \frac{\min(O_g, O_d)}{\max(O_g, O_d)} + \frac{\min(L_g, L_d)}{\max(L_g, L_d)} + 2 \frac{\min(C_g, C_d)}{\max(C_g, C_d)} + 0.5 \frac{\min(BG_g, BG_d)}{\max(BG_g, BG_d)} + 0.5 \frac{\min(BD_g, BD_d)}{\max(BD_g, BD_d)} + \frac{\min(N_g, N_d)}{\max(N_g, N_d)} \right) \quad (3.19)$$

V est donc une mesure de confiance d'un appariement potentiel. V prend la valeur 1 pour un appariement potentiel parfait et tend vers 0 quand la dissimilarité augmente.

L'utilisation de V rend maintenant possible la classification des candidats potentiels associés à un segment, selon leur degré de ressemblance avec ce segment. Puisque V reflète la confiance d'un appariement potentiel, nous pouvons éliminer de la liste de candidats associés à un segment les candidats dont le coefficient V est petit.

Afin de ne pas risquer une élimination abusive de tous les candidats associés à un segment, nous retenons toujours le candidat classé comme meilleur, même si sa mesure V est trop petite ; nous rejetons ensuite tous les candidats potentiels, dont l'écart de leur V par rapport au V du meilleur candidat est supérieur à un écart maximum⁵.

Cette technique permet de restreindre le nombre de candidats potentiels à deux ou trois en moyenne par segment, tout en gardant (dans le 98% des cas) le candidat correct dans la liste. Une étude statistique a montré que, dans 75% des cas, le candidat classé premier correspondait effectivement au candidat correct et dans 15% des cas, le candidat correct était classé à la deuxième place.

3.3 Mise en correspondance des structures 2-D

Dans les paragraphes précédents nous avons présenté :

⁵Dans nos expérimentations l'écart maximum entre les mesures de similarité V de deux candidats potentiels associés à un segment était de l'ordre de 20%.

- une technique pour représenter chaque image séparément par un graphe relationnel qui contient des segments et leurs relations ;
- une méthode géométrique pour établir les appariements potentiels : segment de l'image gauche / segment de l'image droite.

Nous désirons appairier les images stéréo. Cela se ramène à mettre en correspondance les deux descriptions monoculaires, c'est-à-dire à trouver un sous-isomorphisme entre les deux descriptions. La recherche d'un tel sous-isomorphisme est en fait équivalente à la recherche, parmi les appariements potentiels disponibles, d'un ensemble d'appariements tous *compatibles* entre eux. De cette manière, le problème se décompose en deux étapes :

1. établir la notion de compatibilité entre deux appariements potentiels ;
2. rechercher parmi ces appariements des ensembles dont les éléments sont tous mutuellement compatibles.

L'outil mathématique qui semble être le mieux adapté pour résoudre de manière élégante ces deux problèmes est la théorie des graphes. En effet, en représentant les appariements potentiels par de noeuds dans un graphe (que nous allons appeler *graphe de correspondance*), les deux problèmes précédents se transforment respectivement en les deux opérations suivantes :

1. construire les arêtes du graphe ;
2. rechercher dans ce graphe parmi les noeuds trouvés compatibles, un ensemble de noeuds qui soient tous compatibles entre eux.

Nous présentons dans les paragraphes suivants chacun de ces deux aspects.

3.3.1 Construction des arêtes du graphe de correspondance

Nous voulons construire les arêtes dans le graphe de correspondance. Pour ceci, nous allons considérer tous les noeuds par deux et nous allons établir un ensemble de règles permettant de déterminer la compatibilité ou l'incompatibilité entre deux noeuds. Notons par m_{ab} un noeud du graphe de correspondance ; ce noeud représente un appariement potentiel g_a/d_b , c'est-à-dire un segment g_a de l'image gauche qui a été mis en correspondance avec le segment d_b de l'image droite : $m_{ab} = g_a/d_b$. Soient maintenant deux tels noeuds $m_{ij} = g_i/d_j$ et $m_{pq} = g_p/d_q$.

Afin de construire les règles qui vont permettre d'établir la compatibilité ou l'incompatibilité entre les noeuds m_{ij} et m_{pq} , nous allons nous baser sur les similitudes locales entre les deux descriptions monoculaires associées à l'image gauche et à l'image droite, respectivement. Plus précisément, nous allons comparer la relation topographique reliant les deux segments g_i et g_p dans l'image gauche avec la relation topographique reliant les deux segments d_j et d_q dans l'image droite et selon la cohérence ou l'incohérence résultant de la comparaison de ces relations, nous allons pouvoir confirmer ou infirmer la compatibilité entre m_{ij} et m_{pq} .

De cette manière, l'application de l'ensemble de règles sur toutes les paires possibles de noeuds va permettre de construire petit à petit les arêtes du graphe de correspondance. Finalement, nous allons aboutir à un graphe où une arête va signifier la compatibilité entre deux noeuds et l'absence d'une arête entre deux noeuds signifiera que ces noeuds sont incompatibles.

Nous présentons aux paragraphes suivants les règles de mise en correspondance.

3.3.1.1 Règles de compatibilité pour la mise en correspondance

Soient $m_{ij} = g_i/d_j$ et $m_{pq} = g_p/d_q$ deux noeuds du graphe de correspondance. Nous avons :

Règle 1 (colinéarité)

m_{ij} est compatible avec m_{pq} si :

- (($g_i = g_p$) ET (d_j colinéaire avec d_q))
- OU
- (($d_j = d_q$) ET (g_i colinéaire avec g_p))
- OU
- ((g_i colinéaire avec g_p) ET (d_j colinéaire avec d_q))

Règle 2 (connexité)

m_{ij} est compatible avec m_{pq} si :

- (g_i même jonction que g_p) ET (d_j même jonction que d_q)

Règle 3 (ordre)

m_{ij} est compatible avec m_{pq} si :

- (g_i à gauche de g_p) ET (d_j à gauche de d_q)
- OU
- (g_i à droite de g_p) ET (d_j à droite de d_q)

La **Règle 1** reflète la propriété descriptive suivante : deux segments colinéaires dans la scène restent colinéaires dans les deux images. Cette règle est une généralisation de la contrainte de continuité de formes et elle donne la possibilité d'apparier un segment dans une image avec un ensemble de segments colinéaires dans l'autre image. Cette dernière possibilité n'est pas contradictoire avec la contrainte d'unicité car nous pensons que le segment a été fragmenté en plusieurs petits segments colinéaires à cause de défaillances du processus de segmentation. Cette règle permet justement le rassemblement de ces segments colinéaires en un segment plus long et leur mise en correspondance avec un segment unique de l'autre image.

La **Règle 2** reflète la propriété descriptive suivante : si deux courbes sont connectées dans l'espace, leurs projections dans les deux images sont également connectées.

La **Règle 3** est une généralisation de la contrainte d'ordre.

3.3.1.2 Règle d'incompatibilité

De l'application des trois premières règles il ressort que pour quelques paires de noeuds la relation topographique reliant les deux segments de l'image gauche est en contradiction avec la relation topographique reliant les deux segments de l'image droite. Il est donc nécessaire de formuler une quatrième règle d'*incompatibilité* entre deux noeuds :

Règle 4 (incompatibilité)

m_{ij} est incompatible avec m_{pq} si et seulement si g_i, g_p sont reliés par une relation topographique Q_1 (dans l'image gauche) et d_j, d_q sont reliés par une relation topographique Q_2 (dans l'image droite) contradictoire à Q_1 .

Voici quelques cas où on aboutit à une incompatibilité :

Incompatibilités déduites de la Règle 1

- $(g_i = g_p)$ ET $(d_j$ non colinéaire avec $d_q)$
- $(d_j = d_q)$ ET $(g_i$ non colinéaire avec $g_p)$
- $(g_i$ colinéaire avec $g_p)$ ET $(d_j$ non colinéaire avec $d_q)$
- $(d_j$ colinéaire avec $d_q)$ ET $(g_i$ non colinéaire avec $g_p)$
- $(g_i$ colinéaire avec g_p à la tête de $g_p)$ ET $(d_j$ colinéaire avec d_q à la queue de $d_q)$
(Rappelons la signification de tête et de queue associées aux extrémités d'un segment en fonction de son orientation, Fig. 2.20, page 54). La Figure 3.7 illustre un tel cas d'incompatibilité.
- $(g_i$ colinéaire avec g_p à la queue de $g_p)$ ET $(d_j$ colinéaire avec d_q à la tête de $d_q)$
- $(g_i = g_p)$ ET $(d_j$ colinéaire avec $d_q)$ ET
($(longueur(d_j) + longueur(d_q)) > (1.25 \times longueur_estimée(g_i))$)
- $(d_j = d_q)$ ET $(g_i$ colinéaire avec $g_p)$ ET
($(longueur(g_i) + longueur(g_p)) > (1.25 \times longueur_estimée(d_j))$)

Analysons plus précisément les deux derniers critères d'incompatibilité. Pour tout segment d'une image la fonction *longueur_estimée* retourne une estimation de la longueur de ce segment dans l'autre image. La Figure 3.8 illustre comment on estime cette longueur. Soit g_i un segment de l'image gauche dont on veut estimer la longueur dans l'image droite et soient e_1, e_2 les deux lignes épipolaires qui lui sont associées. Sa longueur estimée est donnée par l'écartement maximum entre e_1 et e_2 dans la direction donnée par g_i .

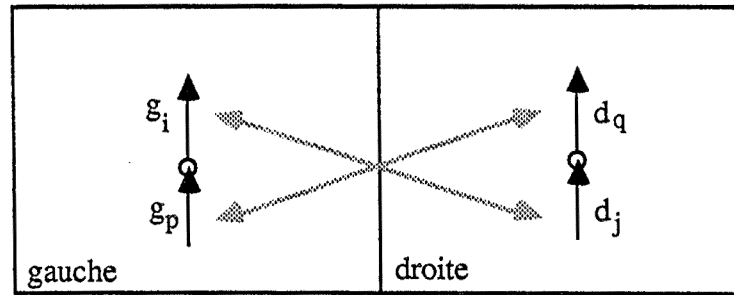


Figure 3.7: g_i/d_j est incompatible avec g_p/d_q parce que g_i et g_p sont colinéaires à la queue de g_i , alors que d_j et d_q sont colinéaires à la tête de d_j .

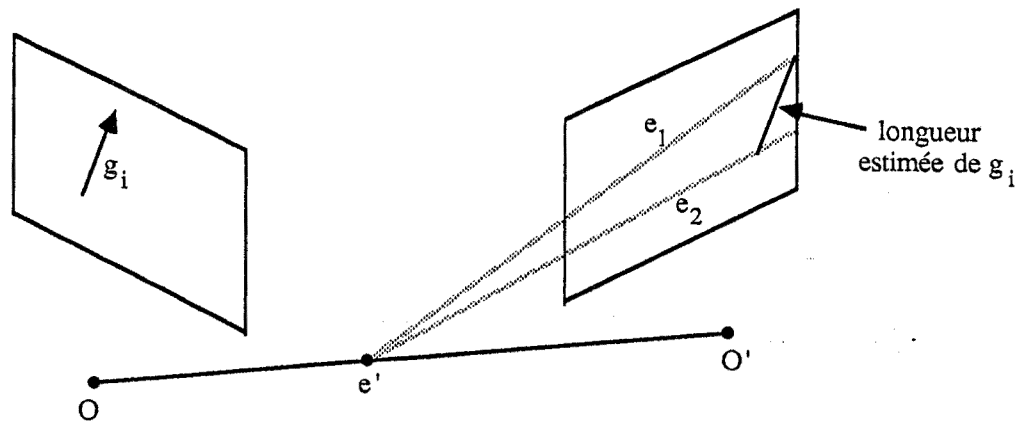


Figure 3.8: Le calcul de la longueur estimée d'un segment g_i de l'image gauche dans l'image droite.

Les deux derniers critères s'énoncent ainsi : deux appariements potentiels sont incompatibles si la somme des longueurs des deux segments d'une image dépasse la longueur estimée du segment de l'autre image multipliée par un coefficient d'erreur d'estimation. Ces deux critères ont été ajoutés afin d'éviter de mettre en correspondance un segment court d'une image avec deux segments longs et colinéaires de l'autre image.

Incompatibilités déduites de la Règle 2

Dans le cas de la deuxième règle, quelques contraintes supplémentaires peuvent être incluses afin de mettre en évidence l'incompatibilité.

En orientant les angles à partir du point de jonction, la valeur de cet angle autorise ou non l'incompatibilité. Mais, ceci ne permet pas de distinguer les quatre cas de jonctions de la Figure 3.9, faisant toutes un angle de $3\pi/2$. En codant l'extrémité des segments participant à la jonction selon le schéma de cette figure (code positif pour les segments participant avec la queue et négatif pour les segments participant avec la tête), il est possible de déduire l'incompatibilité : elle est établie

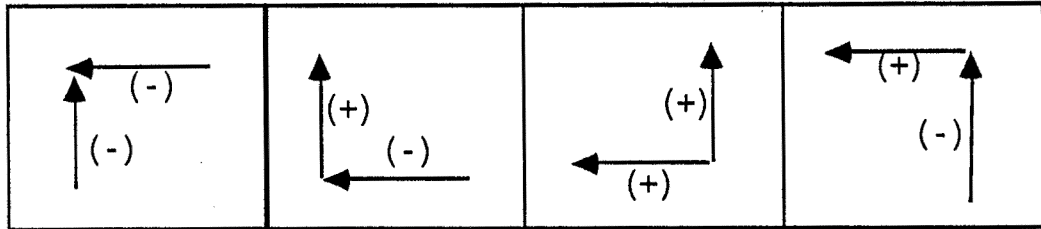


Figure 3.9: Les angles orientés à partir du point de jonction sont tous égaux ($3\pi/2$) pour tous les quatre cas des jonctions. Afin de les distinguer et établir l'incompatibilité, on code l'extrémité des segments participant à la jonction.

aux cas où les extrémités des segments homologues ne reçoivent pas le même code.

Incompatibilités déduites de la Règle 3

Les incompatibilités sont produites lorsqu'il y a une inversion de la relation d'ordre entre les segments de l'image gauche et leurs correspondants de l'image droite. Les deux cas d'incompatibilité sont les suivants (cf. la contrainte d'ordre, page 16) :

- (g_i à gauche de g_p) ET (d_j à droite de d_q)
- (g_i à droite de g_p) ET (d_j à gauche de d_q)

3.3.1.3 Règle de transitivité

Ces quatre règles relient deux appariements potentiels (noeuds) dont les éléments sont de segments voisins et donc directement reliés par une relation topographique. Cependant, s'il n'y a pas de relation topographique reliant les deux segments de l'image gauche et/ou les deux segments de l'image droite, l'application des quatre premières règles sur ces paires de noeuds n'aura aucun effet et par conséquent elle ne pourra pas les rendre ni compatibles ni incompatibles.

Pour établir la compatibilité entre deux noeuds dont les éléments ne sont reliés par aucune relation topographique directe, une cinquième règle est indispensable :

Règle 5 (transitivité)

m_{ij} est compatible avec m_{pq} s'il existe un noeud $m_{rs} = g_r/d_s$ tel que :
 (m_{ij} est compatible avec m_{rs}) ET (m_{rs} est compatible avec m_{pq})

Le but de cette règle est de propager la compatibilité locale à des structures images plus vastes.

Les quatre premières règles ne s'appliquent qu'à un petit pourcentage du nombre total de combinaisons des couples d'appariements possibles. La compatibilité ou incompatibilité de la grande majorité des couples d'appariements est élucidée et établie par la règle de transitivité. Or, une telle approche est relativement dangereuse car on ne maîtrise pas la fermeture transitive ; une utilisation abusive de la règle de

transitivité pourrait par conséquent aboutir à des interprétations complètement erronées. Un exemple simple est illustré sur la Figure 3.10, où g_1/d_1 est compatible avec g_2/d_2 par la règle de connexité et g_2/d_2 est compatible avec g_3/d_3 par la règle d'ordre. Devrait-on en déduire que g_1/d_1 est compatible avec g_3/d_3 , comme la règle de transitivité l'impose ?

L'application des quatre premières règles ne peut pas établir la compatibilité/incompatibilité entre deux noeuds dans les deux cas :

1. **absence partielle de relations topographiques** : les deux segments d'une image sont explicitement reliés par une relation topographique, alors que leurs deux segments correspondants de l'autre image ne sont reliés par aucune relation topographique ;
2. **absence totale de relations topographiques** : il n'y a pas de relation topographique ni entre les deux segments de l'image gauche ni entre leurs deux segments correspondants de l'image droite.

3.3.1.4 Analyse étendue

Alors que dans le cas d'absence totale de relations topographiques aucune décision ne peut être prise d'avance et la règle de transitivité semble être le seul moyen pour éclaircir la compatibilité ou l'incompatibilité entre deux noeuds, toutefois le cas de l'absence partielle de relations topographiques pourrait être éclairci de manière explicite. En effet, lorsqu'il existe une relation entre deux segments d'une image, mais qu'il n'y a pas de relation entre leurs deux segments correspondants de l'autre image, on pourrait effectuer une analyse pour découvrir éventuellement la relation géométrique entre ces deux segments. Cette relation pourrait être absente pour plusieurs raisons : défaillances des processus de segmentation et/ou de groupement, ou simplement parce que les deux segments ne sont pas voisins.

Une telle analyse permettrait d'établir avec certitude uniquement l'incompatibilité entre deux noeuds. Effectivement, décider après une telle analyse que deux noeuds

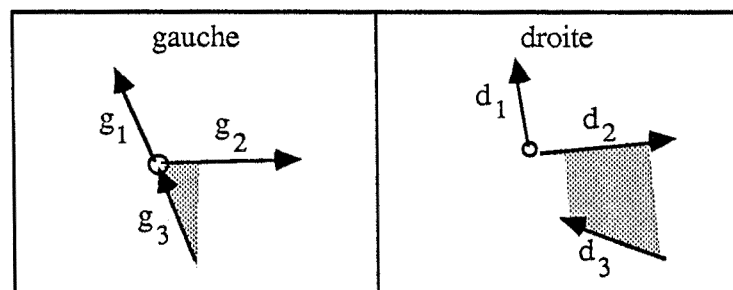


Figure 3.10: (g_1/d_1) est compatible avec (g_2/d_2) par connexité, tandis que (g_2/d_2) est compatible avec (g_3/d_3) par ordre. Devrait-on en déduire que (g_1/d_1) est compatible avec (g_3/d_3) par transitivité ?

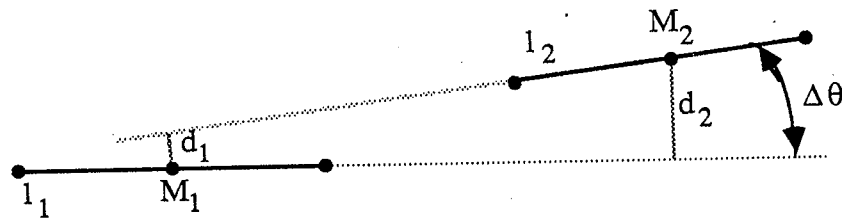


Figure 3.11: Critères géométriques pour déterminer si deux segments sont reliés par colinéarité étendue.

sont compatibles semble être très dangereux car si l'on établit une compatibilité qui s'avère être fautive, alors cette compatibilité pourrait se propager grâce à la règle de transitivité et rendre compatibles des noeuds qui ne devraient pas l'être, produisant ainsi des interprétations globales erronées. De plus, en déterminant après une telle analyse que deux noeuds sont définitivement incompatibles, on évite d'appliquer à ces noeuds la règle de transitivité, écartant ainsi le risque de les rendre compatibles, alors qu'ils ne sont pas en réalité.

Nous présentons à la suite quelques situations d'absence partielle de relations topographiques, où l'incompatibilité entre deux noeuds pourrait s'établir avec certitude. Soient de nouveau $m_{ij} = g_i/d_j$ et $m_{pq} = g_p/d_q$ les deux noeuds et soit Q la relation topographique dans l'image gauche qui relie g_i avec g_p ; il n'y a pas de relation topographique dans l'image droite entre d_j et d_q .

Analyse de colinéarité étendue

Soit $Q = \text{colinéarité ou identité} : ((g_i = g_p) \text{ OU } (g_i \text{ colinéaire avec } g_p))$.

L'analyse doit tester si d_j et d_q sont colinéaires en un sens plus large que celui de colinéarité définie pendant l'analyse monoculaire (§ 2.4.2.2, page 51). Pour que deux segments soient reliés par une colinéarité étendue, les deux conditions suivantes doivent être satisfaites, (cf. Figure 3.11).

- $\epsilon_1 \leq \Delta\theta \leq \pi - \epsilon_1$
- $\max(d_1, d_2) \leq \epsilon_2$

Le premier critère est le même avec le critère de parallélisme appliqué au § 2.4.2.2 dans le cadre de l'analyse monoculaire de deux segments colinéaires. Le deuxième critère teste la distance entre d_j et d_q , d_1 , d_2 étant les distances des points de milieu des segments d_j et d_q aux segments d_q et d_j , respectivement.

Si d_j et d_q ne satisfont pas les deux conditions de colinéarité étendue, alors le couple g_i/d_j est incompatible avec le couple g_p/d_q .

Par contre, si l'analyse de colinéarité étendue constate que d_j et d_q sont colinéaires, une condition supplémentaire est ajoutée afin de trouver l'incohérence éventuelle entre les noeuds m_{ij} et m_{pq} ; elle consiste à comparer les longueurs des segments g_i et g_p avec les longueurs des segments d_j et d_q . La comparaison des longueurs

est effectuée de la façon suivante : on calcule la longueur estimée des segments g_i et g_p dans l'image droite, (la manière dont la longueur estimée est calculée est illustrée sur la Fig. 3.8, page 78) ; si la somme de ces longueurs estimées, multipliée par un coefficient d'erreur d'estimation, est inférieure à la distance entre les deux extrémités les plus éloignées des segments d_j et d_q (reliés par colinéarité étendue), alors les noeuds m_{ij} et m_{pq} sont rendus incompatibles. Cette condition est présentée plus précisément aux deux cas suivants :

- soit ($g_i = g_p$) ;
 m_{ij} est compatible avec m_{pq} si :
 ((distance entre les deux extrémités les plus éloignées de d_j et d_q) >
 ($1.25 \times \text{longueur_estimée}(g_i)$))
- soit (g_i colinéaire avec g_p) ;
 m_{ij} est compatible avec m_{pq} si :
 ((distance entre les deux extrémités les plus éloignées de d_j et d_q) >
 ($1.25 \times (\text{longueur_estimée}(g_i) + \text{longueur_estimée}(g_p))$))

Cette condition supplémentaire a été ajoutée afin d'éviter de mettre en correspondance un segment unique ou deux segments colinéaires d'une image avec deux segments de l'autre image reliés par colinéarité étendue mais étant trop distants l'un de l'autre.

Enfin, si les segments d_j et d_q satisfont toutes les contraintes précédentes, alors l'incompatibilité entre m_{ij} et m_{pq} ne peut pas être établie.

Les Figures 3.12-a et -b illustrent deux cas, où l'analyse de colinéarité étendue rend deux noeuds incompatibles et la Figure 3.12-c présente une situation où cette analyse ne peut pas établir l'incompatibilité entre deux noeuds.

Enfin, en revenant sur l'exemple de la Figure 3.10, l'analyse de colinéarité étendue rend (g_1/d_1) incompatible avec (g_3/d_3) et ainsi on évite de rendre compatibles ces noeuds par la règle de transitivité.

Analyse de connexité étendue

Soit $Q =$ connexité : (g_i même jonction avec g_p)

L'analyse doit tester si d_j et d_q peuvent former une jonction et sous quelles conditions. Pour cela, on trouve le point d'intersection entre d_j et d_q et la configuration locale obtenue est comparée avec celle formée par g_i et g_p . Les contraintes appliquées dans ce cas sont les mêmes avec les contraintes supplémentaires que nous avons introduites pour mettre en évidence l'incompatibilité déduite de la règle de connexité. La Figure 3.13-a illustre un cas, où on pourrait décider de l'incompatibilité entre deux appariements potentiels. Les angles θ_1 et θ_2 orientés à partir des points des jonctions étant égaux, on code les extrémités des segments participant aux jonctions ; on s'aperçoit alors que ce code est négatif pour g_1 (car g_1 participe avec sa tête à la jonction réelle formée par g_1 et g_2), alors que ce code serait positif pour d_1 puisque d_1 participerait avec sa queue à la jonction fictive que d_1 et d_2 auraient formée.

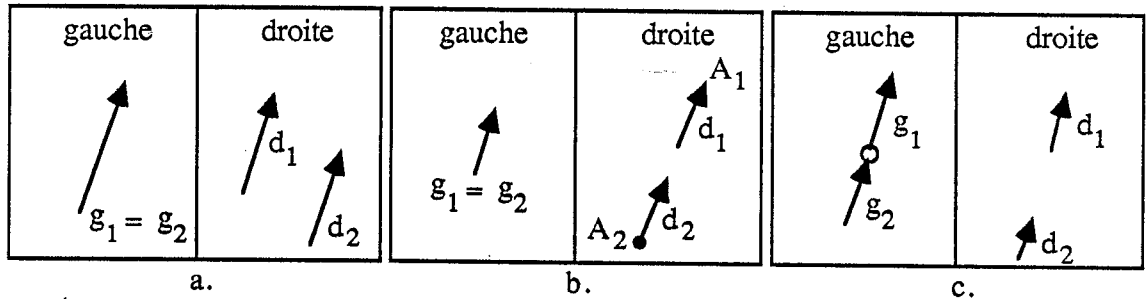


Figure 3.12: Trois cas où l'analyse de colinéarité étendue s'applique. Aux cas (a) et (b) elle rend incompatibles les deux appariements potentiels g_1/d_1 et g_2/d_2 parce que : (a) d_1 et d_2 ne sont pas reliés par colinéarité étendue et (b) la distance entre leurs extrémités les plus éloignées (A_1 et A_2) est trop grande par rapport à la longueur estimée de g_1 dans l'image droite. Au cas (c), cette analyse ne peut pas rendre incompatibles les deux appariements.

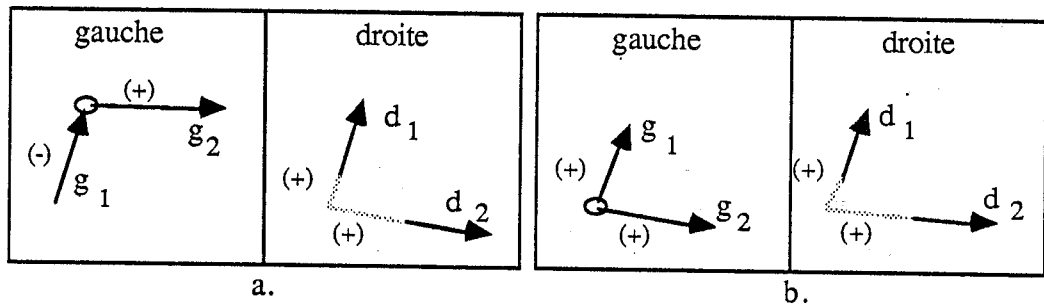


Figure 3.13: Dans l'image gauche g_1 et g_2 forment une jonction, alors que dans l'image droite d_1 et d_2 ne forment pas de jonction. L'analyse de connexité étendue rend incompatibles les couples g_1/d_1 et g_2/d_2 dans le cas (a), tandis que dans le cas (b) elle ne peut pas inférer l'incompatibilité.

Finalement, la Figure 3.13-b présente un cas, où l'analyse de connexité étendue ne peut pas rendre incompatibles les deux appariements.

Analyse d'ordre étendu

Soit $Q = \text{ordre} : ((g_i \text{ à gauche de } g_p) \text{ OU } (g_i \text{ à droite de } g_p))$.

Quand il n'y a pas de relation d'ordre directe entre d_j et d_q , on pourrait établir l'incompatibilité entre m_{ij} et m_{pq} si d_j et d_q satisfont certaines conditions d'ordre. La Figure 3.14 illustre l'analyse effectuée. Soient g_1/d_1 et g_2/d_2 deux appariements potentiels, où g_2 est à gauche de g_1 , tandis que d_1 et d_2 ne sont pas reliés par aucune relation topographique ; on considère la portion du demi-plan situé à droite de d_1 et comprise entre les deux demi-droites perpendiculaires à d_1 issues de ses extrémités, (cf. Figure 3.14-a). Si d_2 se trouve dans cette zone alors g_1/d_1 est incompatible avec g_2/d_2 . Si ce n'est pas le cas, alors les deux appariements ne peuvent pas être rendus incompatibles, (cf. Figure 3.14-b).

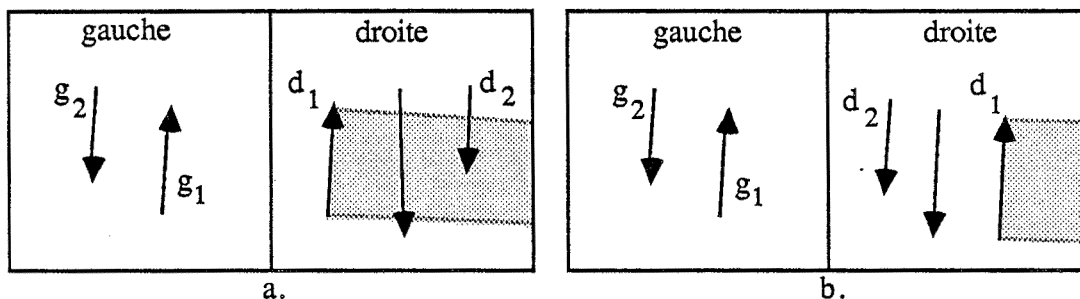


Figure 3.14: Analyse d'ordre étendu : (a) les couples g_1/d_1 et g_2/d_2 sont incompatibles car g_2 est à gauche de g_1 , alors que d_2 se trouve dans la zone-image située à droite de d_1 ; (b) g_1/d_1 et g_2/d_2 ne peuvent pas être rendus incompatibles.

