



Mise en correspondance d'indices images en résolutions multiples

Didier Aubert

► To cite this version:

Didier Aubert. Mise en correspondance d'indices images en résolutions multiples. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG, 1989. Français. NNT: . tel-00332349

HAL Id: tel-00332349

<https://theses.hal.science/tel-00332349>

Submitted on 20 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.

71 5013

THESE

présentée par **Didier AUBERT**

pour obtenir le titre de DOCTEUR

de l'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

(arrêté ministériel du 5 juillet 1984)

(Spécialité : Informatique)

◇◇◇◇◇◇◇◇

MISE EN CORRESPONDANCE D'INDICES IMAGES EN RESOLUTIONS MULTIPLES

◇◇◇◇◇◇◇◇

Date de soutenance : Janvier 1989

Composition du jury : P. JORRAND

S. CASTAN

J. G. POSTAIRE

A. CHEHIKIAN

J. L. CROWLEY

P. STELMASZYK

Président

Rapporteurs

Examineurs

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

Président : Georges LESPINARD

Année 1988

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	LONGEQUEUE 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 Hervé	UFR PGP	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	ENSIMAG	RAMEAU Jean-Jacques	ENSEEG
DEPORTES Jacques	ENSPG	RENAUD Maurice	UFR PGP
DOLMAZON Jean-Marc	ENSERG	ROBERT André	UFR PGP
DURAND Francis	ENSEEG	ROBERT François	ENSIMAG
DURAND Jean-Louis	ENSIEG	SABONNADIÈRE Jean-Claude	ENSIEG
FOGGIA Albert	ENSIEG	SAUCIER Gabrielle	ENSIMAG
FONLUPT Jean	ENSIMAG	SCHLENKER Claire	ENSPG
FOULARD Claude	ENSIEG	SCHLENKER Michel	ENSPG
GANDINI Alessandro	UFR PGP	SILVY Jacques	UFR PGP
GAUBERT Claude	ENSPG	SIRIEYS Pierre	ENSHMG
GENTIL Pierre	ENSERG	SOHM Jean-Claude	ENSEEG
GREVEN Hélène	IUFA	SOLER Jean-Louis	ENSIMAG
GUERIN Bernard	ENSERG	SOUQUET Jean-Louis	ENSEEG
GUYOT Pierre	ENSEEG	TROMPETTE Philippe	ENSHMG
IVANES Marcel	ENSIEG	VEILLON Gérard	ENSIMAG
JAUSSAUD Pierre	ENSIEG	ZADWORNÝ François	ENSERG

**Professeur Université des Sciences
Sociales
(Grenoble II)**

BOLLIET Louis

**Personnes ayant obtenu le diplôme
d'HABILITATION A DIRIGER DES
RECHERCHES**

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

**Chercheurs du C.N.R.S
Directeurs de recherche 1ère Classe**

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

**Directeurs de recherche
2ème Classe**

ALEMANY Antoine
ALLIBERT Colette
ALLIBERT Michel
ANSARA Ibrahim
ARMAND Michel
BERNARD Claude
BINDER Gilbert
BONNET Roland
BORNARD Guy
CAILLET Marcel
CALMET Jacques

COURTOIS Bernard
DAVID René
DRIOLE Jean
ESCUDIER Pierre
EUSTATHOPOULOS Nicolas
GUELIN Pierre
JODD Jean-Charles
KLEITZ Michel
KOFMAN Walter
KAMARINOS Georges
LEJEUNE Gérard
LE PROVOST Christian
MADAR Roland
MERMET Jean
MICHEL Jean-Marie
MUNIER Jacques
PIAU Monique
SENATEUR Jean-Pierre
SIFAKIS Joseph
SIMON Jean-Paul
SUERY Michel
TEODOSIU Christian
VAUCLIN Michel
WACK Bernard

**Personnalités agréées à titre
permanent à diriger des travaux de
recherche (décision du conseil
scientifique)**

E.N.S.E.E.G

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

E.N.S.E.R.G

BOREL Joseph

E.N.S.I.E.G

DESCHIZEAUX Pierre
GLANGEAUD François
PERARD Jacques
REINISCH Raymond

E.N.S.H.G

ROWE Alain

E.N.S.I.M.A.G

COURTIN Jacques

E.F.P.

CHARUEL Robert

C.E.N.G

CADET Jean
COEURE Philippe
DELHAYE Jean-Marc
DUPUY Michel
JOUVE Hubert
NICOLAU Yvan
NIFENECKER Hervé
PERROUD Paul
PEUZIN Jean-Claude
TAIB Maurice
VINCENDON Marc

Laboratoires extérieurs

C.N.E.T

DEVINE Rodericq
GERBER Roland
MERCKEL Gérard
PAULEAU Yves

UNIVERSITE Joseph FOURIER (GRENOBLE I)

Président de l'Université :
M. PAYAN Jean Jacques

Année Universitaire 1987 - 1988

MEMBRES DU CORPS ENSEIGNANT DE SCIENCES ET DE GEOGRAPHIE

PROFESSEURS DE 1ère Classe

ARNAUD Paul
ARVIEU ROBERT
AUBERT Guy
AURIAULT Jean-Louis
AYANT Yves
BARBIER Marie-Jeanne
BARJON Robert
BARNOUD Fernand
BARRA Jean-René
BECKER Pierre
BEGUIN Claude
BELORISKY Elie
BENZAKEN Claude
BERARD Pierre
BERNARD Alain
BERTRANDIAS Françoise
BERTRANDIAS Jean-Paul
BILLET Jean
BOELHER Jean-Paul
BONNIER Jane Marie
BOUCHEZ Robert
BRAVARD Yves
CARLIER Georges
CAUQUIS Georges
CHARDON Michel
CHIBON Pierre
COHEN ADDAD Jean-Pierre
COLIN DE VERDIERE Yves
CYROT Michel
DEBELMAS Jacques
DEGRANGE Charles
DEMAILLY Jean-Pierre
DENEUVILLE Alain
DEPORTES Charles
DOLIQUE Jean-Michel
DOUCE Roland
DUCROS Pierre
FONTAINE Jean-Marc
GAGNAIRE Didier
GERMAIN Jean-Pierre
GIRAUD Pierre
HICTER Pierre
IDELMAN Simon
JANIN Bernard
JOLY Jean-René
KAHANE André, détaché
KAHANE Josette
KRAKOWIAK Sacha

Chimie Organique
Physique Nucléaire I.S.N.
Physique C.N.R.S
Mécanique
Physique Approfondie
Electrochimie
Physique Nucléaire ISN
Biochimie Macromoléculaire Végétale
Statistiques-Mathématiques Appliquées
Physique
Chimie Organique
Physique
Mathématiques Pures
Mathématiques Pures
Mathématiques Pures
Mathématiques Pures
Mathématiques Pures
Géographie
Mécanique
Chimie Générale
Physique Nucléaire ISN
Géographie
Biologie Végétale
Chimie Organique
Géographie
Biologie Animale
Physique
Mathématiques Pures
Physique du Solide
Géologie Générale
Zoologie
Mathématiques Pures
Physique
Chimie Minérale
Physique des Plasmas
Physiologie Végétale
Cristallographie
Mathématiques Pures
Chimie Physique
Mécanique,
Géologie
Chimie
Physiologie Animale
Géographie
Mathématiques Pures
Physique
Physique
Mathématiques Appliquées

LAJZEROWICZ Jeanine
 LAJZEROWICZ Joseph
 LAURENT Pierre-Jean
 LEBRETON Alain
 DE LEIRIS Joël
 LHOMME Jean
 LLIBOUTRY Louis
 LOISEAUX Jean-Marie
 LUNA Domingo
 MACHE Régis
 MASCLE Georges
 MAYNARD Roger
 OMONT Alain
 OZENDA Paul
 PAYAN Jean-Jacques
 PEBAY-PEYROULA Jean-Claude
 PERRIER Guy
 PIERRARD Jean-Marie
 PIERRE Jean-Louis
 RENARD Michel
 RINAUDO Marguerite
 ROSSI André
 SAXOD Raymond
 SENDEL Philippe
 SERGERAERT Francis
 SOUCHIER Bernard
 SOUTIF Michel
 STUTZ Pierre
 TRILLING Laurent
 VALENTIN Jacques
 VAN CUTSEM Bernard
 VIALON Pierre

Physique
 Physique
 Mathématiques Appliquées
 Mathématiques Appliquées
 Biologie
 Chimie
 Géophysique
 Sciences Nucléaires I.S.N.
 Mathématiques Pures
 Physiologie Végétale
 Géologie
 Physique du Solide
 Astrophysique
 Botanique (Biologie Végétale)
 Mathématiques Pures
 Physique
 Géophysique
 Mécanique
 Chimie Organique
 Thermodynamique
 Chimie CERMAV
 Biologie
 Biologie Animale
 Biologie Animale
 Mathématiques Pures
 Biologie
 Physique
 Mécanique
 Mathématiques Appliquées
 Physique Nucléaire I.S.N.
 Mathématiques Appliquées
 Géologie

PROFESSEURS de 2^{ème} Classe

ADIBA Michel
 ANTOINE Pierre
 ARMAND Gilbert
 BARET Paul
 BLANCHI J.Pierre
 BLUM Jacques
 BOITET Christian
 BORNAREL Jean
 BRUANDET J.François
 BRUGAL Gérard
 BRUN Gilbert
 CASTAING Bernard
 CERFF Rudiger
 CHIARAMELLA Yves
 COURT Jean
 DUFRESNOY Alain
 GASPARD François
 GAUTRON René
 GENIES Eugène
 GIDON Maurice
 GIGNOUX Claude
 GILLARD Roland
 GIORNI Alain
 GONZALEZ SPRINBERG Gérardo
 GUIGO Maryse
 GUMUCHAIN Hervé
 GUITTON Jacques

Mathématiques Pures
 Géologie
 Géographie
 Chimie
 STAPS
 Mathématiques Appliquées
 Mathématiques Appliquées
 Physique
 Physique
 Biologie
 Biologie
 Physique
 Biologie
 Mathématiques Appliquées
 Chimie
 Mathématiques Pures
 Physique
 Chimie
 Chimie
 Géologie
 Sciences Nucléaires
 Mathématiques Pures
 Sciences Nucléaires
 Mathématiques Pures
 Géographie
 Géographie
 Chimie

HACQUES Gérard
 HERBIN Jacky
 HERAULT Jeanny
 JARDON Pierre
 JOSELEAU Jean-Paul
 KERCKHOVE Claude
 LONGEQUEUE Nicole
 LUCAS Robert
 MANDARON Paul
 MARTINEZ Francis
 NEMOZ Alain
 OUDET Bruno
 PECHER Arnaud
 PELMONT Jean
 PERRIN Claude
 PFISTER Jean-Claude
 PIBOULE Michel
 RAYNAUD Hervé
 RICHARD Jean-Marc
 RIEDTMANN Christine
 ROBERT Gilles
 ROBERT Jean-Bernard
 SARROT-REYNAULD Jean
 SAYETAT François
 SERVE Denis
 STOECKEL Frédéric
 SCHOLL Pierre-Claude
 SUBRA Robert
 VALLADE Marcel
 VIDAL Michel
 VIVIAN Robert
 VOTTERO Philippe

Mathématiques Appliquées
 Géographie
 Physique
 Chimie
 Biochimie
 Géologie
 Sciences Nucléaires I.S.N.
 Physique
 Biologie
 Mathématiques Appliquées
 Thermodynamique CNRS - CRTBT
 Mathématiques Appliquées
 Géologie
 Biochimie
 Sciences Nucléaires I.S.N.
 Physique du Solide
 Géologie
 Mathématiques Appliquées
 Physique
 Mathématiques Pures
 Mathématiques Pures
 Chimie Physique
 Géologie
 Physique
 Chimie
 Physique
 Mathématiques Appliquées
 Chimie
 Physique
 Chimie Organique
 Géographie
 Chimie

MEMBRES DU CORPS ENSEIGNANT DE L' IUT 1

PROFESSEURS de 1^{ère} Classe

BUISSON Roger	Physique IUT 1
DODU Jacques	Mécanique Appliquée IUT 1
NEGRE Robert	Génie Civil IUT 1
NOUGARET Marcel	Automatique IUT 1
PERARD Jacques	EEA. IUT 1

PROFESSEURS de 2^{ème} classe

BOUTHINON Michel	EEA. IUT 1
CHAMBON René	Génie Mécanique IUT 1
CHEHIKIAN Alain	EEA. IUT 1
CHENAVAS Jean	Physique IUT 1
CHOUTEAU Gérard	Physique IUT 1
CONTE René	Physique IUT 1
GOSSE Jean-Pierre	EEA. IUT 1
GROS Yves	Physique IUT 1
KUHN Gérard, (Détaché)	Physique IUT 1
MAZUER Jean	Physique IUT 1
NICHOULIER Jean	Physique IUT 1
MONLLOR Christian	EEA. IUT 1
PEFFEN René	Métallurgie IUT 1
PERRAUD Robert	Chimie IUT 1
PIERRE Gérard	Chimie IUT 1
TERRIEZ Jean-Michel	Génie Mécanique IUT 1
TOUZAIN Philippe	Chimie IUT 1
VINCENDON Marc	Chimie IUT 1

PROFESSEURS DE PHARMACIE

AGNIUS-DELDORD Claudine	Physique	Faculté La Tronche
ALARY Josette	Chimie Analytique	Faculté La Tronche
BERIEL Hélène	Physiologie et Pharmacologie	Faculté La Tronche
CUSSAC Max	Chimie Thérapeutique	Faculté La Tronche
DEMENGÉ Pierre	Pharmacodynamie	Faculté La Tronche
FAVIER Alain	Biochimie	C.H.R.G.
JEANNIN Charles	Pharmacie Galénique	Faculté Meylan
LATURAZE Jean	Biochimie	Faculté La Tronche
LUU DUC Cuong	Chimie Générale	Faculté La Tronche
MARIOTTE Anne-Marie	Pharmacognosie	Faculté La Tronche
MARZIN Daniel	Toxicologie	Faculté Meylan
RENAUDET Jacqueline	Bactériologie	Faculté La Tronche
ROCHAT Jacques	Hygiène et Hydrologie	Faculté La Tronche
SEIGLE-MURANDI Françoise	Botanique et Cryptogamie	Faculté Meylan
VERAIN Alice	Pharmacie Galénique	Faculté Meylan

MEMBRES DU CORPS ENSEIGNANT DE MEDECINE

PROFESSEURS CLASSE EXCEPTIONNELLE ET 1ère CLASSE

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

PROFESSEURS 2ème CLASSE

BACHELOT Yvan	Endocrinologie	C.H.R.G.
BARGE Michel	Neurochirurgie	C.H.R.G.
BENABID Alim Louis	Biophysique	Faculté La 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.
CHIROSSSEL Jean-Paul	Anatomie-Neurochirurgie	C.H.R.G.
COMET Michel	Biophysique	Faculté La Merci
CONTAMIN Charles	Chirurgie Thoracique et	
	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 et Informatique	Faculté La Merci
	Médicale	
DUPRE Alain	Chirurgie Générale	C.H.R.G.
DYON Jean-François	Chirurgie Infantile	C.H.R.G.
ETERRADOSSI Jacqueline	Physiologie	Faculté La 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 La 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	C.H.R.G.
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 Christian	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 La Merci
SELE Bernard	Cytogénétique	Faculté La Merci
SOTTO Jean-Jacques	Hématologie	C.H.R.G.
STOEBNER Pierre	Anatomie Pathologique	C.H.R.G.
VROUSOS Constantin	Radiothérapie	C.H.R.G.

C'était le seul vrai loup de la portée, le seul qui ressemblât au vieux mâle borgne, et il lui ressemblait trait pour trait, comme une caricature boulotte, à l'exception évidemment de son œil droit, qui était aussi ouvert et vivant qu'un œil peut l'être. Il ne l'avait d'ailleurs pas ouvert tout de suite, le gauche non plus. Pendant de longs jours, il avait vécu dans l'obscurité, ignorant que la lumière pût exister, découvrant le monde avec son odorat, son goût, son toucher.

... Pendant près d'un mois, il passa le plus clair de son temps à dormir. Puis il forçit, ses yeux se dessillèrent et il découvrit la lumière. Elle était très faible, à peine un clair-obscur, mais elle lui permettait de distinguer les frontières du monde. L'espace qui entourait sa mère était limité par des parois sombres, très rapprochées, froides et immobiles. Elles définissaient l'univers, ou du moins ce qu'il croyait être l'univers, et comme il n'avait jamais connu autre chose, le manque de place et l'éclairage insuffisant ne le dérangaient nullement. Une des parois n'était cependant pas comme les autres. Elle était plus claire, la lumière y était plus forte, quelque chose en elle attirait irrésistiblement le regard et l'attention.

... L'interdit maternel et la peur le tenait encore éloigné du mur de lumineux, mais la vie qui bouillonnait en lui le poussait irrésistiblement vers la lumière. Chaque goulée d'air qu'il aspirait, chaque morceau de viande qu'il avalait augmentait sa résolution. Finalement, un jour arriva où le flux qui le portait vers l'avant fut plus fort que ses résistances, et le louveteau gris, arc-bouté sur ses pattes courtaudes, s'enhardit à ramper, les oreilles couchées, le cœur battant, jusqu'au seuil de la tanière.

A l'inverse des parois qu'il avait tenté de franchir auparavant, le mur de lumière semblait reculer au fur et à mesure qu'il s'en approchait. Il s'était attendu à ce que sa petite truffe heurte une surface solide qui l'aurait obligé à rebrousser chemin, mais il ne rencontra que le vide, et l'obstacle qu'il avait imaginé dur et rigide, aussi dur et rigide que les autres frontières de l'univers, se révéla aussi tendre et perméable, aussi facile à traverser que l'espace dans lequel il s'était toujours ébattu. Des sentiments contradictoires l'animaient. La stupéfaction lui coupait les pattes. La crainte de l'inconnu l'incitait à battre en retraite. Mais l'exaltation de l'aventure et l'élan vital qui l'avaient poussé à le tenter l'emportèrent sur tout le reste. Il poursuivit sa progression. La lumière devenait de plus en plus forte. Brusquement, il se trouva à l'entrée du tunnel. Quel choc! Libérant une clarté aveuglante, le mur dans lequel il avait cru s'enfoncer venait de franchir d'un bond une distance inimaginable! Etourdi, ébloui, grisé d'air et d'espace, il trébucha et faillit tomber. Puis ses pupilles s'adaptèrent, et le mur qui s'était si soudainement dérobé devant lui réapparut au loin, immense et bariolé, offrant à son regard un somptueux étagement de couleurs : vert sombre des arbres de la rive, bleus lointains, embrumés, des montagnes, bleu plus clair du ciel, immense et serein.

Croc-Blanc de Jack London, Roman traduit par Philippe Sabathe.

Je tiens à remercier

Monsieur Philippe JORRAND, Directeur de recherche au CNRS, Directeur du LIFIA (Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle), qui me fait l'honneur de présider le jury de cette thèse.

Monsieur J. L. CROWLEY, Professeur à l'ENSIMAG, responsable de l'équipe Vision et Robotique mobile au LIFIA, pour m'avoir accueilli dans son équipe, pour la confiance qu'il m'a témoignée durant cette thèse et pour sa gentillesse.

Messieurs S. CASTAN et J. G. POSTAIRE, respectivement Professeur à l'Université Paul Sabatier et Professeur à l'Université de Lille Flandres Artois, d'avoir bien voulu apprécier le travail de cette thèse.

Monsieur le Professeur A. CHEHIKIAN, Directeur du Laboratoire LTIRF, qui me fait l'honneur de participer au jury de thèse. De plus, il a eu la patience et la gentillesse de m'expliquer de nombreux aspects du traitement du signal.

Monsieur P. STELMASZYK, chef de projet au département recherche à ITMI (Industrie et Technologie de la Machine Intelligente), qui me fait l'honneur de juger le travail de cette thèse. Je tiens aussi à lui décerner la médaille du mérite et de l'amitié pour l'aide constante, aussi bien morale que technique, qu'il m'a apportée durant ma thèse.