3.3.2 Exploration du graphe de correspondance

La mise en correspondance des deux descriptions monoculaires représentant la paire stéréo initiale doit être capable de tenir compte d'incohérences éventuelles entre les deux images. En effet, non seulement un segment d'une image peut avoir plusieurs appariements potentiels dans l'autre image, mais encore les deux descriptions monoculaires peuvent être différentes pour plusieurs raisons :

- un segment visible dans une image peut être absent de l'autre image à cause d'occlusions ou de variations photométriques entre les deux images ou à cause des défaillances du processus d'extraction des segments ;
- une ligne d'une image peut être cassée en plusieurs petits morceaux dans l'autre image ;
- deux lignes séparées dans l'espace forment une jonction ou sont colinéaires dans l'une des deux images à cause d'un alignement accidentel ;
- la contrainte d'ordre n'est pas respectée de temps en temps. Ceci est dû à la présence de surfaces transparentes ou d'objets minces tels que les pieds d'une chaise.

Par conséquent, il n'est pas possible de trouver une correspondance *parfaite* entre les deux images. Nous proposons à la place de rechercher la *meilleure* correspondance disponible.

En termes de graphe, cette recherche est possible en explorant le graphe de correspondance que nous avons construit aux paragraphes précédents afin de trouver un ensemble de noeuds tous compatibles entre eux.

Après l'application des cinq règles de mise en correspondance, nous obtenons un graphe où deux noeuds sont reliés soit par compatibilité soit par incompatibilité. (Si malgré l'application des cinq règles, il y a encore de paires de noeuds qui ne sont pas reliées ni par compatibilité ni par incompatibilité, alors ces paires de noeuds sont rendus incompatibles). En d'autres termes, une arête dans ce graphe relie deux

noeuds mutuellement compatibles et l'absence d'une arête entre deux noeuds signifie leur incompatibilité.

Notons que dans ce graphe la notion de connexité simple [33], [79] n'a pas de sens véritable, puisqu'elle ne prend pas en compte l'incompatibilité sous-jacente. Par contre, notre problème de recherche d'un ensemble de correspondances mutuellement compatibles, correspond à la recherche des *cliques maximales* [8], [17] dans notre graphe de correspondance.

Afin de mettre en évidence ceci, examinons de plus près la notion d'une clique. Une clique dans un graphe est un sous-graphe complet, c'est-à-dire chaque noeud d'une clique est directement connecté à tous les autres noeuds de la clique. Une clique maximale est une clique qui ne peut pas être étendue par l'adjonction d'autres noeuds du graphe. La plus grande clique maximale est la clique maximale contenant le plus grand nombre de noeuds.

Par conséquent, une clique maximale dans ce graphe est en fait un ensemble maximal de noeuds tous compatibles entre eux, ce qui correspond exactement à ce que nous cherchons ! Ainsi notre problème se ramène à rechercher les cliques maximales dans le graphe de correspondance et à sélectionner ensuite parmi ces cliques maximales celle qui est la mieux adaptée pour représenter une correspondance correcte.

La stratégie de mise en correspondance

Prenons un exemple pour montrer d'une part comment le problème de mise en correspondance est représenté par un graphe et d'autre part comment rechercher la meilleure solution possible. La Figure 3.15 montre un ensemble de six segments de l'image gauche (g_1 à g_6) et un ensemble de neuf segments de l'image droite (d_1 à d_9). La Figure 3.16-a présente le graphe relationnel décrivant l'image gauche et la Figure 3.16-b présente le graphe relationnel décrivant l'image droite. Les deux structures à appairer ne sont pas similaires ; en particulier, une partie de la structure de l'image droite n'est pas visible dans l'image gauche (occlusion) et quelques segments longs dans une image ont leurs homologues en deux morceaux dans l'autre image. Nous présentons un processus de mise en correspondance de gauche à droite. (Il pourrait également s'appliquer de droite à gauche).

1. Les différents appariements potentiels sont établis et une mesure de similarité V (décrite au § 3.2) est associée à chaque appariement potentiel. Quinze appariements potentiels ont été établis : m_{13} , m_{16} , m_{17} , m_{23} , m_{26} , m_{27} , m_{34} , m_{45} , m_{49} , m_{53} , m_{56} , m_{57} , m_{61} , m_{62} et m_{68} , où m_{ij} correspond au segment g_i de l'image gauche mis en correspondance avec le segment d_j de l'image droite.
2. Les noeuds individuels du graphe de correspondance sont construits. Comme un noeud représente un appariement potentiel, notre graphe est composé de quinze noeuds.
3. Nous appliquons les trois premières règles de compatibilité. Le Tableau 3.1 présente les noeuds trouvés compatibles. Ce tableau bidimensionnel est sym-

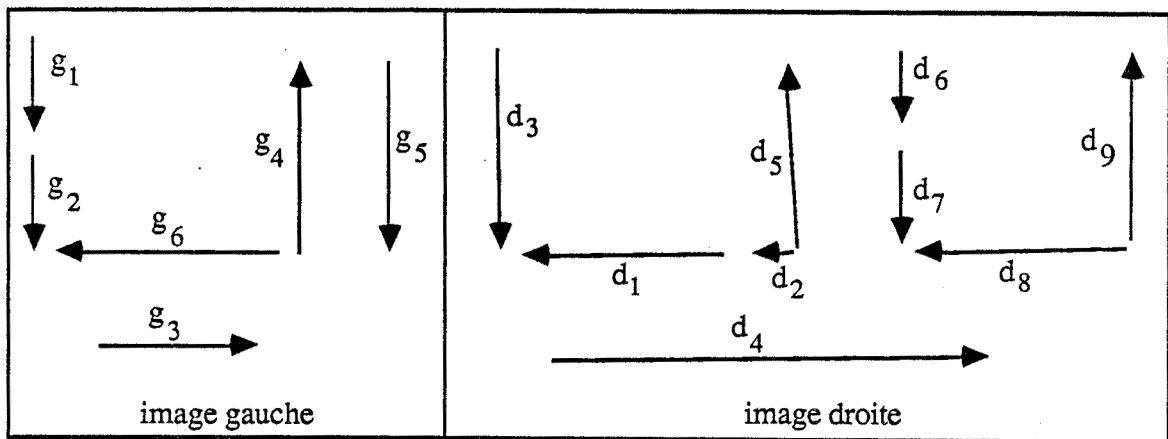


Figure 3.15: Deux structures simples, à mettre en correspondance.

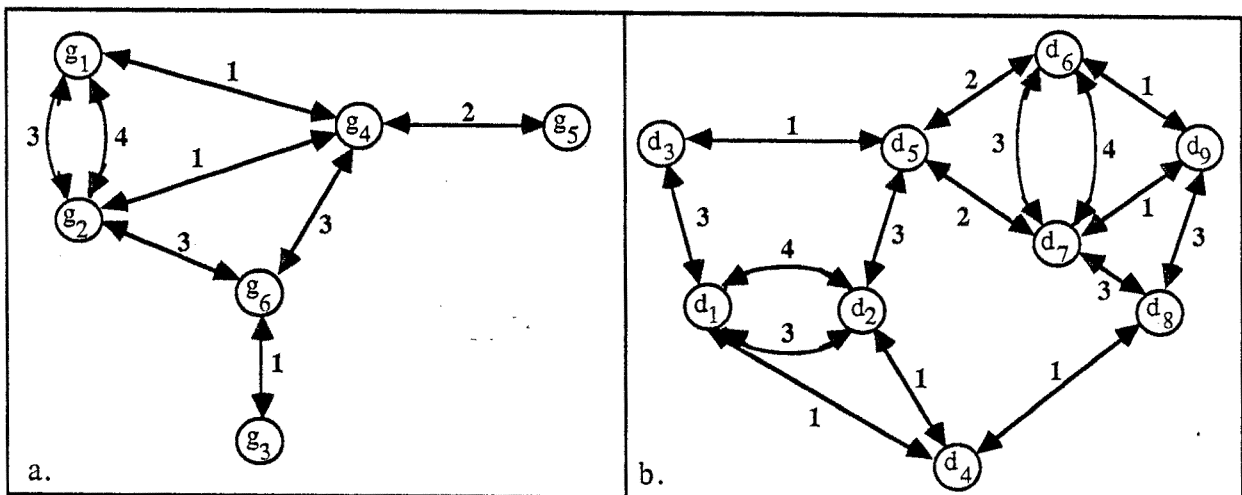


Figure 3.16: (a) Le graphe relationnel associé aux six segments de l'image gauche de la Figure 3.15. (b) Le graphe relationnel décrivant les neuf segments de l'image droite de la même Figure. Les numéros aux arcs des graphes représentent les quatre relations topographiques suivantes : 1-à gauche, 2-à droite, 3-même jonction, 4- colinéarité.

Noeud	13	16	17	23	26	27	34	45	49	53	56	57	61	62	68
13															
16															
17															
23	1														
26															
27		1													
34															
45	3			3											
49		3	3		3	3									
53															
56								3							
57								3			1				
61				2			3								
62							3	2					1		
68						2	3		2						

Table 3.1: Les compatibilités entre les noeuds après l'application des trois premières règles ; le noeud ij représente le segment g_i mis en correspondance avec le segment d_j . Pour chaque couple de noeuds compatibles, le numéro représente la règle qui a établi la compatibilité : 1-règle de colinéarité, 2-règle de connexité, 3-règle d'ordre.

métrique par rapport à sa diagonale principale et il contient, pour chaque couple de noeuds compatibles, le numéro de la règle qui a établi cette compatibilité : 1-règle de colinéarité, 2-règle de connexité et 3-règle d'ordre. Notons ici que sur les 105 combinaisons possibles des couples de noeuds, on a trouvé 19 couples compatibles.

La Figure 3.17 montre les appariements compatibles. Ce graphe est composé d'un noeud isolé (m_{53}) et d'une seule composante connexe contenant tous les autres noeuds. Cette composante connexe ne peut pas être interprétée directement comme une solution de correspondance entre les deux structures. En effet, nous observons qu'à l'intérieur de cette composante connexe il y a des segments d'une image qui sont mis en correspondance simultanément avec plusieurs segments de l'autre image, (par exemple le segment g_1 est mis en correspondance avec les segments d_3 et d_6). La contrainte d'unicité n'est pas respectée. Afin d'arriver à une solution correcte, il faut explicitement tenir compte de noeuds qui sont incompatibles.

4. Nous appliquons alors la règle 4. (En fait elle est établie au cours de l'application des trois premières règles). Douze couples de noeuds ont été trouvés incompatibles : pour cinq d'entre eux l'incompatibilité a été déduite de la règle de colinéarité et pour sept d'entre eux de la règle d'ordre. Les noeuds incompatibles sont indiqués sur le Tableau 3.2 par un numéro comportant un

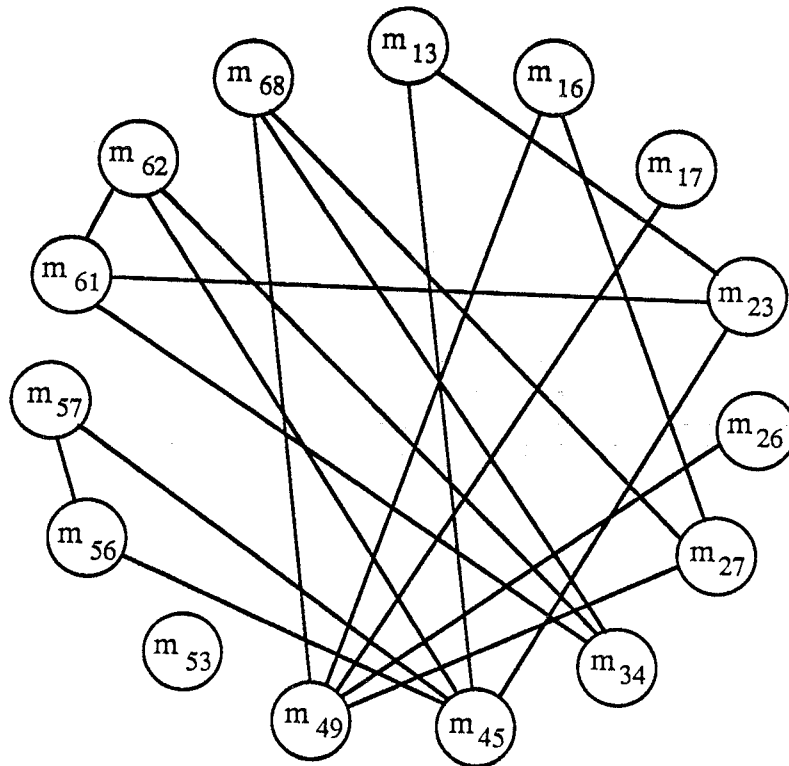


Figure 3.17: Le graphe montrant les compatibilités entre les appariements potentiels après l'application des trois premières règles.

Noeud	13	16	17	23	26	27	34	45	49	53	56	57	61	62	68
13															
16	1														
17	1	1*													
23		1	1												
26	1	1*	1*	1											
27	1		1*	1	1*										
34															
45		3*	3*		3*	3*									
49								1							
53	1			1				3*	3						
56		1	1		1	1			3*	1					
57		1	1		1	1			3*	1					
61					2	2				2					
62					2	2									
68								2				2	1	1	

Table 3.2: Les incompatibilités entre les noeuds après l'application de la règle 4 ; les noeuds directement incompatibles sont montrés par un astérisque et les noeuds qui ne comportent pas cet astérisque ont été trouvés incompatibles après une analyse étendue. Les numéros correspondent soit à l'une des trois premières règles qui sont contredites (cas des numéros comportant l'astérisque), soit à l'analyse étendue qui a provoqué l'incompatibilité : 1-colinéarité étendue, 2-connexité étendue et 3-ordre étendu.

astérisque. Ces numéros correspondent en fait aux règles (1, 2 ou 3) qui sont contredites ce qui permet d'établir l'incompatibilité.

5. Nous appliquons ensuite les analyses appropriées pour établir l'incompatibilité aux cas d'absence partielle de relations topographiques. Le Tableau 3.2 montre tous les noeuds trouvés incompatibles. Cette fois-ci, les numéros sur le tableau correspondent à l'analyse associée à une propriété qui a provoqué cette incompatibilité : 1-colinéarité étendue, 2-connexité étendue et 3-ordre étendu. Aux douze couples de noeuds trouvés incompatibles par la règle 4, trente et un couple de noeuds ont été ajoutés au Tableau 3.2, dont vingt-trois suite à une analyse de colinéarité étendue, sept suite à une analyse de connexité étendue et un suite à une analyse d'ordre étendu. Même après l'application de l'analyse étendue, l'incompatibilité n'a pas pu être établie pour les couples de noeuds suivants : $(m_{13} m_{49})$, $(m_{13} m_{61})$, $(m_{17} m_{68})$, $(m_{23} m_{49})$, $(m_{23} m_{62})$, $(m_{23} m_{68})$, $(m_{45} m_{61})$, $(m_{49} m_{61})$ et $(m_{49} m_{62})$.

L'ensemble de noeuds qui sont incompatibles sont montrés en pointillées sur la Figure 3.18.

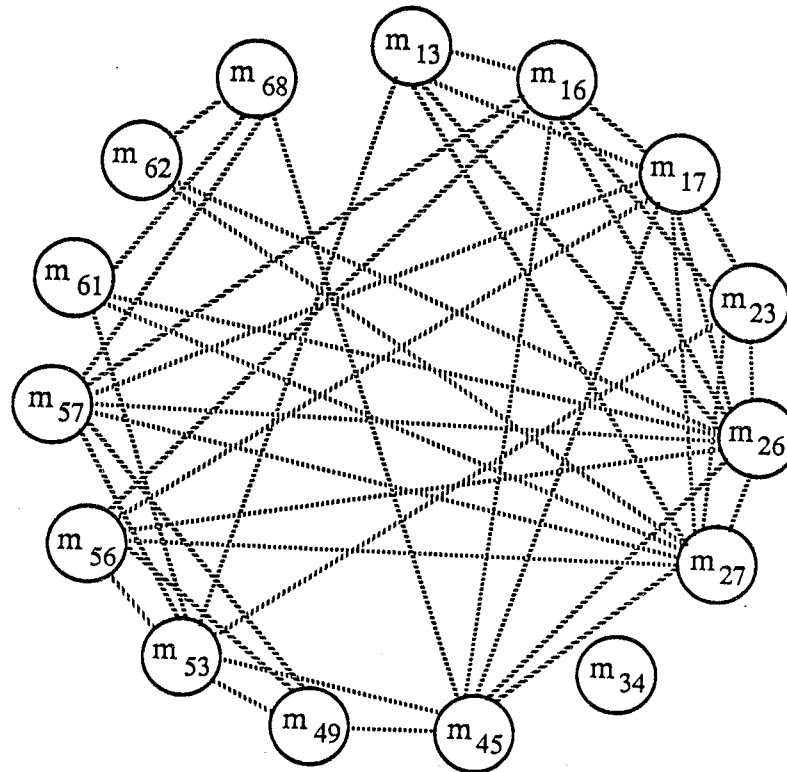


Figure 3.18: Les noeuds incompatibles.

6. Néanmoins, ils restent encore quelques paires de noeuds (43 plus précisément), qui ne sont reliées ni par compatibilité ni par incompatibilité. Alors, on essaie de les rendre compatibles en appliquant la règle de transitivité. Ainsi, de quarante-trois paires de noeuds trente-six deviennent compatibles ; les sept autres couples de noeuds qui restent, contiennent tous le noeud isolé m_{53} ; ce noeud pourrait ne plus être considéré dans le graphe dès l'application des trois premières règles, puisque, comme il n'était pas relié par compatibilité avec aucun autre noeud du graphe, l'application de la transitivité ne pourrait pas le rendre compatible avec aucun autre noeud. Par conséquent, ces sept derniers couples de noeuds sont rendus incompatibles.

Après l'application de la règle de transitivité, on obtient des noeuds qui sont reliés uniquement par compatibilité ou par incompatibilité. Si, malgré l'application de la transitivité, existaient encore de paires de noeuds qui ne seraient reliées ni par compatibilité ni par incompatibilité, alors nous les rendrions incompatibles. Ainsi, nous aboutissons à un graphe où une arête relie deux noeuds mutuellement compatibles et où l'absence d'une arête entre deux noeuds signifie leur incompatibilité. La matrice d'adjacence de ce graphe est présentée au Tableau 3.3. Sur ce tableau, 1 indique que deux noeuds sont mutuellement compatibles et 0 indique qu'ils sont mutuellement incompatibles. L'astérisque sur ce tableau signifie que la compatibilité a été établie pendant l'application des trois premières règles.

Noeud	13	16	17	23	26	27	34	45	49	53	56	57	61	62	68
13															
16	0														
17	0	0													
23	1*	0	0												
26	0	0	0	0											
27	0	1*	0	0	0										
34	1	1	1	1	1	1									
45	1*	0	0	1*	0	0	1								
49	1	1*	1*	1	1*	1*	1	0							
53	0	0	0	0	0	0	0	0	0						
56	1	0	0	1	0	0	1	1*	0	0					
57	1	0	0	1	0	0	1	1*	0	0	1*				
61	1	1	1	1*	0	0	1*	1	1	0	1	1			
62	1	1	1	1	0	0	1*	1*	1	0	1	1	1*		
68	1	1	1	1	1	1*	1*	0	1*	0	1	0	0	0	

Table 3.3: La matrice d'adjacence du graphe de correspondance concernant l'exemple de la Figure 3.15. 1 indique les noeuds compatibles et 0 les noeuds incompatibles. L'astérisque signifie que la compatibilité a été établie au cours de l'application des trois premières règles.

7. Les cliques maximales sont recherchées dans ce graphe. Pour cela, nous appliquons un algorithme décrit par Bolles et Cain [17] et que nous présentons dans l'Annexe B. La Figure 3.19 montre les neuf cliques maximales trouvées : il y a une clique maximale composée de huit noeuds (Fig. 3.19-a), une composée de six noeuds (Fig. 3.19-b), cinq composées de cinq noeuds (Fig. 3.19-c, -d, -e, -f et -g), et deux composées de quatre noeuds (Fig. 3.19-h et -i).

Parmi l'ensemble de ces cliques maximales, celles correspondant au mieux à la réalité sont illustrées sur les Figures 3.19-a (qui est en fait la plus grande clique maximale du graphe) et 3.19-e. Notons ici quelques interprétations bizarres que nous avons obtenues et notamment celles des Figures 3.19-d, -f et -g. En réalité, il s'agit d'anomalies provoquées par le fait que la transitivité a rendu compatibles quelques paires de noeuds, qui ne sont pas réellement compatibles et qu'ensuite cette fausse compatibilité a été propagée par la transitivité à des ensembles de noeuds plus grands. Ainsi, sur la solution de la Figure 3.19-d, l'anomalie a commencé par le fait que les noeuds m_{56} et m_{68} sont devenus compatibles par l'intermédiaire du noeud m_{34} ; quant aux solutions des Figures 3.19-f et -g l'anomalie est due aux compatibilités entre les noeuds m_{16} et m_{17} et les noeuds m_{61} et m_{62} , respectivement, provoquées aussi par l'intermédiaire du noeud m_{34} .

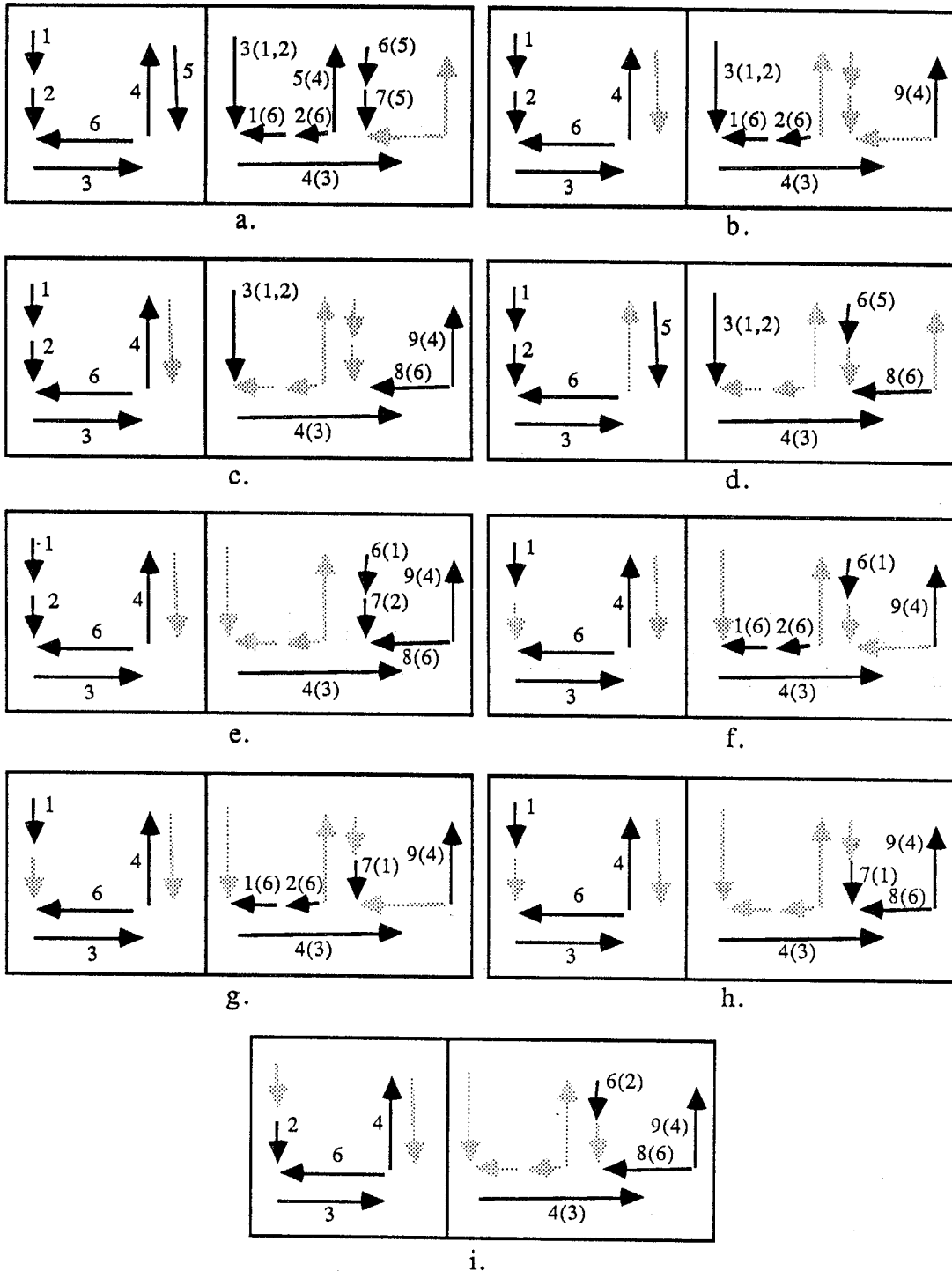


Figure 3.19: Les neuf solutions (cliques maximales) correspondant à l'exemple de la Figure 3.15. Entre parenthèses sont indiqués les segments de l'image gauche correspondants aux segments de l'image droite.

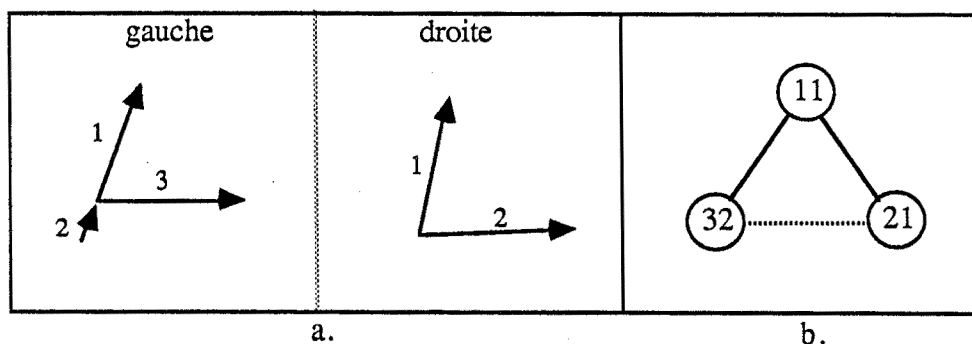


Figure 3.20: (a) Un exemple, où l'algorithme de mise en correspondance aboutit à deux solutions différentes et il retient l'une des deux. (b) Le graphe de correspondance associé à cet exemple (les noeuds incompatibles sont montrés en lignes pointillées).

8. Nous évaluons chaque clique maximale trouvée. Cette évaluation pourrait simplement être le nombre de noeuds composant une clique maximale, c'est-à-dire choisir la plus grande clique maximale. Néanmoins nos expérimentations ont montré qu'un tel choix est relativement mauvais, car il ne tient pas compte du poids local qu'un noeud peut avoir. En effet, un noeud est un appariement potentiel, par conséquent il est possible de le caractériser par une mesure de confiance que nous avons sur cet appariement ; une telle mesure pourrait être la mesure de similarité V que nous avons associée à chaque appariement potentiel (§ 3.2) et qui devient maintenant un bénéfice B_{noeud} que nous associons à chaque noeud. L'évaluation d'une clique maximale est alors la somme des bénéfices de ses noeuds :

$$B_{clique} = \sum B_{noeud} \quad (3.20)$$

La meilleure clique maximale est celle correspondant à la somme la plus élevée.

Ainsi, pour l'exemple de la Figure 3.15, la clique maximale finalement choisie est celle de la Figure 3.19-a.

3.3.3 Quelques cas difficiles

Examinons quelques cas, où la technique de mise en correspondance présentée ci-dessus rencontre des problèmes. Ces cas sont montrés sur les Figures 3.20, 3.21 et 3.22.

En ce qui concerne la Figure 3.20, notre algorithme ne sera pas capable de mettre en correspondance tous les trois segments de l'image gauche avec les deux segments de l'image droite ; il trouvera deux cliques maximales différentes du fait que les appariements g_2/d_1 et g_3/d_2 sont incompatibles et il choisira entre ces deux solutions celle qui aura la somme des bénéfices de ses noeuds la plus élevée.

Dans le cas présenté sur la Figure 3.21, notre algorithme va aussi aboutir à deux solutions identiques ($(m_{11} m_{22})$ et $(m_{21} m_{32})$) du fait que les relations choisies

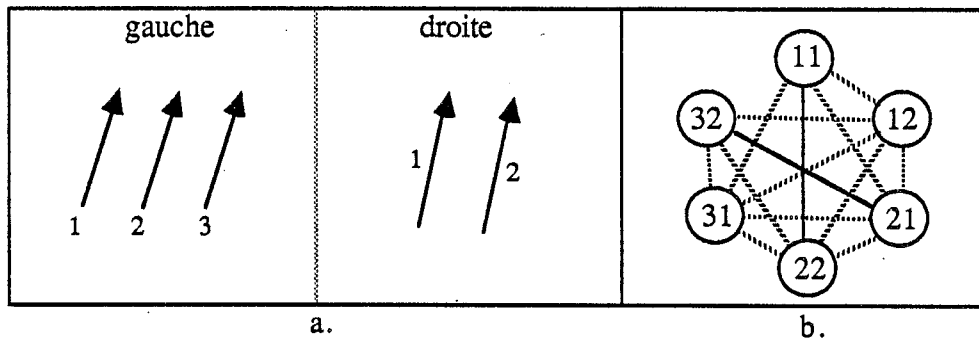


Figure 3.21: Un deuxième exemple, où l'algorithme de mise en correspondance trouve deux solutions identiques et en choisit une. (b) Le graphe de correspondance de cet exemple.

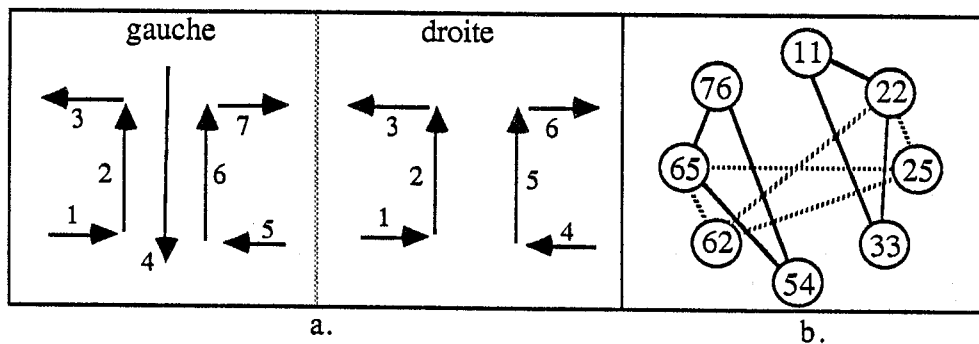


Figure 3.22: Le graphe de correspondance associé à cet exemple montre qu'il y a deux cliques maximales disjointes, qui sont toutes les deux acceptées.

n'autorisent qu'une structure locale deux à deux et il va choisir la meilleure solution selon notre procédure d'évaluation d'une clique.

Quant au cas de la Figure 3.22, entre les deux structures de l'image gauche il existe un segment (g_4), qui est absent entre les deux structures correspondantes de l'image droite. A cause du segment g_4 , les deux structures de l'image gauche ne sont reliées par aucune propriété directe dans le graphe relationnel. (Par contre cette liaison existe bien entre les deux structures correspondantes de l'image droite). Cette inexistence de relations entre les deux structures dans l'une de deux images se traduit en deux cliques maximales différentes dans le graphe de correspondance, c'est-à-dire à deux solutions séparées, qui sont toutes les deux significatives. Il s'agit en réalité de deux cliques maximales qui sont *disjointes* et qui doivent toutes les deux être acceptées.

Deux cliques maximales sont dites disjointes si elles satisfont toutes les conditions suivantes :

- aucun noeud n'est commun entre les deux cliques ;
- aucun segment de l'image gauche participant à la première clique ne participe aussi à la seconde ;

- aucun segment de l'image droite participant à la première clique ne participe aussi à la seconde .

Pour tenir compte de situations comme celle de la Figure 3.22, nous avons ajouté à l'algorithme de mise en correspondance une neuvième étape ; cette étape consiste à accepter non seulement la meilleure clique maximale, mais aussi à considérer itérativement toutes les autres cliques maximales qui sont disjointes de celles déjà retenues et à accepter, parmi cet ensemble, la meilleure clique maximale selon notre procédure d'évaluation d'une clique. Ainsi, pour l'exemple de la Figure 3.22, nous acceptons finalement les deux solutions car elles sont disjointes.

3.3.4 Complexité

La complexité du processus de la construction du graphe est proportionnelle au nombre d'arêtes du graphe. Il est possible de réduire cette complexité en prenant attentivement en considération les contraintes photométriques et géométriques afin d'éliminer un grand nombre d'appariements incorrects. Nous avons vu au § 3.2 comment ceci est rendu possible en utilisant la mesure de similarité V que nous avons associée aux appariements potentiels.

L'énumération de toutes les cliques maximales d'un graphe est un problème NP-difficile. Par conséquent, tout algorithme connu qui recherche de cliques maximales est exponentiel en complexité [8], [17]. Afin d'accélérer la recherche des correspondances entre les images gauche et droite, nous pourrions choisir un algorithme moins complexe que celui de recherche des cliques maximales, comme par exemple un algorithme de prédiction/vérification d'hypothèses. Une autre possibilité pour accélérer cette recherche serait d'implanter un algorithme de recherche des cliques maximales, qui s'arrêterait dès que l'on estimerait qu'à un moment donné une clique contient suffisamment de noeuds pour être significative. Néanmoins, aucune des deux solutions précédentes ne peut garantir que l'interprétation qu'elle fournit soit optimale. **Nous avons donc préféré chercher une correspondance optimale plutôt qu'une solution rapide mais sous-optimale.** Pour cela, nous avons choisi de rechercher explicitement dans le graphe toutes les cliques maximales et de sélectionner ensuite, parmi cet ensemble, les meilleures cliques maximales disjointes comme nous l'avons décrit ci-dessus.

Une façon d'augmenter l'efficacité de l'exploration des cliques maximales consisterait à restreindre au maximum le nombre d'arêtes du graphe. Ceci est possible en détectant explicitement tous les noeuds incompatibles comme nous l'avons décrit. Les noeuds incompatibles restreignent la création de nouvelles arêtes quand on applique la règle de transitivité. Qui plus est, non seulement l'efficacité augmente, mais la probabilité de relier de paires d'appariements potentiels incorrects est considérablement réduite. En effet, la transitivité ne reflète pas des propriétés mesurées dans l'image et des interprétations d'appariements incorrects semblent apparaître lorsque la règle de transitivité est appliquée de façon abusive.

3.4 Résultats expérimentaux

La stratégie de mise en correspondance, que nous avons décrite, a été appliquée à une série de paires d'images stéréo acquises par deux caméras montées sur le robot mobile de l'INRIA. Nous disposons d'ailleurs de données de calibrage du banc stéréoscopique [31]. Les images représentant des scènes d'intérieur sont montrées sur les Figures 3.24 (Bureau 1) et 3.31 (Bureau 2). Les segments extraits de ces paires d'images sont montrés respectivement sur les Figures 3.25 et 3.32. Nous ne montrons pas les graphes relationnels associés à ces images. Il apparaît peu réaliste d'appliquer directement notre stratégie de mise en correspondance aux structures complètes des images. A la place, nous l'appliquons localement à des sous-ensembles d'images (fenêtres) et la mise en correspondance globale est l'ensemble de résultats locaux obtenus dans chacune de ces fenêtres. (La stratégie de mise en correspondance globale est présentée en détail au § 3.4.2). Enfin, nous avons appliqué la même stratégie à deux autres paires d'images, une représentant le toit d'un bâtiment du stade olympique de Munich, (cf. Fig. 3.35 (Toit)) et une autre représentant les branches d'un arbre, (cf. Fig. 3.38 (Arbre)).

3.4.1 Mise en correspondance locale

Nous présentons la stratégie de mise en correspondance appliquée sur deux fenêtres de taille 64×64 pixels. Ces fenêtres sont montrées sur la Figure 3.25 (gauche). La Figure 3.26 (en haut, à gauche) montre 34 segments se trouvant dans la première fenêtre. La Figure 3.26 (en haut, à droite) montre les candidats potentiels dans l'image droite associés à ces segments. Chaque segment de la fenêtre a en moyenne 2 candidats potentiels. Sur ces appariements potentiels, nous appliquons les quatre premières règles et nous construisons le graphe associé : les 77 noeuds du graphe sont reliés par 817 arêtes dont 146 établies par les trois premières règles et 671 par la transitivité, (347 paires d'appariements étaient incompatibles). Nous recherchons ensuite la meilleure clique maximale ; la Figure 3.26 (bas) montre la meilleure correspondance composée de 32 noeuds mutuellement compatibles ; elle a été sélectionnée parmi 18 cliques maximales. Le temps de mise en correspondance était de 14.2 secondes CPU sur un VAX 11/780.

Dans la deuxième fenêtre il y avait 37 segments, (cf. Fig. 3.27 en haut, à gauche). Les 53 appariements potentiels des 37 segments sont montrés sur la Figure 3.27 (en haut, à droite). Le graphe associé était donc composé de 53 noeuds reliés par 511 arêtes (54 étant établies par les trois premières règles et 457 par la transitivité) et 90 paires d'appariements étaient incompatibles. La meilleure correspondance a été choisie après avoir exploré 1357 cliques maximales ; elle contient 2 cliques maximales disjointes, dont une est composée de 18 noeuds et l'autre de 5 noeuds. Le résultat de mise en correspondance est montré sur la Figure 3.27, en bas. Le temps de mise en correspondance pour cet exemple était de 18.5 secondes CPU.

Le Tableau 3.4 récapitule les résultats présentés ci-dessus. Comparant ces résultats, on peut remarquer que le nombre de cliques maximales explorées ne dépend pas du nombre de segments présents dans une fenêtre, mais il dépend essentiellement

MISE EN CORRESPONDANCE LOCALE	Fenêtre 1	Fenêtre 2
Taille de fenêtre	64 × 64	64 × 64
No. segments de l'image gauche	34	37
No. segments de l'image droite	44	34
No. appariements potentiels	77	53
No. noeuds directement compatibles	146	54
No. noeuds compatibles par transitivité	671	457
No. noeuds incompatibles	347	90
No. cliques maximales explorées	18	1357
No. cliques maximales retenues	1	2
No. segments de l'image gauche appariés	32	23
Temps CPU (secondes)	14.2	18.5

Table 3.4: Mise en correspondance locale.

de la structure locale des images, de la justesse de sa détection et du degré de ressemblance des deux images.

3.4.2 Mise en correspondance globale

Considérons une paire d'images stéréo et un découpage de l'image gauche en seize fenêtres de taille 64×64 pixels (la taille des images initiales étant de 256×256 pixels). Ce découpage est montré sur la Figure 3.23. La mise en correspondance globale est effectuée de la façon suivante : la technique de mise en correspondance locale est appliquée de manière séquentielle à chacune de seize fenêtres, l'ordre de leur considération étant indiqué sur la Figure 3.23 par le numéro se trouvant à l'intérieur de chaque fenêtre. Pour chaque fenêtre, nous considérons l'ensemble de segments lui appartenant. Nous sélectionnons ensuite leurs candidats potentiels parmi les segments de l'image droite, nous construisons le graphe associé et nous choisissons la ou les meilleures cliques maximales de ce graphe.

Dès qu'une mise en correspondance partielle est trouvée dans une fenêtre, nous éliminons les segments de l'analyse ultérieure. Ainsi, la recherche des correspondances se simplifie au fur et à mesure de l'exploration des fenêtres suivantes, parce qu'il y a moins de candidats potentiels et donc moins de noeuds dans les graphes de correspondance associés à ces fenêtres. Néanmoins, un segment peut appartenir à plusieurs fenêtres adjacentes ; par conséquent, nous éliminons les interprétations de tout segment se trouvant complètement à l'intérieur d'une fenêtre, mais nous n'appliquons pas cette technique aux segments qui appartiennent à deux fenêtres voisines, afin de ne pas éliminer leurs interprétations correctes quand nous recherchons les correspondances aux fenêtres suivantes.

Notons que par cette approche, comme un segment peut appartenir à plusieurs fenêtres adjacentes, plusieurs interprétations différentes peuvent lui être associées du fait qu'il peut participer à des cliques maximales différentes associées à de fenêtres voisines. Lorsque l'on associe à un segment appartenant à deux fenêtres voisines

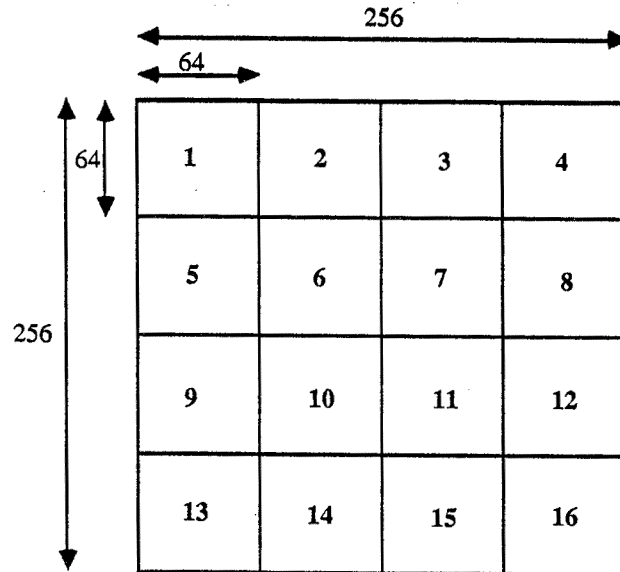


Figure 3.23: Le découpage de l'image gauche en 16 fenêtres et l'ordre de leur considération pendant la mise en correspondance globale.

deux interprétations différentes, on parle alors de présence d'un *conflit*. Lors de la détection d'un tel conflit, il est résolu en acceptant comme interprétation finale pour le segment concerné celle qui participe à la clique maximale dans laquelle nous avons le plus de confiance, c'est-à-dire la clique maximale qui est la meilleure entre les deux cliques (au sens de la somme des bénéfiques). Il faut noter ici que malgré le nombre important de segments appartenant à plusieurs fenêtres à la fois (de l'ordre de 25% environ du nombre total de segments détectés), pour la grande majorité de ces segments il n'y a pas de conflit : ces segments reçoivent les mêmes interprétations malgré leur appartenance à plusieurs cliques maximales différentes ! (Une étude expérimental sur dix paires d'images stéréoscopiques a montré qu'en moyenne 10% seulement de segments appartenant à deux fenêtres adjacentes sont concernés par des conflits).

Cette stratégie de mise en correspondance globale a été appliquée aux quatre paires d'images des Figures 3.24, 3.31, 3.35 et 3.38.

Dans le premier exemple (Bureau 1), il y avait 415 segments dans l'image gauche et 425 segments dans l'image droite, (cf. Fig. 3.25). L'algorithme de mise en correspondance a exploré au total 122818 cliques maximales, il a retenu 19 (pour les seize fenêtres) et il a apparié 289 segments de l'image gauche avec 298 segments de l'image droite, (rappelons qu'un segment long d'une image peut être mis en correspondance avec plusieurs segments courts et colinéaires de l'autre image). La Figure 3.28 montre le résultat de mise en correspondance. L'algorithme a résolu 8 conflits, ce qui signifie que, parmi les 94 segments de l'image gauche qui appartenaient au moins à deux fenêtres adjacentes, l'algorithme a attribué seulement à 8 segments des interprétations doubles qu'il a ensuite élucidées. Notre algorithme a apparié environ 70% des segments en 150.4 minutes CPU (construction du graphe,

BUREAU 1	Recherche exhaustive	Recherche non exhaustive
Temps CPU (minutes)	150.4	4.4
No. conflits	8	10
No. segments de l'image gauche appariés	289	282
No. segments de l'image droite appariés	298	291
No. appariements faux	4	7
No. appariements corrects non établis	26	34
No. cliques maximales explorées	122818	3988
No. cliques maximales retenues	19	18

Table 3.5: Résultats de mise en correspondance globale pour la paire d'images de Bureau 1. La première colonne correspond à une recherche exhaustive des cliques maximales et la deuxième colonne à une recherche non exhaustive.

résolution des conflits et recherche des cliques maximales). Parmi le nombre total d'appariements établis, il y avait 4 appariements qui n'étaient pas corrects, c'est-à-dire un pourcentage 1.4% du nombre total d'appariements. Finalement, parmi les 126 segments de l'image gauche qui n'étaient pas appariés, 26 segments avaient leur correspondant dans l'image droite que notre algorithme pour de raisons diverses n'a pas pu appairer. La Figure 3.29 montre l'ensemble de segments non appariés.

Le temps (150.4 minutes CPU) de mise en correspondance globale pour cet exemple est notamment dû au temps de mise en correspondance locale dans l'une de seize fenêtres ; plus précisément, pour la fenêtre contenant la chaise entre les deux bureaux de la Figure 3.24 gauche, le graphe de correspondance contenait 142 noeuds, notre système a exploré plus de 100000 cliques en retenant finalement une et le temps de cette mise en correspondance locale était de 141.8 minutes CPU !

Il est intéressant de comparer le résultat de mise en correspondance de la Figure 3.28, obtenu en 150.4 minutes CPU où une recherche exhaustive des cliques maximales a été effectuée, avec le résultat de mise en correspondance montré sur la Figure 3.30 établi en 4.4 minutes CPU ; ce dernier résultat a été obtenu en restreignant le nombre de cliques maximales explorées par fenêtre à 500, c'est-à-dire que dès que le nombre de cliques maximales obtenues dans une fenêtre dépassait 500, on arrêta la recherche et on choisissait la meilleure clique parmi ces 500. Les résultats comparatifs entre la mise en correspondance avec la recherche exhaustive et la recherche non exhaustive sont présentés au Tableau 3.5.

Dans le deuxième exemple (Bureau 2), commençant avec 351 segments dans l'image gauche et 315 segments dans l'image droite, (Figure 3.32), l'algorithme, après avoir exploré exhaustivement 12447 cliques maximales, il a retenu 22 cliques et il a apparié 197 segments de l'image gauche avec 199 segments de l'image droite, (cf. Fig. 3.33). Il a résolu 4 conflits (84 segments étaient communs à de fenêtres adjacentes) et le temps de calcul était de 7.4 minutes CPU. Parmi les appariements établis, 2 n'étaient pas corrects et parmi les 154 segments de l'image gauche non appariés, 36 auraient dû être appariés. La Figure 3.34 montre tous les segments

non appariés. On constate facilement que la grande majorité de ces segments sont des segments soit trop courts pour être significatifs, soit visibles dans une image et absents dans l'autre.

Dans le troisième exemple (Toit), l'algorithme a trouvé une interprétation composée de 61 noeuds, commençant avec 100 segments dans l'image gauche et 107 segments dans l'image droite (Figure 3.36). La Figure 3.37(haut) montre le résultat de la mise en correspondance ; sur cette figure les segments appariés portent le même numéro. Parmi ces appariements, il y a 8 qui sont faux et de 39 segments de l'image gauche non appariés, il y a 15 qui auraient dû être appariés. La Figure 3.37(bas) montre l'ensemble de segments qui n'ont pas été appariés.

Dans le quatrième exemple (Arbre), commençant avec 380 segments dans l'image gauche et 379 segments dans l'image droite (Figure 3.39), l'algorithme a apparié 264 segments de l'image gauche avec 267 segments de l'image droite, (Fig. 3.40)(haut) ; il a résolu 3 conflits et le temps de calcul était de 17.4 minutes CPU. Les segments non appariés pour cette paire d'images sont montrés sur la Figure 3.40(bas).

Le Tableau 3.6 résume tous les résultats présentés ci-dessus.

A titre de comparaison de ces résultats, nous présentons, pour les deux derniers exemples (Images de Toit et d'Arbre), les résultats de mise en correspondance obtenus par l'algorithme de système stéréo binoculaire développé par N. Ayache [2] (Figures 3.41-haut et 3.41-bas, respectivement). Nous constatons que globalement nos résultats sont meilleurs dans la mesure où le nombre d'appariements obtenu est nettement supérieur. Cependant, notre système est moins rapide, d'une part parce qu'il effectue un traitement supplémentaire (la construction des deux graphes relationnels décrivant les deux images) et d'autre part car la recherche dans le graphe de correspondance est exhaustive, tandis que l'algorithme d'exploration du graphe de correspondance utilisé par N. Ayache est du type prédiction-vérification d'hypothèses, c'est-à-dire non exhaustif.

3.5 Discussion

Dans ce chapitre, nous avons proposé une nouvelle technique afin de résoudre le problème de mise en correspondance d'une paire stéréo. Elle consiste à extraire de chaque image de structures locales 2-D et ensuite à appairer les structures similaires entre les deux images. La mise en correspondance est effectuée en découpant les images en un ensemble de fenêtres sur chacune desquelles on construit un graphe (l'espace d'interprétation) et on recherche les cliques maximales les mieux adaptées pour représenter la correspondance entre les deux images.

Un problème qui surgit alors est comment découper les images initiales en fenêtres. D'une part, les fenêtres choisies doivent être assez petites pour que le graphe de correspondance associé à chacune de ces fenêtres contienne relativement peu de noeuds et que sa construction et son exploration soient effectuées assez rapidement. Néanmoins, si la taille des fenêtres est trop petite, alors on risque de ne pas avoir suffisamment de structures et d'aboutir à des erreurs de mise en correspondance ; ce résultat erroné va très probablement se propager ensuite aux fenêtres

MISE EN CORRESPONDANCE GLOBALE	Bureau 1	Bureau 2	Toit	Arbre
Taille image	256 × 256	256 × 256	240 × 240	240 × 240
Taille fenêtres locales	64 × 64	64 × 64	80 × 80	60 × 60
No. segments de l'image gauche	415	351	100	380
No. segments de l'image droite	425	315	107	379
No. appariements potentiels	833	590	288	660
No. cliques maximales explorées exhaustivement	122818	12447	3560	27144
No. cliques maximales retenues	19	22	11	19
No. segments de l'image gauche appariés	289	197	61	264
No. segments de l'image droite appariés	298	199	60	267
No. appariements faux	4	2	8	12
No. appariements corrects non établis	26	36	15	27
No. segments de l'image gauche communs au moins à deux fenêtres	94	84	29	104
No. conflits	8	4	7	3
Temps CPU (minutes) recherche exhaustive	150.4	7.4	1.1	17.4
Temps CPU (minutes) recherche non-exhaustive	4.4	2.0	0.3	3.1

Table 3.6: Mise en correspondance globale.

suivantes, parce que, comme nous l'avons déjà décrit, nous ne considérons plus les interprétations établies quand nous recherchons la correspondance dans les fenêtres suivantes. D'autre part, en choisissant des fenêtres assez grandes, on minimise les risques précédents, mais on pénalise considérablement le temps de calcul de mise en correspondance globale, puisque la taille des différents graphes de mise en correspondance locale augmente.

Un autre aspect du problème lié au découpage des images en fenêtres est le suivant : quand on effectue ce genre de découpage d'une image, on découpe en réalité le graphe relationnel représentant cette image et par conséquent, on coupe un certain nombre de relations liant les segments voisins de l'image. Or, ce découpage doit être effectué de telle façon que, le nombre de liaisons coupées dans le graphe relationnel soit minimum.

En récapitulant, il n'est pas très convenable de découper les images en fenêtres égales, comme nous l'avons fait. Il faut plutôt les découper d'une manière qui tienne compte à la fois du nombre de noeuds composant les graphes des mises en correspondance locales et du nombre de liaisons coupées dans les graphes relationnels.

Une deuxième question que soulève notre technique de mise en correspondance concerne l'évaluation des solutions trouvées : comment sélectionner parmi elles celle qui est la plus significative ? Parmi l'ensemble de cliques maximales explorées, nous considérons comme meilleure celle qui satisfait notre critère de meilleure correspondance. Après avoir exploré plusieurs critères d'évaluation, nous avons finalement retenu celui que nous avons présenté ci-dessus et qui semblait fournir dans la grande majorité des cas la solution la plus significative ; cependant, rien ne nous garantit finalement que la clique maximale que nous avons retenue comme meilleure est effectivement celle qui est la plus significative. Une vérification serait possible seulement dans la phase de reconstruction de la scène ou si nous avons à notre disposition un modèle quelconque de la scène traitée.

La méthode de mise en correspondance que nous avons présentée ressemble beaucoup aux méthodes de mise en correspondance de modèles géométriques avec des données soit 2-D [17], [43], soit 3-D [18]. Cependant, par rapport à ces méthodes, notre méthode comporte une différence essentielle : aucune hypothèse de rigidité quant aux structures à apparier n'a été utilisée.

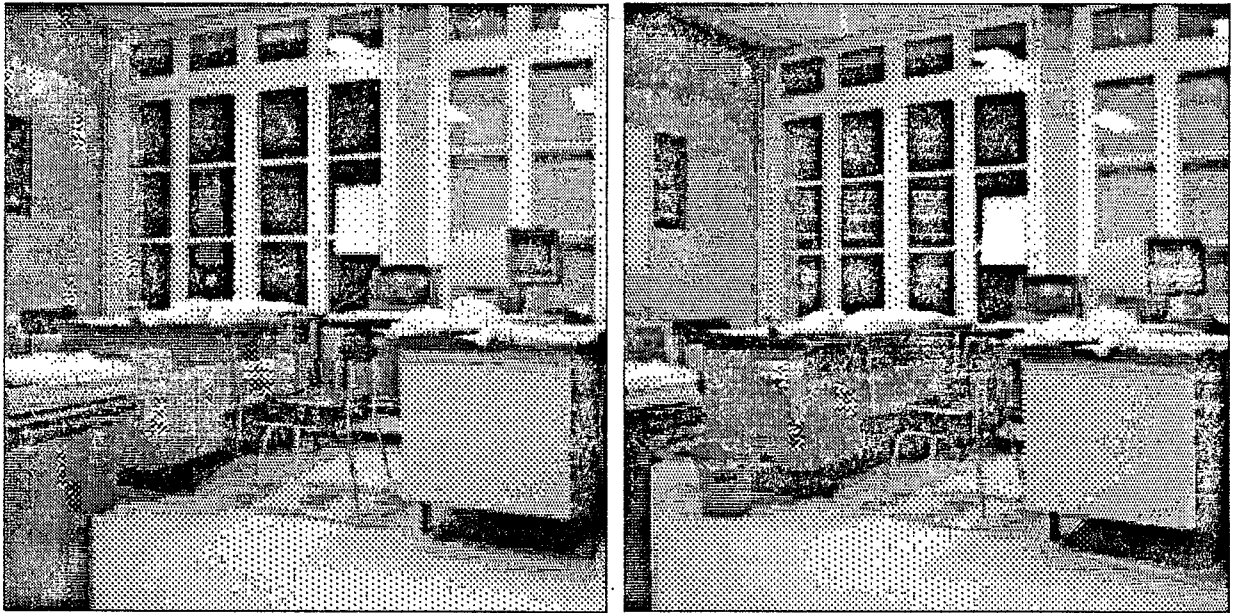


Figure 3.24: La première paire d'images.

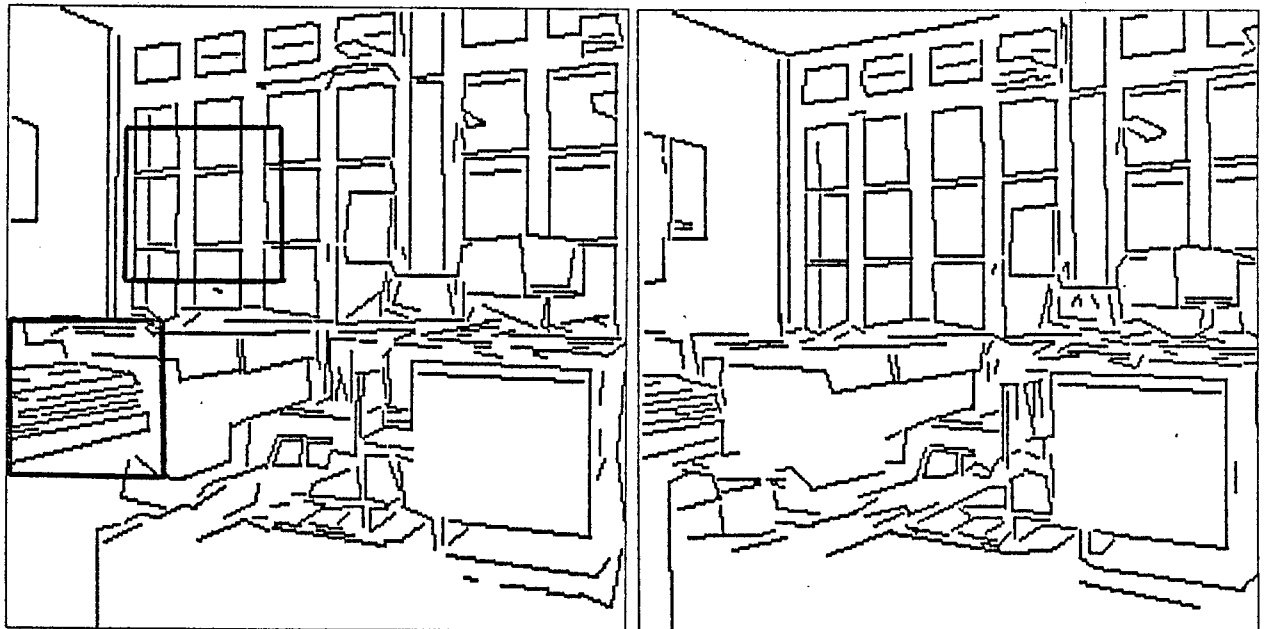


Figure 3.25: Les segments extraits de la première paire d'images. Sur la figure de gauche sont montrées les deux fenêtres sur lesquelles la mise en correspondance locale est présentée.

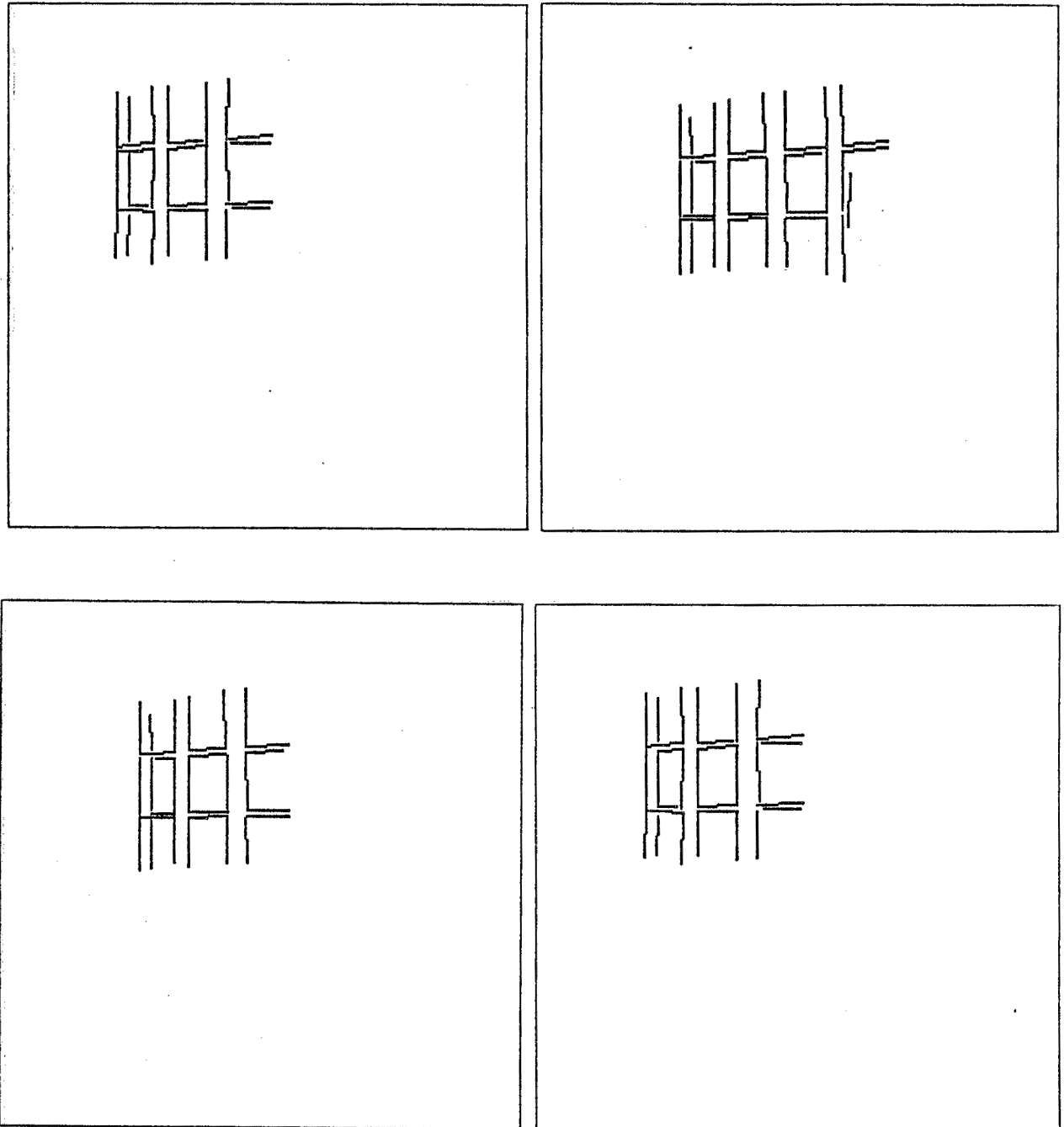


Figure 3.26: Premier exemple de mise en correspondance locale ; en haut : les segments se trouvant dans la première fenêtre et leurs candidats potentiels dans l'image droite ; en bas : le résultat de mise en correspondance (32 segments).