Monsieur G. MEZIN, Président Directeur Général de la société ITMI, pour la confiance qu'il m'a témoignée en m'intégrant dans sa société.

Madame S. EUDES (Froggy) et son époux qui ont admirablement "subi" la rébarbative "corvée" des corrections des fautes de français. Ils ont également assuré de multiples fois, ainsi que Monsieur F. RAMPARANY et son épouse, la fastidieuse tâche de m'éviter de manger au restaurant universitaire.

Madame M. E. STELMASZYK, pour avoir accepté de rester seule de nombreux week-end, pendant que son mari m'aidait dans mes recherches et dans la rédaction de ma thèse.

Les membres du LIFIA, et en particulier ceux de l'équipe VISION : J. CROWLEY, Y. DEMAZEAU, C. DISCOURS, J. L. GORDILLO, R. HORAUD, A. LUX, F. RAMPARANY,

ainsi que les membres de l'équipe VISION dans le département Recherche à ITMI : Y. DESCOTTE, L. LEFORT, R. PESTY, H. PHAM, T. SKORDAS, J. N. SOULIER, P. STELMASZYK, R. ZIGMANN. Ils m'ont bien conseillé, fourni les bons documents au moment opportun et ont créé une ambiance formidable tout au long de ma thèse. Je tiens aussi à souligner l'aide qu'ils m'ont offert en m'aidant à coller les nombreuses photos qui illustrent ma thèse.

Les secrétaires du LIFIA et d'ITMI, pour leurs aides et leurs merveilleux sourires.

Le Laboratoire LIFIA et la société ITMI, pour l'environnement matériel et financier qu'ils m'ont offert en vue de la réalisation de cette thèse.

Macintosh, Macwrite et Macdraw pour leurs convivialités.

Une partie des travaux présentés a été supporté financièrement par des projets Européens de recherche Esprit P940 et Mithra.

I INTRODUCTION	1
I.1 Position du problème	1
I.2 Notre approche et nos apports	2
I.3 Organisation du mémoire	3
 II REPRESENTATION DE L'IMAGE	 5
II.1 Introduction	5
II.2 Les indices images	6
II.2.1 Extraction des indices images	6
II.2.1.1 Les points caractéristiques	7
II.2.1.2 Les contours	9
II.2.1.3 Les régions	18
II.2.2 Caractéristiques des indices	19
II.2.2.1 Attributs des indices images	19
II.2.2.2 Robustesse des attributs	21
II.2.3 Conclusion	22
II.3 Structures d'indices	23
II.4 Les relations entre les indices	24
II.4.1 Choix du critère relationnel	24
II.4.2 Détermination des relations	26
II.5 Conclusion	27
 III CORRESPONDANCE	 28
III.1 Introduction	28
III.2 Correspondance au niveau pixel	29
III.2.1 Introduction	29
III.2.2 Méthodes de corrélation	30
III.2.3 Méthodes de différences d'images	31
III.2.4 Méthodes d'analyse du gradient spatio-temporel.	32
III.2.5 Conclusion	33

IV.3.1.1 La détection	71
IV.3.1.2 Limitation de la détection	73
IV.3.1.3 Amélioration de la détection	74
IV.3.1.4 Résultats	75
IV.3.1.5 Caractéristiques sur les pics	80
IV.3.1.6 Conclusion	82
IV.3.2 Les crêtes	83
IV.3.2.1 La détection	83
IV.3.2.2 Résultats	85
IV.4 Les arbres de pics	87
IV.4.1 La détection	87
IV.4.2 Limitations	88
IV.4.3 Caractéristiques	88
IV.4.4 Résultats	90
IV.4.5 Conclusion	91
IV.5 Relations	91
IV.6 Conclusion	93

V APPARIEMENT PAR SIMILITUDE D'ARBRES DE PICS

V.1 Introduction	94
V.2 Mise en correspondance	94
V.2.1 Zone de recherche	95
V.2.2 Appariement par similarité	96
V.2.2.1 Caractéristiques discriminantes	96
V.2.2.2 Fonction de dissemblance	97
V.2.2.3 Génération des appariements	97
V.3 Applications et évaluations des résultats	98
V.4 Détermination du domaine d'application	105
V.5 Conclusion	109

VI APPARIEMENT DE PICS AVEC HERITAGE ET CLIQUE MAXIMALE	110
VI.1 Introduction	110
VI.2 Principes de la méthode	110
VI.2.1 Introduction	110
VI.2.2 Coarse to fine	111
VI.2.3 Présentation de la méthode	113
VI.3 Mise en correspondance	115
VI.3.1 Zone de recherche	115
VI.3.1.1 Contrainte épipolaire	116
VI.3.1.2 Zone définie par disparité	116
VI.3.2 Création des hypothèses d'appariement	118
VI.3.3 Compatibilité entre les hypothèses	118
VI.3.3.1 Compatibilité relationnelle	119
VI.3.3.2 Evaluation de la compatibilité	122
VI.3.3.3 Recherche de la plus grande clique maximale	122
VI.4 Applications et résultats	124
VI.5 Conclusion	134
 VII CONCLUSION	 136

BIBLIOGRAPHIE

ANNEXE 1

ANNEXE 2

ANNEXE 3

I INTRODUCTION

I.1 Position du problème

La reconstruction tridimensionnelle de l'environnement perçu par un ensemble de caméras requiert la mise en correspondance des données acquises. L'appariement est une étape importante qui conditionne la validité des processus ultérieurs. Dans cette thèse, nous développons ce thème de "mise en correspondance" et proposons des techniques originales.

L'image étant composée de pixels, la quantité d'informations est énorme (une grande partie de celles-ci étant redondantes ou inutiles). Par conséquent, il est souvent intéressant d'acquérir une représentation de l'image explicitant les informations utiles face au but à atteindre. Cette phase initiale de représentation de l'image requiert une attention particulière car elle va déterminer les possibilités de réalisation, l'efficacité et la validité des traitements ultérieurs.

D'une représentation de l'image sous forme d'une matrice de pixels, on passe à une représentation sous forme d'indices images et souvent, à une représentation sous forme de graphe d'indices image dans lequel les arcs sont des relations contextuelles entre certains indices. Il s'agit maintenant d'apparier les indices entre des images.

L'un des problèmes capitaux de la mise en correspondance concerne l'ambiguïté d'appariement. En effet, pour un indice donné dans une image, il est généralement possible de lui associer de nombreux indices sur l'image suivante. Pour restreindre ce problème, il faut pouvoir se donner les moyens de limiter l'ensemble potentiel des correspondants et/ou rechercher le correspondant le plus vraisemblable.

Ainsi, une première solution est l'utilisation d'indices très discriminants, en considérant qu'il n'existe généralement qu'un seul choix possible de correspondant.

Une autre possibilité, fréquemment utilisée dans les systèmes de vision actuels, consiste à utiliser les informations contextuelles contenues dans les images, c'est à dire prendre en compte les arcs des graphes de représentation des images. Cette approche permet de trouver un

ensemble cohérent d'hypothèses d'appariement selon le critère contextuel.

Dans cette dernière solution, les coûts d'appariement sont souvent élevés. Pour limiter ce problème, il est souhaitable d'utiliser un partitionnement efficace des images. Efficace signifiant qu'il réduit le coût et engendre un nombre limité d'erreurs d'appariement.

L.2 Notre approche et nos apports

Nous proposons des techniques de mise en correspondance robustes et fiables pour la reconnaissance d'objets ou pour la reconstruction tridimensionnelle. Notre approche consiste à utiliser une technique multirésolutions dans laquelle l'information est répartie de manière différente à divers niveaux. Ainsi, l'information globale est généralement détectée dans les images de faible résolution, et l'information locale (à savoir les détails) se retrouve dans les niveaux supérieurs.

Avant de développer nos techniques de correspondance, nous nous proposons de généraliser la problématique d'un système de mise en correspondance et de mettre en évidence les différentes étapes d'un tel système. Cette généralisation, illustrée par de nombreux exemples, contribue à la compréhension des algorithmes de correspondance existants.

Nos applications utilisent comme primitives à apparier les extrema locaux dans les images de Laplacien. L'apport de notre travail dans l'extraction de ce type d'indices consiste en une amélioration considérable de leur robustesse au moyen de calculs basés sur l'analyse de la courbure et de la concavité locale.

Un algorithme de mise en correspondance structurelle est alors proposé. Il regroupe les informations contenues dans différents niveaux de résolution en vue de créer une structure. La mise en correspondance de ces structures fait apparaître une combinatoire faible, une simple fonction de coût permettant de quantifier et de mettre en évidence les appariements. Les expérimentations indiquent le domaine d'application d'un tel système, ainsi que ses principales limitations.

Ces expérimentations nous amènent à proposer une technique de mise en correspondance qui exploite l'information contextuelle présente dans les images. Cette technique utilise un schéma

de type "coarse to fine". L'apport majeur de notre approche réside dans l'introduction d'un critère global de mise en correspondance, qui améliore considérablement la robustesse des techniques de "coarse to fine" en limitant les problèmes de faux appariements et de propagations erronées de ceux-ci à travers les divers niveaux de résolution. En contre-partie, l'utilisation d'un schéma de type "coarse to fine", permet de limiter le problème de l'explosion combinatoire qui caractérise souvent les techniques basées sur l'utilisation de critères globaux d'appariement.

I.3 Organisation du rapport

Notre présentation s'articule autour des cinq chapitres suivant :

Dans le chapitre II, nous présentons une synthèse sur les techniques générales permettant d'extraire une représentation de l'image adaptée à l'application envisagée. Cette phase de prétraitement de l'image fournit les primitives à appairer.

Diverses techniques permettent de mettre en correspondance des indices images. Cependant, une étude détaillée fait ressortir un schéma général découpé en trois phases qui sont:

- le calcul des zones de recherche des appariements
- la création d'un ensemble d'hypothèses d'appariement sur un critère de similitude
- la recherche d'un sous-ensemble d'hypothèses mutuellement compatibles.

Cette synthèse fait l'objet du chapitre III.

Au chapitre IV, nous détaillons les représentations des images adoptées. Deux types de représentation ont été utilisés :

- les arbres de pics
- les pics et leurs relations spatiales

L'utilisation d'indices de types différents nous conduit à proposer deux systèmes différents d'appariement.

Le chapitre V présente un système basé sur la représentation des indices sous la forme d'arbres de pics. L'appariement est effectué uniquement sur un critère local de similarité.

Les limites propres à cette technique d'appariement nous amènent à envisager un système utilisant les primitives pics et un critère global de mise en correspondance. Un schéma "coarse to fine" permet de limiter les coûts en temps de calcul. Cette réalisation fait partie du chapitre VI.

II REPRESENTATION DE L'IMAGE

Ce chapitre fait la synthèse des méthodes permettant de passer de la représentation de l'image sous forme d'une matrice de pixels à une représentation plus adaptée à l'application.

II.1 Introduction

Une image numérique de N lignes et M colonnes est constituée de $N*M$ pixels. Ainsi, le nombre de données pixels peut être considérable si les images traitées sont de taille importante. Les conséquences en sont des temps de traitement longs et une occupation de mémoire importante. Les principales caractéristiques d'un pixel sont sa position et son niveau de luminance (ou des niveaux chromatiques dans les images couleurs). Ainsi les informations sur l'image qu'il véhicule sont pauvres et très sensibles aux variations dans la scène. Il est préférable alors, dans certaines réalisations, de travailler à partir de données plus denses en informations contenues dans l'image en procédant à un changement de niveau de représentation.

D'une représentation image de pixels, on passe à une représentation beaucoup plus dense en indices visuels. Ceux-ci concentrent certaines propriétés locales sur l'image. En vue de prendre en compte des informations contextuelles de l'image, des relations entre les indices peuvent alors être créées. Donc de la matrice de pixels initiale, on passe ainsi à une représentation sous forme d'un graphe d'indices (cf figure II.1).

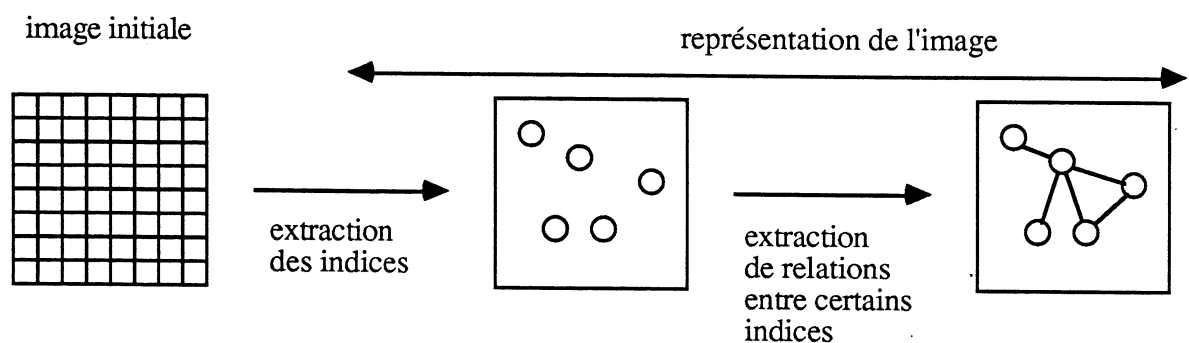


Figure II.1: Changement de niveau de représentation.

L'acquisition de cette représentation de l'image passe par une phase d'extraction des indices, puis parfois par un regroupement de ceux-ci sous forme de structures et en dernier lieu par une phase de création des relations entre eux. Ces trois phases sont traitées dans leur ordre chronologique dans les chapitres suivants.

II.2 Les indices images

En vue de limiter la quantité des données à traiter et d'accéder directement à des informations pertinentes et suffisantes pour l'analyse de la scène, on met en oeuvre une phase de prétraitement de l'image qui consiste à extraire des **indices images**. Ces indices correspondent à la projection d'**indices de scène** tels les sommets des objets, leur contour ou leur élément de surface. Ainsi, les techniques permettant d'extraire les indices images sont basées sur la connaissance des conséquences qu'ont, par projection perspective sur l'image, les propriétés caractérisant les indices scènes. Par exemple une arête sur un objet peut être définie comme étant l'intersection de deux surfaces non coplanaires de cet objet. Ces deux surfaces reçoivent donc dans la scène un éclairage différent. Par projection, les deux régions résultantes ont des niveaux d'intensité différents. L'indice image de type contour correspond alors à la zone de discontinuité en intensité séparant les deux régions.

Des travaux réalisés jusqu'à présent, il ressort 3 types d'indices images (classification de [Bolles 82]):

- les points caractéristiques
- les contours
- les régions

Nous présentons ci-dessous les techniques d'extraction de ces indices, leurs avantages et inconvénients; puis dans une seconde section, les attributs qu'il est possible d'affecter à chaque type d'indice.

II.2.1 Extraction des indices images

Cette section est une présentation de la caractérisation de chaque type d'indice image, et des techniques résultantes permettant de les extraire. Elle fournit aussi une énumération, pour chaque type d'indice, des causes d'extraction de "faux" indices (en terme de projection d'indices ayant une signification physique) ou de non extraction de "vrais" indices.

II.2.1.1 Les points caractéristiques

Certains points dans la scène possèdent des caractéristiques particulières. C'est le cas des sommets d'un objet, des points de maximum de courbure ou d'inflexion le long de leur contour 3D, ou des points de courbure maximale ou d'inflexion sur une surface. Il peut être intéressant, selon l'application à traiter, de baser l'appariement sur la projection de tels points.

L'une des premières questions à laquelle nous devons répondre est la suivante: comment extraire de tels points des images?

Ces points dans l'image, se caractérisent par une grande variation en luminance ou le long des contours par rapport au voisinage. La détection de grandes variations locales en luminance a donné naissance à des opérateurs d'extraction travaillant directement sur la luminance de l'image, celle des variations locales sur les contours à des opérateurs travaillant sur les contours préalablement extraits de l'image. Ces deux catégories d'algorithmes d'extraction sont présentées ci-dessous. Pour une comparaison portant sur les résultats de diverses techniques se référer à [Dreschler 83].

Les opérateurs d'extraction travaillant directement sur la luminance de l'image.

Dans ce cas, la détection des points d'intérêts est effectuée par une évaluation locale de la fonction intensité de l'image. Ainsi, le bruit et l'échantillonnage peuvent perturber la détection des points d'intérêts.

[Moravec 77][Moravec 81] s'intéresse à l'extraction des maximaux locaux en variance dans l'image. La variance directionnelle est mesurée sous un petit masque de taille spécifiée (4x4 à 8x8). On calcul alors la somme des carrés des différences de pixels adjacents dans chacune des quatre directions (horizontale, verticale, et les deux diagonales), la variance représentant le minimum des quatre valeurs. Chaque point de l'image maximal local par rapport au voisinage, et supérieur à un seuil fixé est alors extrait.

[Crowley 81] convolue l'image avec un opérateur de différence gaussienne. De l'image de passe-bande obtenue, il extrait les pics qui sont les maximaux locaux du laplacien de l'image. Dans le chapitre suivant nous détaillerons davantage cette réalisation.

Les opérateurs d'extraction travaillant sur les contours extraits.

Dans ce second cas, des contours sont extraits dans l'image. Ils correspondent, comme nous le verrons dans la section suivante, aux crêtes maximales du gradient de l'image ou

aux passages à zéro du laplacien de l'image. Le long de ces contours sont alors détectés les points de courbure maximale ou les points d'inflexion. Une autre technique consiste à extraire les points d'intersection de segments, ce qui donne souvent les sommets des objets [Rueb 87]. Par ces méthodes, la détection des points est donc sensible aux erreurs de segmentation et diffère selon l'utilisation d'un gradient ou d'un laplacien. Les erreurs se produisent surtout aux environs des points d'intérêts.

Ayant extrait les points caractéristiques, on est amené à répondre à la question suivante qui est: Les points extraits sont-ils effectivement les images de points caractéristiques de la scène?

Les techniques utilisées montrent que les variations de luminance et le bruit dans les images provoquent la détection de tels indices. De même, des phénomènes dus à la projection perspective engendrent des points de courbure maximale là où il n'y a pas de point physique correspondant (cf. figure II.2 tirée de [williams 81]).

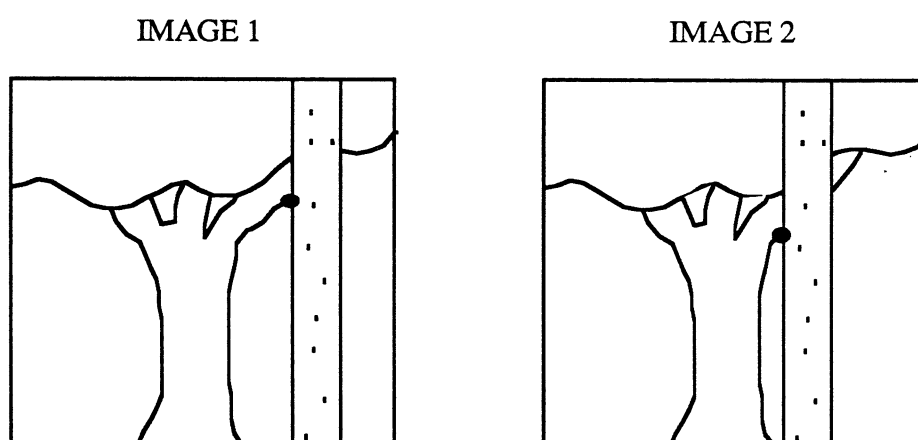


Figure II.2: Dans ces deux images successives, contenant un arbre et un poteau télégraphique, le point noir représente un "faux" point caractéristique (dans le sens de point ayant une signification physique).

La déformation perspective de certaines formes engendre des courbures maximales le long des contours. Par exemple, la base d'un cylindre sera perçue comme une ellipse, ce qui donne deux points de courbure maximale, alors que ceux-ci n'existent pas en réalité. Par contre, un faible contraste empêche la détection de certains indices visuels de type point.

Les indices de type "point caractéristique" possèdent une quantité d'informations réduite pour décrire l'image. De ce fait, on lui préfère souvent d'autres types d'indices tels que les contours.

II.2.1.2 Les contours

Il existe d'autres caractéristiques de forme: les contours, les segments de droites, les arcs de cercles. Un contour dans la scène sépare des surfaces de position, de réflectance ou de texture différentes, ce qui se traduit dans l'image par une zone de transition séparant des régions de luminance ou de texture différentes.

La définition la plus employée caractérisant les contours dans les images est la suivante: *un contour dans une image est une zone de transition séparant deux zones d'intensité homogène moyenne différentes.*

Les contours constituent le type d'indice mentionné dans le plus grand nombre de travaux du fait de leur représentation géométrique riche par rétroprojection. Ceux-ci sont extraits soit à partir d'un gradient, soit à partir d'un laplacien. Il convient tout d'abord de calculer le gradient ou le laplacien soit sur la totalité de l'image, soit localement suivant la direction donnée par les calculs précédents.

Le gradient:

Il permet de mettre en évidence de fortes variations d'intensité lumineuse. En posant $I(i,j)$ la fonction image, le gradient de la fonction image est défini comme suit:

$$\overrightarrow{\text{GRAD}(I(i,j))} = \begin{pmatrix} \frac{\delta I(i,j)}{\delta i} \\ \frac{\delta I(i,j)}{\delta j} \end{pmatrix}$$

Dans le cas d'images numériques, le gradient est approximé par des filtres à réponse impulsionnelle finie. Ils sont souvent limités à de trop petites tailles en raison du temps de calcul et leurs coefficients sont souvent des multiples de 2 pour des facilités d'implémentation. Les plus usités sont :

Roberts :

$$F = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

composante -45° composante $+45^\circ$ du gradient

$$\text{module} = \sqrt{[F(I(i,j))]}^2 + [G(I(i,j))]}^2$$

$$\arg = \frac{\pi}{4} + \arctg \frac{F(I(i,j))}{G(I(i,j))}$$

Prewitt :

Il convolue un filtre de dérivée première $[1 \ 0 \ -1]$ avec un filtre passe-bas $[1 \ 1 \ 1]$, ce qui donne les filtres F et G ci-dessous.

$$F = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad G = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

composante 90° composante 0° du gradient

$$\text{module} = \sqrt{[F(I(i,j))]}^2 + [G(I(i,j))]}^2$$

$$\arg = \arctg \frac{F(I(i,j))}{G(I(i,j))}$$

Son filtre étant séparable, il permet de limiter le nombre d'opérations lors de la convolution entre le filtre et l'image. Au lieu de convoluer l'image avec le filtre F (ou le filtre G), les lignes sont convoluées par le filtre passe-bas $[1 \ 1 \ 1]$ et les colonnes par le filtre $[1 \ 0 \ -1]$. La convolution par le filtre passe-bas permet de limiter les hautes fréquences dans l'image, mais on lui préfère le filtre de Sobel, qui procède à une convolution avec un filtre passe-bas plus performant.

Sobel :

Sobel [Duda 73] convolue un filtre de dérivée première $\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$ avec un filtre passe-bas $\begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$, ce qui lui donne les filtres F et G ci-dessous.

$$F = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

composante 90°

$$G = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

composante 0°

du gradient

$$\text{module} = \sqrt{[F(I(i,j))]^2 + [G(I(i,j))]^2}$$

$$\text{arg} = \text{arctg} \frac{F(I(i,j))}{G(I(i,j))}$$

Comme pour le Prewitt, ce filtre possède l'avantage d'être séparable. Par contre comme ceux présentés précédemment, il n'est pas obligatoirement optimal. De nouveaux filtres ont alors été définis à partir de critères d'optimalité.

Canny:

Canny [Canny 86] modélise un contour par un échelon unidimensionnel infini d'amplitude A superposé à un bruit blanc gaussien de moyenne nulle et d'écart type σ . Ensuite, il définit des critères de "bon filtre" pour la détection des contours, et recherche le filtre à support borné répondant à ceux-ci. Les trois critères qu'il utilise sont:

Une bonne détection: Ce critère revient à maximiser le rapport **signal** sur **bruit**.

Une bonne localisation: Les points détectés étiquetés comme contour doivent être aussi près que possible du contour réel.

Une réponse unique pour un seul contour: Le premier critère semble répondre à ce problème, cependant sa formulation mathématique montre qu'il n'en est rien. Il permet d'éliminer les effets d'écho provoqués par le bruit.

Canny approche, pour des raisons de facilité d'implémentation, la solution obtenue en une dimension par la dérivée d'une gaussienne (cf. figure II.3).

$$g'(r) = -\frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

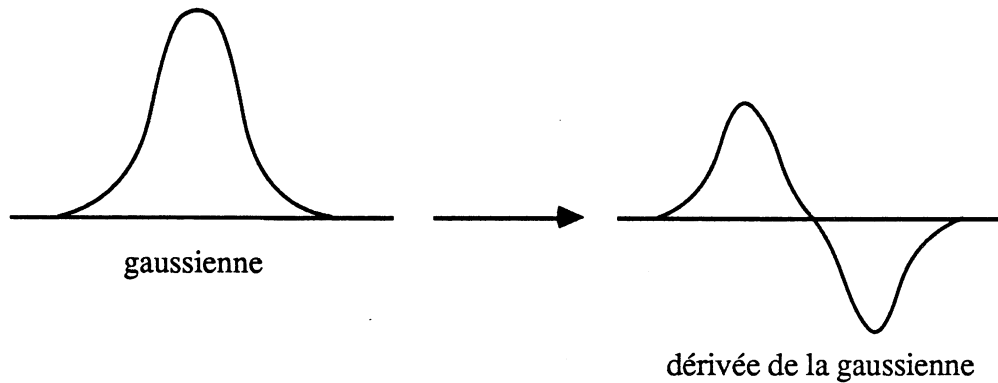


Figure II.3: Dérivée d'une gaussienne.

Le gradient en deux dimensions correspond à:

$$\overrightarrow{\text{GRAD}}(I(i,j)) = \begin{pmatrix} \frac{\delta}{\delta i} (g(j)*g(i)*I(i,j)) \\ \frac{\delta}{\delta j} (g(j)*g(i)*I(i,j)) \end{pmatrix} = \begin{pmatrix} g(j)*g'(i)*I(i,j) \\ g'(j)*g(i)*I(i,j) \end{pmatrix}$$

Deriche [Deriche 87] améliore les critères d'optimalité de Canny en introduisant des fonctions à support infini. La solution qu'il obtient en une dimension est de la forme:

$$f(x) = -cx \exp(-\alpha|x|) \text{ avec } c \text{ et } \alpha \text{ réels positifs constants.}$$

Il implémente f de manière récursive. Il en résulte est un coût moindre en nombre d'opérations que le calcul de la convolution d'une image avec une fonction à support étendu, et cela requiert un nombre constant d'opérations par points de l'image quelle que soit la taille du filtre utilisé.

Le gradient en deux dimensions obtenu est:

$$\overrightarrow{\text{GRAD}}(I(i,j)) = \begin{pmatrix} F(j)*f(i)*I(i,j) \\ f(j)*F(i)*I(i,j) \end{pmatrix}$$

avec F une primitive de f.

$$F(x) = -c (1 + \alpha|x|) \exp(-\alpha|x|)$$

Jacquot [jacquot 88] obtient des résultats comparables à ceux de Deriche en proposant une architecture matérielle simple. Cette approche doit permettre l'extraction des contours à cadence vidéo.

Le Laplacien :

$$\Delta I(x,y) = \delta^2 I(x,y) / \delta^2 x + \delta^2 I(x,y) / \delta^2 y$$

Les images étant digitalisées, le laplacien est approximé par des masques. Parmi ceux-ci on peut noter:

Les laplaciens classiques :

$$M = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad \text{où} \quad M = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Robinson :

$$M = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

Ces techniques présentent l'avantage de ne nécessiter qu'une seule convolution (au lieu de deux pour les techniques de gradient). Bien que les temps de calcul soient ainsi réduits, il faut souligner l'absence de l'information sur l'orientation des points de contraste extraits. De plus, le bruit de haute fréquence est accentué par le laplacien.

Marr et Hildreth :

Inspiré par les travaux de Wilson et Bergen en neurophysiologie sur les systèmes visuels des mammifères, Marr et Hildreth [Marr 80] ont développé une technique de détection de contour, à diverses résolutions, basée sur la convolution de l'image avec des laplaciens de gaussien de tailles différentes. Partant de la constatation [Marr 76a][Marr 76b] que dans les images naturelles les changements d'intensités apparaissent dans un intervalle d'échelle important et qu'un filtre ne peut être optimal à toutes les échelles, il faut utiliser divers filtres, chacun étant optimal à une fréquence donnée. Les réponses obtenues par l'application sur l'image de chacun de ces filtres doivent ensuite être combinées pour extraire toutes les informations pertinentes et utiles de l'image (au sens de l'application).

A une échelle donnée, pour détecter les changements d'intensité dans l'image, il convient tout d'abord de trouver un filtre qui permette de réduire l'intervalle d'échelle dans lequel sont perçus les changements d'intensité. Le filtre optimal adopté est une gaussienne G .

$$G(i,j) = \frac{1}{2\pi\sigma^2} \exp \left(-\frac{i^2 + j^2}{2\sigma^2} \right)$$

où l'écart type σ permet en le faisant varier, de changer d'échelle.

Ce filtre passe-bas a de plus l'avantage de réduire le bruit de dérivée seconde introduit lors de la seconde étape. En effet, pour détecter les changements d'intensité dans l'image, Marr et Hildreth vont rechercher les passages à zéro de la dérivée seconde du produit de convolution entre la gaussienne et l'image.

$$f(i,j) = \Delta(G(i,j) * I(i,j))$$

Or d'après la règle de dérivation appliquée à la convolution,

$$f(i,j) = \Delta(G(i,j)) * I(i,j)$$

$$\Delta(G(i,j)) = \frac{1}{2\pi\sigma^4} \left(\frac{i^2 + j^2}{\sigma^2} - 2 \right) \exp \left(-\frac{i^2 + j^2}{2\sigma^2} \right)$$

ΔG est approximé par l'opérateur de Wilson. Ce dernier correspond à une différence de deux gaussiennes d'écart type différent (Marr et Hildreth ont choisi une différence d'écart type de 2). Cet opérateur de Wilson se dénomme dans la littérature DOG (DOG : Difference Of Gaussian) (cf. figure II.4)

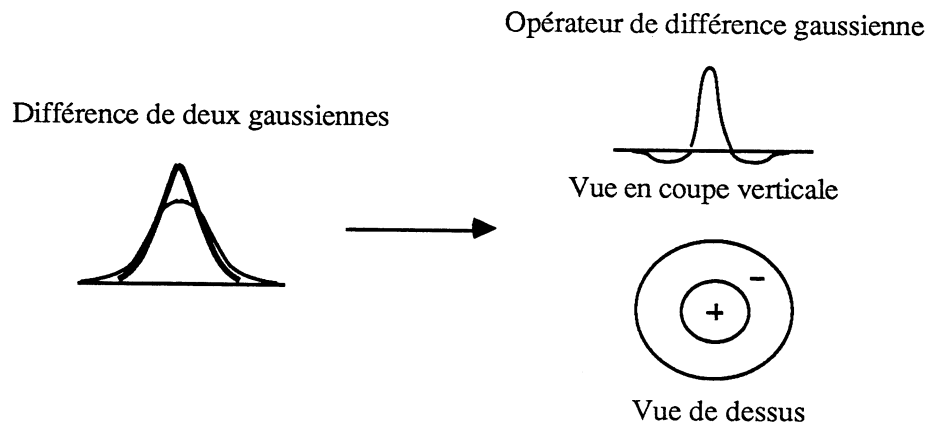


Figure II.4: Vue en coupe verticale et horizontale de l'opérateur de différence gaussienne.

Crowley :

En s'appuyant sur des résultats expérimentaux prouvant que l'homme et certains mammifères séparent les images en diverses bandes de fréquences, Crowley développe une technique permettant d'obtenir une pyramide multi-résolution d'images. Cette théorie, connue sous le nom de "théorie multi-canaux" peut être mise en évidence en mesurant l'adaptation d'un seuil de sensibilité sur des fonctions sinusoidales de diverses fréquences. Pour calculer le laplacien à diverses résolution, Crowley aboutit à un opérateur, DOLP (DOLP : Difference Of Low Pass), similaire à celui utilisé par Marr et Hildreth. L'avantage majeur de son approche est le développement d'un moyen rapide de calculer la transformée de DOLP [Crowley 84].

Après l'application d'un de ces opérateurs sur l'image, il convient d'extraire les points de contours. Ceux-ci correspondent à un maximum sur les crêtes pour les gradients et à un passage à zéro pour les laplaciens (cf. figure II.5).

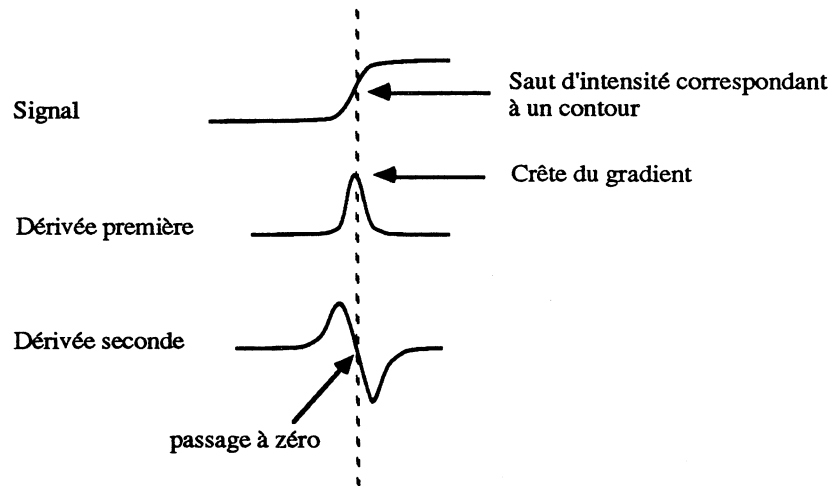


Figure II.5: Détection des contours dans le gradient et le laplacien de l'image.

La recherche des crêtes, contrairement à la recherche des passages à zéro, fournit un nuage de points autour des zones de contraste. Il faut donc amincir ces zones pour obtenir des contours d'épaisseur un pixel. Un seuillage est ensuite nécessaire pour limiter les points de contraste dûs au bruit.

Le résultat de ces opérations conduit à une image binaire composée de points de contraste. Certains travaux utilisent directement les points de contraste [Marr 77], d'autres au contraire travaillent à partir d'une représentation dérivée plus synthétique tels les segments de droites et/ou les arcs de cercle. Il doivent utiliser une étape supplémentaire qui consiste à chaîner les points de contraste.

Le chaînage des points de contraste produit les contours dans l'image. Une élimination des petites chaînes (seuil sur la longueur) ou une prolongation des chaînes intéressantes (par hystérésis par exemple [Giraudon 86a]) est couramment utilisée pour réduire les effets de bruit et de mauvais contraste. Cependant, un trop fort seuillage peut éliminer des contours intéressants; c'est donc souvent à partir du type de scène à traiter que le seuil est déterminé empiriquement ou statistiquement. Une autre méthode consiste à extraire les contours à diverses fréquences (techniques de multi-résolution), et à les améliorer en tenant compte des informations à toutes les résolutions.

L'ultime étape vise alors à acquérir des segments de droites ou des arcs de cercles à partir des contours. Un aperçu des différentes techniques d'obtention de segments est fourni par [Lux 85].

La définition d'un contour dans l'image n'implique pas que le contour extrait corresponde physiquement à un contour dans la scène: une variation de luminance, les ombres, les reflets, le bruit dans l'image (malgré le seuillage vu précédemment et l'élimination des petites chaines), provoquent l'apparition de tels indices. Certaines formes, comme les sphères ou les cylindres, possèdent un contour par projection perspective qui n'existe pas en réalité dans la scène. Des cassures de contours sont possibles à cause d'instabilités dues aux algorithmes de détection, aux reflets, aux bruits, aux ombres ou au faible contraste. Cela crée plusieurs indices contours au lieu d'un seul initialement. Par contre des contours ne sont pas détectés si le contraste est trop faible, (lors de la fusion d'indices due à des alignements accidentels par projection perspective) ou si la longueur de la pente de la transition est supérieure à la taille de l'opérateur utilisé (cf. figure II.6 extraite de [Huertas 86]).

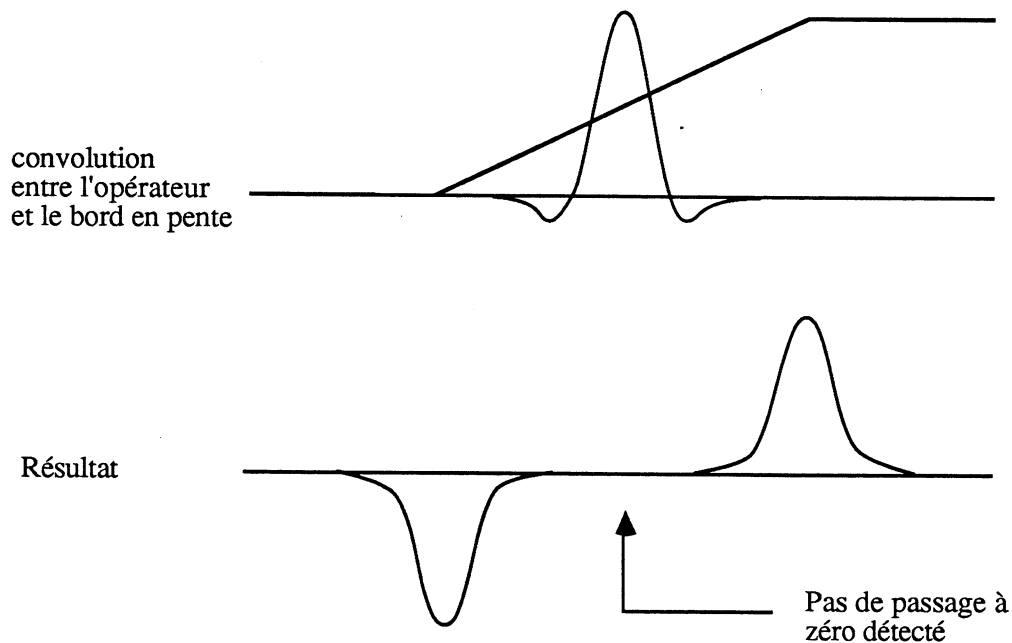


Figure II.6: La taille du contour en pente est plus grande que celle de l'opérateur. De ce fait aucun passage à zéro n'est détecté. La détection d'un contour en pente requiert que la taille de l'opérateur soit supérieure à celle du contour en pente.

Bien que riche en informations sur l'image, cet indice est supplanté dans certaines réalisations par un indice encore plus riche, l'indice de type région.

II.2.1.3 Les régions

Un autre indice de scène correspond aux surfaces des objets. C'est une donnée plus riche que les contours car elle possède non seulement les informations contours (frontières des régions), mais aussi les paramètres intrinsèques aux régions comme le niveau de luminance ou la texture. Les surfaces étant généralement homogènes dans la scène, elles se projettent dans l'image sous forme de régions d'intensité ou de texture homogène. Les régions sont alors souvent définies comme un ensemble de points connexes possédant les mêmes caractéristiques (intensité lumineuse, couleur, texture...).

Il existe deux techniques duales pour extraire les régions des images :

La recherche des contours. la recherche des régions.

Ces techniques sont duales en ce sens qu'une région définit un contour par sa frontière et qu'un contour fermé définit une région. Toutefois, les résultats obtenus par ces deux méthodes sont différents.

La recherche des contours (critère de discontinuité)

Chaque ligne de contraste fermée définit une région (cf. l'extraction des contours ci-dessus). Elle est basée sur la recherche de discontinuités au sein de l'image. Le bruit du capteur ou les conditions d'éclairage, influent peu sur la détection des contours. Par contre, il est parfois difficile et coûteux en temps de détecter des contours fermés.