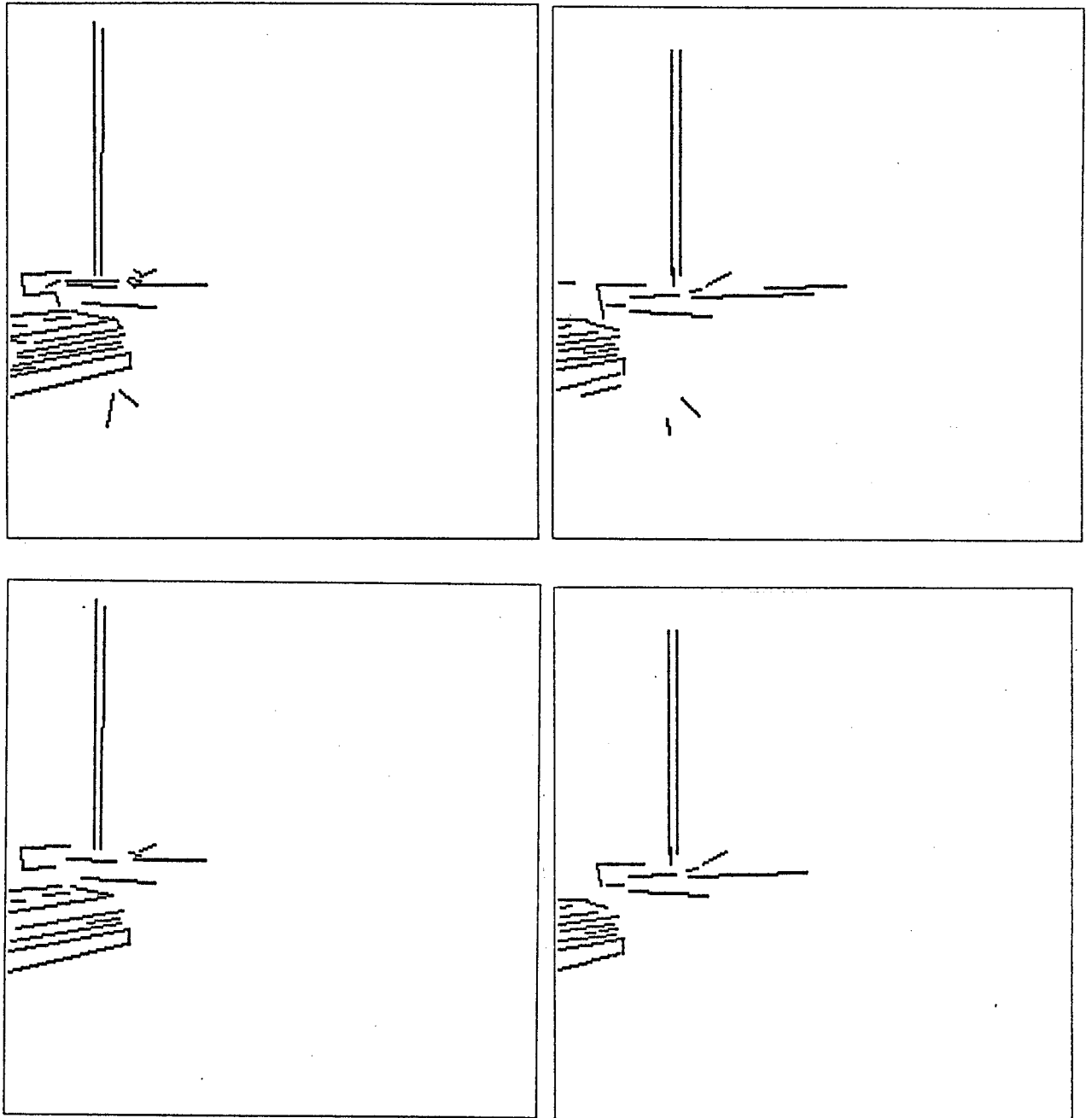


Figure 3.27: Deuxième exemple de mise en correspondance locale ; en haut : les segments se trouvant dans la deuxième fenêtre et leurs candidats potentiels dans l'image droite ; en bas : le résultat de mise en correspondance (23 segments).



Figure 3.28: Le résultat de l'appariement des segments extraits de la première paire d'images.

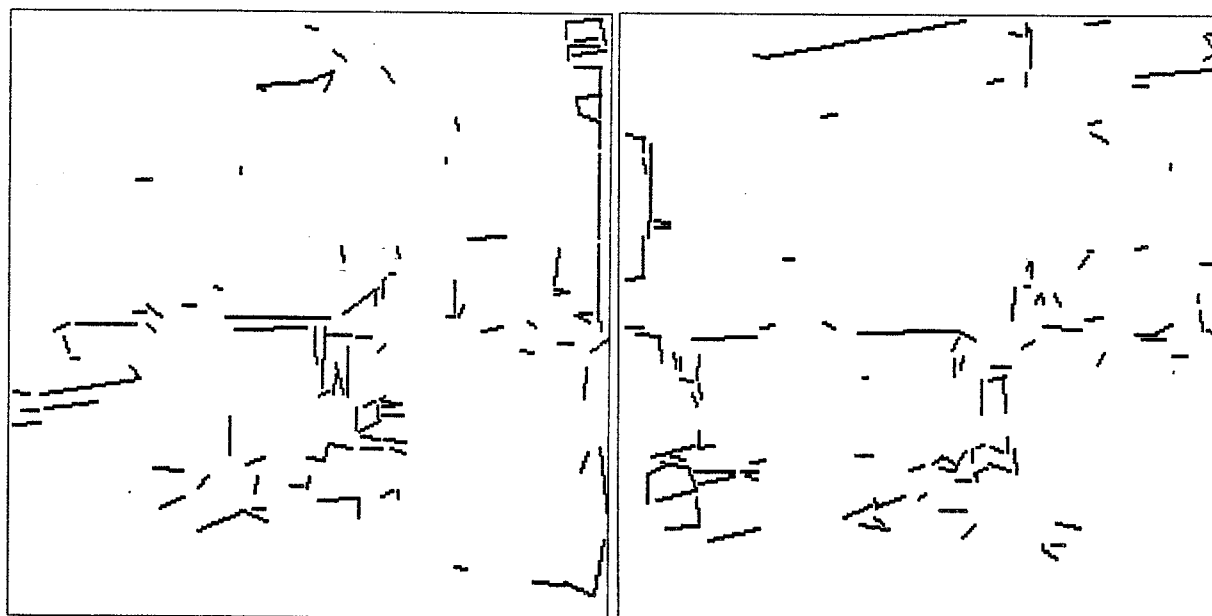


Figure 3.29: Les segments non appariés pour la première paire d'images.



Figure 3.30: Le résultat de mise en correspondance de la première paire d'images obtenu par une recherche non exhaustive des cliques maximales.

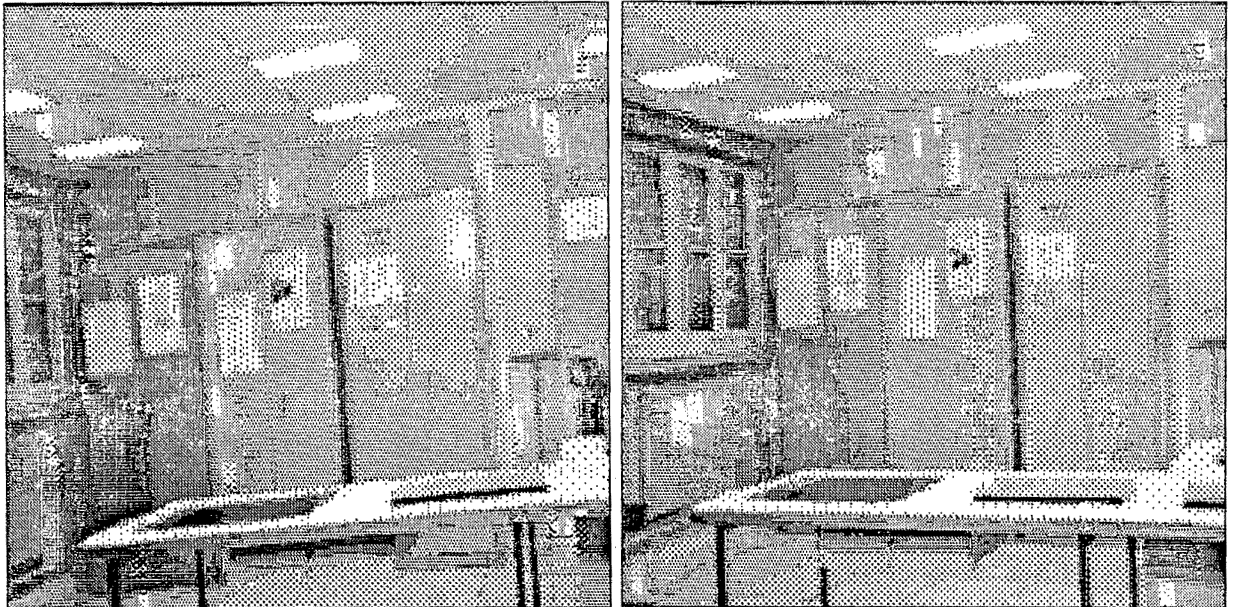


Figure 3.31: La deuxième paire d'images.

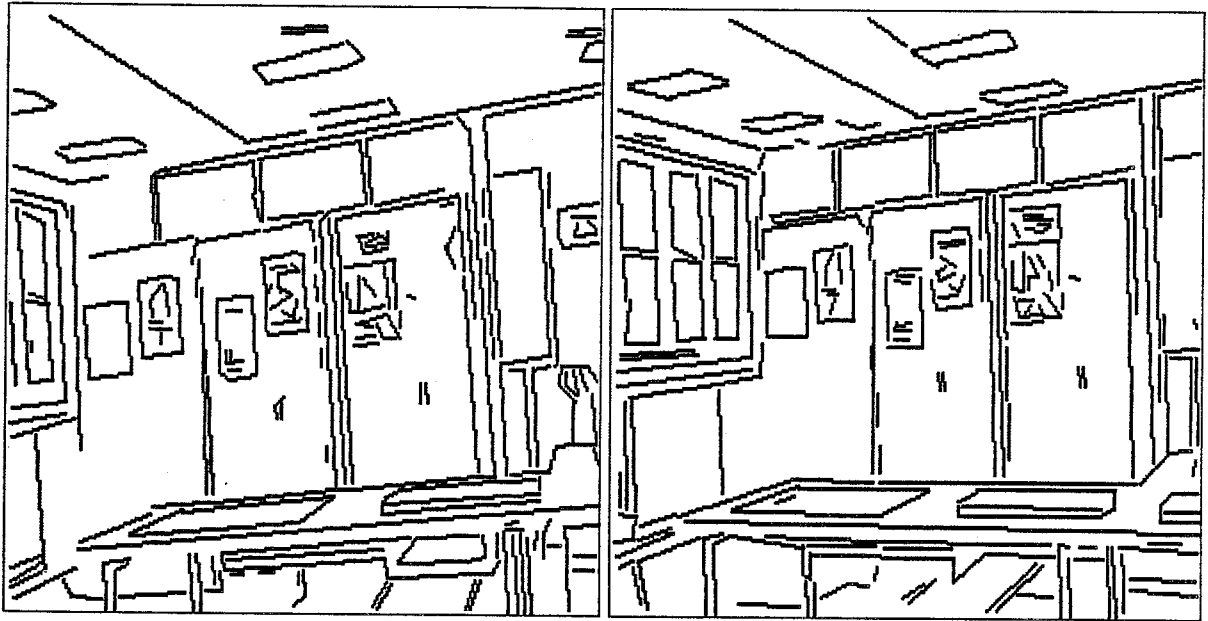


Figure 3.32: Les segments extraits de la deuxième paire d'images.

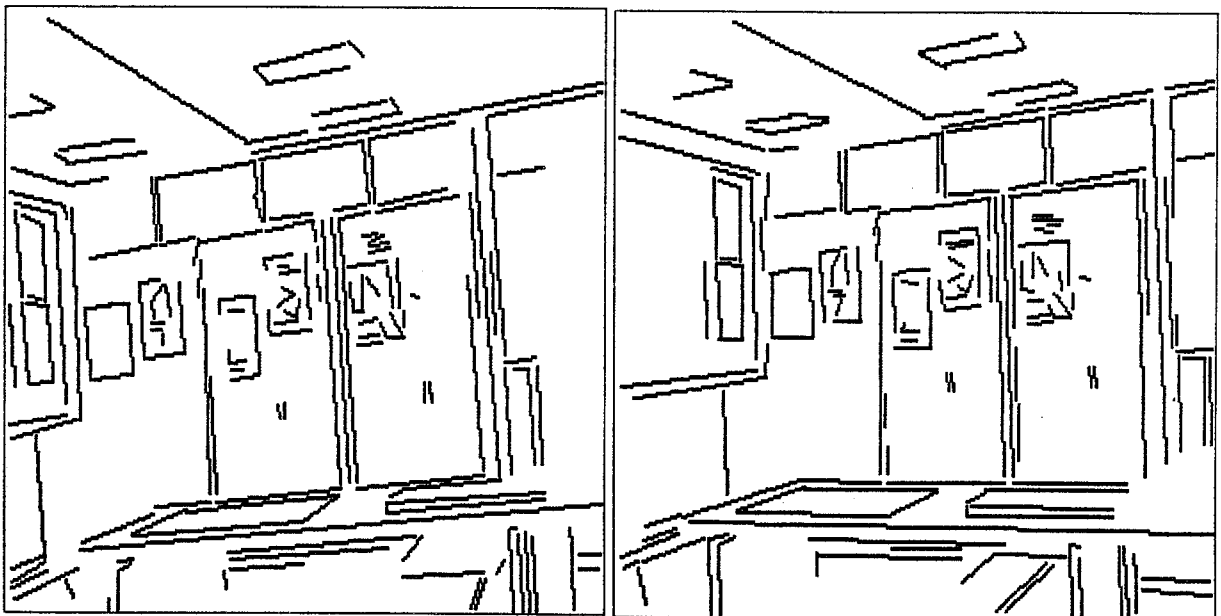


Figure 3.33: Le résultat de l'appariement des segments extraits de la deuxième paire d'images.

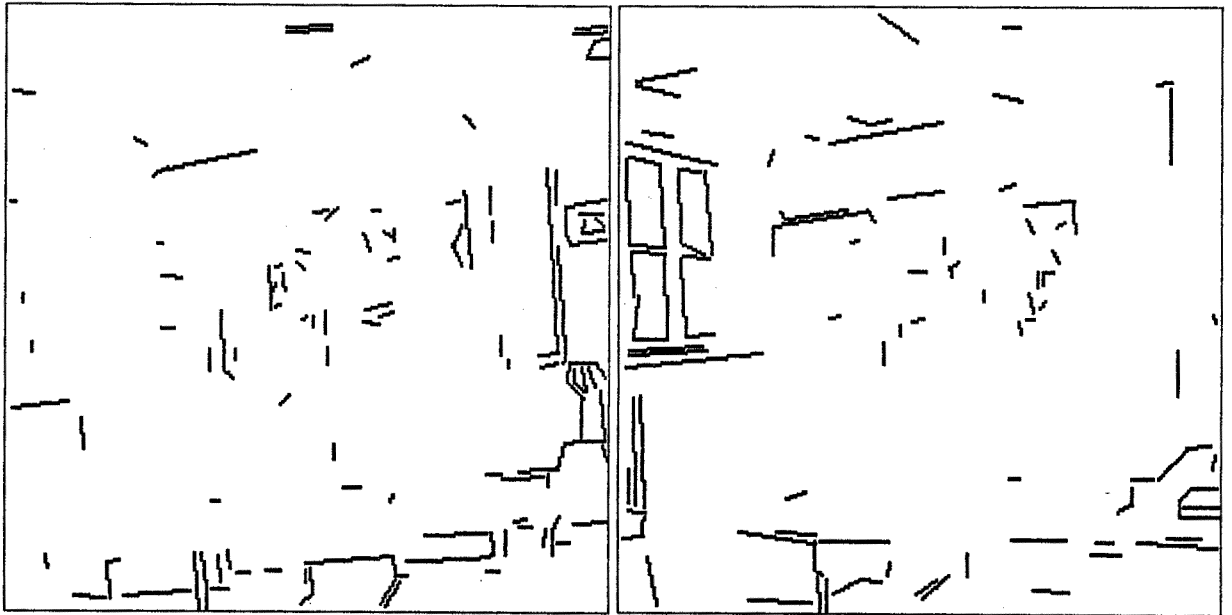


Figure 3.34: Les segments non appariés pour la deuxième paire d'images.

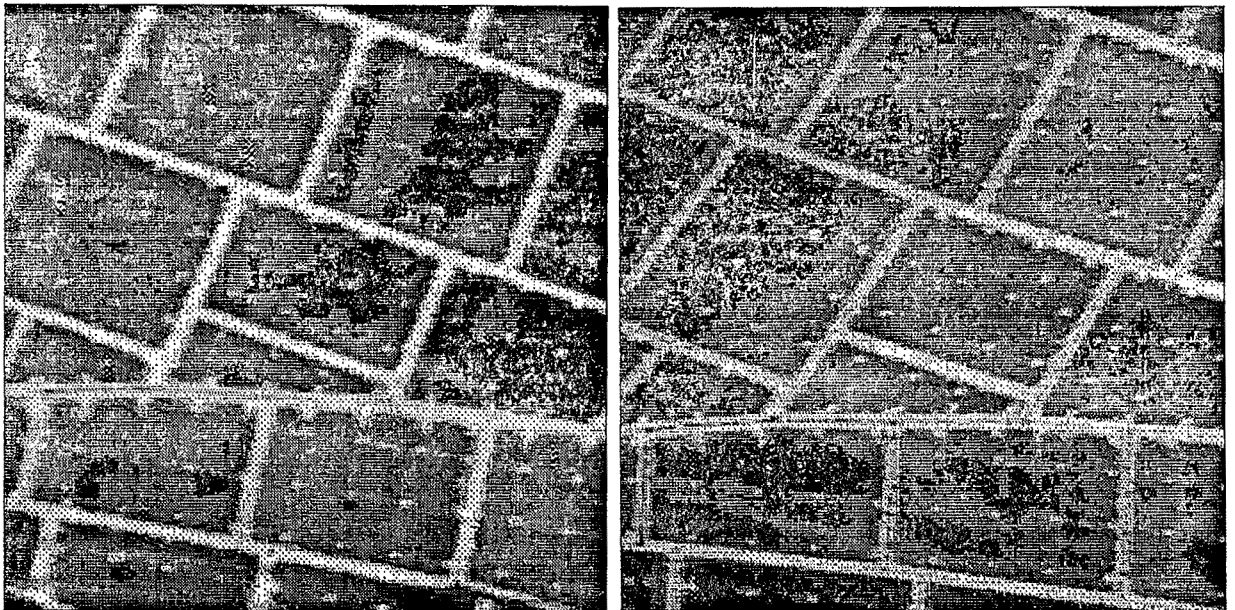


Figure 3.35: La troisième paire d'images.

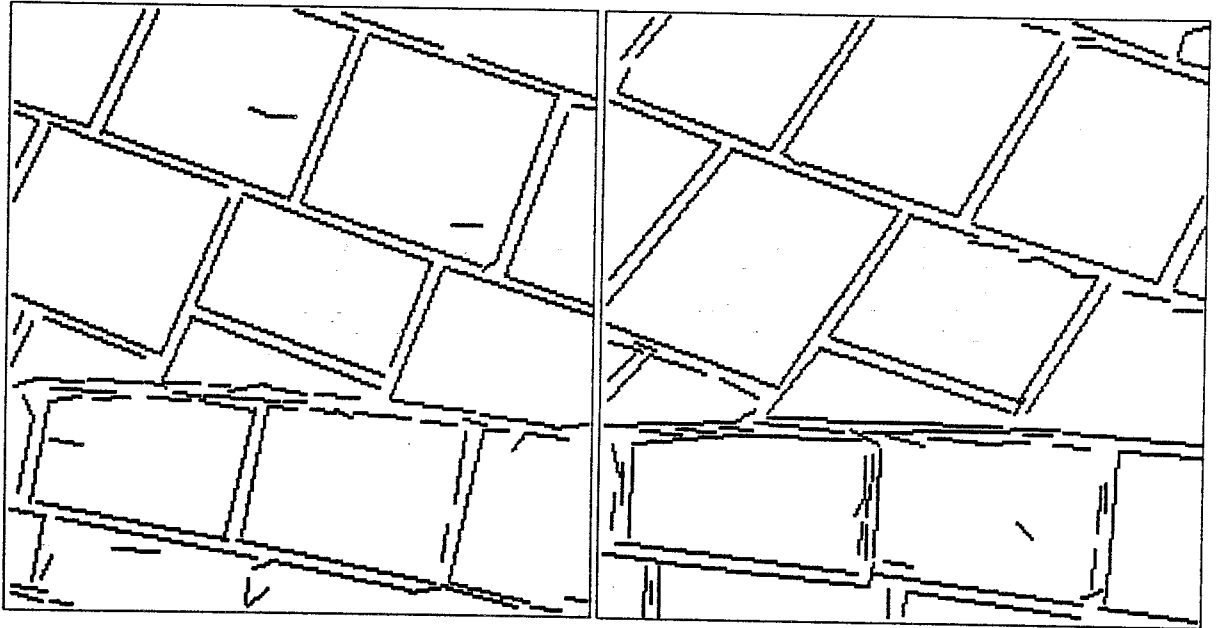


Figure 3.36: Les segments extraits des images de Toit.

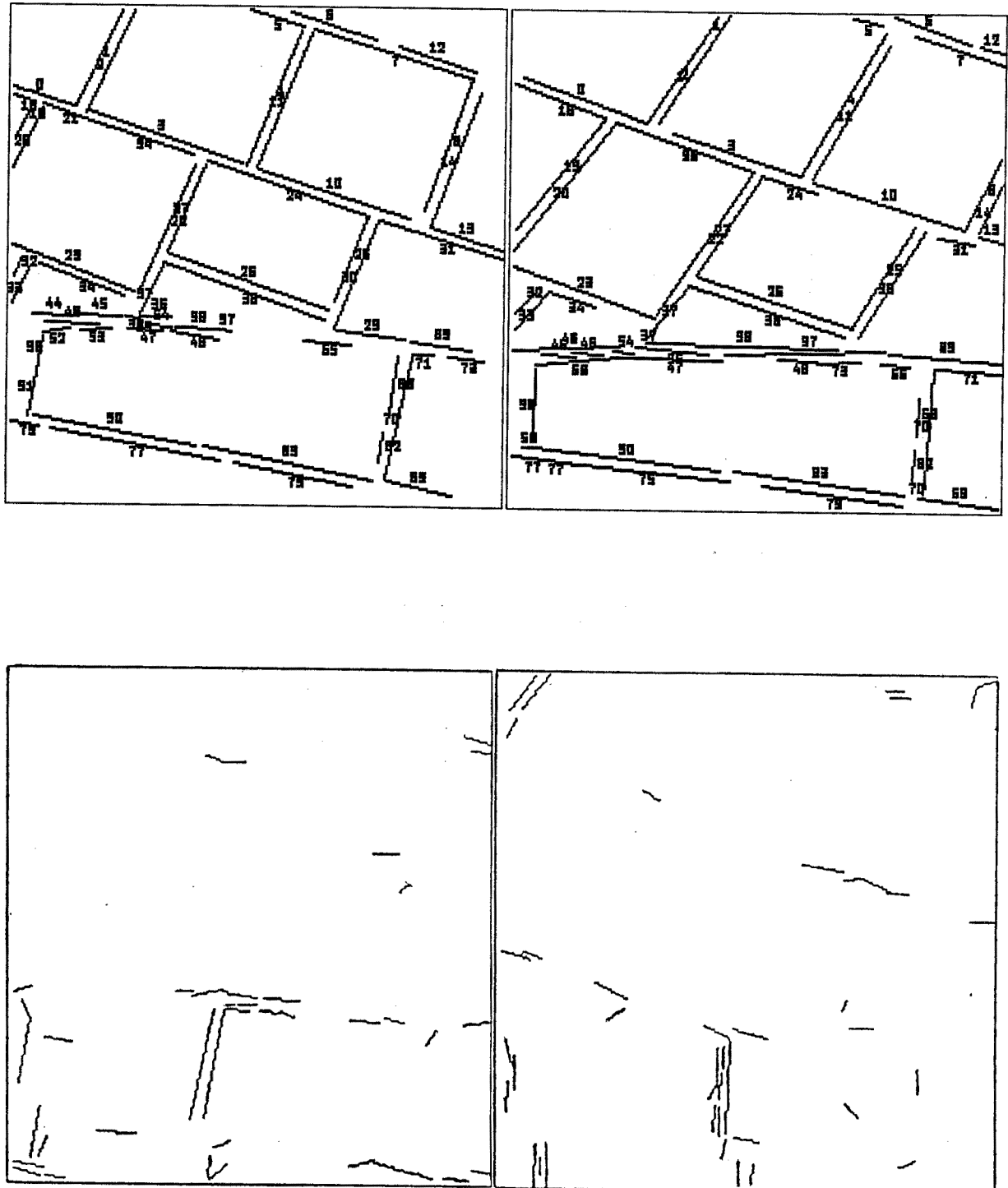


Figure 3.37: En haut : le résultat de mise en correspondance pour les images de Toit. Les segments qui se correspondent portent le même numéro.
En bas : les segments non appariés.

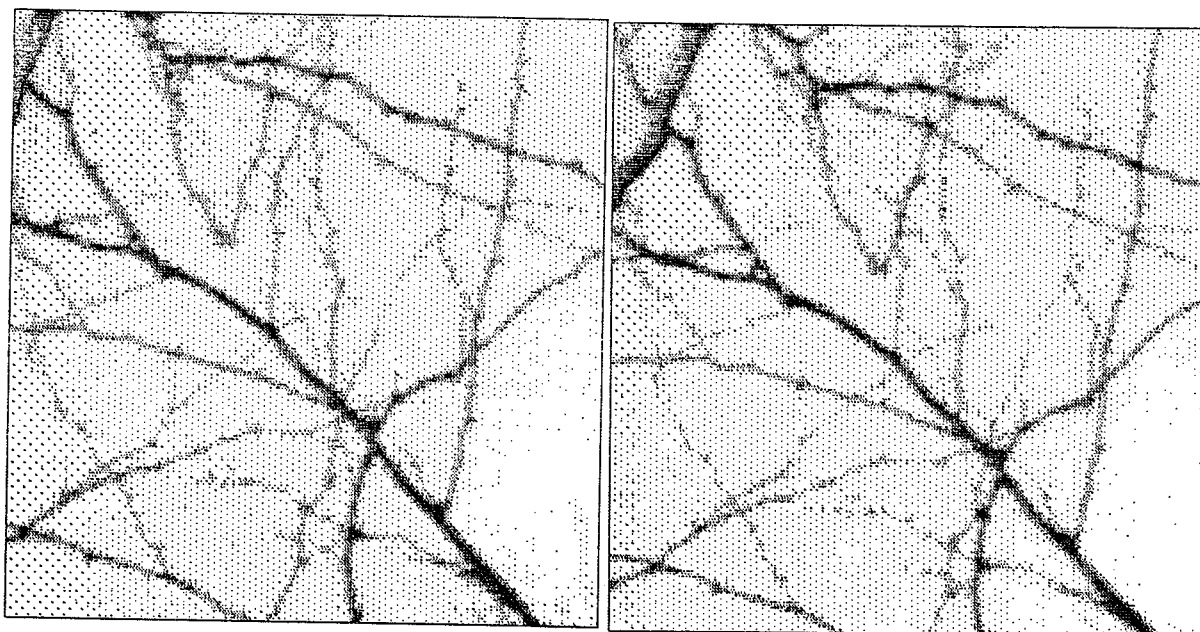


Figure 3.38: La quatrième paire d'images.