La recherche des régions (critère de continuité)

Un bon aperçu des diverses méthodes de recherche de régions est présenté dans [Rosenfeld 82]. Il en ressort que cette technique peut être découpée en deux classes :

fusion de régions (connue dans la littérature sous le vocable de "region merging"
ou "region growing")

cassure de régions (connue dans la littérature sous le vocable de "region splitting")

Dans la première classe [Gambotto 85], les techniques sont de type "local" vers "global". Elles démarrent avec les pixels et les étiquettes en fonction des caractéristiques choisies. Les pixels voisins ayant même étiquetage seront fusionnés au sein d'une même région. Cette méthode permet de détecter de petits détails dans l'image. Par contre elle est très sensible au bruit.

Dans la seconde classe, les techniques sont de type "global" vers "local". Elles démarrent avec l'image entière et segmentent cette dernière. La segmentation peut être obtenue par l'application de divers seuils, ceux-ci seuils pouvant être calculés à l'aide de l'histogramme de l'image. Cette méthode permet de détecter des formes globales dans l'image, mais il est difficile d'obtenir les petits détails.

En vue d'accroître la qualité de la recherche des régions et la vitesse de traitement, [Horowitz 74] fait coopérer les deux classes de méthodes. Ce procédé donne des algorithmes connus sous le nom de division-et-fusion ("split-and-merge method").

[Wrobel 86a] [Wrobel 86b] [Monga 87] ont choisi de faire coopérer les techniques de recherche de contours et de recherche des régions. Ainsi, la segmentation en région tient compte des points de contrastes obtenues par une extraction préalable, ce qui permet d'améliorer le partitionnement de l'image.

Bien que l'indice visuel de type région soit le plus riche en information sur l'image, le découpage en régions entre images successives n'est pas stable et souvent coûteux en temps d'exécution: On note l'apparition de régions due aux reflets, aux ombres, aux cassures d'une région initiale à cause des phénomènes lumineux, ou la disparition de régions due à la fusion de plusieurs régions si le contraste est faible entre elles, ou à cause d'ombres et de reflets qui peuvent créer une continuité d'intensité entre plusieurs régions.

II.2.2 Caractéristiques des indices

Cette section donne un inventaire des divers attributs couramment associés à chaque type d'indices et traite ensuite de leur robustesse.

II.2.2.1 Attributs des indices images

Lorsqu'on exécute du traitement sur des indices, c'est en fait sur les caractéristiques de ces derniers qu'il est appliqué. Dire qu'on compare deux droites n'a de sens que si nous sous-entendons que ce test s'effectue par exemple sur leur caractéristique longueur ou sur leur orientation. Il existe deux types de caractéristiques :

- les caractéristiques d'ordre géométrique
- les caractéristiques d'ordre photométrique.

A chaque indice choisi, est associé un ensemble de caractéristiques. Nous citons, dans cette partie, diverses caractéristiques couramment utilisées dans la littérature selon le type d'indice choisi :

Pour les points caractéristiques :

- * le niveau d'intensité
- * la valeur des voisins
- * le signe du point
- * l'étiquetage

Pour les segments, les arcs de cercles, les contours :

- * la longueur
- * le rayon de courbure
- * l'intensité moyenne à droite et à gauche, contraste
- * le nombre de jonctions
- * l'orientation de segment
- * les coordonnées des points extrémité

Pour les régions :

- * la valeur d'intensité moyenne de la région
- * l'aire
- * les mesures de la texture
- * le centre de gravité
- * la longueur du contour de la région
- * la couleur
- * le nombre de trous dans la région

Chaque caractéristique est définie sur un domaine de valeur qui lui est propre. Par exemple, l'orientation d'un segment appartient à l'intervalle d'angle $[0, 360]$, ou l'intensité d'un point est comprise entre 0 et 255 en valeur entière, si on possède 256 niveaux de gris. Il faut préciser que la représentation du domaine de valeur est différente suivant l'application. Ainsi, au lieu de choisir l'intervalle d'angle $[0, 360]$, il est possible d'utiliser $[0, 2\pi]$, ou $[-\pi, +\pi]$. Cette représentation est fortement liée aux outils en notre possession pour les manipuler et aux avantages qu'offre une représentation par rapport à une autre.

II.2.2.2 Robustesse des attributs

Lors de l'exposé des problèmes d'existence ou non des indices à cause de phénomènes perturbateurs, ont été passées sous silence les modifications que ces phénomènes engendrent sur les indices. Ce choix est volontaire car nous pouvons parler de l'existence d'une entité abstraite, ce concept ne se basant pas sur la notion de représentation de l'entité; mais une modification d'une entité abstraite ne peut être entendue que dans le sens d'une modification dans la représentation de cette entité. Toute perturbation est donc perçue par le biais d'une modification d'une ou plusieurs valeurs des attributs (ou caractéristiques) associées à l'indice.

Pour pouvoir parler de modification de la valeur d'une caractéristique, il faut pouvoir comparer la valeur courante par rapport à une valeur modèle. Cette valeur modèle est soit la valeur que devrait avoir la caractéristique extraite de l'image n'ayant subi aucune perturbation, soit la valeur que possède la caractéristique dans une image précédente de la même scène. Etant donné que nous nous intéressons dans cette étude uniquement aux appariements entre deux images successives, la variation de la valeur d'une caractéristique est, pour nous, perçue comme une distance séparant les deux valeurs successives de la caractéristique.

Sur deux images successives, les caractéristiques de type photométrique sont sensibles à la variation de luminance et au bruit aléatoire dans les images. Par exemple, la valeur de l'intensité moyenne d'une région varie si la partie de la scène qui l'a engendrée est éclairée différemment ou si elle est vue sous un angle différent. La modification des formes, due à la projection perspective entre deux images successives peut aussi engendrer des changements des caractéristiques de type photométrique. Par exemple, la modification de forme d'une région engendre une variation sur le nombre de pixels qui la constitue et de ce fait change la valeur moyenne de l'intensité de la région. Le résultat est semblable si l'algorithme d'extraction des indices n'est pas stable avec le mouvement entre chaque prise de vue.

Réciproquement, la variation de luminance et le bruit aléatoire engendrant des changements d'intensité des pixels de l'image, se traduisent lors de l'extraction des indices par une perturbation des caractéristiques d'ordre géométrique. Par exemple, une modification de contraste ou d'ombre entre deux images peut masquer une partie d'un contour; ainsi sa longueur s'en trouve changée. De même le changement de forme entre deux images successives due à la projection perspective ou à l'instabilité d'un algorithme d'extraction engendre des perturbations sur ce type de caractéristique.

II.2.3 Conclusion

La transformation matrice de pixels à indices images permet de passer d'un ensemble de données à un ensemble plus restreint dans lequel chaque donnée (ou indice image) est une concentration d'informations (les caractéristiques) directement exploitables. Les indices sont moins sensibles que les pixels aux phénomènes lumineux, car une variation d'intensité de pixels n'entraîne pas obligatoirement l'absence de détection d'un indice image. Cependant, nous avons remarqué que les techniques utilisées ne permettent pas toujours la détection de tous les "vrais" indices de l'image. Inversement, des indices images ne correspondant à aucune propriété intrinsèque de la scène peuvent être détectés.

Nous avons ensuite mis l'accent sur le fait que les indices sont des concepts (entités abstraites); et que par conséquent, c'est par le biais de leurs caractéristiques qu'ils sont appréhendés. Ces attributs peuvent ne pas être stables à cause de phénomènes lumineux, d'occlusion, de bruit, de variations de formes dues à la projection perspective.

Ainsi, les indices possèdent chacun des avantages et des inconvénients par rapport aux autres, et plus ou moins de caractéristiques.

La robustesse de leurs caractéristiques face aux phénomènes perturbateurs et la quantité d'informations qu'ils possèdent diffère.

Le nombre d'indices extrait de l'image et la taille de mémoire qu'il faut utiliser pour les stocker varient selon le type de scène à traiter et le type d'indice choisi.

Ils diffèrent aussi sur la manière de les utiliser, sur leur facilité d'emploi et des traitements qu'on leur fera subir. Certains traitements seront possibles avec un type d'indice et impossibles avec un autre. Ce qui explique que selon le type d'image à traiter, le choix du type d'indice et de ses caractéristiques puisse faciliter ou non les traitements ultérieurs. C'est un problème similaire à celui énoncé par [Marr 82] avec la représentation des nombres. Pour déceler qu'un nombre est un multiple de 10, il est plus facile de travailler en base 10 qu'en base 2. Par contre si on veut repérer qu'il est un multiple de 2, la représentation binaire est plus commode. Le choix des indices est donc un compromis entre les avantages et les inconvénients de chacun, orienté par le but à atteindre et le type de scène perçu.

Dans certaines applications, des données plus riches en informations sur l'image et moins nombreuses sont préférées aux trois types simples d'indices présentés précédemment. Ce sont les structures d'indices.

II.3 Structures d'indices

Les structures d'indices sont des regroupements locaux d'indices, où chacun d'eux est lié à d'autres indices du groupement. Le but est de regrouper tous les indices appartenant à une zone de l'image correspondant à la projection d'une même entité de la scène. Par exemple, les indices correspondant par projection à des parties d'un même objet de la scène peuvent être regroupés au sein d'une même structure. Le problème majeur de cette technique est le découpage de l'image en zones caractérisant chacune la projection d'une entité de la scène.

Le découpage de l'image peut être obtenu par des contraintes très importantes sur la scène. Lévine [Levine 83] s'intéresse au problème du suivi d'objets souples se déformant dans le temps: les cellules sanguines. Chaque forme cellulaire est divisée en régions polygonales. Ces régions constituent le type d'indice simple qu'il a choisi. Pour chaque cellule, les régions sont rattachées par des liens d'adjacences. Ainsi, chaque cellule est représentée par une structure de région. Pour mener à bien cette réalisation, les contraintes très fortes de non recouvrement des cellules et de fort contraste avec le fond sont nécessaires. L'utilisation de telles contraintes limite l'application des structures d'indices à un type spécifique de scènes.

[Asada 86][Herault 87][Crowley 84] développent une représentation structurelle des formes de l'image en multi-résolution, ce qui permet une création des structures d'indices indépendante du type de scènes analysées.

[Asada 86] et [Herault 87] recherchent des maximums de courbure ou des points d'inflexion le long d'une courbe. Pour cela, ils modélisent la fonction courbe et lui applique une dérivée première ou seconde. Les pics maximum positifs ou minimum négatifs sont alors extraits à divers niveaux de résolution. Les pics correspondant à travers les résolutions sont ensuite liées fournissant des arbres de pics. Ces arbres représentent une caractéristique locale des formes dans l'image comme , une extrémité, un coin, une jointure lisse, un saut, une bosse [Asada 86]. Cependant en faible résolution, les points extraits le sont sur des contours ovoïdes. Ainsi, ils n'ont pas réellement de signification sur une des caractéristiques de la forme. Par contre, la forme ovoïde prise globalement présage de la présence d'un objet dans la scène. C'est donc une information intéressante que la technique de [Crowley 84] permet d'exploiter.

Nous présentons succinctement la méthode de Crowley, car elle sera détaillée dans le chapitre IV. L'image initiale est convoluée avec divers filtres de différence gaussienne d'écart type différent. Cela permet l'obtention d'une pyramide de laplacien de l'image. Dans chaque image les extrema locaux du laplacien, les pics, sont extraits. Afin de simplifier dans un premier temps la compréhension, on peut considérer que les pics représentent dans les plus faibles résolutions globalement une forme. En plus haute résolution, ils représentent des coins ou des extrémités de formes. Les pics sont ensuite liés par des liens de parenté (père-fils) à travers les niveaux de résolution, en vue de l'obtention des structures "arbres de pics". Cependant, un arbre de pics ne représente pas obligatoirement une forme ou une partie de forme dans l'image. Comme pour l'indice de type région, il peut se produire des fusions ou des cassures d'arbres de pics.

Les structures d'indices permettent de restreindre le nombre de données sur l'image. De même, les structures regroupant des indices, elles sont plus riches en informations que chaque indice pris séparément, donc plus discriminantes. De plus, les liens entre indices donnent une idée sur la morphologie de chaque forme perçue.

Cependant, les problèmes cités ci-dessus font que ce type de représentation est peu usité. On lui préfère souvent une représentation prenant en compte des liens contextuels entre tous les indices de l'image. C'est cette dernière que nous présentons.

II.4 Les relations entre les indices

En plus des données d'ordre géométrique et photométrique, il est possible d'extraire de l'image des relations caractérisant des liens contextuels entre les indices. L'intérêt en est la conservation d'informations contextuelles contenues dans l'image en vue de son exploitation ultérieure. Nous obtenons un graphe dont les noeuds sont des indices et les arcs des relations liant deux indices.

II.4.1 Choix du critère relationnel

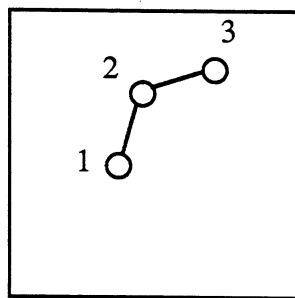
Ce choix est motivé par plusieurs contraintes:

Le graphe souhaité:

Selon l'application, il est préférable d'acquérir le graphe le plus riche possible en informations contextuelles. Plus nombreux sont les arcs dans le graphe, plus riche est l'information contextuelle acquise. Un graphe complet constitue la représentation la plus

riche en informations contextuelles pour le type de relation choisi.

De même, selon l'algorithme désiré pour résoudre l'application, il est souhaitable, voire indispensable, d'avoir comme donnée un type particulier de graphe, par exemple un graphe connexe ou un graphe complet. Or certains types de relations ne permettent pas d'obtenir de tels graphes. Par exemple, la relation "voisin de" n'existe pas entre deux indices hors de la zone de définition du voisinage. Ainsi, le graphe obtenu ne peut être complet. Il peut par contre être connexe (mais pas toujours) dans ce cas (cf figure II.8).



— : Définit la relation "voisin de " entre deux indices.
La distance entre les indices 1 et 3 est trop importante pour pouvoir posséder une telle relation entre eux.
Le graphe n'est donc pas complet.

Figure II.8: Problème du choix du type de relation en fonction du type de graphe que l'on souhaite obtenir.

Les contraintes de temps:

Selon le type de relations choisies, on observe des variations importantes dans leur temps de calcul. Par conséquent, les contraintes de temps de certaines applications rendent illicites l'emploi de certaines relations.

Les indices choisis:

Selon le type d'indice choisi, il sera possible ou non de définir certaines relations (cf. figure II.7).

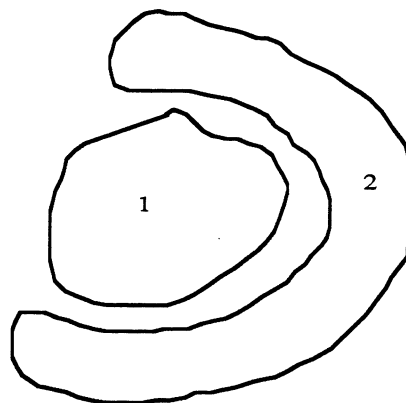


Figure II.7: Si les indices choisis sont de type région, et que nous définissons les relations topologiques: "à droite de", "à gauche de", alors comment formuler la relation liant les régions 1 et 2

Le choix d'un type de relation conduit à la création d'un vocabulaire afin d'étiqueter les relations entre deux indices. [Horaud 87] , par exemple, définit le vocabulaire "à droite de", "à gauche de", "colinéaire à", "même jonction". Il faut alors créer les outils permettant la détermination de chaque terme du vocabulaire.

II.4.2 Détermination des relations

La détermination des termes étant spécifique à chaque type d'indices et à chaque application, nous n'allons pas en énumérer les diverses réalisations. Par contre, il est important de mettre l'accent sur le fait que pour un même vocabulaire, les relations entre les indices diffèrent suivant la manière de les déterminer. Nous étayons ci-dessous ce propos en donnant l'exemple d'une réalisation particulière.

[Saint Vincent 86] définit un vocabulaire contenant une unique entité: "voisin de". La notion de voisinage suggère, qu'en deçà d'une certaine distance, deux indices sont voisins et qu'au delà ils ne le sont pas. Ainsi, il faut déterminer une limite de voisinage. Pour cela, il applique sur l'image une grille rectangulaire. L'image est alors découpée en plusieurs fenêtres. Soit T la taille d'une fenêtre, alors deux indices I_1, I_2 sont dits voisins si la distance euclidienne les séparant est inférieure ou égale à T . Par contre, si cette distance est supérieure à $2T$ ils ne peuvent être voisins. Entre T et $2T$, I_1 et I_2 seront voisins ou non, selon leur position relative par rapport à la grille. En théorie cela revient à définir, bien que plus long en temps de calcul, un cercle autour de I_1 dont le rayon représente la zone maximale de voisinage. Chaque indice à l'intérieur de ce cercle est alors voisin de I_1 . Nous remarquons que si nous faisons varier ce rayon, le nombre d'indices dits "voisin de" I_1 varie (cf. figure II.9).

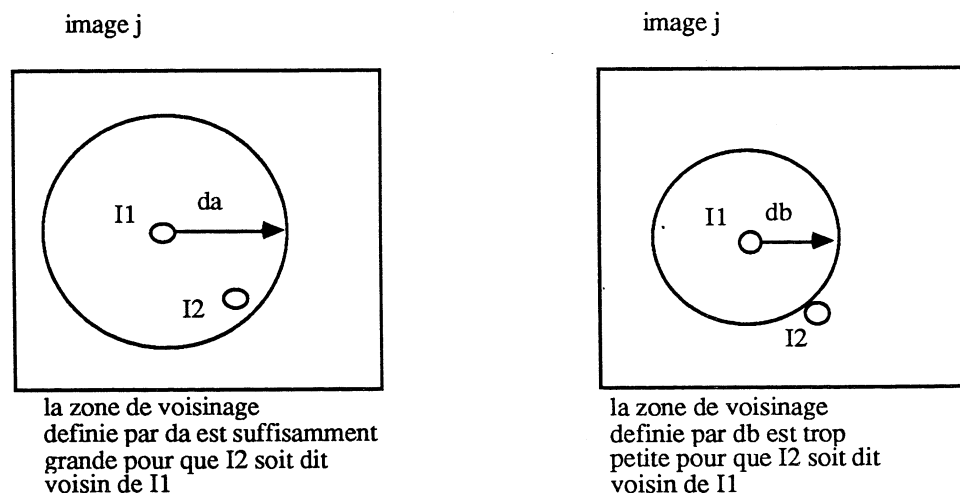


Figure II.9: Sensibilité de la valeur des relations selon la manière dont elles sont déterminée.

Par conséquent, pour un même vocabulaire de relation, le graphe diffère selon la manière dont sont déterminées les relations.

II.5 Conclusion

Chaque pixel d'une image numérique est caractérisé par ses coordonnées (i,j) et sa valeur en intensité. Pour des raisons de quantité de données, de robustesse face aux perturbations, de facilité d'emploi quant au but à atteindre, l'image est transformée sous forme d'un graphe d'adjacence G dont les noeuds sont les indices visuels choisis et les arcs des relations entre ces derniers. Cette transformation peut être présentée sous le formalisme suivant:

$$G = (I, R) \quad \text{avec } I \text{ et } R \text{ deux ensembles disjoints}$$

I (noeuds dans le graphe) représente l'ensemble des indices extraits de l'image. Un indice image est un groupement connexe de pixels ayant une propriété commune. Ce groupement est réduit à un seul pixel dans le cas de points caractéristiques.

$$I = \{i_1, \dots, i_n\} \quad \text{où } n \text{ est le nombre d'indices extraits de l'image.}$$

Soit C l'ensemble des caractéristiques choisi. Ces caractéristiques sont d'ordre géométrique ou photométrique.