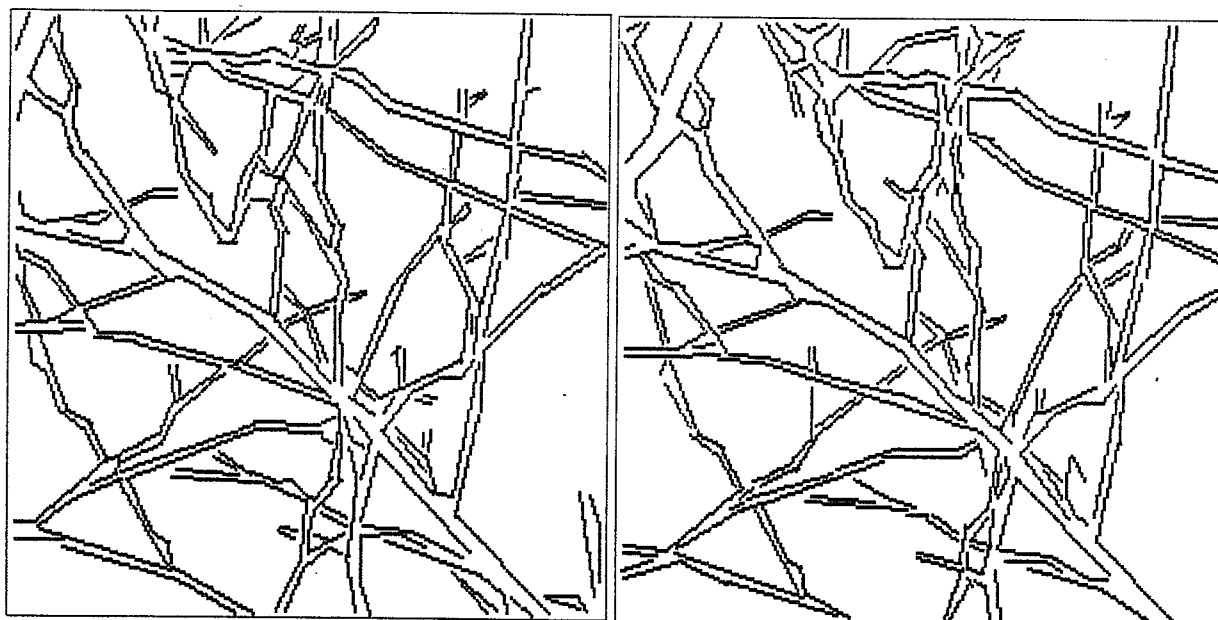


Figure 3.39: Les segments extraits des images d'Arbre.

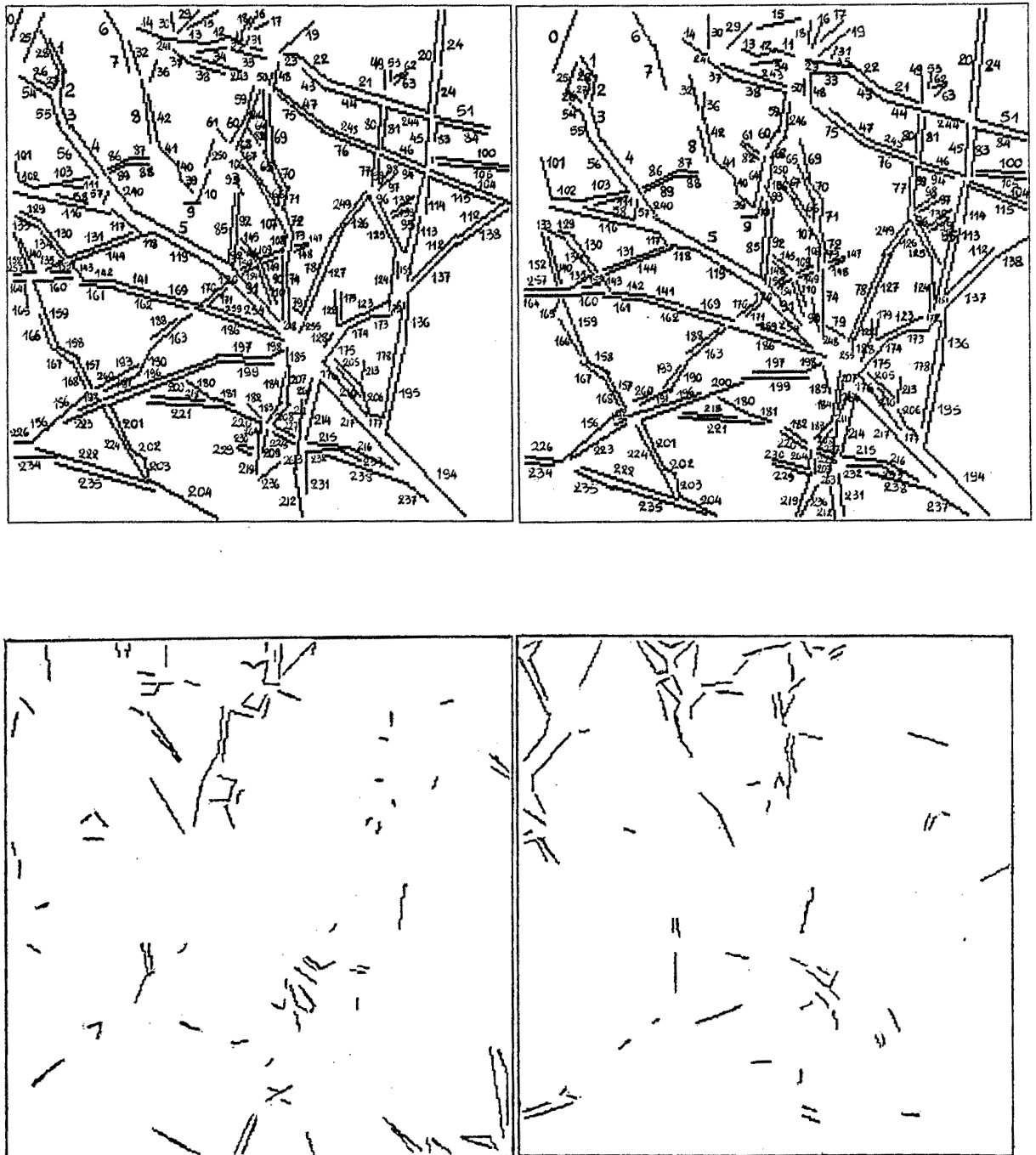


Figure 3.40: En haut : le résultat de mise en correspondance pour les images d'Arbre.
En bas : les segments non appariés.

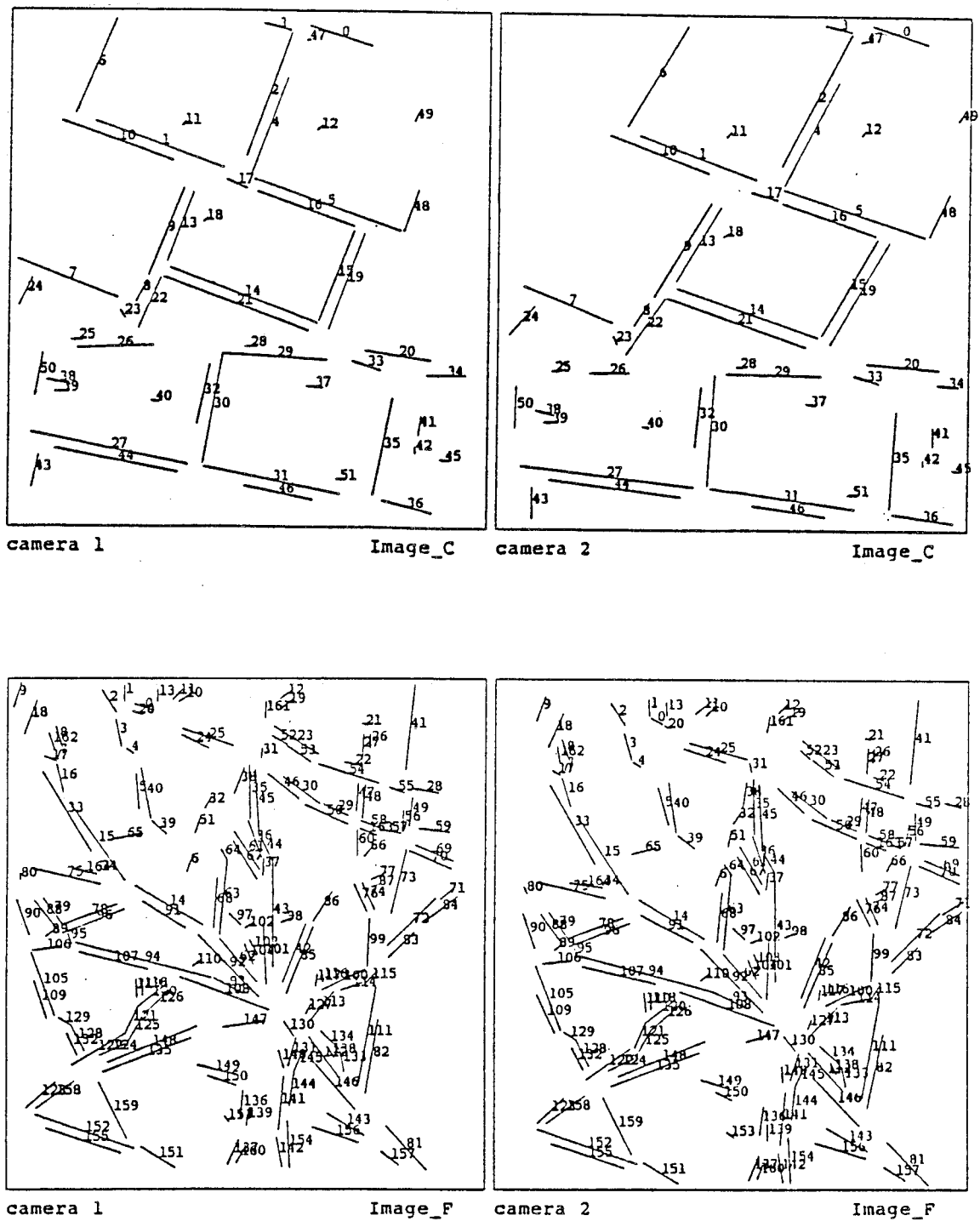


Figure 3.41: Le résultat de mise en correspondance pour les images de Toit et d'Arbre obtenu par l'algorithme de stéréovision binoculaire de N. Ayache [2]-pages 122,124.

En haut : le résultat de l'appariement pour les Images de Toit.

En bas : le résultat de l'appariement pour les Images d'Arbre.

Chapitre 4

La reconstruction de surfaces

Le but d'un système de vision stéréo doit être d'une part de retrouver les caractéristiques 3-D de la scène et d'autre part de reconstruire les surfaces de la scène qui contiennent ces caractéristiques afin de déterminer avec précision sa géométrie.

La sortie de notre algorithme de mise en correspondance est une liste de segments appariés entre les images gauche et droite. Disposant maintenant de cette liste, l'étape suivante consiste à construire un "premier" modèle tridimensionnel de la scène qui est constitué d'un ensemble de segments 3-D obtenus par l'extraction de la profondeur des segments 2-D appariés.

4.1 Extraction de la profondeur

Avant d'étudier comment il est possible de reconstruire dans l'espace un couple de segments homologues, examinons d'abord comment on peut extraire la profondeur d'un couple de points appariés.

4.1.1 Extraction de la profondeur pour un couple de points homologues

L'extraction de la profondeur de deux points appariés est effectuée à l'aide d'une méthode géométrique appelée *triangulation*. La géométrie de la triangulation stéréo est illustrée schématiquement sur la Figure 4.1 pour le cas de points 2-D projetés dans des images monodimensionnelles. Connaissant les paramètres optiques des deux caméras, leur relation spatiale et le fait que le point p de l'image gauche corres-

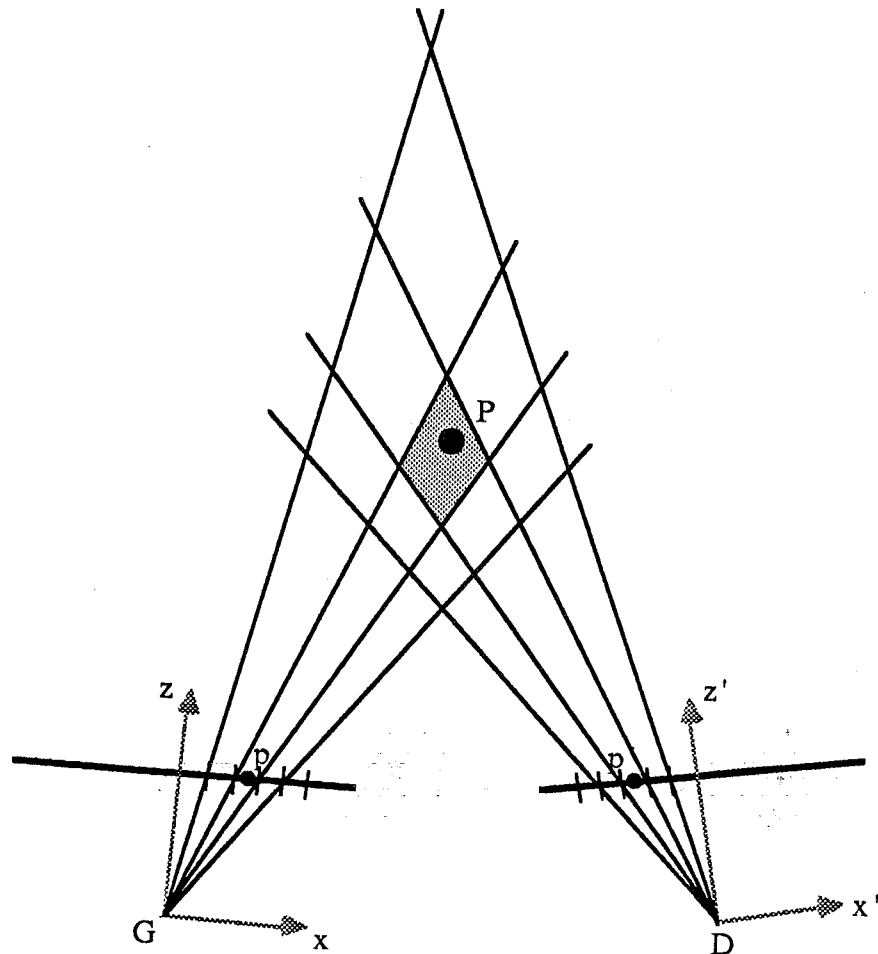


Figure 4.1: Incertitude sur la position du point P résultant du processus de triangulation.

ponde au point p' de l'image droite, le point physique P se localisera à l'intersection des deux demi-droites Gp et Dp' (triangulation).

Regardons comment ce processus de triangulation peut être appliqué à notre système stéréo. Considérons comme repère de référence le repère associé à la caméra gauche. Soient (X, Y, Z) les coordonnées du point physique P , que nous désirons retrouver, par rapport à ce repère. Essayons d'exprimer les coordonnées de P en fonction des coordonnées de ses projections p et p' dans les deux images.

L'équation (3.5) relie les coordonnées (X, Y, Z) de P avec les coordonnées (x, y) de p :

$$X - xZ = 0 \quad (4.21)$$

$$Y - yZ = 0 \quad (4.22)$$

Afin de relier les coordonnées de P avec les coordonnées (x', y') de p' , nous utilisons l'équation (3.6) en remplaçant X' , Y' et Z' par leurs expressions données

par l'équation (3.4). Nous obtenons ainsi :

$$(x' \sin \phi + \cos \phi)X + (-x' \cos \phi + \sin \phi)Z = -b_x + x'b_z \quad (4.23)$$

$$(y' \sin \phi)X + Y - (y' \cos \phi)Z = y'b_z \quad (4.24)$$

Les quatre équations (4.21) à (4.24) expriment les coordonnées de P en fonction des coordonnées de ses projections p et p' dans les deux images et des paramètres connus ϕ , b_x et b_z du capteur. A cause du fait que les images sont imprécises, les deux rayons Gp et Dp' (cf. Fig 4.1) ne se rencontrent pas exactement. Ainsi les trois coordonnées X , Y et Z sont obtenues en résolvant le système d'équations linéaires (4.21) à (4.24) aux moindres carrés. De cette manière on obtient une position estimée de P dans l'espace par rapport au repère de la caméra gauche.

Des équations (4.21) à (4.24), on observe que la position spatiale de P dépend :

- de la position dans les images des deux projections p et p' ;
- des paramètres du capteur stéréo.

A cause de l'échantillonnage et de la quantification des données de départ et du bruit inhérent à tout système de formation d'images, les points homologues p et p' sont localisés dans les images avec une certaine erreur. De plus, les équations (4.21) à (4.24) supposent que les paramètres du capteur sont connus très précisément, c'est-à-dire que le capteur est parfaitement calibré, ce qui malheureusement n'est jamais vrai. Ces incertitudes sur la localisation des points images et sur les paramètres du capteur vont engendrer une erreur sur la position estimée de P .

Les erreurs sur la position de P provoquées par la quantification des images sont illustrées sur la Figure 4.1, où les positions des points p et p' sont connues à un pixel près. La position estimée de P peut se situer partout dans la zone hachurée qui entoure la position réelle de P .

Afin de modéliser la zone d'incertitude associée au point P , nous avons utilisé la théorie du Filtre de Kalman étendu (FKE). En ce qui concerne la présentation de cette théorie, ainsi que de ses nombreuses applications à des problèmes de vision 3-D, nous renvoyons le lecteur intéressé aux travaux d'Ayache et Faugeras [3], [4] et de Lustman [52]. Ici nous nous contentons d'appliquer cet outil mathématique à notre cas précis.

Voici une présentation impropre du FKE : on désire estimer un vecteur de paramètres inconnus (l'état d'un système asservi), dont on connaît une estimée initiale et une matrice de covariance associée à cette estimée et représentant son incertitude. Le vecteur de paramètres inconnus est lié à des mesures bruitées (les observations de l'état du système) par un ensemble d'équations non linéaires (décrivant l'évolution de l'état du système). L'utilisation du FKE permet d'améliorer récursivement l'estimée initiale et la matrice de covariance qui lui est associée chaque fois qu'une nouvelle mesure arrive.

Dans notre cas, le vecteur de paramètres inconnus est la position tridimensionnelle du point physique P dont on connaît par la triangulation une position estimée initiale ; nous lui associons une première matrice de covariance assez grossière

représentant le fait que cette première estimée est connue avec une grande imprécision. Nous disposons alors de deux mesures qui sont les deux projections du point P dans les plans images et des équations (4.21) à (4.24) qui lient ces mesures avec les paramètres inconnus. (Observons que les paramètres inconnus X , Y et Z sont liés par ces quatre équations aux mesures x, y et x', y' de façon non linéaire – par contre le système est linéaire en X , Y et Z). Dans l'Annexe C, nous présentons l'application des équations du FKE afin d'extraire les paramètres recherchés de P : nous avons négligé les erreurs dues au calibrage et nous avons supposé que les deux mesures p et p' sont localisées avec une précision d'un pixel.

Néanmoins, compte tenu du fait qu'on ne dispose finalement que de deux mesures, il est évident que l'utilisation du FKE ne peut pas fournir une estimation sur la position du point P meilleure que celle calculée par la triangulation. Par contre, la matrice de covariance associée à la position de P devient plus précise par rapport à la matrice de covariance de départ.

Cette matrice de covariance peut être utilisée pour définir une région ellipsoïdale autour de la position estimée de P représentant son incertitude. Les longueurs des axes de cet ellipsoïde sont déterminées par les racines carrées des trois valeurs propres de la matrice de covariance, tandis que la direction de ces axes est donnée par les trois vecteurs propres correspondant aux trois valeurs propres, respectivement.

4.1.2 Extraction de la profondeur pour un couple de segments homologues

A partir d'un couple de segments appariés (g, d) , il est possible de reconstruire un segment S dans l'espace. La Figure 4.2 montre comment cette reconstruction est effectuée. Chacun des segments g et d détermine un plan (appelé *plan d'interprétation*) avec le centre optique de la caméra correspondante. Le segment S de l'espace doit être contenu dans chacun de ces deux plans d'interprétation et par conséquent S est déterminé par l'intersection de ces deux plans.

La géométrie de reconstruction illustrée sur cette figure suppose que les extrémités des segments g et d se correspondent parfaitement. En réalité, ceci n'est pas vrai et afin de reconstruire correctement le segment S dans l'espace, on doit d'abord déterminer le long des deux segments homologues des parties qui se correspondent réellement. Nous allons appeler ces parties *zones de correspondance maximum* de deux segments homologues.

La détermination des zones de correspondance maximum de deux segments homologues est réalisée de la manière suivante : sur chaque segment, une zone est déterminée par l'intersection du segment avec les deux lignes épipolaires associées à son segment correspondant de l'autre image. On trouve ensuite sur les deux segments homologues les parties maximales autorisées pouvant se correspondre parfaitement. La Figure 4.3 présente les zones de correspondance maximum associées à deux segments appariés g et d . A_1, A_2 sont les deux extrémités de g et B_1, B_2 sont les deux extrémités de d . L'intersection de g avec les deux lignes épipolaires gauches associées à d détermine deux points B'_1, B'_2 . De façon similaire, dans l'image droite

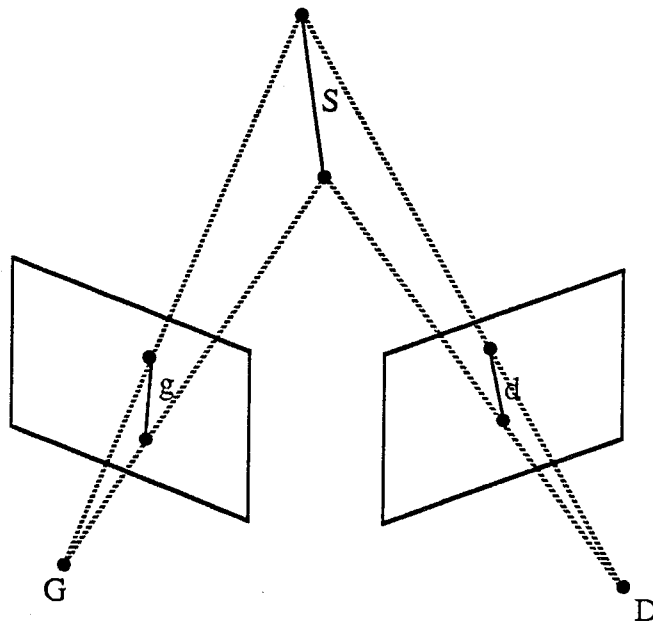


Figure 4.2: La géométrie de reconstruction d'un segment S de l'espace à partir d'un couple de segments homologues (g, d) . S est déterminé par l'intersection des deux plans d'interprétation associés à g et à d , respectivement.

sont déterminés les points A'_1, A'_2 . Les zones de correspondance maximum des deux segments g et d sont les parties $A_1B'_2$ et A'_1B_2 , respectivement dont les extrémités sont en correspondance.

Examinons maintenant le cas où la direction des segments g et d est presque parallèle à la direction des lignes épipolaires. D'une part, il n'est pas possible de calculer de points homologues sur ce couple de segments et par conséquent il est impossible de déterminer des zones de correspondance maximum (cf. Fig. 3.4, page 68). D'autre part, les deux plans d'interprétation associés aux deux segments sont confondus et donc l'intersection des deux plans est le plan tout entier. Il en découle que la reconstruction de ce genre de segments est impossible et la seule information éventuellement exploitable est le fait de leur appartenance à un plan d'interprétation.

4.1.3 Reconstruction des segments 3-D en pratique

Commençant avec un couple de segments appariés, les étapes suivantes permettent de reconstruire un segment 3-D et de modéliser l'incertitude qui lui est associée.

1. On détermine sur les deux segments appariés les zones de correspondance maximum.
2. On reconstruit par triangulation les extrémités correspondantes de ces zones. De cette manière on obtient deux points dans l'espace qui représentent les deux extrémités du segment 3-D.

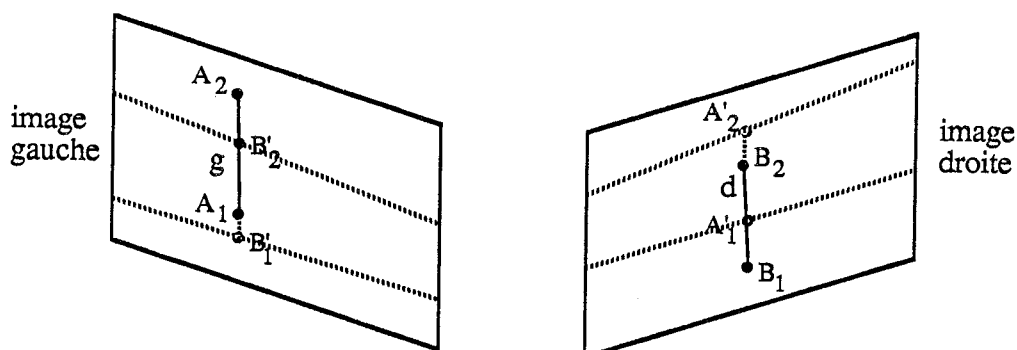


Figure 4.3: Les zones de correspondance maximum associées à deux segments homologues g , d : la partie $A_1B'_2$ du segment $g(A_1A_2)$ correspond parfaitement à la partie A'_1B_2 du segment $d(B_1B_2)$.

3. A chaque point 3-D ainsi reconstruit, on lui associe à l'aide du FKE une matrice de covariance représentant l'incertitude sur sa position.

Notons ici que par rapport à la modélisation des incertitudes concernant les points images (§ 4.1.1), il y a une différence essentielle : les extrémités des zones de correspondance maximum sont localisées avec une précision au dessous du pixel. Ceci est dû au fait que les segments, qui sont terminés par ces points, approximent aux moindres carrés un ensemble de points de contour, par conséquent ces segments sont détectés avec une précision au dessous du pixel. Qui plus est, au § 4.1.1 nous avons supposé que le modèle du bruit associé aux points images consistait à considérer que ces points sont connus à un pixel près et ceci dans les deux directions de l'image. Ici, nous avons utilisé comme erreur, l'erreur quadratique obtenue lors de l'approximation polygonale des chaînes de contour.

Il en découle un premier modèle de la scène du type *fil de fer* constitué d'un ensemble de segments 3-D où chaque segment est représenté par :

- les coordonnées 3-D de ses deux extrémités ;
- deux matrices de covariance représentant l'incertitude sur la position de ses extrémités.

4.1.4 Résultats expérimentaux

Nous exposons ici le résultat de la reconstruction des segments 3-D pour les images représentant des scènes d'intérieur montrées sur les Figures 3.24 et 3.31. Pour ces images nous disposons de données de calibrage du banc stéréo relativement précises. A la liste des segments appariés par l'algorithme de mise en correspondance, nous appliquons le processus de reconstruction que nous avons décrit. Tous les résultats que nous allons présenter restent quand même à un niveau assez "subjectif", puisque

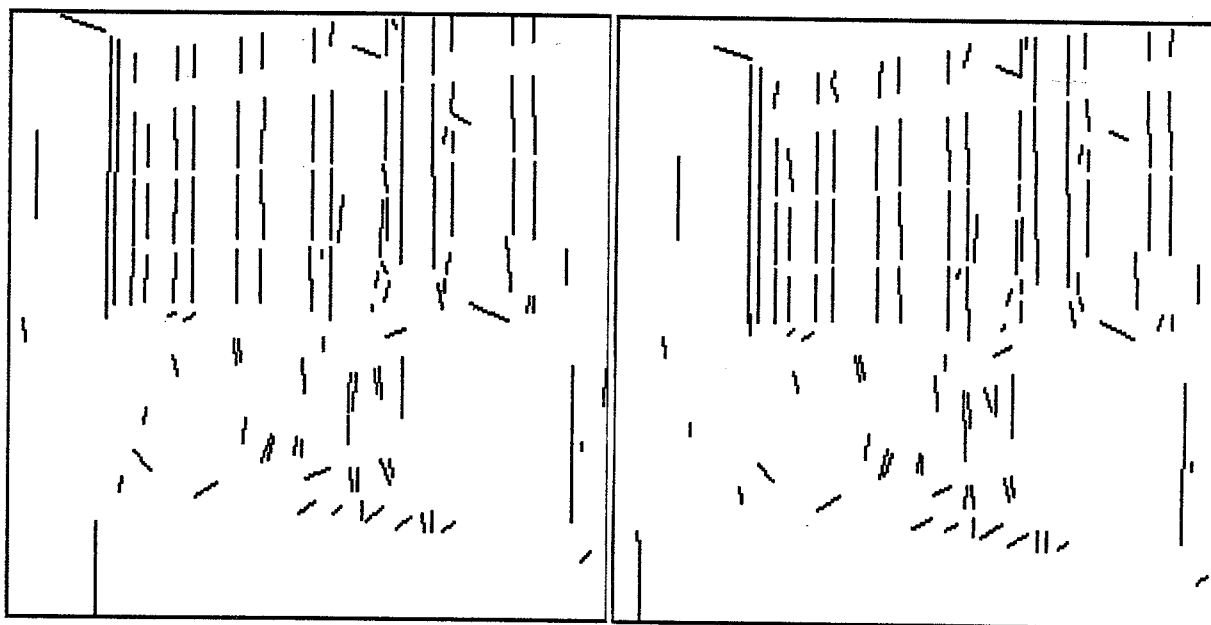


Figure 4.4: Les zones de correspondance maximum des segments appariés pour les images de Bureau 1. Tous les segments horizontaux appariés n'ont pas été reconstruits car leur direction est parallèle à la direction des lignes épipolaires.

n'ayant pas à notre disposition aucun modèle géométrique des scènes traitées connu à priori, nous n'avons pas pu procéder à une vérification de ces résultats.

Dans le premier exemple (Bureau 1), 289 segments de l'image gauche ont été appariés avec 298 segments de l'image droite (cf. Fig. 3.28, page 106). La Figure 4.4 montre les zones de correspondance maximum des segments des images gauche et droite appariés. On constate qu'une partie considérable des segments appariés n'est pas directement utilisable pour la reconstruction, puisque beaucoup de segments ont la direction des lignes épipolaires. Ainsi, on ne reconstruit finalement que 138 segments 3-D.

Il n'est pas facile de visualiser les segments 3-D, ni les incertitudes qui leur sont associées. Afin de donner une idée sur la variation de la profondeur de la scène, nous avons considéré une caméra fictive se situant au "plafond" de la scène et ayant un modèle identique à celui de la caméra gauche, nous avons ensuite exprimé toutes les mesures 3-D par rapport au repère associé à cette caméra et nous avons effectué une projection perspective des segments 3-D sur le plan image de cette caméra. La Figure 4.5-gauche présente le résultat de cette projection, où pour mieux comprendre, nous avons ajouté manuellement quelques informations sémantiques concernant cette scène (en pointillées, deux bureaux et deux murs). Nous montrons également la position de la caméra gauche (G) et la position du nouveau repère de visualisation (R). Sur cette figure, les segments 5, 11 et 19 proviennent en fait de l'appariement de segments images correspondant à des reflets. La Figure 4.5-droite présente l'incertitude en profondeur associée à trois segments de la scène (les plus éloignés par rapport à la caméra gauche). Cette incertitude est de l'ordre de 50 cm

RECONSTRUCTION DES SEGMENTS 3-D	Bureau 1	Bureau 2
No. segments 3-D reconstruits	138	112
variation en x de la scène (cm)	-240.9 / 133.6	-245.6 / 150.8
variation de l'incertitude en x (cm)	0.15 / 0.91	0.18 / 0.98
variation en y de la scène (cm)	-225 / 96.7	-140.7 / 93.6
variation de l'incertitude en y (cm)	0.13 / 0.88	0.13 / 0.76
variation en z de la scène (cm)	143.2 / 887.2	207.6 / 856.7
variation de l'incertitude en z (cm)	1.9 / 66.9	1.61 / 50

Table 4.1: Résultats de la reconstruction des segments 3-D pour les scènes de Bureau 1 et de Bureau 2.

environ sur 7 mètres de profondeur.

Pour cette scène, la variation de l'incertitude en z associée aux extrémités des segments 3-D en fonction de la variation de la profondeur de ces extrémités est montrée sur la Figure 4.6 : l'incertitude en z s'accroît très rapidement quand la profondeur augmente. Quant à la variation de l'incertitude en fonction des variations en x et en y, elle est très faible (entre 0.15 et 0.91 cm pour les variations en x et entre 0.13 et 0.88 cm pour les variations en y).

Dans le deuxième exemple (Bureau 2), 141 segments 3-D ont été reconstruits à partir de 197 segments de l'image gauche appariés avec 199 segments de l'image droite. Les zones de correspondance maximum des segments homologues sont illustrées sur la Figure 4.7. (Comparez ces zones de correspondance maximum avec l'ensemble de segments appariés et montrés sur la Figure 3.33, page 108).

La Figure 4.8-gauche montre une vue aérienne de la scène observée par une caméra fictive considérée être située au "plafond" de la scène. Sur cette figure on constate que la plupart des segments 3-D appartiennent aux armoires et aux murs du fond de la scène. (Les segments 0, 1, 5, 8 et 9 qui créent des confusions appartiennent en fait au plafond de la scène). On s'aperçoit également d'une erreur de mise en correspondance (segment 19). La Figure 4.8-droite présente l'incertitude en profondeur associée aux extrémités de trois segments de la scène.

Enfin, la Figure 4.9 présente la variation de l'incertitude en profondeur associée aux extrémités des segments 3-D en fonction de la variation de la profondeur de ces extrémités. Sur cette figure il y a un gros nuage de points dont l'incertitude en profondeur est à peu près égale (il s'agit en réalité de segments appartenant aux armoires et aux murs du fond de la scène). On s'aperçoit également qu'il y a quelques segments ayant la même profondeur que les segments précédents mais dont l'incertitude en profondeur est beaucoup plus grande ; ceci est dû au fait que les segments images, dont ces segments 3-D sont construits, ont été localisés avec une erreur beaucoup plus importante.

Le Tableau 4.1 résume tous les résultats présentés ci-dessus.

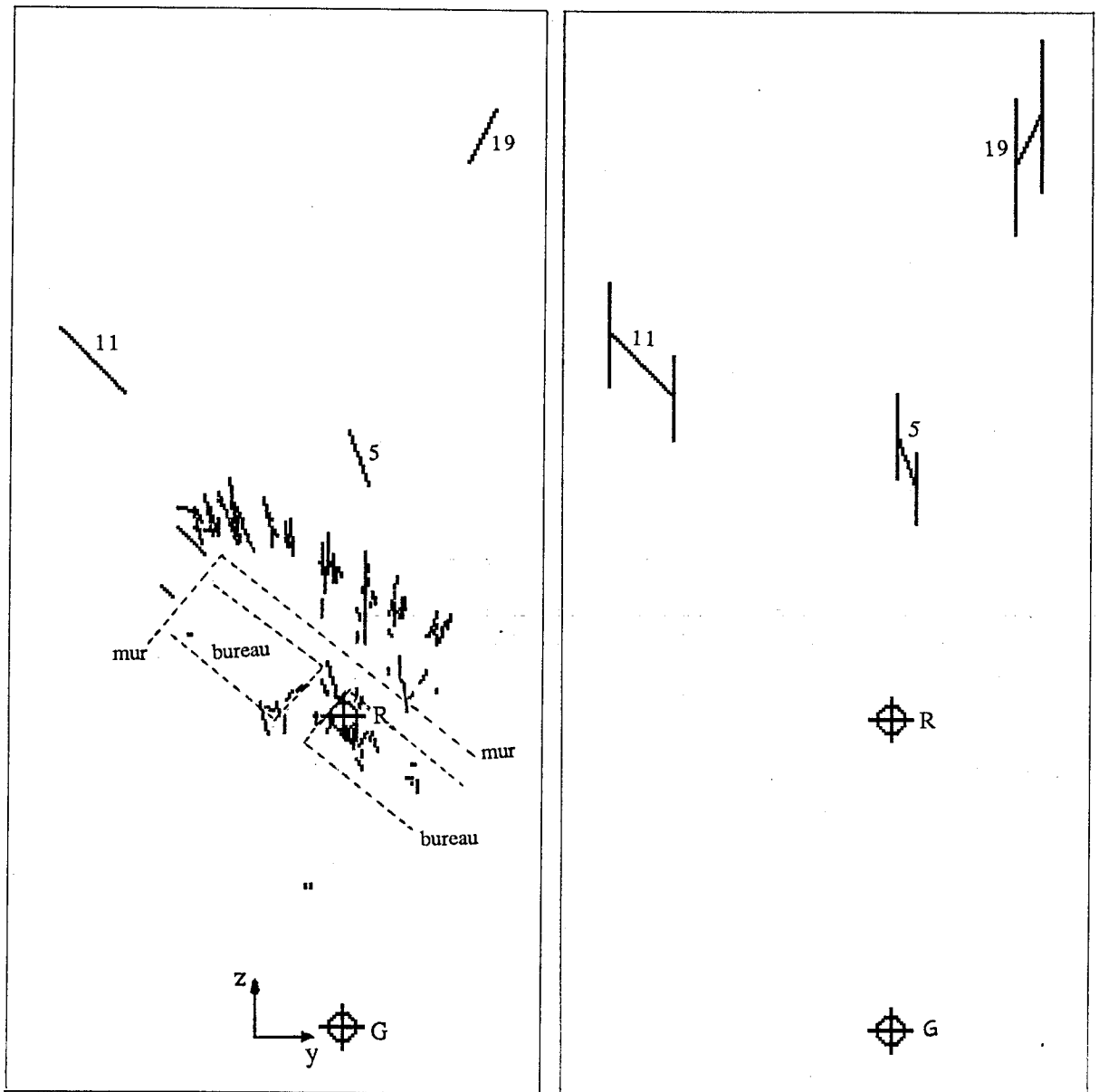


Figure 4.5: A gauche : les segments 3-D pour les images de Bureau 1 vus par une caméra fictive (R) considérée être située au "plafond" de la scène. A droite : l'incertitude en profondeur associée à trois segments 3-D, les plus éloignés par rapport à la caméra gauche (G).

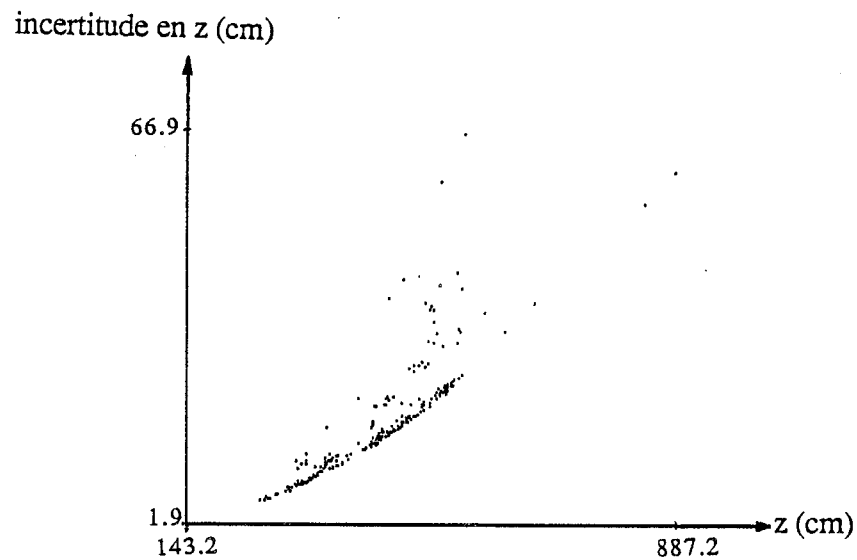


Figure 4.6: L'incertitude en profondeur en fonction de la variation de la profondeur pour les extrémités des segments 3-D (Bureau 1).

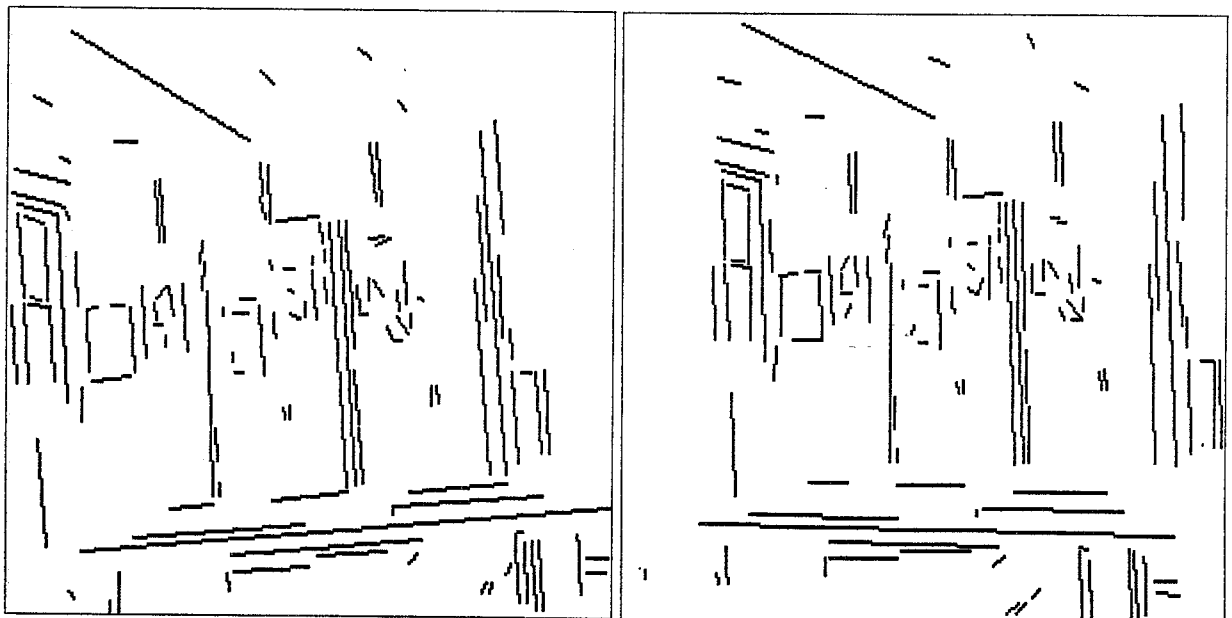


Figure 4.7: Les zones de correspondance maximum des segments appariés pour les images de Bureau 2. Notons ici aussi l'absence de tous les segments parallèles aux lignes épipolaires.

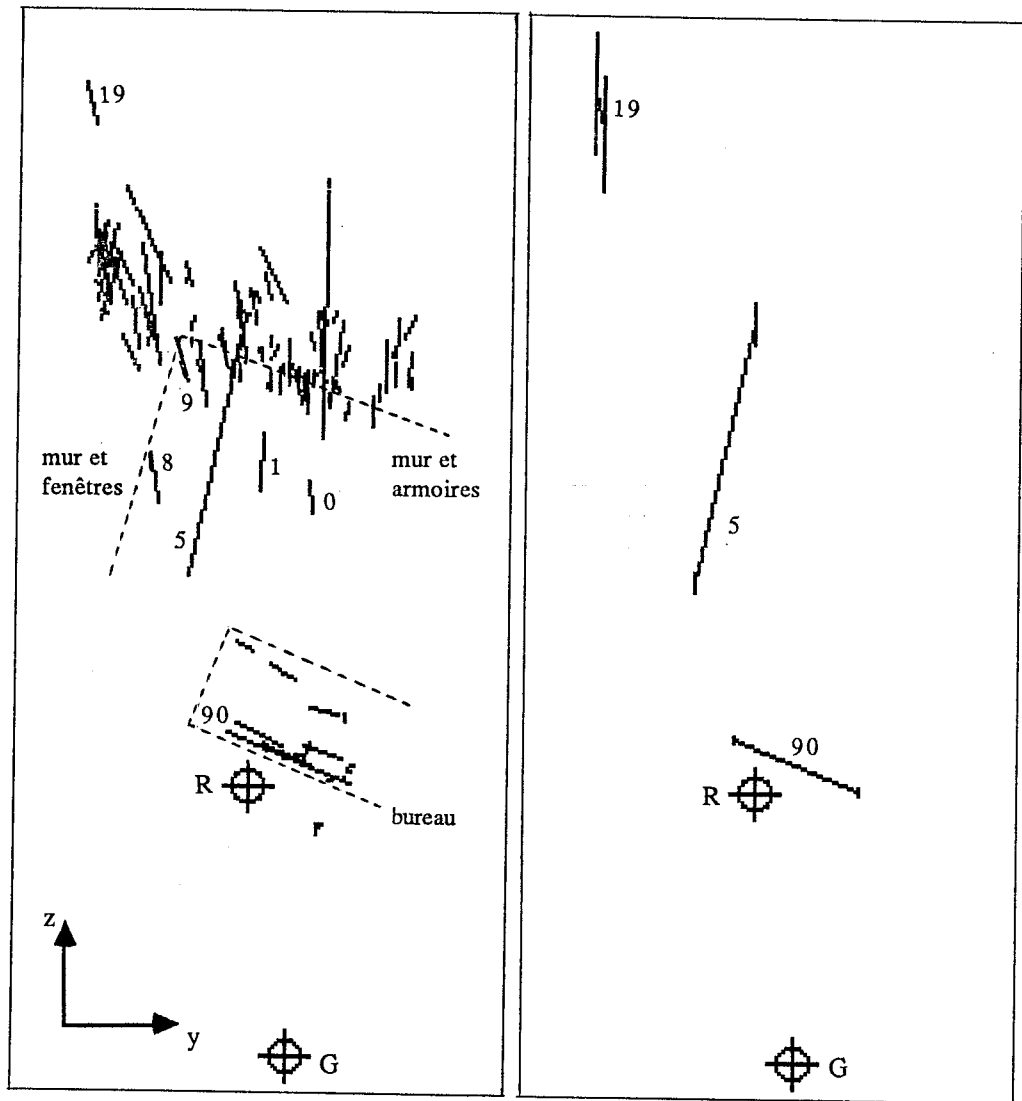


Figure 4.8: A gauche : vue aérienne des segments 3-D pour les images de Bureau 2. A droite : l'incertitude en profondeur associée à trois segments 3-D.

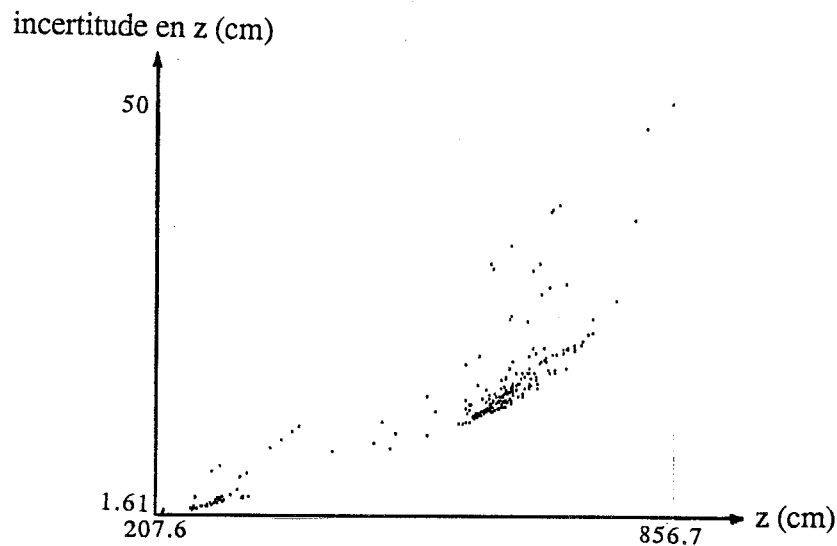


Figure 4.9: L'incertitude en profondeur en fonction de la variation de la profondeur pour les extrémités des segments 3-D (Bureau 2).

4.2 Reconstruction des surfaces

Cette section est une présentation de quelques idées que nous avons quant à la reconstruction de surfaces plutôt qu'une description d'un système complet de reconstruction. Ces idées vont être présentées en conjonction avec quelques exemples et quelques résultats expérimentaux préliminaires illustrant la validité de la méthode envisagée.

Notre but est de pouvoir retrouver des éléments de surfaces de scène qui contiennent les segments 3-D que nous avons construits. Pour diminuer la complexité du problème, nous proposons de décrire la scène en l'approximant en termes d'un ensemble de facettes planes et afin d'arriver à une telle description, nous suggérons de combiner à la fois l'information tridimensionnelle portée par les segments 3-D et l'information monoculaire contenue dans les deux images stéréo.

Par rapport aux systèmes de reconstruction stéréo existants, nous proposons ici d'exploiter en plus l'information provenant du groupement des caractéristiques de l'image gauche et/ou de l'image droite et d'intégrer cette information au processus de reconstruction. La présence de structures proéminentes dans les images reflète la géométrie et la topologie de la scène traitée et l'utilisation "intelligente" de ces structures (détectées de la manière décrite au Chapitre 2) peut servir énormément à la création et à la propagation de bonnes hypothèses concernant les surfaces de scène.

4.2.1 Le rôle du groupement à l'extraction de surfaces planes

Supposons que nous disposions d'un ensemble de segments formant une structure locale dans une image ; utilisant la liste des segments appariés dans les images gauche et droite et la description monoculaire associée à l'autre image, nous pouvons rapidement vérifier si cette structure a été mise en correspondance avec une structure

similaire de l'autre image. Si c'est le cas, faisons l'hypothèse que l'ensemble de segments composant les deux structures correspondantes appartient à une surface plane de la scène. Afin de trouver cette surface nous utilisons les segments 3-D résultant de la reconstruction des segments composant les deux structures homologues : les extrémités de ces segments forment un "nuage" de points dans l'espace ; il est possible maintenant de trouver la surface plane qui passe au mieux (au sens des moindres carrés) par cet ensemble de points 3-D grâce à une technique de calcul d'un plan par la méthode des composantes principales (voir [8], [68]). Afin de confirmer l'hypothèse de surface plane, voici quelques critères dont nous devons tenir compte :

Contraintes d'acceptation d'une surface plane initiale

- L'erreur quadratique associée à l'extraction de ce plan doit être petite ;
- les segments 3-D ne doivent pas couper le plan en question, c'est-à-dire ils doivent être parallèles au plan ;
- la distance des segments 3-D au plan doit être petite ;
- le plan trouvé doit couper la région d'incertitude (ellipsoïde) associée aux extrémités des segments 3-D.

Si un parmi ces quatre critères n'est pas satisfait, alors l'hypothèse de surface plane est rejetée. Par contre, si tous les critères sont satisfaits, on infère qu'une surface plane a été détectée. Ce plan est représenté par son vecteur normal et par sa distance à l'origine (le centre optique de la caméra gauche). Lors de la détection d'un plan, les segments 3-D qui ont donné "naissance" à ce plan sont projetés sur le plan et leur position 3-D est ainsi améliorée et mise à jour.

Etudions par la suite comment il est possible de propager l'hypothèse de départ afin d'inclure d'autres segments 3-D au plan déjà formé. Voici deux remarques importantes quant à la démarche à suivre :

- les segments 3-D candidats à être inclus dans ce plan ont une grande probabilité d'être spatialement proches de ceux déjà inclus.
- deux segments spatialement proches restent voisins lorsqu'ils se projettent dans une image.

Utilisant donc la description monoculaire de l'image gauche et/ou de l'image droite, nous pouvons nous munir d'une liste de segments images voisins des projections images des segments 3-D qui appartiennent déjà à ce plan. Cette liste des segments voisins peut se scinder à l'aide de la liste des segments appariés et de la liste des segments reconstruits en trois nouvelles listes :

1. une liste composée de segments qui ont été appariés et ensuite reconstruits dans l'espace ;

2. une liste composée de segments appariés mais non reconstruits à cause de limitations inhérentes à la géométrie du capteur (segments parallèles aux lignes épipolaires) ;
3. une liste composée de segments non appariés.

Ne tenant pas compte de la troisième liste, voici comment on peut envisager d'inclure éventuellement au plan déjà extrait les segments appartenant aux deux premières listes.

En ce qui concerne la première liste, cette liste peut rapidement se traduire en une liste de segments 3-D représentant les candidats à être inclus directement dans le plan. Pour chaque segment 3-D appartenant à cette liste, la technique suivante peut être utilisée pour rajouter (éventuellement) incrementalement un nouveau segment dans le plan :

Contraintes d'ajout d'un segment dans un plan

- le segment 3-D doit être presque parallèle au plan ;
- la distance des deux extrémités du segment au plan doit être petite ;
- le plan doit couper la région d'incertitude associée à chaque extrémité du segment.

Lorsqu'un segment satisfait toutes ces contraintes, alors il est rajouté au plan et d'une part on met à jour ses paramètres en le projetant sur ce plan et d'autre part on remet à jour les paramètres du plan lui même.

Quant aux segments appartenant à la deuxième liste, une technique de prédiction/vérification peut être utilisée pour leur reconstruction. Décrivons la succinctement : soit g un segment de l'image gauche appartenant à cette liste et soit (q) le plan d'interprétation qui lui est associé. La direction de (q) est donnée par son vecteur normal \vec{n} :

$$\vec{n} = \frac{p_1 \times p_2}{\|p_1\| \|p_2\|} \quad (4.25)$$

où $p_1 = (x_1, y_1, 1)$ et $p_2 = (x_2, y_2, 1)$ sont les coordonnées des deux extrémités du segment g dans le repère associé à la caméra gauche (rappel : la distance focale $f = 1$). Soit (π) le plan de l'espace extrait grâce à quelques segments images reconstruits et se trouvant dans la même structure locale que g . On peut maintenant supposer avec une grande probabilité que g est l'image d'un segment S de l'espace qui appartient aussi au plan (π) . S est déterminé par l'intersection du plan (π) et du plan d'interprétation (q) associé à g . (Cf. Figure 4.10). On peut utiliser maintenant l'image droite pour vérifier si cette hypothèse est correcte ou non. Soit d le segment de l'image droite homologue au segment g . S est projeté dans l'image droite et si cette projection et le segment d se chevauchent suffisamment, alors l'hypothèse est validée, sinon elle est rejetée.

Notons que par cette technique, on pourrait aussi envisager de reconstruire des segments ou plus généralement des structures visibles dans une image mais absentes

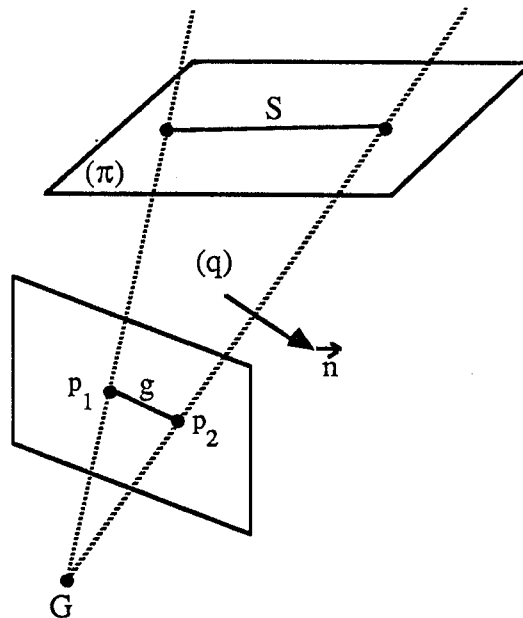


Figure 4.10: La reconstruction d'un segment g de l'image gauche sous l'hypothèse que le segment S de l'espace, dont g est l'image, appartient à un plan connu (π) . S est déterminé par l'intersection du plan d'interprétation (q) associé à g et du plan (π) .

de l'autre, sous l'hypothèse qu'elles proviennent de structures 3-D appartenant à une surface de scène déjà extraite.

Nous avons donc présenté une méthode qui fait passer une surface plane par un ensemble initial de segments, et qui rajoute ensuite de manière itérative d'autres segments dans ce plan. Deux problèmes considérables se posent :

1. comment initialiser le processus d'extraction d'une surface plane, c'est-à-dire quel est l'ensemble de segments qu'il faut utiliser pour trouver un plan initial fiable ?
2. dans le cas où le plan qu'on a supposé exister ne satisfait pas les quatre contraintes d'acceptation d'une surface plane initiale, comment faut-il partitionner l'ensemble de points 3-D (segments 3-D) ayant fourni ce faux plan pour que des plans corrects soient extraits ?

Tandis qu'au deuxième problème nous ne pouvons proposer aucune solution efficace, le premier peut être résolu de manière assez élégante. En effet, voici deux cas précis de formation d'un plan dans l'espace :

- par deux droites sécantes ;
- par deux droites parallèles.

Ces deux conditions se traduisent dans les images par :

- deux droites concourantes de l'espace, se rencontrent aussi dans les images ;
- deux droites parallèles de l'espace sont presque parallèles dans les images (à condition d'une perspective relativement faible).

En fait, nous voulons faire un choix judicieux d'un ensemble initial de segments susceptibles d'appartenir au même plan. Précédemment, nous avons supposé que cet ensemble est fourni par les deux descriptions monoculaires associées aux deux images stéréo ; en effet, suite aux deux remarques ci-dessus, voici quelques structures locales des images que l'on pourrait utiliser pour cette phase d'initialisation :

- des jonctions images à deux segments, c'est-à-dire commencer avec un angle de l'image gauche et son homologue de l'image droite. Deux conditions supplémentaires sont nécessaires dans ce cas : premièrement, ces angles doivent être formés par des segments suffisamment longs pour qu'ils ne soient pas dus au bruit ; deuxièmement, ces angles ne doivent pas provenir d'une détection de jonctions en T, puisque ce genre de jonctions indique des endroits où il y a très probablement une occlusion d'une surface par une autre.
- commencer avec une paire de segments presque parallèles, suffisamment longs et délimitant des régions homologues dans les deux images.

Les Figures 4.11-a,-b illustrent respectivement ces deux cas de formation d'une surface plane.

Ici, nous nous sommes alors servis de structures images relativement simples pour initialiser le processus d'extraction de surfaces planes. De structures plus riches que les précédentes comme par exemple de parallélogrammes ou d'autres polygones fermés pourraient s'avérer extrêmement utiles à la reconstruction : la présence de telles structures dans les deux images stéréo due à une configuration particulière de prise de vue est très peu probable ; aussi quand ces structures sont explicitement détectées, elles pourraient engendrer une hypothèse initiale très fiable quant à l'existence d'une surface plane de la scène. De plus, lorsqu'il y a des segments non appariés mais appartenant à ces structures, alors il est presque certain que ces segments sont coplanaires avec le reste de la structure et par conséquent peuvent être reconstruits de la manière illustrée sur la Figure 4.10.

4.2.2 Caractérisation d'une surface plane

Etudions maintenant comment il est possible de caractériser une surface plane qui a été extraite de la manière décrite ci-dessus.

Une surface plane est représentée par :

- ses paramètres géométriques intrinsèques (son vecteur normal et sa distance à l'origine) ;
- l'ensemble de segments 3-D lui appartenant.