Soit V l'ensemble des valeurs possibles de chaque caractéristique.

$$\begin{aligned} \forall j = 1 \dots n, i_j &\in C \times V \\ i_j &= ((C_x, V_{xj}), \dots, (C_y, V_{yj})) \\ &\text{avec } C_x, \dots, C_y \text{ appartenant à } C \\ &\text{et } V_{xj}, \dots, V_{yj} \text{ appartenant à } V \end{aligned}$$

R (arêtes dans le graphe) est l'ensemble des relations entre certains indices.

$$R = \{r_{a,b}, \dots, r_{x,y}\} \quad \text{où } r_{s,t} \text{ représente une relation liant l'indice } s \text{ à l'indice } t.$$

$$r_{s,t} = N_{rs,t} \quad \text{où } N_{rs,t} \text{ est le nom de la relation entre l'indice } s \text{ et l'indice } t.$$

$$N_{rs,t} \in V_o \quad \text{où } V_o \text{ est l'ensemble du vocabulaire désignant les relations.}$$

III CORRESPONDANCE

Ce chapitre fait une synthèse de l'état de l'art dans le domaine de l'appariement entre indices images.

III.1 Introduction

Initialement la vision fut utilisée pour remplacer l'homme dans des tâches fastidieuses et automatiques. C'est le cas de divers contrôles de pièces dans la production automatisée. Cette opération a pour but de vérifier si la pièce est conforme à des normes (techniques de calculs dimensionnels) ou si elle se présente correctement devant la prochaine unité de la chaîne. Dans ce type de scène, la source lumineuse est choisie de manière adéquate ainsi que sa position, ce qui limite les problèmes dus aux phénomènes lumineux. De plus, le système de vision est positionné de telle manière que les pièces passent obligatoirement dans le champ de vision à une distance voulue. Les seules inconnues du problème sont alors :

- l'instant d'arrivée de la pièce.
- la position de la pièce.
- les défauts de la pièce.

Ainsi, pour contrôler une pièce, la première tâche est de repérer son arrivée. Pour cela il convient de détecter un changement survenu dans la scène visualisée, puis de déterminer sa position et sa qualité. Le contrôle de la position ou celui de la qualité nécessite la comparaison entre la forme perçue et un modèle a priori de cette forme. C'est une correspondance image-modèle. Ce domaine abondamment exploré [Bolles 79][Bolles 86][Ayache 84][Souvignier 83][Hebert 83] est très contraint car il fait intervenir des connaissances a priori sur l'objet qu'on voit et sous quelles conditions il doit être perçu. Il fait maintenant place à des travaux qui vont dans le sens d'une généralisation en vue d'une pluralité des tâches à accomplir.

Actuellement, une partie importante des recherches en vision a pour but une coopération entre l'homme et le système de vision. Celui-ci ne cherche pas à se substituer à l'homme, mais l'aide dans les tâches qu'il doit accomplir. Les grands axes de recherche sont:

- la compréhension de scène
- l'imagerie médicale
- la poursuite de cibles.

La correspondance ne s'effectue plus entre un modèle et une forme, mais entre des formes sur des images temporellement successives ou prises au même instant à des positions différentes (cas de la stéréovision). Le problème majeur est d'apparier des formes correspondant à un même élément dans la scène. C'est une correspondance image-image.

L'exposé ci-dessous traite de manière non exhaustive, les principales approches envisagées à ce jour dans le cadre de la correspondance image-image. Il est possible de séparer les méthodes d'appariement d'images en deux classes selon le support de traitement.

Les approches appartenant à la première classe, réalisent l'appariement d'images directement à partir des supports bruts des images (les pixels).

Les approches de la seconde classe possèdent une phase initiale d'extraction d'indices images. La correspondance s'effectue alors entre ces indices.

L'exposé ci-dessous suit cette décomposition. Dans la partie "correspondance au niveau pixel", différentes approches sont présentées succinctement. L'accent est surtout mis sur les réalisations mettant en œuvre des indices images. Ces méthodes appartiennent à la partie "correspondance d'indices".

III.2 Correspondance au niveau pixel

III.2.1 Introduction

L'image est un support borné représentant un échantillonnage du signal vidéo perçu. Elle est donc constituée de points élémentaires appelés les pixels. Ceux-ci se caractérisent par leur position dans l'image (i,j) , et par un niveau d'intensité lui-même échantillonné. Par conséquent, la valeur d'un pixel appartient à un intervalle $[0,N]$ où N est de 255 dans la

plupart des systèmes de vision actuels. Les pixels peuvent aussi être caractérisés par des composantes chromatiques si les images acquises sont en couleur.

Si deux images d'une même scène sont prises à des positions ou des temps légèrement différents, la plupart des informations contenues dans la première image existeront dans la seconde mais seront portées par des pixels différents. Ainsi, apparier deux images revient à trouver la position des informations communes. Cette section est une présentation succincte des méthodes réalisant directement l'appariement d'images à partir du support brut de celles-ci.

III.2.2 Méthodes de corrélation

La corrélation de deux fonctions fournit une réponse maximale si elles sont identiques dans l'intervalle sur lequel est appliqué la mesure. Les images pouvant être considérées comme des fonctions de $R^2 \rightarrow R$, il est possible d'appliquer la corrélation sur des parties d'images. Une fenêtre sur une image correspond à une fenêtre sur une autre image, si le contenu de celles-ci est identique, donc si le calcul de la corrélation entre les deux fenêtres est maximal.

Soit F une fenêtre de référence et I une image de taille $N \times M$, la similitude entre F et une portion de l'image correspond à un maximum de C (la normalisation est nécessaire pour que le maximum de C corresponde effectivement à un maximum de similitude).

$$C(s,t) = \frac{\sum_{i=1}^N \sum_{j=1}^M F(i,j).I(i+s, j+t)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^M F^2(i,j) \sum_{i=1}^N \sum_{j=1}^M I^2(i+s, j+t)}}$$

Une revue des diverses mesures de corrélation et des performances de chacune est présentée dans [Burt 82]. La similarité peut aussi être détectée en recherchant le minimum de la fonction différence S_1 ou le minimum de la fonction différence au carré S_2 (fonctions citées dans [Negre 83]).

$$S_1(s,t) = \sum_{i=1}^N \sum_{j=1}^M |I(i+s, j+t) - F(i,j)|$$

$$S_2(s,t) = \sum_{i=1}^N \sum_{j=1}^M (I(i+s, j+t) - F(i,j))^2$$

Les résultats obtenus par les méthodes de corrélation sont très sensibles aux distorsions dues à la perspective (problème étudié par Castan [Castan 84]), aux occlusions, aux variations lumineuses, aux ombres, aux reflets et aux bruits. Ces perturbations posent le problème de la détermination de la taille des fenêtres. En effet, ces dernières ne doivent pas être trop petites car insuffisamment discriminantes ni trop grandes car trop sensibles aux phénomènes perturbateurs.

Le coût des calculs de corrélation est important. Ainsi, au lieu de réaliser une convolution entre deux images, il est parfois préférable de réaliser un produit dans le domaine de Fourier. Deux fenêtres identiques par produit de la transformée de Fourier donneront alors une constante quelle que soit la fréquence. Une autre manière de restreindre le coût consiste à travailler uniquement dans des zones d'intérêt, qui sont des fenêtres centrées sur des indices image. Cette technique s'apparentant aux appariements aux niveaux indices ne fait pas partie de cette section. Elle sera présentée ultérieurement lors du calcul de similitude entre indices.

III.2.3 Méthodes de différences d'images

La différence pixels à pixels de deux images correspondant à une même scène acquise à des instants temporellement espacés fait apparaître, dans le cas d'une caméra stationnaire, des zones d'intérêt différentes. Dans une zone où aucun changement n'est intervenu, la valeur résultant de l'opération différence est nulle. Les zones de valeur non nulles correspondent, en théorie, à des objets qui se sont déplacés entre les deux prises de vue [Jain 81]. La détection de ces zones permet donc de localiser les formes en mouvement entre les deux images. Cependant, au cours du temps la lumière de la scène varie; ce qui entraîne des variations dans les reflets, les ombres, le niveau d'intensité perçu des objets de la scène. De plus, du bruit

aléatoire existe dans les images. Ces phénomènes engendrent donc des variations dans des régions théoriquement stables au cours du temps, ce qui se traduit par des valeurs non nulles par différence.

Par conséquent, malgré l'emploi du filtrage et du seuillage qui permettent d'accroître la robustesse de cette technique [Stelmaszyk 85][Stelmaszyk 88a], il faut constater que la différence d'images ne donne pas toujours des résultats satisfaisants sur des scènes dans lesquelles la lumière n'est pas contrôlée (cas des scènes réelles). De plus, la contrainte sur l'absence de mouvement de la caméra limite considérablement le champ d'application de cette technique.

III.2.4 Méthodes d'analyse du gradient spatio-temporel

Lorsque la caméra ou les objets de la scène sont en mouvement, différentes parties de l'image se déplacent et se déforment au cours du temps. Les méthodes d'analyse du gradient spatio-temporel visent à retrouver le déplacement de chaque pixel (c'est à dire son correspondant dans l'image suivante) en utilisant les changements en intensité en chaque point de l'image dans le temps et dans l'espace.

Soit l'intensité lumineuse dénotée par $E(x,y,t)$. En supposant que les pixels de chaque forme ont une valeur d'intensité constante au cours du temps, on peut écrire:

$$\frac{dE}{dt} = 0$$

$$\text{or } \frac{dE}{dt} = \frac{\delta E}{\delta x} \frac{dx}{dt} + \frac{\delta E}{\delta y} \frac{dy}{dt} + \frac{\delta E}{\delta t}$$

$$\text{donc } \vec{G} \cdot \vec{V} = -C$$

où $\vec{G}$ est le gradient spatial au point (x,y)

$\vec{V}$ est la vitesse du point (x,y)

C est la variation en intensité du pixel (x,y)

ou une formulation équivalente:

$$\Delta E = - \vec{G} \cdot \vec{D}$$

où ΔE est la différence d'intensité au point (x,y)
entre les deux images consécutives
 $\vec{G}$ est le gradient spatial au point (x,y)
 $\vec{D}$ est le vecteur déplacement du point (x,y)

Or dans cette formulation, il y a deux inconnues qui sont D_x et D_y qui ne peuvent être déterminés au moyen d'une seule équation. D'où l'utilisation de contraintes supplémentaires telles qu'un déplacement supposé constant des pixels constituant une forme en mouvement [Fennema 79], ou une faible variation de déplacement dans le voisinage de chaque pixel [Horn 81] [Yachida 83]. Ces contraintes permettent d'obtenir des critères à minimiser en vue de résoudre un ensemble d'équations [Yachida 83], ou de définir des accumulations de déplacement possible (dénomé dans la littérature sous le vocable de "clustering"). Par exemple grâce à la transformée de Hough qui donne des pics correspondant au mouvement des formes [Fennema 79].

Cependant, dans le cas de scènes réelles, le bruit et les variations d'intensité lumineuse infirment l'hypothèse selon laquelle l'intensité E est constante. Sans oublier les coûts de calcul généralement importants, les restrictions sur le mouvement des objets et les importantes erreurs au voisinage d'occlusion d'objets qui rendent ces techniques souvent mal aisées d'emploi sur des scènes réelles.

III.2.5 Conclusion

Les techniques présentées ci-dessus sont très sensibles aux phénomènes naturels de variation de luminance et aux déformations dues aux mouvements. Elles sont souvent coûteuses en temps de calcul car le nombre de pixels est important. Or les pixels ne sont pas tous utiles à l'application traitée: les informations nécessaires peuvent être condensées sous forme d'indices images; le nombre de données est alors faible et chaque donnée est utile à l'application courante. En outre, cette technique permet d'exploiter des connaissances a priori sur la géométrie fournie par le système de prise de vue.

III.3 Correspondance d'indices

III.3.1 Introduction

Les méthodes de mise en correspondance basées sur des indices visuels se construisent suivant le même schéma. Un module de représentation restreinte de l'image fournit le support de données voulu sur lequel est effectué l'appariement. La partie correspondance est composée:

- d'un module de création d'hypothèses d'appariement
- d'un module de détermination de zones restreintes de recherche des correspondants
- souvent d'un module de vérification de compatibilité entre les hypothèses.

C'est au sein de ces modules que les différences apparaissent. Nous développons donc en détail le contenu de chacun.

III.3.2 Création des hypothèses d'appariement (critère local)

III.3.2.1 Introduction

La première étape de la mise en correspondance des indices de deux images consiste à créer un ensemble H d'hypothèses d'appariement entre les indices visuels.

Soit un indice visuel i de la première image et un indice visuel j de la seconde image, il y a création d'une hypothèse d'appariement entre i et j si et seulement si i et j sont "semblables".

Cette définition pose le problème de la quantification de la ressemblance. Comme nous l'avons vu précédemment, la mesure de la ressemblance va s'appuyer sur la valeur des caractéristiques des indices. Elle est réalisée au moyen d'une fonction de similarité recevant en entrée les caractéristiques de deux indices visuels, et fournissant en sortie une mesure de cette similarité. Seront alors supposés appariés les indices dont la similarité est "suffisamment grande". Comme nous venons de le lire, la création d'hypothèses d'appariement se décompose en plusieurs phases :

- la création d'une fonction de similarité.
- la création ou le rejet d'hypothèses.

Chacune de ces parties fait l'objet dans ce chapitre d'un développement plus approfondi.

III.3.2.2 Fonction de similarité

Le but de ce module est la quantification de la ressemblance de deux indices. Nous définissons ci-dessous une fonction de ressemblance f_r répondant au problème.

La ressemblance peut être perçue par le biais d'une mesure de corrélation [Moravec 77][Hannah 84] ou des mesures S_1, S_2 introduites dans le paragraphe III.2.3. Celles-ci permettent de localiser une fenêtre centrée sur un indice dans l'image de droite similaire à celle définie dans l'image de gauche autour de l'indice à apparier. Elles s'appuient donc sur une comparaison des valeurs en intensité du voisinage des deux indices correspondants, d'où les problèmes définis en III.2.3. Ainsi, on lui préfère souvent une mesure de similarité basée uniquement sur les caractéristiques des indices.

La ressemblance (ou la dissemblance) peut aussi être perçue comme une mesure de distance séparant deux indices. Cette mesure peut être calculée soit localement pour chaque caractéristique, soit globalement en divisant la somme des mesures de distance obtenues pour chaque caractéristique par le nombre total de caractéristiques. Cette mesure locale entre les indices I et J s'écrit donc:

$$\forall i = 1 \dots n, \quad |V_{iI} - V_{iJ}| \quad (\text{où } n \text{ désigne le nombre de caractéristiques})$$

et devient de manière plus global:

$$\sum_{i=1}^n \frac{|V_{iI} - V_{iJ}|}{n}$$

Deux remarques nous amènent à enrichir la quantification définie ci-dessus:

Tout d'abord, il faut souligner que l'**impact de chaque caractéristique** est différent dans la formulation ci-dessus. La raison en est la différence entre les domaines de variation de chaque caractéristique. Par exemple:

- $|V_{aI} - V_{aJ}|$ aura comme valeur maximale j si V_a appartient à l'intervalle (i,j).

$|V_{bI} - V_{bJ}|$ aura comme valeur maximale y si V_b appartient à l'intervalle (x,y).

Cette différence dans les domaines de variation entraîne une discrimination entre les caractéristiques basées sur un critère n'ayant aucun rapport avec la mesure de la ressemblance. De ce fait, on normalise chaque différence pour égaliser son impact. Ainsi [Levine 83], calcule le rapport $| (V_{iI} - V_{iJ}) / (V_{iI} + V_{iJ}) |$, ce qui donne pour chaque i un résultat appartenant à $(0,1)$. Toutefois, on remarque que dans son cas il n'y a pas d'invariance en translation. Effectivement :

soit la translation t en valeur pour chaque caractéristique.

$$| (V_{iI} + t) - (V_{iJ} + t) | = | V_{iI} - V_{iJ} |$$

par contre

$$| (V_{iI} + t) + (V_{iJ} + t) | \neq | V_{iI} + V_{iJ} |$$

Donc le rapport est différent par translation des valeurs.

Cette non invariance peut être souhaitée dans certaines applications, comme nous le verrons ultérieurement, mais à exclure dans d'autres. Pour cette raison, Price [Price 82] divise chaque différence par une quantité fixe, dépendant uniquement de la caractéristique considérée.

$$\forall i, \text{ il calcule } | (V_{iI} - V_{iJ}) / Q_i |$$

Q_i peut être par exemple la borne supérieure des valeurs possibles pour la caractéristique i , ce qui donne dans ce cas pour chaque résultat une valeur appartenant à $(0,1)$.

Une autre normalisation souvent utilisée est la distance normalisée de Mahalanobis formulée ainsi:

$$(\vec{V}_I - \vec{V}_J)^T (C_I + C_J)^{-1} (\vec{V}_I - \vec{V}_J)$$

où chaque composante de $\vec{V}_K$ est la valeur d'une des caractéristiques de l'indice K .

La taille de ce vecteur est fonction du nombre de caractéristiques choisies.

C_K désigne la matrice de covariance sur les mesures des valeurs des caractéristiques de l'indice k .

Elle permet de prendre en compte la covariance sur les mesures des valeurs des caractéristiques. Dans le cas où ces mesures sont décorrélées, la formulation devient similaire à celles présentées précédemment:

$$\left(\sum_{i=1}^n \frac{(V_{iI} - V_{iJ})^2}{\sigma_i^2} \right) \quad \text{où } \sigma_i \text{ est la variance sur la mesure de la caractéristique } i.$$

La seconde remarque concerne la **robustesse** qui diffère selon les caractéristiques. La confiance accordée à des caractéristiques robustes doit être plus élevée que celle accordée à des caractéristiques non robustes. L'impact doit donc être différent selon les caractéristiques et lié à la confiance que l'on accorde à une caractéristique. De ce fait à chaque caractéristique c_i est affecté un poids P_i . Dans le cas des mesures locales, la quantification de la ressemblance de deux indices I et J peut s'écrire sous la forme suivante [Kass 86]:

$$fr = (r_1, \dots, r_n) \quad \text{avec } r_i = \frac{|V_{iI} - V_{iJ}|}{Q_i} \cdot P_i$$

Dans le cas global [Price 82], cette quantification devient:

$$f_r = \frac{1}{n} \sum_{i=1}^n \frac{|V_{iI} - V_{iJ}|}{Q_i} \cdot P_i$$

III.3.2.3 Création ou rejet d'hypothèses

Nous avons vu précédemment comment quantifier la similitude entre deux indices. Il faut maintenant déterminer si la valeur obtenue permet de penser que les deux indices forment un appariement plausible.

Comme le précise [Kass 86], pour que deux indices I et J soient considérés comme semblables, il faut que *quel que soit i, la valeur de la caractéristique c_i de l'indice I de la première image soit "proche" de la valeur de la caractéristique c_i de l'indice J de la seconde image*. Donc quel que soit i, il faut que $|V_{iI} - V_{iJ}| = \epsilon_i$, avec ϵ_i "petit". Pour tout i, ϵ_i "petit" entraîne que la valeur de similitude est elle même faible. Afin de supposer un appariement, il faut comparer la valeur de similitude à un seuil.

Le seuil doit être choisi de telle façon que la quasi-totalité des hypothèses "valides" soit retenue et que le nombre de mauvaises hypothèses soit limité. La variation des valeurs des caractéristiques est due à des phénomènes perturbateurs (mouvement de la caméra, mouvement des objets, changement de luminance, bruit, algorithmes, etc...). S'il est possible de quantifier l'influence des perturbations sur la valeur de chaque caractéristique, on peut déterminer exactement la valeur maximale de similarité au dessus de laquelle tout appariement est invalide. Cependant, cette quantification ne semble pas toujours possible. On lui préfère souvent une maximalisation globale des variations possibles. Dans ce cas, le seuil obtenu permet d'accepter toutes les bonnes hypothèses, mais aussi un nombre important d'hypothèses erronées. D'où l'emploi parfois d'un seuil déterminé statistiquement sur divers types d'images, représentant un bon compromis entre "bonnes" hypothèses et "mauvaises" hypothèses retenues:

Ainsi, en deçà du seuil, il est permis de penser que les deux indices "sont susceptible de" correspondre. Il faut insister particulièrement sur l'emploi du terme "sont susceptible de" correspondre. En effet il est possible d'avoir des couples d'indices ne correspondant pas à l'appariement réel, mais dont la valeur de ressemblance est en deçà du seuil. Pour cette raison les appariements créés ne sont que des hypothèses.

III.3.2.4 Conclusion

La création d'hypothèses peut être proposée sous le formalisme suivant:

Soit I_1 l'ensemble des indices de la première image.

$$I_1 = \{i_1, \dots, i_n\}$$

Soit I_2 l'ensemble des indices de la seconde image.

$$I_2 = \{j_1, \dots, j_m\}$$

Soit $f_r : I^2 \rightarrow V_h$ (où V_h est l'ensemble des valeurs possibles de ressemblance) une fonction de mesure de similitude qui à partir d'un indice i de I_1 et d'un indice j de I_2 retourne la quantification de leur ressemblance.

$$f_r(i,j) = v_{i,j}$$

Si la mesure de similitude $v_{i,j}$ est inférieure au seuil fixé, il y a alors création d'une hypothèse d'appariement h_p appartenant à H (où H représente l'ensemble des hypothèses d'appariement)

$$H = \{h_1, \dots, h_q\}$$

avec h_p appartenant à H inclus dans le produit scalaire $I_1 \times I_2 \times V_h$.

$$h_p = (i, j, v_{i,j}).$$

III.3.3 Zone de recherche

III.3.3.1 Introduction

Soit N le nombre d'indices dans la première image et M le nombre d'indices dans la seconde. D'après la section précédente, il y a $N \times M$ calculs de ressemblance effectués, ce qui peut être coûteux en temps. En outre, plus on teste d'appariements, plus il y a de chances de créer effectivement des hypothèses d'appariement, ce qui entraîne une augmentation du nombre des éléments de l'ensemble des hypothèses H ; d'où d'une part, un coût en place mémoire pouvant être important pour stocker l'ensemble des appariements possibles et d'autre part, une augmentation du coût en temps de calcul lors de la recherche ultérieure des appariements "valides" dans H . Il est donc nécessaire de réduire le nombre de calculs de ressemblance.

Pour cela, il faut restreindre la zone de recherche des appariements. Celle-ci doit contenir un ensemble restreint d'indices incluant l'indice "valide". Or un point dans une image est la projection d'un point de la scène. De ce fait, la détermination de cette zone requiert les moyens de quantifier une zone restreinte dans la scène de position possible de l'indice scène correspondant. Des connaissances a priori sur la scène ou sur les appariements précédents permettent de calculer ces zones.

III.3.3.2 Les méthodes

Nous avons découpé les méthodes de restriction de la zone de recherche en deux classes. La première englobe les méthodes basées sur des connaissances a priori; la seconde regroupe les méthodes utilisant les appariements précédents.

III.3.3.2.1 Connaissances a priori

Les connaissances a priori varient selon le type de scène étudiée.

- D'après la classification de [Jain 81], il existe 4 types de scènes différents:

	caméra fixe	caméra mobile
objet statique	SOSC	SOMC
objet mobile	MOSC	MOMC

SOSC (static object, static camera): Elle correspond à la mise en correspondance entre les formes dans une image et les modèles a priori des objets perçus. Ce type de scène ne fera pas l'objet de cette étude, car constitue pas une mise en correspondance entre images.

SOMC (static object, mobile camera): Dans ce type de scène, la caméra est en mouvement et les objets sont statiques. Soit un pixel (i,j) dans la première image. Il a été obtenu par projection perspective d'une partie de la scène sur le plan de la rétine. Si nous supposons que les pixels sont des points au sens mathématique, c'est à dire sans surface, un point dans l'image correspond à la projection perspective d'un point dans la scène. Or pour un pixel (i,j) dans l'image, il existe une infinité de points dans la scène pouvant engendrer ce pixel par projection perspective (en raison de la perte d'informations lorsqu'on passe d'un espace 3D à un espace 2D). Ces points se situent sur la droite dans l'espace de la scène passant par le centre optique O et le pixel $P(i,j)$ (cf. figure III.1).

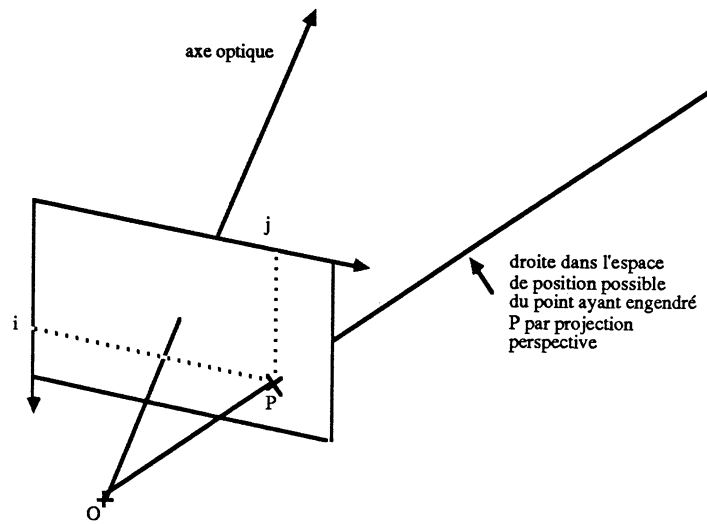


Figure III.1: Représentation de la droite dans l'espace 3D englobant tous les points se projetant en P.

Si on déplace la caméra de manière à ce que le point de la scène ayant engendré le pixel (i,j) se situe toujours dans le champ de vision, il sera représenté par un pixel dans la nouvelle image. Or nous ne savons pas où se situe le point sur la droite, ce qui implique que la zone de recherche du correspondant est obligatoirement la projection de la droite sur le second plan image.

Pour cela il faut connaître d'une part, la calibration pour déterminer les paramètres intrinsèques et extrinsèques de la caméra et d'autre part, le mouvement de la caméra pour déterminer la position du second plan image. La projection perspective d'une droite restant une droite, la zone de recherche du correspondant se situe sur une ligne appelée "épipolaire" (cf. figure III.2).

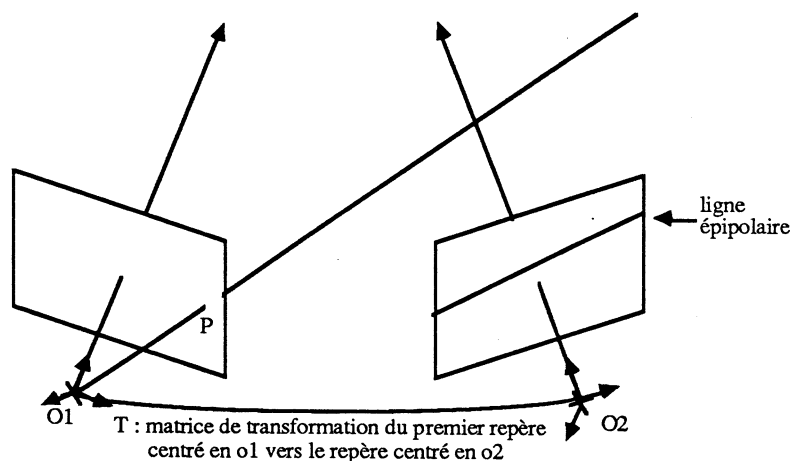


Figure III.2: Représentation de la ligne épipolaire

La zone de recherche peut être encore réduite à une portion de la ligne épipolaire [Skordas 88]. En effet, considérons un point P dans la scène perçu via un point p dans une image. La droite passant par le centre focal O et le point p définit la droite de position possible dans la scène. Soit une seconde image de centre focal O' , et un point p' sur la droite épipolaire tel que la droite $O'p'$ soit parallèle à OP . En parcourant la droite OP , les points obtenus par projection se situent tous d'un seul côté du point p' le long de la ligne épipolaire (p' correspond à la projection d'un point situé à l'infini sur la droite OP). Le point p' est donc une borne restreignant la zone de recherche le long de l'épipolaire (cf. figure III.3).

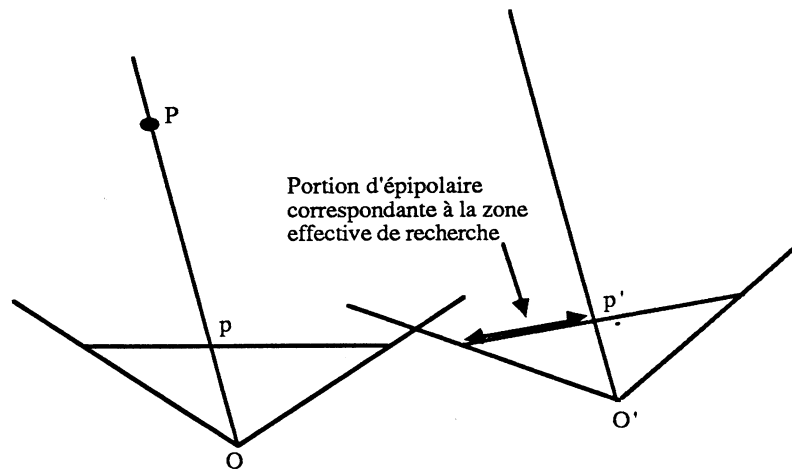


Figure III.3: Vue de dessus du plan épipolaire défini par p , O et O' avec une délimitation de la zone de recherche.

Cependant le pixel n'est pas un point au sens mathématique mais une petite surface dans l'image; De plus la droite de rétro-projection est imprécise à cause d'incertitudes persistant sur sa position après calibration. Ainsi, c'est un cône et non une droite qui représente la zone dans l'espace des positions possibles. Par projection, la zone de recherche est un cône autour de la ligne épipolaire. Par la suite, nous considérerons le pixel comme un point et l'incertitude sur la position nulle, car nous présentons uniquement les principes d'obtention des zones de recherche. Il est toutefois indispensable pour le calcul de celles-ci, de tenir compte des incertitudes et de la taille du pixel.

La connaissance a priori nécessaire dans ce type de scène pour calculer la zone de recherche porte donc sur le mouvement de la caméra.

MOSC (mobile object, static camera): La caméra est fixe, mais les objets peuvent être mobiles et leurs mouvements indépendants. Soit un point dans la scène appartenant à un objet en mouvement. Si on connaît le déplacement maximal possible des objets pendant l'intervalle de prise de vue, alors la zone de position possible dans la scène d'un point sur un objet appartient à une sphère dont le centre est la position précédente et le rayon correspond au déplacement maximal prévu du point dans l'espace.

Comme précédemment, on ne connaît pas la position du point dans l'espace, mais on sait qu'il appartient à la droite de l'espace passant par le centre optique et le pixel projection dans l'image. Donc chaque point sur cette droite est un candidat possible, et sa zone de position possible se situe dans la sphère avec pour centre le point et pour rayon le déplacement maximal pendant l'intervalle de prise de vue. La zone de position possible dans la scène est donc un cylindre avec pour axe la droite et pour rayon le déplacement maximal. Il est donc possible dans ce type de scène de déterminer une zone de recherche si on connaît le déplacement maximal pendant l'intervalle de temps séparant deux acquisitions. Cette zone (ellipse) est la projection perspective du volume de position possible dans la scène sur le plan de la rétine (cf. figure III.4).

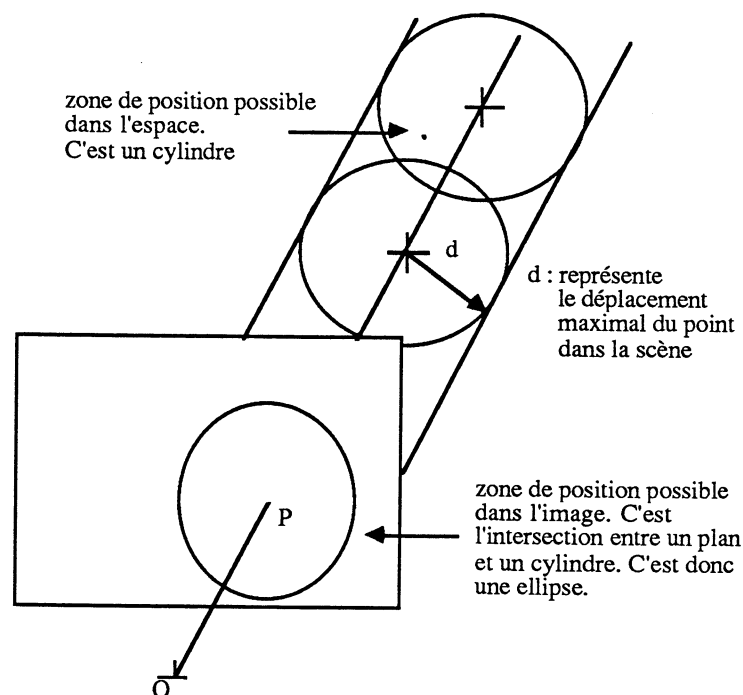


Figure III.4: Zone de recherche dans le cas de scène MOSC.

D'autres informations telles que la direction, le sens de déplacement, les variations maximales de direction ou les variations maximales de vitesse et d'accélération peuvent encore restreindre cette zone de recherche [Aubert 86].

MOMC (mobile object, mobile camera): La caméra et les objets sont en mouvement. Comme dans le cas précédent, la connaissance du déplacement maximal possible entre deux prises de vue des objets permet de déterminer le cylindre autour de la droite de rétroprojection du pixel considéré. Pour calculer la zone de recherche, il faut projeter le volume de position possible dans la nouvelle image. Il est par conséquent indispensable de connaître aussi le mouvement de la caméra (cf. figure III.5).

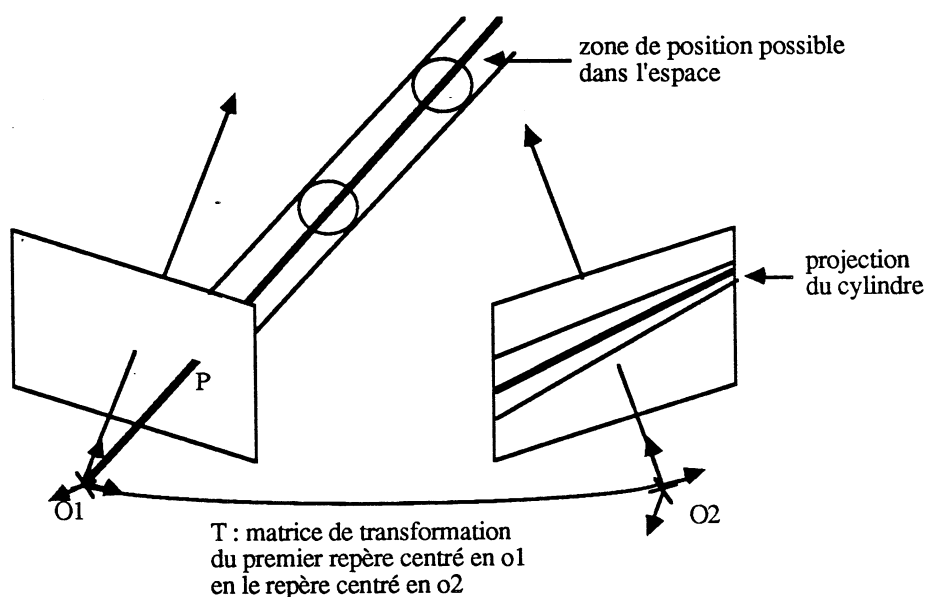


Figure III.5: Zone de recherche dans le cas de scène de type MOMC.

Il faut remarquer que le cas de la vision stéréo appliqué à une scène dynamique fait partie de ce type de scène. Néanmoins, le temps séparant l'acquisition de deux ou plusieurs images est nul. Ainsi la zone de position possible dans l'espace se réduit à la ligne épipolaire comme dans le cas des scènes de type SOMC.

En conclusion, la zone de recherche peut être calculée moyennant des connaissances a priori sur la dynamique de la scène et sur le modèle de projection utilisé. C'est le cas dans la plupart des scènes traitées actuellement (vision stéréo, scène de trafic routier, scène industrielle: convoyeur). Mais il est aussi possible de définir des zones de recherche restreintes grâce aux informations extraites d'appariements précédents.

III.3.3.2.2 Informations fournies par des appariements précédents

Cas d'une série d'images successives:

Soit une suite d'images espacées dans le temps. Après chaque acquisition d'une nouvelle image, la mise en correspondance est lancée pour appairer les indices respectifs de l'image courante avec ceux de l'image précédente. Ainsi lors de la phase d'appariement entre l'image j et l'image $j-1$ se sont déjà produites $j-2$ mises en correspondance d'images dont il est intéressant de tenir compte car elles permettent de déterminer de manière relativement précise la localisation des indices perçus. Deux cas sont à distinguer pour analyser ces techniques:

- les objets sont dynamiques
- les objets sont statiques.

Dans le premier cas, les scènes sont du type MOSC et MOMC. L'appariement entre les deux premières images définit, par triangulation un segment de position possible le long de la droite de rétroprojection passant par le point apparié et le centre focal de la seconde image. Ce segment correspond à l'intersection entre le cylindre de position possible dans l'espace généré au moment de l'acquisition de la seconde image et la droite de rétroprojection construite à partir du point apparié de la seconde image. Chaque point de ce segment est potentiellement animé d'un mouvement maximal pendant l'intervalle de temps séparant l'acquisition de la seconde et de la troisième image. Donc chaque point de ce segment est le centre d'une sphère de positions possibles. L'union de toutes ces sphères forme un volume. Celui-ci est en fait une portion du cylindre de position possible généré s'il n'y avait pas eu prise en compte de l'appariement précédent au moment de l'acquisition de la troisième image. La zone de recherche dans la troisième image est donc limitée à la projection de cette portion de cylindre (cf. figure III.6). Le mécanisme de détermination du volume de positions possibles pour les appariements suivants est identique à celui présenté ci-dessus.

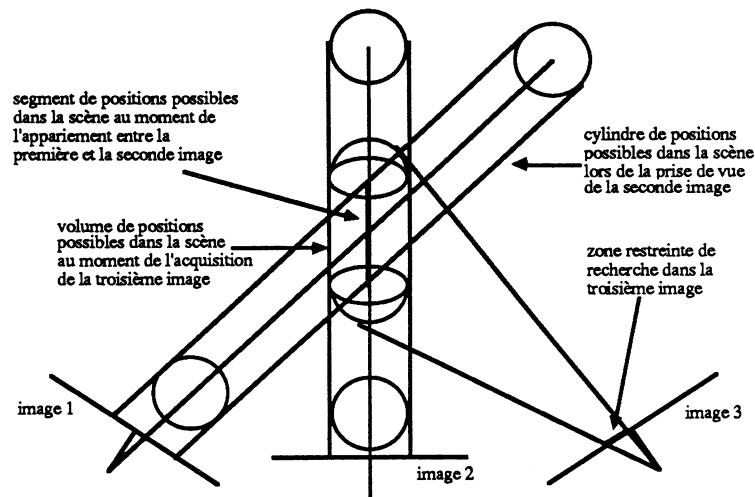


Figure III.6: Zone de recherche dans le cas de scène MOSC et MOMC avec prise en compte des appariements précédents.

Dans le second cas (scène obligatoirement de type SOMC), les appariements successifs permettent une localisation et une restriction de la sphère d'incertitude autour du point 3D obtenu par triangularisation (le pixel n'est pas un point mais une petite surface, d'où par triangularisation imprécisions de mesure et de localisation dans l'espace 3D). [Ramparany 88] a utilisé une technique similaire en restreignant par appariements successifs des ellipsoïdes de positions possibles des segments 3D visualisés.

Moyennant une connaissance sur le mouvement de la caméra, on peut projeter ce volume de positions possibles dans l'image j . La donnée des appariements précédents permet donc de substituer à la zone épipolaire, une zone beaucoup plus restreinte autour de la projection du point 3D (cf. figure III.7).

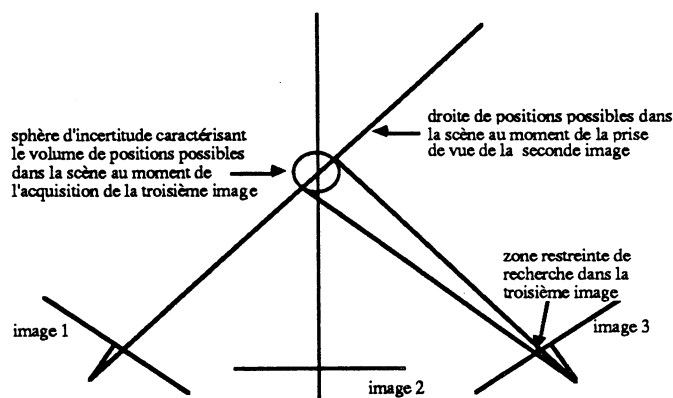


Figure III.7: Zone de recherche restreinte par l'utilisation d'appariements précédents dans le cas de scènes de type SOMC.