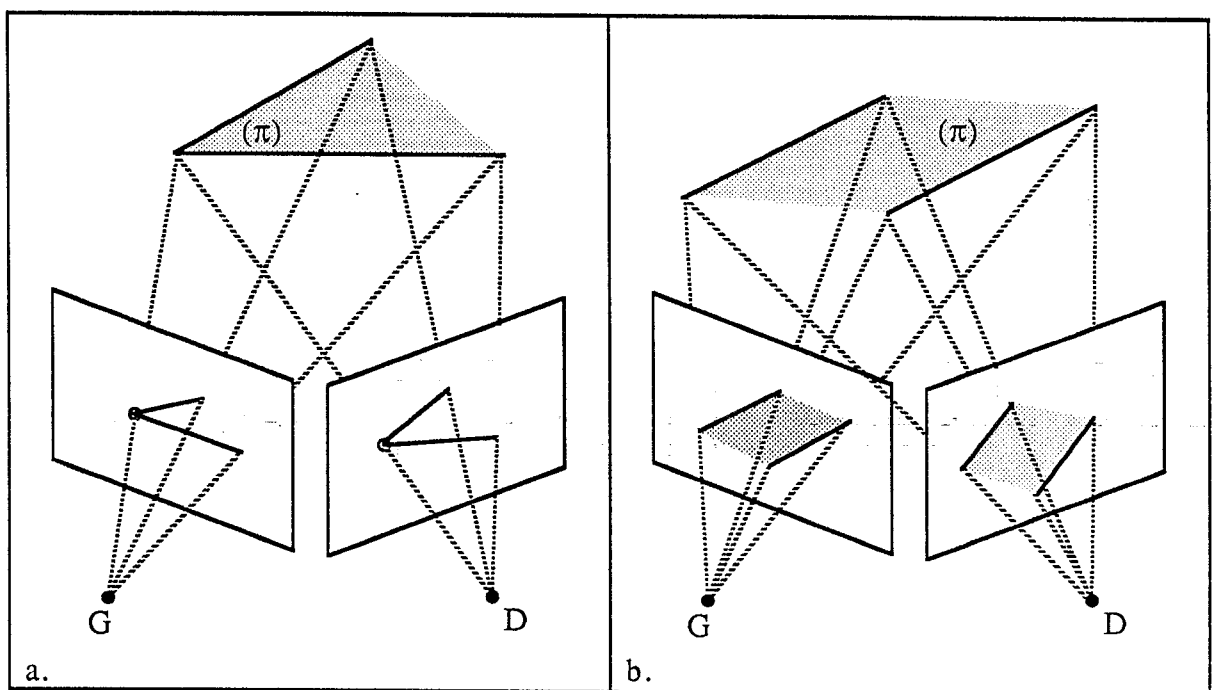


Figure 4.11: Extraction de surfaces planes de scène utilisant l'information du groupement : (a) le plan (π) de l'espace provient de la reconstruction de deux angles images homologues. (b) le plan (π) est construit à partir de deux régions images homologues (à partir de deux segments presque parallèles délimitant ces régions.)

Comme nous l'avons déjà mentionné, tous les segments 3-D ayant été trouvés appartenir à une surface plane sont ensuite projetés sur cette surface. L'enveloppe convexe des extrémités des segments ainsi projetés représente en première approche la frontière physique de la surface (plus précisément, cette frontière est déterminée par l'ensemble de segments faisant partie de l'enveloppe convexe).

En théorie, si la surface plane détectée est une surface opaque correspondant à une surface réelle de la scène, aucune partie de la scène cachée par cette surface ne doit être visible dans les images stéréo. Voici comment il est possible d'effectuer rapidement ce test de visibilité : la projection de l'enveloppe convexe de la surface sur les deux images stéréo représente grossièrement dans chaque image la zone occupée par cette surface. Si à l'intérieur de cette zone il existe un segment dont le correspondant 3-D se trouve derrière la surface (par rapport à l'origine), alors trois hypothèses sont possibles : la surface ne correspond à aucune surface réelle de la scène, la surface est transparente ou bien le segment se trouvant derrière cette surface provient d'un appariement erroné. En rejetant l'hypothèse de surface transparente et si un nombre suffisant de segments images enfreint l'hypothèse de l'opacité de la surface, alors cette surface ne correspond à aucune surface réelle de la scène et elle devrait être rejetée.

Malheureusement, dans la pratique la validation d'une surface plane détectée s'avère beaucoup moins évidente. Même en négligeant le fait qu'une surface réelle peut avoir de trous à son intérieur, certaines des mesures 3-D, résultant de la reconstruction des segments images appariés, sont trop bruitées pour les confier et rejeter une surface détectée, lorsque ces mesures enfreignent l'hypothèse de l'opacité de la surface.

4.2.3 Résultats expérimentaux

N'ayant pas complètement achevé la reconstruction de surfaces, nous allons présenter quelques résultats expérimentaux préliminaires qui semblent être assez encourageants.

Compte tenu de la grosse incertitude que possèdent certaines mesures 3-D, (par exemple pour les images de Bureau 2, (Figure 3.31, page 107), cette incertitude est de l'ordre de 30 à 50 cm pour de mesures se trouvant entre 5 et 8 mètres de l'origine !), il nous paraît peu réaliste d'essayer de retrouver les surfaces de la scène à partir de ces mesures. A la place, nous proposons d'éliminer toutes les mesures dont l'incertitude en position est supérieure à un seuil et de **ne reconstruire finalement qu'un modèle partiel mais assez précis de la scène traitée**. Ce modèle correspond en réalité aux objets et surfaces de scène qui sont les plus proches de l'origine.

De toutes les paires d'images stéréo que nous avons traitées, les images de Bureau 2 sont en fait le mieux calibrées. Par conséquent, nous allons utiliser uniquement cette paire d'images pour illustrer nos expérimentations.

La Figure 4.12 montre la zone de la scène qu'on peut reconstruire relativement avec confiance. Il y a 68 segments 3-D dont l'incertitude en profondeur est inférieure à 15 cm. Le champ de profondeur de ces segments varie entre 2 et 6 mètres. Notons

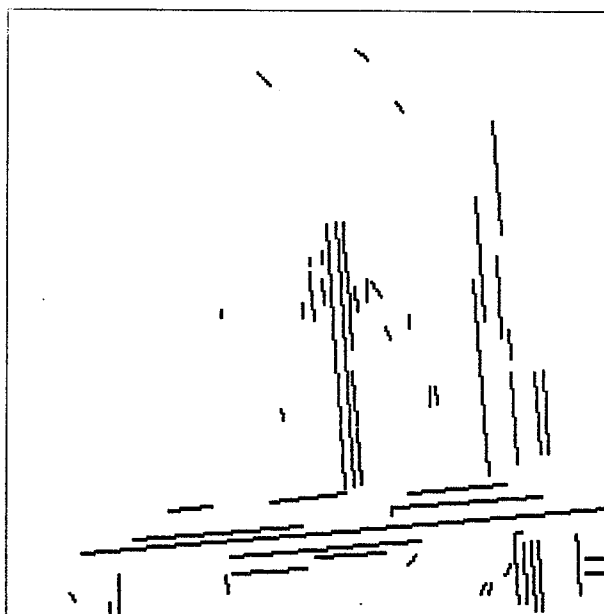


Figure 4.12: Bureau 2 : les segments 3-D utiles pour la reconstruction de surfaces.

que tous les segments se trouvant aux fenêtres et une grande partie de segments appartenant aux armoires et au mur du fond de la scène ont été éliminés car leur incertitude était largement supérieure au seuil de 15 cm.

Ces 68 segments 3-D ainsi que les deux descriptions monoculaires des deux images constituent maintenant l'entrée au processus de reconstruction de surfaces planes. La Figure 4.13 présente trois surfaces planes que l'on a pu reconstruire pour cette scène ; il s'agit en fait de trois plans qui contiennent le plus grand nombre de segments 3-D.

Le premier plan (correspondant au bureau de la scène) contient 12 segments 3-D et il a été extrait en commençant par les segments parallèles 89 et 90. Les segments 3-D 86, 88, 112 et 113 proviennent en fait de segments images qu'on a pu reconstruire dans l'espace. Ces quatre segments ont été inclus dans ce plan car ils ont satisfait les trois contraintes d'ajout d'un segment 3-D dans un plan. Quant aux six autres segments (153 à 158), ils ont été reconstruits dans ce plan en utilisant le plan d'interprétation associé à leurs segments homologues d'une image et l'hypothèse de leur appartenance à ce plan a été ensuite vérifiée dans l'autre image.

Le deuxième plan correspond à une armoire de la scène et il a été initialement détecté en utilisant les segments parallèles 30 et 46. Il contient 12 autres segments dont 10 (les segments 22, 24, 37, 81, 116, 117, 118, 119, 123 et 124) proviennent de segments images reconstruits dans l'espace et les deux autres (les segments 141 et 142) proviennent de segments images non reconstruits directement ; ces segments donc ont été supposés et vérifiés appartenir à ce plan.

Quant au troisième plan, il correspond au plafond de la scène et il contient 13 segments 3-D dont seulement trois (les segments 0, 1 et 9) proviennent de segments images directement reconstruits dans l'espace.

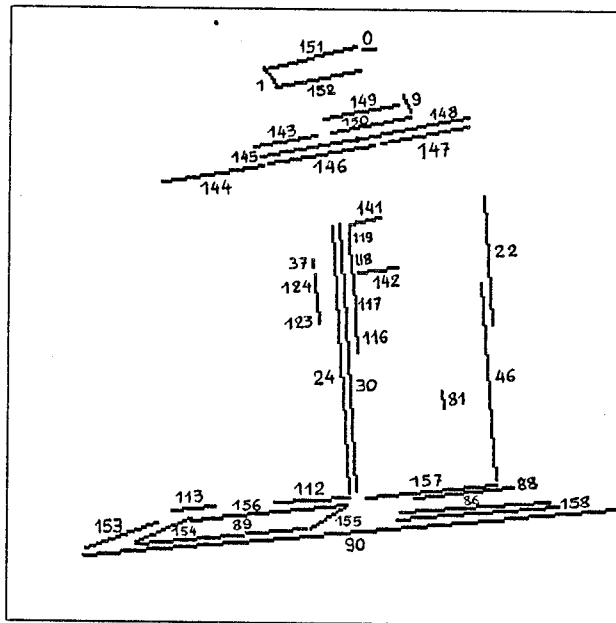


Figure 4.13: Trois plans détectés pour la scène de Bureau 2.

Le Tableau 4.2 récapitule tous les résultats présentés ci-dessus.

BUREAU 2 SURFACES PLANES	plan 1 (bureau)	plan 2 (armoie)	plan 3 (plafond)
équation du plan	$0.11x + 0.99y + 0.0z = 35.48$	$0.21x - 0.03y + 0.98z = 606.8$	$0.05x + 0.99y + 0.05z = -102.2$
No. segments 3-D	12	14	13
segments utilisés pour la détection initiale	89, 90	30, 46	0, 1
segments appartenant au plan	89, 90, 86, 88, 112, 113, 153, 154, 155, 156, 157, 158	30, 46, 22, 24, 37, 81, 116, 117 118, 119, 123, 124, 141, 142	0, 1, 9, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152

Table 4.2: Résultats d'extraction de trois surfaces planes pour la scène de Bureau 2.

Chapitre 5

Conclusions générales

Dans cette thèse, nous avons présenté un système de vision stéréoscopique. Ce système est constitué de trois grandes parties :

- le processus de groupement des caractéristiques images (extraction de structures locales 2-D) ;
- le processus de mise en correspondance des deux images (appariement de structures similaires entre les deux images) ;
- le processus de reconstruction des surfaces de la scène (reconstruction de structures appariées).

Le processus de groupement consistait à extraire des caractéristiques images (segments de droite) et à grouper ces caractéristiques en structures locales 2-D composées de segments colinéaires, de segments connexes et de segments appartenant à une même région image. Sur la base de ces groupements, nous avons ensuite décrit chaque image séparément en termes d'un graphe relationnel représentant les caractéristiques images et leurs relations structurelles.

Nous avons ainsi abouti à deux descriptions monoculaires représentant la paire stéréoscopique initiale. La mise en correspondance a été réalisée en recherchant les structures similaires 2-D entre les deux images, ce qui est équivalent à trouver un isomorphisme partiel entre les deux descriptions monoculaires des images. Pour ceci, nous avons construit un graphe (l'espace d'interprétation) et nous avons recherché les cliques

maximales les mieux adaptées pour représenter la correspondance entre les deux images.

Quant au processus de reconstruction des surfaces, après avoir reconstruit dans l'espace les structures appariées, nous avons fait l'hypothèse que ces structures faisaient partie d'éléments de surfaces planes de la scène et nous avons essayé de rajouter ensuite de manière itérative d'autres caractéristiques images appariées. De cette façon il résulte une restitution partielle de la structure tridimensionnelle de la scène.

Notre système stéréo est essentiellement basé sur la notion de groupement des caractéristiques. Les approches précédentes se sont concentrées à des caractéristiques plutôt individuelles et n'exploraient pas ou sous-utilisaient les contraintes disponibles avec les relations entre ces caractéristiques. Elles étaient donc obligées d'utiliser d'heuristiques et d'hypothèses du genre continuité du monde afin de mettre en correspondance ces caractéristiques et de retrouver les surfaces de scène qui les contiennent. Grâce au groupement, la méthode de mise en correspondance que nous avons utilisée s'est affranchie de l'hypothèse de continuité du monde qui en réalité est fréquemment violée, surtout lorsque l'on traite de scènes d'intérieur. Qui plus est, le groupement a permis l'émission de bonnes hypothèses concernant les surfaces auxquelles appartiennent les mesures de profondeur obtenues ; finalement, le groupement a également permis la reconstruction des caractéristiques images ayant la direction des lignes épipolaires qui autrement seraient éliminées.

Cependant, nous n'avons exploré ici qu'une voie élémentaire de groupement et nous avons constaté que malgré son caractère "fruste", son rôle est très important tant pour la mise en correspondance que pour la récupération des surfaces de la scène. En améliorant donc le processus de groupement et en détectant et en privilégiant des structures locales 2-D encore plus riches et complexes que celles que nous avons utilisées, il serait possible d'accroître considérablement les performances globales de notre système.

Quant à la mise en correspondance, nous avons déjà mentionné que parmi les problèmes importants à traiter et à régler est celui du découpage des images en fenêtres pour appliquer le processus de mise en correspondance locale et également le problème de la sélection parmi les cliques maximales trouvées de celle qui est la plus significative.

Enfin, concernant les améliorations à apporter au processus de reconstruction des surfaces, nous avons vu que ce processus est fortement pénalisé par le fait que l'on n'obtient qu'une carte creuse et assez bruitée de la profondeur ; par conséquent, on ne peut reconstruire avec confiance que des zones relativement petites de la scène. Afin d'améliorer donc les performances de ce processus, il faut augmenter la densité des données 3-D et diminuer les incertitudes qui leurs sont associées. Compte tenu du fait que notre système met en correspondance et reconstruit des segments de droite fournissant ainsi une carte creuse de la profondeur, on devrait le combiner avec d'autres systèmes complémentaires utilisant des caractéristiques images différentes, telles que la texture, la couleur et/ou l'ombrage. Quant à la précision des données tridimensionnelles résultant de la triangulation des caractéristiques appariées, elle

peut être améliorée en combinant les données stéréo avec des mesures provenant d'autres capteurs et/ou d'autres vues stéréo [4].

Comme nous l'avons déjà souligné, une description monoculaire riche des images pourrait justement aider à éviter le besoin d'une carte dense et précise de la profondeur. Dans ce contexte, notre système pourrait être vu comme un système complémentaire avec celui développé par Quan, Mohr et Thirion [72], qui recherche dans les images monoculaires des propriétés invariantes par projection perspective et qui les met ensuite en correspondance dans une séquence de vues stéréo créant ainsi de manière incrémentale un modèle géométrique de la scène.

Une dernière extension envisageable de notre système serait sa parallélisation à l'aide de processeurs spécialisés. Une étude récente [40] a montré qu'il y a des parties importantes de ce système qui pourraient être entièrement implantées en parallèle ; ce parallélisme pourrait intervenir tant à un niveau numérique (extraction des contours, chaînage, extraction des segments de droite, recherche des candidats potentiels des segments de droite, etc.) qu'à un niveau symbolique (parallélisation de génération des descriptions monoculaires, de génération et d'exploration du graphe de correspondance, etc.).

En conclusion, nous avons développé un système stéréo capable à reconstituer des caractéristiques individuelles 3-D, mais aussi des primitives surfaciques de scène. Ce système pourrait constituer l'entrée d'un système de vision de plus haut niveau en lui fournissant certaines des informations nécessaires afin d'effectuer de tâches complexes, telles que la reconnaissance de formes complexes, l'inspection ou le contrôle dimensionnel de pièces industrielles ou bien le guidage pour la navigation d'un robot mobile qui se déplace dans un environnement quelconque.



Annexe A

Modélisation du capteur stéréo

Tous les systèmes de calibrage d'une caméra fournissent une matrice de transformation perspective permettant le passage des coordonnées 3-D aux pixels (modélisation des paramètres internes d'une caméra). Pour la stéréo, il faut connaître en plus la position relative entre les deux caméras. Pour ceci, les systèmes de calibrage stéréo fournissent en fait la matrice de calibrage de chaque caméra par rapport à un repère absolu lié au système des caméras ; cette matrice permet le passage d'un point de l'espace, dont les coordonnées sont exprimées par rapport au repère absolu, directement aux pixels des caméras.

Soit le système stéréo de la Figure A.1, où M est le repère absolu, G est le repère lié à la caméra gauche (situé à son centre optique), L le repère lié au plan image gauche, D le repère lié à la caméra droite (situé à son centre optique) et R le repère lié au plan image droit. Notons A_{IJ} la matrice permettant le passage du système de coordonnées lié à I au système de coordonnées lié à J (ex. A_{MG} est la matrice de passage de M à G). Les matrices A_{ML} et A_{MR} sont les deux matrices de calibrage fournies par un système de calibrage stéréo. Pour notre cas, ce système est celui de G. Toscani [31]. A_{ML} et A_{MR} ont la forme de la matrice suivante :

$$A = \begin{pmatrix} a_u r_{1x} + u_0 r_{3x} & a_u r_{1y} + u_0 r_{3y} & a_u r_{1z} + u_0 r_{3z} & a_u t_x + u_0 t_z \\ a_v r_{2x} + v_0 r_{3x} & a_v r_{2y} + v_0 r_{3y} & a_v r_{2z} + v_0 r_{3z} & a_v t_y + v_0 t_z \\ r_{3x} & r_{3y} & r_{3z} & t_z \end{pmatrix} \quad (\text{A.26})$$

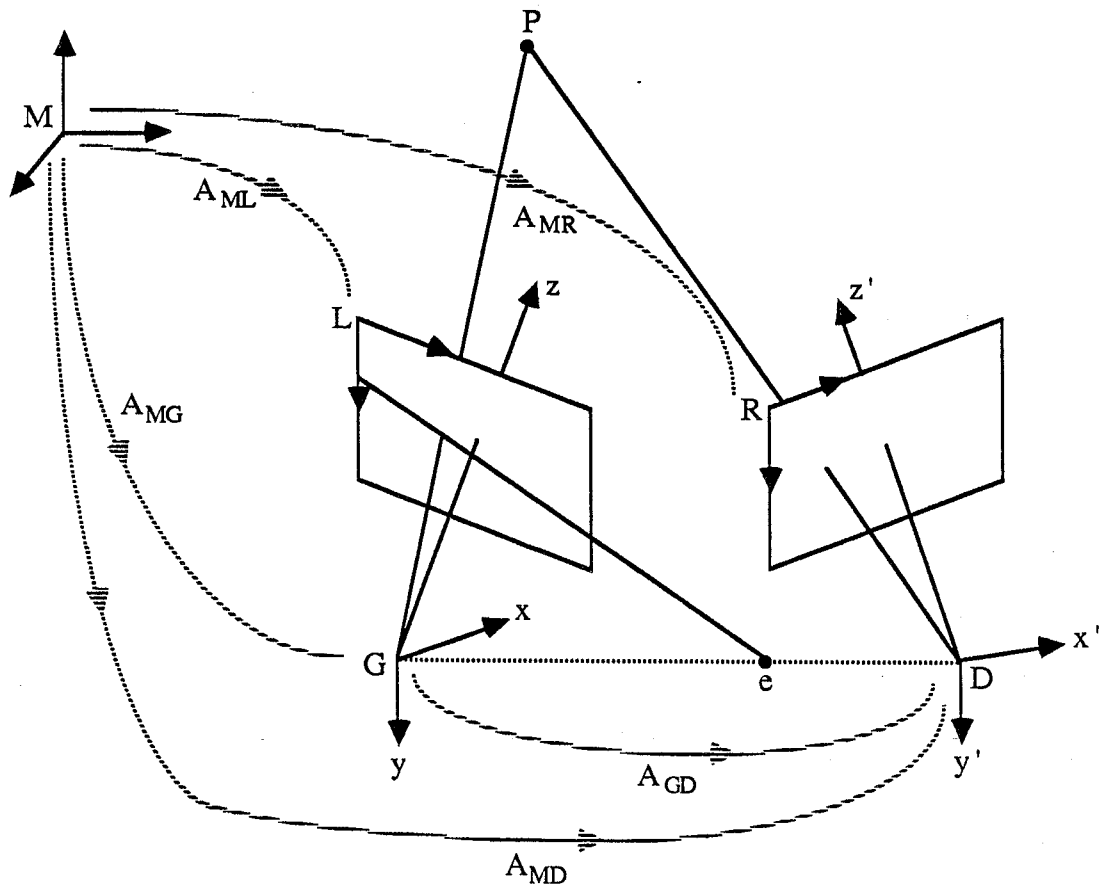


Figure A.1: Les matrices de passage pour le calibrage d'un système stéréo.

où a_u, a_v, u_0, v_0 , sont les quatre paramètres internes d'une caméra et

$$rot = \begin{pmatrix} r_{1x} & r_{1y} & r_{1z} \\ r_{2x} & r_{2y} & r_{2z} \\ r_{3x} & r_{3y} & r_{3z} \end{pmatrix}$$

et $\vec{t} = (t_x, t_y, t_z)^T$ sont respectivement la rotation et la translation qu'il faut faire subir au système de coordonnées lié à G ou à D pour arriver au système de coordonnées lié à M .

A partir d'une telle matrice il est possible de retrouver exactement les valeurs des quatre paramètres internes et de rot et \vec{t} (voir [52], pages 29-33). Ainsi, les matrices A_{MG} , A_{GL} , A_{MD} et A_{DR} peuvent être retrouvées. Connaissant ces matrices, la matrice A_{GD} de l'équation (3.3) est obtenue par :

$$A_{GD} = A_{MD} \times A_{MG}^{-1} = A_{MD} \times A_{GM}$$

Par exemple pour les images de Bureau 1 (Fig. 3.24, page 103), les deux matrices de calibrage sont les suivantes :

$$A_{ML} = \begin{pmatrix} 20.88 & 0.15 & 6.43 & -109.3 \\ 0.53 & 20.05 & 6.90 & -368.37 \\ 0.0 & 0.0 & 0.03 & 1.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix} A_{MR} = \begin{pmatrix} 18.76 & 0.22 & 7.61 & -498.85 \\ -0.18 & 18.36 & 6.69 & -307.05 \\ 0.0 & 0.0 & 0.03 & 1.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

A partir de ces deux matrices, on trouve que la matrice A_{GD} est :

$$A_{GD} = \begin{pmatrix} -0.99 & 0.0 & 0.09 & -24.97 \\ 0.0 & 1.0 & 0.0 & 0.30 \\ -0.09 & 0.0 & 0.99 & 0.78 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

d'où on infère que $\phi = 5.5^\circ$, $b_x = -24.97$ cm, $b_y = 0.3 \approx 0$ et $b_z = 0.78$ cm. (Rappelons que ϕ et b_x, b_y, b_z sont respectivement l'angle de rotation et le vecteur de translation qu'il faut faire subir au système de coordonnées associé à D pour obtenir G).

Nous allons maintenant essayer d'exprimer tous les paramètres du capteur stéréo par rapport au repère associé à la caméra gauche. Afin de faciliter les notations, dans ce qui suit nous notons avec un prime les coordonnées des points lorsqu'elles sont exprimées par rapport à D et sans prime lorsqu'elles sont exprimées par rapport à G . La matrice A_{GD} est donnée par l'équation (3.3) ; la matrice A_{DG} est donc la suivante :

$$A_{DG} = A_{GD}^{-1} = \begin{pmatrix} \cos \phi & 0 & -\sin \phi & -b_x \cos \phi + b_z \sin \phi \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & \cos \phi & -b_x \sin \phi - b_z \cos \phi \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.27})$$

Les coordonnées du centre optique D de la caméra droite exprimées dans le repère G sont :

$$\begin{pmatrix} X'_D \\ Y'_D \\ Z'_D \\ 1 \end{pmatrix} = A_{GD} \begin{pmatrix} X_D \\ Y_D \\ Z_D \\ 1 \end{pmatrix} \quad (\text{A.28})$$

Mais (X'_D, Y'_D, Z'_D) sont les coordonnées de D par rapport au repère associé à D . Par conséquent $(X'_D, Y'_D, Z'_D) = (0, 0, 0)$. De l'équation (A.28) on obtient alors :

$$\begin{pmatrix} X_D \\ Y_D \\ Z_D \\ 1 \end{pmatrix} = A_{GD}^{-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (\text{A.29})$$

d'où on trouve que :

$$\begin{pmatrix} X_D \\ Y_D \\ Z_D \end{pmatrix} = \begin{pmatrix} -b_x \cos \phi + b_z \sin \phi \\ 0 \\ -b_x \sin \phi - b_z \cos \phi \end{pmatrix} = \begin{pmatrix} U \\ 0 \\ W \end{pmatrix} \quad (\text{A.30})$$

Soit $P = (X_P, Y_P, Z_P)$ un point de l'espace. L'équation du plan épipolaire PGD est la suivante :

$$(-WY_P)x + (WX_P - UZ_P)y + (UY_P)z = 0$$

L'équation de la base stéréo GD est :

$$\frac{x}{U} = \frac{z}{W}$$

L'équation du plan image gauche est :

$$z = f$$

Trouvons les coordonnées de l'épipoles gauche $e = (X_e, Y_e, Z_e)$; e est l'intersection de la base et du plan image gauche. Ses coordonnées sont alors :

$$e = (X_e, Y_e, Z_e) = \left(\frac{U}{W}, 0, 1\right)$$

Il est facile à démontrer maintenant que, pour n'importe quel point p' de l'image droite, la ligne épipolaire gauche qui lui est associée passe par l'épipoles gauche $e = (U/W, 0, 1)$. Effectivement, considérons l'équation (3.10). En la combinant avec l'équation (A.30), on obtient :

$$y'Wx + (b_zx' - b_x)y - Uy' = 0$$

d'où, en remplaçant (x, y) par les coordonnées images de e (qui sont $(U/W, 0)$), on aboutit à :

$$y'W\frac{U}{W} + (b_zx' - b_x)0 - Uy' = 0 \Rightarrow 0 = 0 \quad \text{c.q.d.}$$

Annexe B

L'algorithme des cliques maximales

L'algorithme que nous décrivons énumère toutes les cliques maximales (i.e. de sous-graphes complets) d'un graphe. Une description plus détaillée de cet algorithme se trouve dans [16].

L'algorithme est une procédure récursive de trois paramètres : C, P et PuS. L'ensemble C représente une clique qu'il faut étendre en une clique maximale. L'ensemble P contient tous les candidats pour étendre la clique C, c'est-à-dire que tout élément de P est directement connecté avec tous les noeuds de C. Les éléments de P peuvent ne pas être connectés entre eux mais tous sont directement connectés avec tous les noeuds de C. L'ensemble PuS est une union de P et de l'ensemble contenant tous les "suspects" qui pourraient être utilisés pour étendre C mais qui arbitrairement ont été exclus. L'algorithme est le suivant :

```
RECURSIVE PROCEDURE
maxCliques (set C, P, PuS)
  BEGIN "maxCliques"
    setElements X, Y
    X_chooseOneElementFrom (PuS);
    For Each Y IN intersect (P, complement (neighbors (X))) DO
      BEGIN "chooseElement"
        P_deleteElementFromSet (Y, P);
        IF emptySet (intersect (PuS, neighbors (Y))) THEN
```

```

List-C-as-Maximal-Clique
ELSE   maxCliques (addElementToSet (Y, C),
                intersect (P, neighbors (Y)),
                intersect (PuS, neighbors (Y)));
END "chooseElement"
END "maxCliques"
```

Les procédures qui sont désignées par des noms longs réalisent les opérations standards sur les ensembles : intersection, union, choix d'un élément, suppression d'un élément, ajout d'un élément, etc. La procédure `neighbors ()` retourne l'ensemble de noeuds qui sont directement connectés au noeud représenté par l'argument de cette procédure. Afin d'énumérer toutes les cliques maximales d'un graphe, l'appel suivant doit être effectué : `maxCliques (emptySet, setOfAllNodes, setOfAllNodes)`.

Annexe C

Le filtre de Kalman étendu appliqué à notre système

C.1 Présentation du filtre de Kalman étendu

Nous introduisons brièvement le formalisme du filtre de Kalman étendu. Une description détaillée se trouve dans [2], [4] et [52].

Soit $\alpha \in \mathbb{R}^p$ un vecteur inconnu que l'on désire estimer. Ce vecteur est lié à des mesures $m_i \in \mathbb{R}^n$ par un ensemble d'équations (non nécessairement linéaires) de la forme :

$$f(m_i, \alpha) = 0 \quad (\text{C.31})$$

$f()$ est une application de \mathbb{R}^p dans \mathbb{R}^n . On ne mesure pas, en fait, les vecteurs m_i mais une estimation bruitée de ceux ci :

$$\hat{m}_i = m_i + v_i$$

où v_i est une erreur aléatoire. Nous supposons que v_i est un bruit blanc, c'est-à-dire que v_i est une variable aléatoire centrée de covariance connue :

$$\begin{aligned} E[v_i] &= 0 \\ E[v_i v_i^T] &= \Lambda_i \quad \geq 0 \\ E[v_i v_j^T] &= 0 \quad \forall i \neq j \end{aligned}$$

Le filtre de Kalman étendu s'applique en principe à des équations linéaires. Lorsque les équations (C.31) ne sont pas linéaires, on les linéarise en appliquant

le développement de Taylor au premier ordre autour d'une estimée $\hat{\alpha}_{i-1}$ du vecteur inconnu α :

$$\hat{\alpha}_{i-1} = \alpha + w_i$$

où w_i est une erreur aléatoire pour laquelle on fait les mêmes hypothèses que pour v_i , à savoir $E[w_i] = 0$ et $E[w_i w_i^T] = S_{i-1}$, où S_{i-1} est une matrice définie positive donnée. Linéarisant l'équation (C.31) autour de $(\hat{m}_i, \hat{\alpha}_{i-1})$, on obtient :

$$f(m_i, \alpha) = 0 \approx f(\hat{m}_i, \hat{\alpha}_{i-1}) + \frac{\partial f}{\partial m}(m_i - \hat{m}_i) + \frac{\partial f}{\partial \alpha}(\alpha - \hat{\alpha}_{i-1}) \quad (\text{C.32})$$

Les dérivées partielles $\partial f / \partial m$ et $\partial f / \partial \alpha$ sont estimées au point $(\hat{m}_i, \hat{\alpha}_{i-1})$. L'équation (C.32) se réécrit :

$$q_i = M_i \alpha + u_i \quad (\text{C.33})$$

avec

$$\begin{aligned} q_i &= -f(\hat{m}_i, \hat{\alpha}_{i-1}) + \frac{\partial f}{\partial \alpha} \hat{\alpha}_{i-1} \\ M_i &= \frac{\partial f}{\partial \alpha} \\ u_i &= \frac{\partial f}{\partial m}(m_i - \hat{m}_i) \end{aligned}$$

Les statistiques du second ordre sur u_i sont obtenues facilement de celles sur v_i :

$$\begin{aligned} E[u_i] &= 0 \\ W_i &\stackrel{\text{def}}{=} E[u_i u_i^T] = \frac{\partial f}{\partial m} \Lambda_i \frac{\partial f^T}{\partial m} \end{aligned}$$

L'équation (C.33) est maintenant une équation de mesure linéaire où q_i est la nouvelle mesure. Les équations du filtre de Kalman qui permettent d'obtenir une nouvelle estimation $(\hat{\alpha}_i, S_i)$ de α à partir de $(\hat{\alpha}_{i-1}, S_{i-1})$ sont les suivantes :

$$\begin{aligned} \hat{\alpha}_i &= \hat{\alpha}_{i-1} + k_i(q_i - M_i \hat{\alpha}_{i-1}) \\ k_i &= S_{i-1} M_i^T (W_i + M_i S_{i-1} M_i^T)^{-1} \\ S_i &= (I - k_i M_i) S_{i-1} \end{aligned}$$

L'estimée $\hat{\alpha}_{i-1}$ est donc corrigée à l'étape i d'un terme proportionnel à l'erreur courante $q_i - M_i \hat{\alpha}_{i-1}$ (appelée *innovation*). Le facteur de proportionnalité k_i est appelé *gain de Kalman*.