Cas de deux images successives perçues à divers niveaux de résolution:

Cette technique, introduite dans les travaux de [Marr 80][Grimson 85][Moravec 77], consiste à utiliser les appariements effectués entre deux images à des niveaux de faible résolution en vue de restreindre les zones de recherche en plus haute résolution. Les deux images successives à apparier sont convoluées avec des opérateurs de tailles différentes de type passe-bas ou de type passe-bande. Le résultat pour chaque image est une pyramide d'images de passe-bas (respectivement une pyramide d'images de passe-bande) à divers niveaux de résolution. Les appariements aux faibles résolutions sont grossiers mais rapides et robustes car les données sont éparpillées et peu denses [Grimson 85]. De plus, chaque indice étant présent à divers niveaux de résolutions successifs, à chaque indice dans une image correspond une zone de descendance en plus haute résolution (ou zone d'héritage) dans laquelle se situe l'indice homologue (ou indice fils). Ainsi, les mises en correspondance en faible résolution permettent de restreindre, à des zones de descendance, l'espace de recherche des appariements dans les plus hautes résolutions. Cette approche peut être formulée de la manière suivante:

soit I_1 un indice de la première image.

(i, j, K) sont les coordonnées dans l'image 1 d'un pixel de I_1 et K est le niveau de résolution de l'image.

soit I_2 l'indice de la seconde image mis en correspondance avec I_1 .

(l, m, K) sont les coordonnées dans l'image 2 d'un pixel de I_2 et K est le niveau de résolution de l'image (avec $l = i + d_i$, $m = j + d_j$).

L'appariement entre I_1 et I_2 définit une disparité (D). Celle-ci correspond à la différence de position des indices associés entre les deux images. Elle est proportionnelle à la position de l'indice scène par rapport à la caméra.

$$D = (d_i, d_j) \text{ où } d_i = l_i - l_j \text{ et } d_j = m_j - m_i$$

Soit un indice $I(X, Y, K-1)$ situé dans la zone de descendant de l'indice I_1 . Cet indice hérite de la disparité définie lors de l'appariement de son père I_1 avec I_2 , ce qui détermine une zone de recherche dans la seconde image où se situe son correspondant.

Soit $I'(X', Y', K-1)$ l'indice de la seconde image correspondant à l'indice I .

Alors I' se situe dans la zone :

$$X + d_j - V_a \leq X' \leq X + d_j + V_a$$

$$Y + d_i - V_a \leq Y' \leq Y + d_i + V_a$$

où V_a représente la variation en disparité due au changement de résolution et à l'imprécision sur la mesure de la disparité (cf. figure III.8).

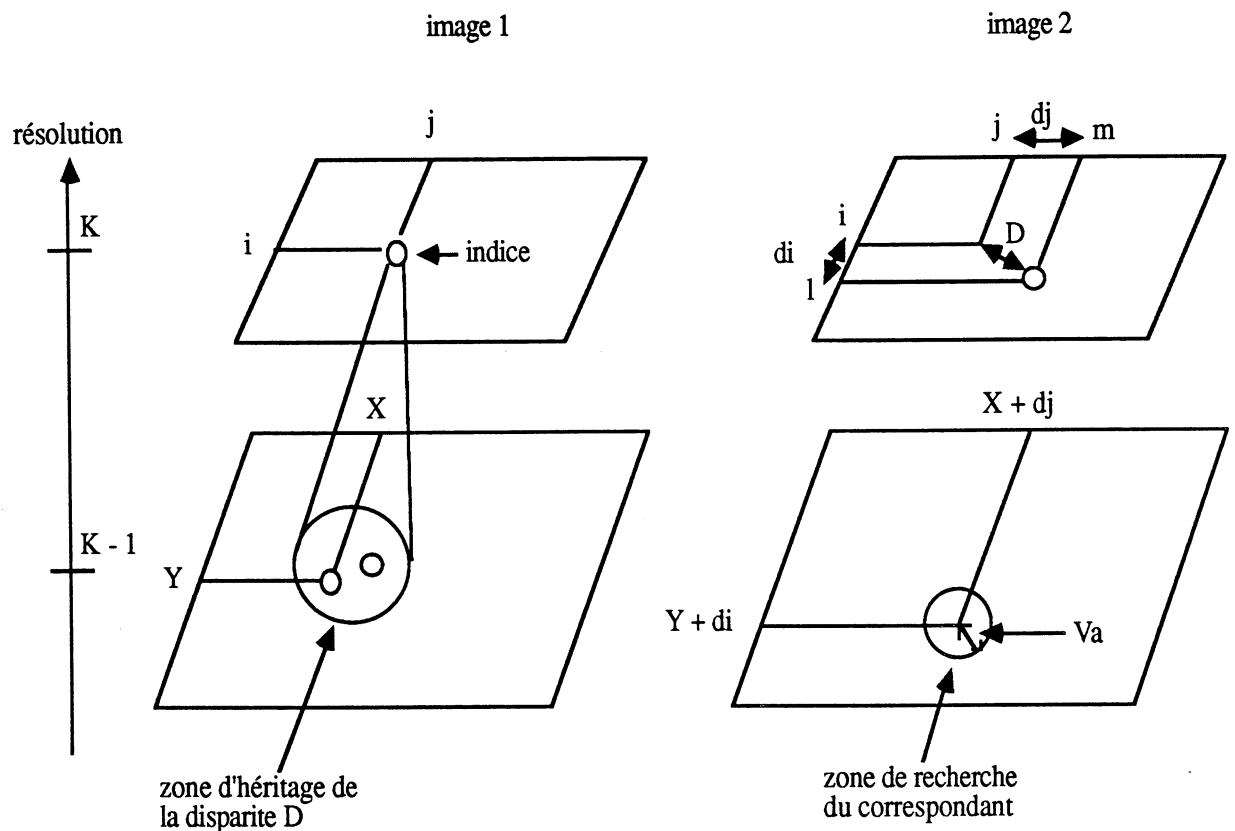


Figure III.8: Limitation des zones de recherche par héritage de disparité dans un schéma multi-résolutions.

Il faut remarquer que ces techniques ne requièrent aucune connaissance a priori.

III.3.3.3 Conclusion

La détermination des zones de recherche nécessite souvent des connaissances sur la position de la caméra, et sur le mouvement des objets voire des deux à la fois.

La connaissance de la position de la caméra ou de son déplacement n'implique pas de contrainte sur la scène étudiée. Elle est fournie à partir de la calibration pour toute la durée de l'étude lorsque les caméras ont une position fixe les unes par rapport aux autres (c'est le cas de la stéréovision) et par la calibration et le mouvement du bras du robot dans le cas où la caméra y est fixée. Il faut noter toutefois que les composants du mouvement du bras du robot sont entachés d'erreurs; de ce fait la position de la caméra est moins précise que dans le cas stéréo. Ces erreurs de position doivent être quantifiées ou du moins maximisées pour que les calculs des zones de recherche soient valides.

Par contre, la connaissance du mouvement des objets ou de certaines de ces composantes est très restrictive sur le type de scène à étudier. Le système construit à partir de telles contraintes est alors dépendant de la scène à analyser.

Il faut remarquer que dans les applications, il est fréquent de combiner différentes méthodes de calcul de zones de recherche. Ainsi, [Grimson 85] utilise la contrainte épipolaire et l'héritage des disparités obtenues à de plus faibles résolutions. Pour un indice, la zone de recherche de son correspondant est l'intersection de celles fournies par les diverses méthodes (cf. figure III.9).

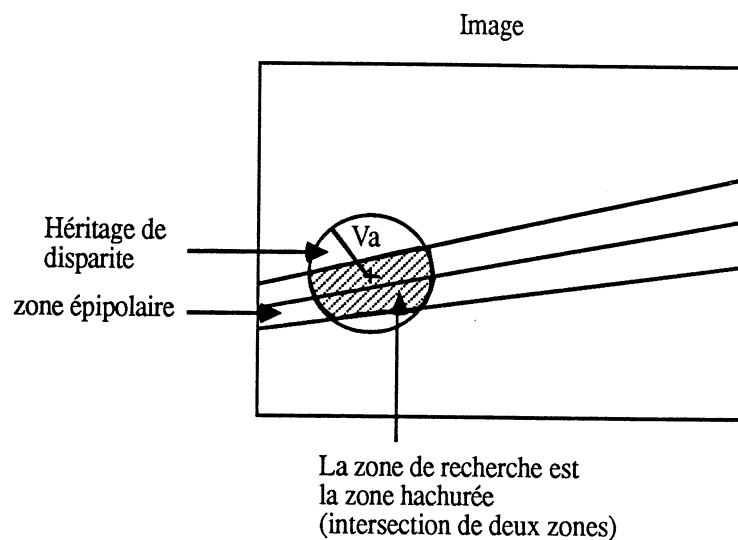


Figure III.9: Combinaison de zones de recherche.

La restriction de la zone de recherche des correspondants permet de limiter pour chaque indice i de la première image le nombre de calculs de similitude à M_i (nombre d'indices dans la zone de recherche de l'indice i) au lieu de M (nombre total d'indices dans la seconde image). Ainsi, le nombre total de calcul de similitude est de:

$$\sum_{i=1}^N M_i$$

(où N est le nombre d'indices de la première image)

Ce coût est d'autant plus faible et permet une limitation du nombre des hypothèses d'appariement d'autant plus grande que les zones de recherche sont restreintes. Cependant, au nombre de déterminations de similarité il faut rajouter N déterminations de zones de recherche (un par indice de la première image). Ce surcoût peut être important, mais le nombre limité d'hypothèses ainsi obtenu est compensé lors de la phase ultérieure de vérification globale des appariements.

III.3.4 Compatibilité entre les hypothèses (critère global)

III.3.4.1 Introduction

Il est fréquent de rencontrer des systèmes utilisant uniquement un critère local pour mettre en correspondance des indices. C'est le cas par exemple des techniques de suivi d'indices visuels [Stelmaszyk 88b] ou des algorithmes de "coarse to fine" [Hannah 84][Moravec 77]. Ils s'appuient sur le fait que les zones de recherche des correspondants sont suffisamment petites pour n'avoir que très rarement plusieurs appariements possibles. Or un critère uniquement local repose sur la quantification de la ressemblance. Que faire alors si dans une même zone de recherche, les indices pouvant fournir un appariement plausible se ressemblent? C'est un cas d'ambiguïté d'appariement.

Pour limiter ce phénomène, une méthode simple consiste à utiliser l'effet "pulling" [Marr 79] [Grimson 81] [Xu 87]. Lorsqu'il existe deux indices semblables dans une même zone de recherche, si l'un des deux est mis en correspondance dans une autre zone, il est supprimé (pull); le second est alors considéré comme le bon match car l'ambiguïté a été levée (cf. figure III.10).

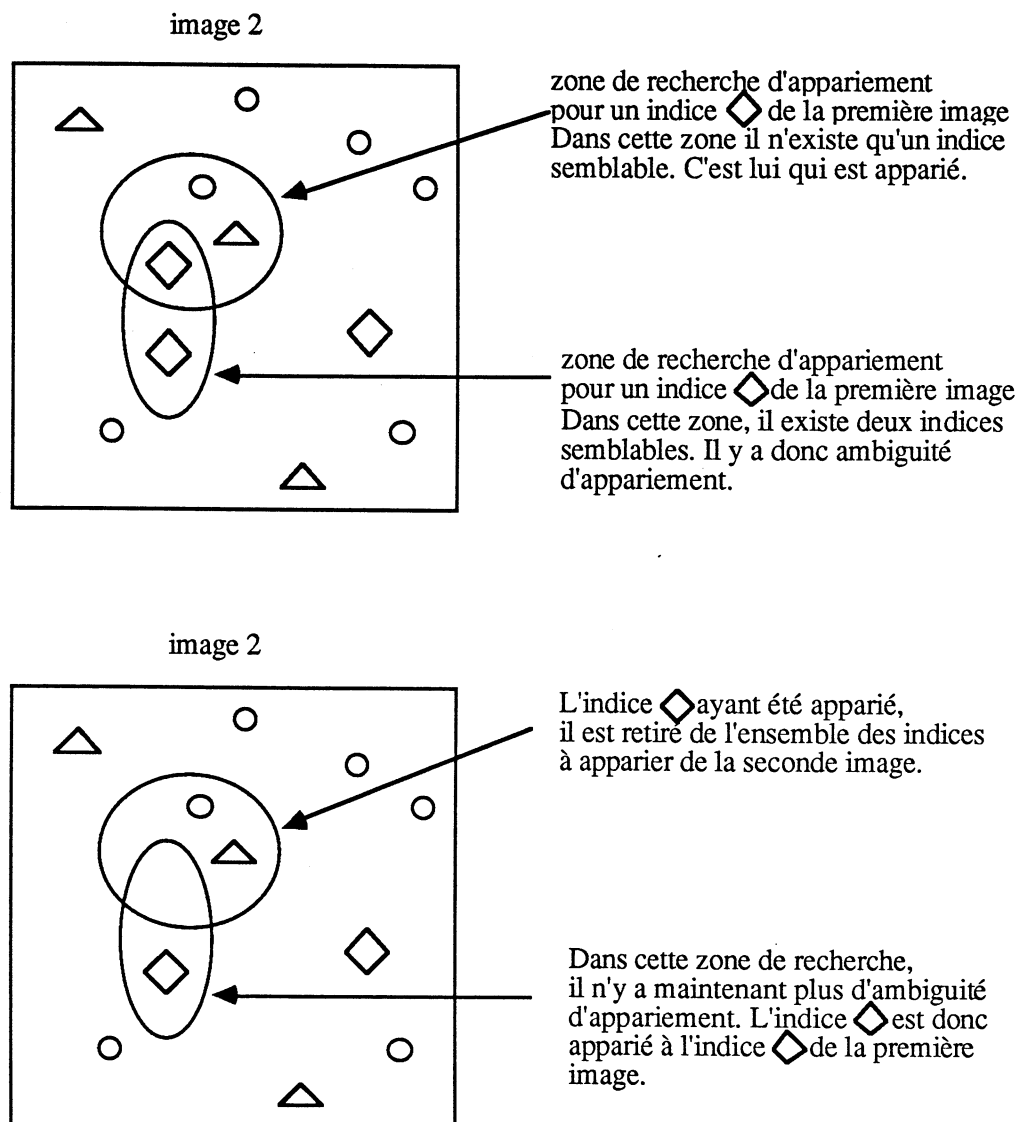


Figure III.10: Exemple sur l'application de l'effet "pulling".

Mais que faire si aucun des 2 indices ne peut être mis en correspondance sans ambiguïté? De plus si dans la zone de recherche il n'existe qu'un indice semblable à celui à appairer, il constituera le bon appariement. Or ce procédé engendre des erreurs car le bon appariement peut être absent à cause de phénomène lumineux, d'une occlusion ou tout simplement car il se situe en dehors de l'image. Cette méthode n'est donc pas satisfaisante dans le cas général.

Une mise en correspondance globale prenant en compte des informations contextuelles dans l'image permet de lever, dans la mesure du possible, les ambiguïtés d'appariement. Les informations contextuelles sont fournies par des relations entre les indices au sein de l'image. Ainsi, lorsque deux indices constituent deux solutions possibles d'un appariement, l'ambiguïté

est levée en testant la compatibilité relationnelle de chacune des deux hypothèses avec l'ensemble des autres hypothèses. Le "bon" sous-ensemble d'hypothèses est alors le plus grand sous-ensemble d'hypothèses mutuellement compatibles selon le critère relationnel choisi.

En pratique, les hypothèses retenues ne sont pas toujours mutuellement compatibles, car la définition de la compatibilité n'est pas toujours possible entre chacune des hypothèses. Cependant, ne pas avoir de compatibilité entre elles, n'implique pas forcément que celles-ci soient incompatibles. Le problème de l'appariement global se décompose donc en deux étapes:

trouver un critère de compatibilité entre les hypothèses pour pouvoir les comparer

chercher le sous-ensemble maximal d'hypothèses compatibles selon ce critère.

Ces deux points sont traités ci-dessous.

III.3.4.2 Critère de compatibilité

Soit A le graphe représentant la première image (image de gauche par exemple).

Soit B le graphe représentant l'image suivante (image de droite).

Soient les lettres G et D pour désigner respectivement les indices de l'image de gauche et ceux de l'image de droite.

Soient h_{ia} et h_{jb} deux hypothèses d'appariement telles que :

$$h_{ia} = (G_i, D_a, v_1)$$

$$\text{et } h_{jb} = (G_j, D_b, v_2)$$

où G_i et G_j sont deux noeuds dans le graphe A

D_a et D_b sont deux noeuds dans le graphe B

v_1 et v_2 sont respectivement les valeurs associées aux hypothèses h_{ia} et h_{jb} .

On définit une compatibilité entre h_{ia} et h_{jb} si:

$R(G_i, G_j)$ et $R(D_a, D_b)$ sont définissables et comparables

et

$R(G_i, G_j)$ est similaire à $R(D_a, D_b)$

où R désigne la relation entre deux indices (c'est un arc du graphe).

Il faut toutefois remarquer qu'une autre définition de la compatibilité est aussi couramment utilisée. Elle est basée sur le fait que si le (ou les) lien(s) contextuel(s) entre G_i et G_j est (sont) comparable(s) à celui (ou ceux) entre D_a et D_b , alors certaines mesures (par exemple la disparité) définies sur l'hypothèse h_{ia} se conservent sur h_{jb} .

Ainsi h_{ia} est compatible avec h_{jb} si:

$R(G_i, G_j)$ et $R(D_a, D_b)$ sont définissables et comparables

et

$M(G_i, D_a)$ est similaire à $M(G_j, D_b)$

où M désigne une mesure sur l'hypothèse.

Deux termes de relation différents n'impliquent pas forcément que ces relations soient incompatibles. Ainsi, comme ce fut le cas pour la quantification de la ressemblance entre les indices, la détermination de la compatibilité entre deux hypothèses nécessite de faire une mesure de similarité entre les relations. Celle-ci dépend étroitement des relations utilisées et des applications. De ce fait, nous n'en donnons pas une formulation absolue, mais l'illustrons par deux exemples pris dans la littérature.

Dans le premier exemple[Skordas 88], la formulation de la compatibilité employée est la première introduite précédemment. Le vocabulaire des relations choisi entre chaque indice d'une image (ici des segments de droite) est le suivant :

$Vo = \{ \text{collinéaire_avec, même_jonction_que, à_gauche_de, à_droite_de} \}$

soit deux hypothèses d'appariement:

$h_{ia} = \{G_i, D_a\}$ et $h_{jb} = \{G_j, D_b\}$

Dans cette application, la compatibilité topologique entre h_{ia} et h_{jb} est alors définie par les règles suivantes :

règle 1 : (colinéarité) : h_{ia} est compatible avec h_{jb} si

(($i=j$) et ($a \neq b$) et (D_a collinéaire_avec D_b))

ou

(($i \neq j$) et ($a=b$) et (G_i collinéaire_avec G_j))

(ces deux premières conditions prennent en compte le cas de cassure de segment)

ou

(($i \neq j$) et ($a \neq b$) et (G_i collinéaire_avec G_j) et (D_a collinéaire_avec D_b))

règle 2 : (connectivité) : h_{ia} est compatible avec h_{jb} si

(($i \neq j$) et ($a \neq b$) et (G_i même_jonction_que G_j) et (D_a même_jonction_que D_b))

règle 3 : (ordre) : h_{ia} est compatible avec h_{jb} si

(($i \neq j$) et ($a \neq b$) et (((G_i à_gauche_de G_j) et (D_a à_gauche_de D_b)) ou

((G_i à_droite_de G_j) et (D_a à_droite_de D_b)))

Pour ces trois premières règles, la similarité est définie par l'identité sur le vocabulaire des relations. Les deux hypothèses sont donc compatibles si $R(G_i, G_j) \equiv R(D_a, D_b)$.

règle 4 : (transitivité) : h_{ia} est compatible avec h_{jb} si il existe un h_{kc} tel que

(h_{ia} est compatible avec h_{kc}) et (h_{kc} est compatible avec h_{jb})

Cette dernière règle permet de propager la compatibilité lorsque les relations ne sont pas définissables ou comparables. Cependant elle ne reflète pas de propriétés extraites de l'image, ce qui conduit par une utilisation abusive à l'obtention d'appariements incorrects [Skordas 88].