C.2 Application du filtre de Kalman à notre système

Etudions maintenant comment ce formalisme s'applique à notre système. Nous disposons de deux mesures, une dans l'image gauche $p(x, y)$ et une autre dans l'image droite $p'(x', y')$, les points p, p' étant appariés. Le vecteur inconnu α est les coordonnées 3-D du point de l'espace P résultant de la triangulation de p et p' . $\alpha = (X \ Y \ Z)^T$ (exprimé dans le repère lié à la caméra gauche). Trouvons l'équation $f(m, \alpha)$ reliant les mesures avec le vecteur inconnu.

- 1ère mesure : le point $p(x, y)$.

$m_1 = (x \ y)^T$. Les équations (3.5) fournissent la fonction $f(m, \alpha)$:

$$f(m_1, \alpha) = 0 = \begin{pmatrix} Zx - X \\ Zy - Y \end{pmatrix} \quad (\text{C.34})$$

De l'équation (C.34) on obtient les dérivées partielles :

$$\frac{\partial f}{\partial \alpha} = \begin{pmatrix} -1 & 0 & x \\ 0 & -1 & y \end{pmatrix} \quad (\text{C.35})$$

$$\frac{\partial f}{\partial m} = \begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix} \quad (\text{C.36})$$

- 2ème mesure : le point $p'(x', y')$.

$m_2 = (x' \ y')^T$. Il faut relier les coordonnées de p' avec le vecteur inconnu α . Pour ceci, utilisant les équations (3.6) et remplaçant X' , Y' et Z' par leurs valeurs données par (3.4) nous obtenons que :

$$x' = \frac{X \cos \phi + Z \sin \phi + b_x}{-X \sin \phi + Z \cos \phi + b_z}$$

$$y' = \frac{Y}{-X \sin \phi + Z \cos \phi + b_z}$$

La fonction $f(m, \alpha)$ est donc :

$$f(m_2, \alpha) = 0 = \begin{pmatrix} (-X \sin \phi + Z \cos \phi + b_z)x' - (X \cos \phi + Z \sin \phi) - b_x \\ (-X \sin \phi + Z \cos \phi + b_z)y' - Y \end{pmatrix} \quad (\text{C.37})$$

et les dérivées partielles sont :

$$\frac{\partial f}{\partial \alpha} = \begin{pmatrix} -x' \sin \phi - \cos \phi & 0 & x' \cos \phi - \sin \phi \\ -y' \sin \phi & -1 & y' \cos \phi \end{pmatrix} \quad (\text{C.38})$$

$$\frac{\partial f}{\partial m} = \begin{pmatrix} -X \sin \phi + Z \cos \phi + b_z & 0 \\ 0 & -X \sin \phi + Z \cos \phi + b_z \end{pmatrix} \quad (\text{C.39})$$

Voici maintenant les étapes de mise en oeuvre du filtre de Kalman :

1. Nous supposons que les points p et p' sont détectés avec une erreur d'un pixel. La matrice Λ_i est donc :

$$\Lambda_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

2. nous commençons par une estimée $\hat{\alpha}_0$ du vecteur α ; $\hat{\alpha}_0$ est donnée par la triangulation des points p et p' .

3. Nous associons à $\hat{\alpha}_0$ une matrice de covariance initiale S_0 très grande (par ex. 2 mètres dans chaque direction) :

$$S_0 = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

4. Par la première mesure $m_1 = p$:
calcul des $\partial f / \partial \alpha$ et $\partial f / \partial m$, autour de $(m_1, \hat{\alpha}_0)$ par les équations (C.35) et (C.36), respectivement.
calcul des $q_1, M_1, W_1, k_1, \hat{\alpha}_1, S_1$.
5. Par la deuxième mesure $m_2 = p'$:
calcul des $\partial f / \partial \alpha$ et $\partial f / \partial m$, autour de $(m_2, \hat{\alpha}_1)$ à l'aide des équations (C.38) et (C.39), respectivement.
calcul des $q_2, M_2, W_2, k_2, \hat{\alpha}_2, S_2$.

Ainsi nous trouvons une estimation sur la position du point P dans l'espace ($\hat{\alpha}_2$) et une matrice de covariance (S_2) représentant l'incertitude que nous avons sur la position de P .

Bibliographie

- [1] R. D. Arnold and T. O. Binford. Geometric constraints in stereo vision. In *SPIE Vol. 238 - Image Processing for Missile Guidance*, pages 281–292, 1980.
- [2] N. Ayache. *Construction et Fusion de Représentations Visuelles 3D - Applications à la robotique Mobile*. PhD thesis, Titre du Docteur es Sciences, Université de Paris Sud, Centre d'Orsay, France, May 1988.
- [3] N. Ayache and O. D. Faugeras. Building, registering, and fusing noisy visual maps. In *Proc. First International Conference on Computer Vision*, pages 73–82, London, England, June 1987.
- [4] N. Ayache and O. D. Faugeras. *Maintaining Representations of the Environment of a Mobile Robot*. Technical Report 789, INRIA Rocquencourt, France, February 1988.
- [5] N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *International Journal of Computer Vision*, 1(2):107–131, 1987.
- [6] N. Ayache and F. Lustman. Fast and reliable passive trinocular stereovision. In *Proc. First International Conference on Computer Vision*, pages 422–427, London, Great Britain, June 1987.
- [7] H. H. Baker and T. O. Binford. Depth from edge and intensity based stereo. In *Proc. 7-th International Joint Conference on Artificial Intelligence*, pages 631–636, Vancouver, Canada, August 1981.
- [8] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice Hall Inc., 1982.
- [9] S. T. Barnard. A stochastic approach to stereo vision. In *Proc. 5th AAAI-86*, pages 676–680, Philadelphia, USA, August 1986.

- [10] S. T. Barnard and M. A. Fischler. Computational stereo. *ACM Computing Surveys*, 14:553–572, December 1982.
- [11] S. T. Barnard and W. B. Thompson. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(4):333–340, July 1980.
- [12] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.
- [13] S. D. Blostein and Th. S. Huang. Error analysis in stereo determination of 3-D point positions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(6):752–765, November 1987.
- [14] J. D. Boissonnat, O. D. Faugeras, and E. Le Bras-Mehlman. *Representing Stereo Data with the Delaunay Triangulation*. Technical Report 788, INRIA Rocquencourt, France, February 1988.
- [15] J.D. Boissonnat. Geometric structure for three-dimensional shape representation. *ACM Transactions on Graphics*, 3(4):266–286, October 1984.
- [16] R. C. Bolles. Robust feature matching through maximal cliques. In *Proc. SPIE Technical Symposium on Imaging Applied to Automated Industrial Inspection and Assembly*, Washington, D.C., USA, April 1979.
- [17] R. C. Bolles and R. A. Cain. Recognizing and locating partially visible objects; the Local-Feature-Focus method. *International Journal of Robotics Research*, 1(3):57–82, 1982.
- [18] R. C. Bolles and R. Horaud. 3DPO: A Three-Dimensional Part Orientation System. *International Journal of Robotics Research*, 5(3):3–26, Fall 1986. Also in *Three-Dimensional Machine Vision*, T. Kanade, editor, Kluwer Academic Publisher, 1987, pages 399–450.
- [19] K. L. Boyer and A. C. Kak. Structural stereopsis for 3-D vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10(2):144–166, March 1988.
- [20] P. Burt and B. Julesz. A disparity gradient limit for binocular fusion. *Science*, 208:615–617, May 1980.
- [21] S. Castan and J. Shen. A stereo algorithm taking into account the perspective distortions. In *Proc. 7th International Conference on Pattern Recognition*, pages 441–443, Montreal, Canada, 1984.
- [22] J. P. Cocquerez and A. Gagalowitz. Mise en correspondance de régions dans une paire d'images stéréo. In *Proc. Colloque Image du MARI-CESTA*, pages 165–173, Paris, France, May 1987.
- [23] J. H. Connell and M. Brady. Generating and generalizing models of visual objects. *Artificial Intelligence*, 31:159–183, 1987.

- [24] J. L. Crowley and A. C. Parker. A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(2):156–170, March 1984.
- [25] J. L. Crowley and F. Ramparany. Mathematical tools for representing uncertainty in perception. In *Proc. of the 1987 AAAI Workshop on Spatial Reasoning and Multi-Sensor Fusion*, St. Charles, Illinois, USA, October 1987.
- [26] Y. Demazeau. *Niveaux de représentation en vision par ordinateur : indices d'images et indices de scène*. PhD thesis, Institut National Polytechnique de Grenoble, France, December 1986.
- [27] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.
- [28] H. F. Durrant-Whyte. Consistent integration and propagation of disparate sensor observations. *International Journal of Robotics Research*, 6(3):3–24, Fall 1987.
- [29] R. D. Eastman and A. M. Waxman. Disparity functionals and stereo vision. In *Proc. DARPA Image Understanding Workshop*, pages 245–254, Miami Beach, Florida, USA, December 1985.
- [30] O. D. Faugeras. Vision artificielle en dimension trois. In *Actes du 2ème Atelier Scientifique TIPI*, pages I:1–94, Aussois, Savoie, France, April 1988.
- [31] O. D. Faugeras and G. Toscani. The Calibration Problem for Stereo. In *Proc. Computer Vision and Pattern Recognition*, pages 15–20, Miami Beach, Florida, USA, June 1986.
- [32] D. B. Gennery. *Modelling the environment of an exploring vehicle by means of stereo vision*. PhD thesis, Stanford Artificial Intelligence Laboratory, AIM 339, June 1980.
- [33] M. Gondran and M. Minoux. *Graphes et Algorithmes*. Eurolles, 1985.
- [34] W. E. L. Grimson. Binocular shading and visual surface reconstruction. *Computer Vision, Graphics, and Image Processing*, 28:19–43, 1984.
- [35] W. E. L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(1):17–34, January 1985.
- [36] W. E. L. Grimson. *From Images to Surfaces*. MIT Press, Cambridge, Massachusetts, 1981.
- [37] M. J. Hannah. Bootstrap stereo. In *Proc. DARPA Image Understanding Workshop*, pages 201–208, April 1980.

- [38] M. J. Hannah. *Computer Matching of Areas in Stereo Images*. PhD thesis, Stanford Artificial Intelligence Laboratory, AIM 239, July 1974.
- [39] M. J. Hannah. *Description of SRI's Baseline Stereo System*. Technical Report 342, SRI International, October 1984.
- [40] L. Herault. *Parallélisation d'Algorithmes pour la Stéréovision*. Technical Report, CEA-CENG, Grenoble, France, May 1988. Workshop on *OCCAM and the transputers*.
- [41] M. Herman and T. Kanade. Incremental reconstruction of 3D scenes from multiple, complex images. *Artificial Intelligence*, 30:289–341, 1986.
- [42] W. Hoff and N. Ahuja. Extracting surfaces from stereo images: An integrated approach. In *Proc. First International Conference on Computer Vision*, pages 284–294, London, Great Britain, June 1987.
- [43] R. Horaud and Th. Skordas. Model-based strategy planning for recognizing partially occluded parts. *IEEE Computer Magazine*, 58–65, August 1987. Special issue on *CAD-Based Robot Vision*.
- [44] M. Ito and A. Ishii. Three-view stereo analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4):524–532, July 1986.
- [45] A. Izaguirre, P. Pu, and J. Summers. A new development in camera calibration: Calibrating a pair of mobile cameras. *International Journal of Robotics Research*, 6(3):104–116, Fall 1987.
- [46] M. Kass. Computing visual correspondence. In Alex P. Pentland, editor, *From Pixels to Predicates*, chapter 4, pages 78–92, Ablex Publishing Corporation, Norwood, New Jersey, 1986.
- [47] M. Kass. Linear image features in stereopsis. In *Proc. 5th AAAI-86*, pages 707–713, Philadelphia, USA, August 1986.
- [48] D. J. Kriegman, E. Triendl, and T. O. Binford. A mobile robot: Sensing, planning and locomotion. In *Proc. IEEE International Conference on Robotics and Automation*, pages 402–408, Raleigh, North Carolina, USA, April 1987.
- [49] H. S. Lim and T. O. Binford. Stereo correspondence: Features and constraints. In *Proc. DARPA Image Understanding Workshop*, pages 373–380, Miami Beach, Florida, USA, December 1985.
- [50] P. Long-Limozin. *Vision Stéréoscopique Appliquée à la Robotique*. PhD thesis, Université de Nice, France, October 1986.
- [51] D. Lowe. Visual recognition from spatial correspondence and perceptual organization. In *Proc. 9-th International Joint Conference on Artificial Intelligence*, pages 953–959, Los Angeles, USA, August 1985.

- [52] F. Lustman. *Vision Stéréoscopique et Perception du Mouvement en Vision Artificielle*. PhD thesis, Université de Paris-Sud Centre d'Orsay, France, December 1987.
- [53] D. Marr. *Vision*. W. H. Freeman, San Francisco, 1982.
- [54] D. Marr and T. Poggio. *A Theory of Human Stereo Vision*. Technical Report AI Memo 451, MIT Artificial Intelligence Laboratory, November 1977.
- [55] H. A. Martins, J. R. Birk, and R. B. Kelley. Camera models based on data from two calibration planes. *Computer Vision, Graphics, and Image Processing*, 17:173–180, 1981.
- [56] L. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, RA-3(3):239–248, June 1987.
- [57] J. E. W. Mayhew and J. P. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–385, 1981.
- [58] E. S. McVey and J. W. Lee. Some accuracy and resolution aspects of computer vision distance measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4:646–649, November 1982.
- [59] G. Medioni and R. Nevatia. Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31:2–18, 1985.
- [60] R. Mohan, G. Medioni, and R. Nevatia. Stereo error detection, correction, and evaluation. In *Proc. First International Conference on Computer Vision*, pages 315–324, London, Great Britain, June 1987.
- [61] R. Mohr and B. Wrobel. La correspondance stéré vue comme une recherche de chemin optimal. In *Proc. 4ème Congrès Reconnaissance des Formes et Intelligence Artificielle*, pages 71–79, Paris, France, January 1984.
- [62] O. Monga. A new segmentation method and its application to stereo vision. In *Proc. SPIE Technical Symposium on Optics, Electro-optics, and Sensors*, Orlando, Florida, USA, May 1987.
- [63] O. Monga and B. Wrobel. Segmentation d'images : vers une méthodologie. *Traitement du Signal*, 4(3):169–193, 1987. Special issue on *Computer Vision*.
- [64] H. P. Moravec. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. Technical Report CMU-RI-TR-3, Carnegie-Mellon University, September 1980.
- [65] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(2):139–154, March 1985.

- [66] T. Pavlidis and S. L. Horowitz. Segmentation of plane curves. *IEEE Trans. on Computers*, C-23(8):860–870, August 1974.
- [67] M. Pietikainen and D. Harwood. Depth from three camera stereo. In *Proc. Computer Vision and Pattern Recognition*, pages 2–8, Miami Beach, Florida, USA, June 1986.
- [68] J. P. Planquart. *Segmentation d'images Multi-senseurs en vision Tridimensionnelle par ordinateur*. PhD thesis, Institut National Polytechnique de Grenoble, France, February 1987.
- [69] G. F. Poggio and T. Poggio. The analysis of stereopsis. *Annual Review of Neuroscience*, 7:379–412, 1984.
- [70] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.
- [71] S. B. Pollard, J. Porril, J. E. W. Mayhew, and J. P. Frisby. *Disparity Gradient, Lipschitz Continuity, and Computing Binocular Correspondence*. Technical Report AIVRU-010, University of Sheffield, Great Britain, 1986.
- [72] L. Quan, R. Mohr, and E. Thirion. Utilisation d'invariants par perspective en vision monoculaire. In *Actes du 2ème Atelier Scientifique TIPI*, pages XXVII:1–17, Aussois, Savoie, France, April 1988.
- [73] G. V. S. Raju, T. O. Binford, and S. Shekhar. Stereo matching using Viterbi algorithm. In *Proc. DARPA Image Understanding Workshop*, pages 766–776, Los Angeles, USA, February 1987.
- [74] G. Reynolds and J. R. Beveridge. Searching for geometric structure in images of natural scenes. In *Proc. DARPA Image Understanding Workshop*, pages 257–271, Los Angeles, California, USA, February 1987.
- [75] A. Robert de Saint Vincent. *Perception et Modélisation de l'Environnement d'un Robot Mobile : Une Approche par Stéréovision*. PhD thesis, Université Paul Sabatier de Toulouse, France, November 1986.
- [76] T. Skordas and R. Horaud. Groupement robuste des caractéristiques image. In *Actes 6ème Congrès en Reconnaissance des Formes et Intelligence Artificielle*, pages 563–568, Antibes, France, November 1987.
- [77] R. C. Smith and P. Cheeseman. On the representation and estimation of spatial uncertainty. *International Journal of Robotics Research*, 5(4):56–68, Winter 1986.
- [78] P. Stelmaszyk, D. Aubert, and J. L. Crowley. Mise en correspondance structurelle en vision multirésolution. In *Actes du 2ème Atelier Scientifique TIPI*, pages XXXVI: 1–9, Aussois, Savoie, France, April 1988.

- [79] R. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal of Computation*, 1(2):146–160, June 1972.
- [80] D. Terzopoulos. Integrating visual information from multiple sources. In Alex B. Pentland, editor, *From Pixels to Predicates*, chapter 6, pages 111–142, Ablex Publishing Corporation, Norwood, New Jersey, 1986.
- [81] D. Terzopoulos. Multilevel computational processes for visible surface reconstruction. *Computer Vision, Graphics, and Image Processing*, 24:52–96, 1983.
- [82] M. Thonnat. Semantic interpretation of 3-D stereo data. In *Proc. 8th International Conference on Pattern Recognition*, pages 1051–1054, Paris, France, October 1986.
- [83] R. Y. Tsai. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, August 1987.
- [84] R. Weiss and M. Boldt. Geometric grouping applied to straight lines. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 489–495, Miami Beach, Florida, USA, June 1986.
- [85] A. P. Witkin and J. M. Tenenbaum. On perceptual organisation. In Alex B. Pentland, editor, *From Pixels to Predicates*, chapter 7, pages 149–169, Ablex Publishing Corporation, Norwood, New Jersey, 1986.
- [86] A. P. Witkin and J. M. Tenenbaum. On the role of structure in vision. In Rosenfeld, editor, *Human and Machine Vision*, pages 481–543, Academic Press, New York, 1983.
- [87] B. Wrobel-Dautcourt. *Perception de la distance par mise en correspondance de régions entre des images stéréoscopiques*. PhD thesis, Institut National Polytechnique de Lorraine, France, March 1988.
- [88] A. L. Yuille and T. Poggio. *A Generalized Ordering Constraint for Stereo Correspondence*. Technical Report AI Memo 777, MIT Artificial Intelligence Laboratory, May 1984.



Liste des Figures

1.1	La géométrie d'un système stéréoscopique.	6
1.2	La géométrie épipolaire d'un système stéréo.	13
1.3	Violation de la règle d'unicité.	15
1.4	Définition du gradient de disparité.	17
1.5	La contrainte d'ordre.	18
1.6	La zone interdite associée à un point P de l'espace.	18
1.7	La géométrie d'un système stéréo trinoculaire.	23
1.8	Le modèle physique du plateau fin et flexible pour la reconstruction de surfaces.	27
1.9	Ambiguïté quant à la reconstruction d'une surface.	31
1.10	L'architecture globale de notre système stéréo.	34
2.1	L'architecture globale du système de description monoculaire d'une image.	39
2.2	La structure locale que nous voulons associer à un segment-image.	42
2.3	La configuration locale d'une jonction en T.	42
2.4	Détection des T sur des situations difficiles.	43
2.5	Critères géométriques pour la détection d'une jonction en T.	44
2.6	Solutions proposées pour les T.	45
2.7	Fragmentation d'un segment lors de la détection d'un T.	45
2.8	Jonctions déformées par la segmentation.	47
2.9	Situations où le groupement devient difficile.	47
2.10	Critères géométriques pour la détection d'un angle.	48
2.11	Exemple de détection des angles.	49
2.12	Formation d'une jonction à partir des angles.	50
2.13	Le groupement aux cas difficiles.	50
2.14	Le groupement en présence d'une structure répétitive.	51

2.15	Critères géométriques pour la détection de deux segments colinéaires.	52
2.16	Un exemple de détection des segments colinéaires.	52
2.17	Un segment-image et la détection des deux bandes adjacentes qui lui sont associées.	53
2.18	Les segments à gauche et à droite d'un segment.	53
2.19	Relation ambiguë entre deux segments.	54
2.20	La tête et la queue d'un segment.	54
2.21	La structure locale associée à un segment-image.	55
2.22	Un ensemble de six segments d'une image avec leurs orientations.	56
2.23	Le graphe relationnel qui décrit les six segments.	56
2.24	Le résultat du groupement à l'intérieur d'une fenêtre.	57
2.25	L'image d'un bureau.	59
2.26	Les segments de droite extraits de l'image d'un bureau.	60
2.27	Les jonctions en T.	60
2.28	Les jonctions à deux segments.	61
2.29	Les jonctions à trois segments.	61
2.30	Les jonctions à quatre segments ou plus.	62
2.31	Les segments colinéaires.	62
3.1	La géométrie du capteur stéréo.	65
3.2	Une vue de dessus d'un plan épipolaire.	66
3.3	La bande épipolaire associée à un segment.	67
3.4	La bande épipolaire pour un segment horizontal.	68
3.5	Recherche des candidats potentiels d'un segment.	69
3.6	Détermination de la projection \vec{v} d'un vecteur \vec{V} dans l'image.	70
3.7	Un exemple d'incompatibilité déduite de la première règle.	78
3.8	Le calcul de la longueur estimée d'un segment.	78
3.9	Incompatibilité déduite de la troisième règle.	79
3.10	Un exemple où la règle de transitivité conduit à un résultat ambigu.	80
3.11	Critères géométriques pour déterminer si deux segments sont reliés par colinéarité étendue.	81
3.12	Exemples d'incompatibilité par colinéarité étendue.	83
3.13	Exemples d'incompatibilité par connexité étendue.	83
3.14	Exemples d'incompatibilité par ordre étendu.	84
3.15	Deux structures simples, à mettre en correspondance.	86
3.16	Les deux graphes relationnels décrivant les deux structures.	86
3.17	Le graphe montrant les compatibilités entre les appariements potentiels après l'application des trois premières règles.	88
3.18	Les noeuds incompatibles.	90
3.19	Les neuf solutions de l'exemple.	92
3.20	Premier problème de mise en correspondance.	93
3.21	Deuxième problème de mise en correspondance.	94
3.22	Le graphe de correspondance associé à cet exemple montre qu'il y a deux cliques maximales disjointes, qui sont toutes les deux acceptées.	94

3.23	Exemple de découpage d'une image en fenêtres.	98
3.24	La première paire d'images.	103
3.25	Les segments extraits de la première paire d'images.	103
3.26	Premier exemple de mise en correspondance locale.	104
3.27	Deuxième exemple de mise en correspondance locale.	105
3.28	Le résultat de l'appariement des segments extraits de la première paire d'images.	106
3.29	Les segments non appariés pour la première paire d'images.	106
3.30	Le résultat de mise en correspondance de la première paire d'images obtenu par une recherche non exhaustive des cliques maximales.	107
3.31	La deuxième paire d'images.	107
3.32	Les segments extraits de la deuxième paire d'images.	108
3.33	Le résultat de l'appariement des segments extraits de la deuxième paire d'images.	108
3.34	Les segments non appariés pour la deuxième paire d'images.	109
3.35	La troisième paire d'images.	109
3.36	Les segments extraits des images de Toit.	110
3.37	Les segments appariés et ceux non appariés pour les images de Toit.	111
3.38	La quatrième paire d'images.	112
3.39	Les segments extraits des images d'Arbre.	112
3.40	Les segments appariés et ceux non appariés pour les images d'Arbre.	113
3.41	Le résultat de mise en correspondance pour les images de Toit et d'Arbre obtenu par l'algorithme de N. Ayache.	114
4.1	Incertitude sur la position d'un point de l'espace résultant du proces- sus de triangulation.	116
4.2	Reconstruction d'un segment 3-D à partir d'un couple de segments homologues.	119
4.3	Les zones de correspondance maximum associées à deux segments homologues.	120
4.4	Les zones de correspondance maximum des segments appariés pour les images de Bureau 1.	121
4.5	Les segments 3-D et leur incertitude en profondeur pour les images de Bureau 1.	123
4.6	L'incertitude en profondeur en fonction de la variation de la pro- fondeur pour les extrémités des segments 3-D (Bureau 1).	124
4.7	Les zones de correspondance maximum des segments appariés pour les images de Bureau 2.	124
4.8	Les segments 3-D et leur incertitude en profondeur pour les images de Bureau 2.	125
4.9	L'incertitude en profondeur en fonction de la variation de la pro- fondeur pour les extrémités des segments 3-D (Bureau 2).	126
4.10	Un segment g de l'image gauche et sa reconstruction dans l'espace utilisant le plan d'interprétation qui lui est associé.	129

4.11	Extraction de surfaces planes de scène utilisant l'information du groupement.	131
4.12	Bureau 2 : les segments 3-D utiles pour la reconstruction de surfaces.	133
4.13	Trois plans détectés pour la scène de Bureau 2.	134
A.1	Les matrices de passage pour le calibrage d'un système stéréo.	140

A U T O R I S A T I O N de S O U T E N A N C E

VU les dispositions de l'article 15 Titre III de l'arrêté du 5 juillet 1984 relatif aux études doctorales

VU les rapports de présentation de Messieurs MOHR Roger, Professeur

RICHETIN Marc, Professeur

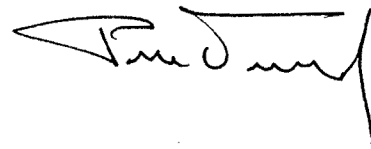
Monsieur SKORDAS Thomas

est autorisé(e) à présenter une thèse en soutenance en vue de l'obtention du diplôme de DOCTEUR de L'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE, spécialité "INFORMATIQUE".

Fait à Grenoble, le 23 Septembre 198

Georges LEFEBVRE
Président
de l'Institut National Polytechnique
de Grenoble

AO. le Vice-Président,



Résumé

Nous décrivons un système de Vision Stéréoscopique. Le but d'un tel système est de mettre en correspondance deux images d'une même scène acquises de deux points de vue différents et de reconstituer ensuite la structure tridimensionnelle (3D) de la scène. Notre approche se résume ainsi :

Des segments de droite sont extraits de l'image de gauche et de l'image de droite. Ces segments sont groupés convenablement afin de former des structures locales 2D. Sur la base de ces groupements, nous décrivons chaque image par un graphe relationnel (que nous appelons *description monoculaire*) où les noeuds représentent les segments avec leurs propriétés géométriques et les arcs représentent des relations structurelles entre les segments.

Nous aboutissons ainsi à deux descriptions monoculaires représentant la paire stéréo initiale. La mise en correspondance est obtenue en trouvant des structures similaires entre les deux descriptions, ce qui est équivalent à la recherche d'un isomorphisme. Dans ce but, utilisant des contraintes dictées par la géométrie du capteur stéréo, il est possible d'établir, pour un segment d'une image, une liste de candidats potentiels associés dans l'autre image. Ces appariements potentiels sont représentés par des noeuds dans un graphe (que nous appelons *graphe de correspondance*). Les arêtes de ce graphe représentent des similarités photométriques, géométriques et topologiques locales entre les deux images. Des correspondances entre les deux images sont équivalentes à des ensembles de noeuds mutuellement compatibles. Ces ensembles sont fournis par les cliques maximales de ce graphe. La meilleure clique maximale, choisie selon une procédure d'évaluation, est considérée comme étant la meilleure correspondance disponible.

Enfin, nous procédons à la reconstruction de surfaces 3D de la scène. Après avoir déterminé la position 3D des segments appariés, nous nous servons de deux descriptions monoculaires afin de reconstruire d'éléments de surfaces planes de la scène qui contiennent les segments 3D.

Mots clés :

Stéréovision – Groupement d'indices
Description monoculaire – Stéréocorrespondance
Recherche dans un graphe – Cliques maximales
Perception 3D.