Dans l'exemple suivant [St-Vincent 86], c'est la seconde formulation de la compatibilité qui est utilisée. Le choix du vocabulaire des relations est "voisin de", et la mesure sur les hypothèses est la disparité (Disp) entre les deux indices appariés.

Ainsi, h_{ia} est compatible avec h_{jb} si:

G_i est voisin de G_j et D_a est voisin de D_b

et

$|\text{Disp}(G_i, D_a) - \text{Disp}(G_j, D_b)| = \epsilon$ avec ϵ petit.

Cependant, cette définition est infirmée lorsque des formes voisines représentent la projection d'objets très éloignés en profondeur. D'où des risques d'erreurs d'appariement dans les voisinages de grandes variations en disparité.

III.3.4.3 Méthodes visant à limiter les ambiguïtés d'appariement

Cette section présente les principes généraux permettant d'extraire le "bon" ensemble d'hypothèses en éliminant grâce à des critères relationnelles les ambiguïtés d'appariement lorsque c'est possible (cf. figure III.11). La relaxation, la prédiction-vérification et la recherche de cliques maximales sont trois techniques couramment utilisées.

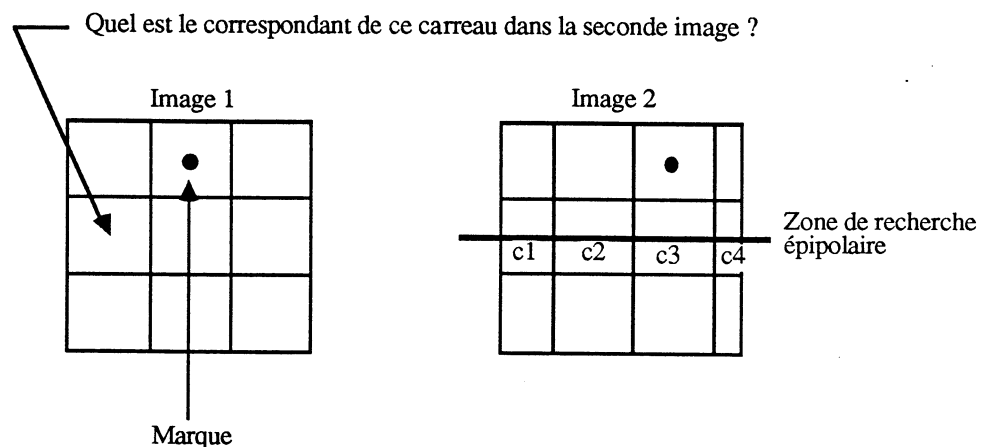


Figure III.11: La marque introduit une contrainte contextuelle qui permet de définir que le correspondant est le carreau c2. Sans elle, il est impossible de choisir entre les carreaux c1, c2, c3 et c4.

Relaxation [Marr 79][Medioni 84]

Cette technique initialement introduite en vision pour résoudre le problème de la recherche d'un étiquetage cohérent [Rosenfeld 76] fut ensuite utilisée dans divers types d'applications tels que la mise en correspondance entre indices. Chaque hypothèse (ou objet au sens de Rosenfeld) est reliée à d'autres appariements potentiels par le biais de relations d'exclusion ou de compatibilité. A chaque hypothèse (ou label au sens de Rosenfeld), est affectée la mesure de

similarité des deux indices associés.

$$h_i = \{i_1, i_2, v_i\} \quad \text{où } v_i = f_r(i_1, i_2)$$

f_r est la fonction de ressemblance

Le principe d'un processus de relaxation consiste à faire varier lors d'une itération de l'algorithme, les valeurs affectées aux hypothèses en fonction des valeurs des hypothèses auxquelles elles sont liées. Si une hypothèse a une forte valeur à l'itération courante, elle augmentera la valeur des hypothèses qu'elle confirme (relation de compatibilité), par contre elle diminuera la valeur des hypothèses qu'elle infirme (relation d'exclusion). L'arrêt de l'algorithme est obtenu après un nombre d'itérations fixe ou lorsqu'il n'y a plus d'évolution significative à chaque itération. La solution est alors formée par les hypothèses ayant été davantage confirmées, c'est à dire celles dont le champ valeur est supérieur à un certain seuil.

Nous remarquons que dans ce type de méthode, la compatibilité ou l'incompatibilité n'a pas besoin d'être définie entre chaque paire d'hypothèses. Chaque "bonne" hypothèse n'est donc pas forcément définie comme compatible avec toutes les autres "bonnes" hypothèses.

Prédiction-vérification [Ayache 85] [St -Vincent 86]

Inclus dans l'ensemble H des hypothèses d'appariements, il existe un sous ensemble d'appariements "valide". Un appariement peut-être considéré comme "valide" si il est cohérent avec un nombre important d'autres appariements "valide". La méthode de prédiction-vérification est basée sur ce principe. Pour un élément de H, élément PREDIT "valide", Il faut alors VERIFIER si cette supposition est licite. Pour cela, on recherche toutes les hypothèses de H consistantes, selon un critère de compatibilité relationnelle, avec l'élément prédite "valide". S'il en existe, soit H_c l'ensemble de ces hypothèses. La recherche des appariements consistants est alors poursuivie récursivement avec chaque élément de H_c . Le processus s'arrête lorsqu'on ne peut plus trouver d'hypothèses consistantes. Ainsi, pour chaque hypothèse prédite "valide", on peut lui associer le nombre d'hypothèses consistantes propagées (pouvoir de prédiction d'une hypothèse : [Ayache 85]). On partitionne alors l'ensemble H en sous ensembles d'appariements incompatibles selon la contrainte d'unicité de Marr. Dans chacun de ces sous ensembles, l'appariement retenu est celui dont le pouvoir de prédiction est le plus élevé. Diverses techniques de choix et de tri entre les hypothèses permettent de limiter le coût de tels algorithmes.

Nous remarquons encore que dans ce type de méthode, il n'est pas nécessaire de définir totalement la compatibilité ou l'incompatibilité entre chaque paire d'hypothèses. Chaque "bonne" hypothèse est seulement confirmée par un nombre important d'autres bonnes hypothèses.

Recherche de cliques maximales [Bolles 79][Skordas 88]

Intuitivement, l'ensemble des "bonnes" hypothèses est le plus grand sous-ensemble d'hypothèses mutuellement compatibles de l'ensemble des hypothèses H . Pour résoudre de manière optimale l'appariement, il est donc intéressant de se donner les moyens de trouver un tel sous-ensemble. Ce problème, transposé dans la théorie des graphes, correspond exactement à la recherche de la plus grande clique maximale d'un graphe G dont les noeuds sont les hypothèses d'appariements et les arcs définissent la compatibilité entre deux noeuds. L'absence d'arcs entre deux noeuds met en évidence l'incompatibilité des deux hypothèses. L'obtention d'un tel graphe nécessite de pouvoir déterminer la compatibilité entre tous couples d'hypothèses. D'après le schéma de détermination de la compatibilité (section III.3.4.2), cela implique la connaissance des relations liant tout couple d'indices dans chaque image. Les graphes modélisant les images doivent donc être complets. Alors, à partir des deux graphes complets modélisant deux images, il est possible de créer un graphe associé G . Une clique (ou sous-graphe complet) dans le graphe G représente donc un sous-ensemble H_s de H d'hypothèses mutuellement compatibles. S'il n'existe aucune autre hypothèse de H compatible avec celle de H_s , alors H_s est représenté dans le graphe par une clique maximale. Ainsi, l'ensemble des appariements "valides" correspond à la plus grande clique maximale dans le graphe. En cas de litiges (plusieurs plus grandes cliques maximales), est choisie celle dont la somme des valeurs de ressemblance de ces noeuds est la plus grande. On peut introduire ici des caractéristiques d'indices non introduites dans la fonction de ressemblance (car insuffisamment robustes), mais intéressantes car très discriminantes (par exemple la longueur des segments).

Il faut remarquer que la recherche des cliques dans un graphe est un problème NP-complet [Ballard 82]. De ce fait tous les algorithmes déterministes permettant de résoudre ce problème ont un coût en temps exponentiel résultant du nombre de noeuds et d'arcs. Ce type d'algorithme semble donc difficilement exploitable. Cependant, comme le précise [Medioni 84], ce problème peut être contourné si on possède les moyens de contrôler le nombre des données en entrée.

III.3.4.4 Conclusion

A partir de l'ensemble des hypothèses d'appariement et des relations entre les indices, les méthodes globales présentées ci-dessus permettent de déterminer le sous-ensemble H_s de H des hypothèses mutuellement compatibles (ou du moins compatibles avec un grand nombre d'autres hypothèses de H_s). L'ensemble des hypothèses de H_s est considéré comme constituant l'appariement entre les deux images traitées. Cette approche apporte une solution élégante pour lever, dans la majorité des cas, les ambiguïtés d'appariements.

III.4 Conclusion

Les techniques de mises en correspondance d'images peuvent être séparées en deux classes. Les méthodes appariant les images au niveau pixel et celle basée sur l'appariement des indices extraits dans les images. Les techniques au niveau pixels sont souvent sensibles aux variations photométriques, et coûteuses en temps de calcul. De ce fait, nous nous sommes intéressés aux techniques d'appariement entre indices images

L'appariement d'indices s'effectue en trois phases. Tout d'abord, on crée des hypothèses d'appariement sur un critère local de ressemblance entre indices. La similarité entre deux indices est mise en évidence en mesurant la distance séparant les valeurs des caractéristiques de ces indices. Une distance faible met en évidence deux indices semblables qui constituent alors une hypothèse d'appariement. Afin d'optimiser les temps de calcul et de limiter les erreurs, il convient, lorsque cela est possible, de restreindre les zones de recherche. Ceci a pour but de limiter la taille de l'ensemble des hypothèses d'appariement. Néanmoins cette restriction ne permet généralement pas de s'affranchir du problème des ambiguïtés de mise en correspondance. Elles doivent être levées par une prise en compte globale des contraintes contextuelles dans les images.

IV REPRESENTATION MULTIRESOLUTION DE L'IMAGE

Ce chapitre souligne l'intérêt d'une représentation multi-résolutions de l'image. Il présente ensuite, au sein de cette représentation, les indices et les structures d'indices que nous avons choisis et les techniques concernant leurs extractions. Enfin, il décrit les relations utilisées entre les indices.

IV.1 Introduction

Considérons un système ayant pour but, à un instant donné, de lire les inscriptions écrites sur une pancarte accrochée à un arbre. Ce système utilise une caméra comme moyen d'appréhender l'environnement extérieur. Pour atteindre son but, il doit d'abord être capable de repérer un arbre sur une image. Si les opérations d'extraction d'indices travaillent en haute résolution (niveau élevé de détails), l'arbre sera perçu comme un ensemble de petites régions représentant les feuilles ou les espaces entre les feuilles, et d'autres représentant la texture de son tronc. Cet ensemble de petites régions est de plus noyées parmi d'autres régions représentant les herbes, les fleurs, les nuages. Il est donc difficile de le repérer. Or il est possible de caractériser un arbre uniquement par un tronc et un feuillage pris globalement. Il est préférable par conséquent de travailler en faible résolution (niveau faible en détails) pour le détecter. Si le robot a besoin de faire une distinction entre les arbres, il peut après les avoir repérés globalement dans l'image, les percevoir de façon plus précise en augmentant la résolution. Il en est de même lorsque l'on doit distinguer la pancarte; il est alors inutile de s'intéresser à la texture sur le tronc et aux inscriptions. Par contre, la lecture des inscriptions doit se faire en haute résolution (vue locale ou en détail de l'image), car dans ce cas, nous devons distinguer les détails de chaque lettre pour les reconnaître. Ainsi, au sein d'une même image, les informations utiles ne se situent pas toujours au même niveau de résolution. D'où l'intérêt de construire une représentation de l'image dans laquelle les informations sont distribuées à diverses résolutions.

Une autre raison justifiant l'utilisation des techniques multi-résolutions réside dans le gain apporté en coût de calcul. En effet, une opération appliquée à toute l'image est souvent très coûteuse. Si l'opération est effectuée en faible résolution, elle est appliquée à une faible

quantité de données (car peu de détails); d'où un faible coût en temps de calcul. Les résultats en faible résolution peuvent alors guider les opérations dans les plus hautes résolutions, ce qui limite les coûts totaux de calcul.

Diverses techniques permettent de créer une telle représentation [Moravec 77] [Marr 80] [Burt 81] [Crowley 81]. Parmi celles-ci, nous utilisons la transformée de DOLP (Difference Of Low Pass) définie par Crowley qui est un moyen efficace pour l'obtention d'une représentation multi-résolutions de l'image.

Il faut ensuite définir et extraire de ces images les indices que nous voulons manipuler. Nous serons amené à utiliser un indice de type point caractéristique "les pics", et une structure d'indices "les arbres de pics". Chacun d'eux fait l'objet d'un exposé détaillé dans la suite de ce chapitre.

Puis, nous décrivons les relations choisies pour représenter les informations contextuelles des images.

IV.2 Pyramide multi-résolutions d'images

IV.2.1 Principe de l'opérateur

Le but est l'obtention d'une représentation de l'image initiale à divers niveaux de résolution. Initialement, une pyramide multi-résolutions était obtenue en découpant l'image à un niveau k en carré de quatre pixels. L'image de résolution inférieure était alors composée de pixels résultant de l'intensité moyenne à l'intérieur de chaque carré de l'image du niveau k (se référer par exemple à Moravec [Moravec 77]). Ainsi d'un niveau de résolution au suivant la taille de l'image diminue par un facteur de 2 (cf. figure IV.1 par [Rosenfeld 84]).

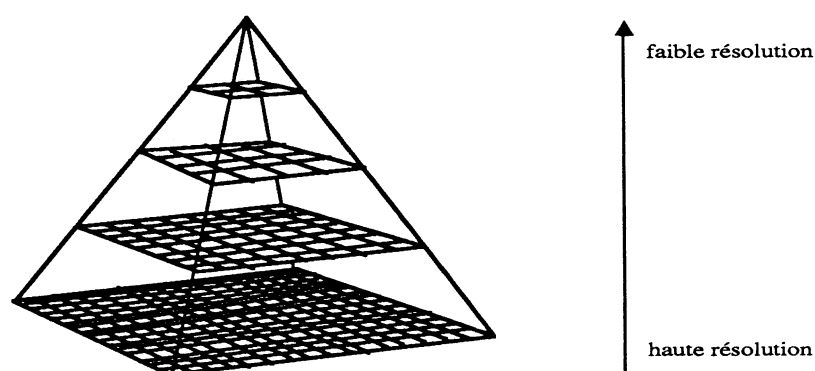


Figure IV.1: Structure pyramidale d'images en multi-résolution.

Cependant, en procédant ainsi, on crée des repliements de spectre ("aliasing"). Pour cette raison, une autre technique est généralement utilisée. Elle consiste à convoluer l'image initiale I avec des gaussiennes de tailles différentes. Plus grande est la taille de la gaussienne, plus importante est l'élimination des hautes fréquences. Cette opération permet donc d'abaisser la résolution de l'image. Elle permet aussi, en respectant la condition du Théorème de Whittaker-Shannon (la fréquence d'échantillonnage doit être au moins égale à deux fois la fréquence maximale du signal) de rééchantillonner le signal pour obtenir la pyramide d'images sans créer de repliement de spectre.

Pour obtenir une pyramide de laplaciens, il faut définir les opérateurs de passe-bande qui seront convolués à l'image initiale I pour donner les passe-bandes d'images B_k . Ces opérateurs sont des laplaciens de gaussiennes. Ils sont souvent approximatés en soustrayant un filtre passe-bas gaussien avec sa propre copie de taille inférieure par un facteur S .

Soit un filtre passe-bas gaussien (g_{k-1}) d'écart type σ .

Soit le même filtre passe-bas gaussien (g_k) d'écart type $S\sigma$.

On définit alors un opérateur passe-bande b_k par la différence entre g_{k-1} et g_k

$b_k = g_{k-1} - g_k$, où k est le niveau de résolution (cf figure IV.2).

(avec $b_0 = 1 - g_0$). Cet opérateur est connu dans la littérature sous le nom informel de chapeau mexicain.

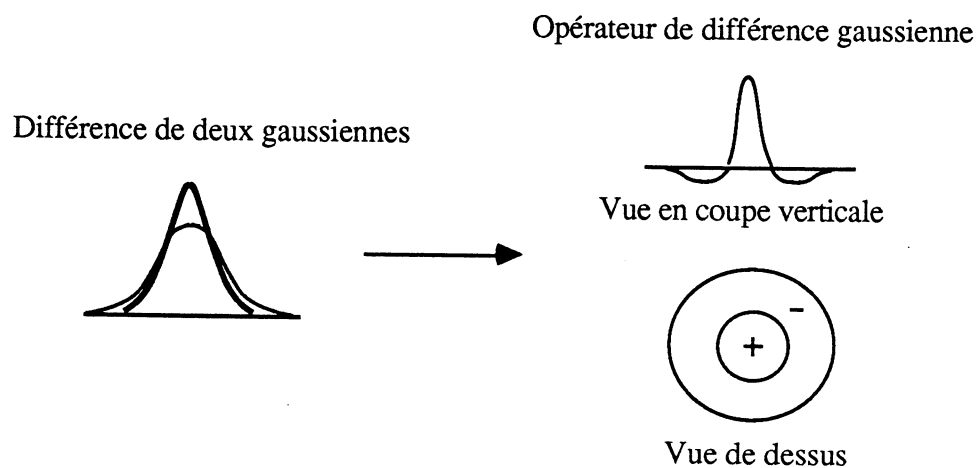


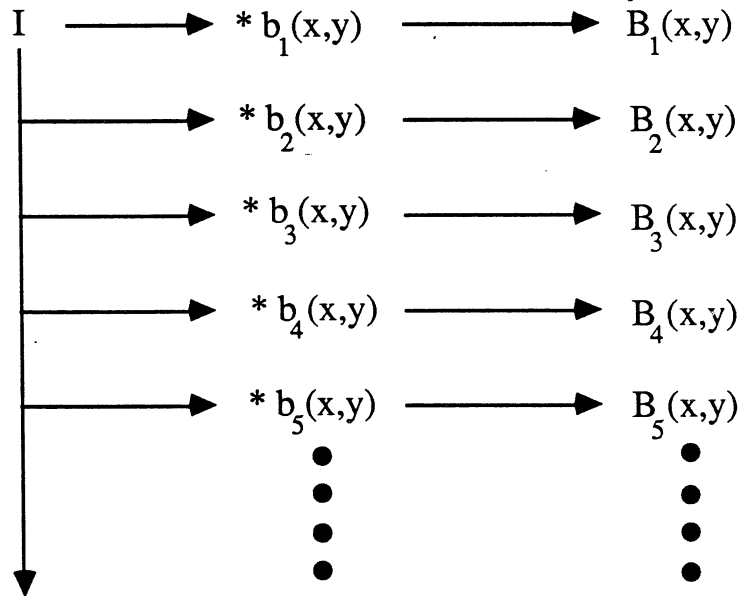
Figure IV.2: Vue en coupe verticale et horizontale de l'opérateur de différence gaussienne.

En faisant croître successivement l'écart type des gaussiennes par un facteur S , on définit une succession d'opérateurs passe-bande de taille de plus en plus grande. Chaque opérateur est alors appliqué à l'image initiale $I(x,y)$ et fournit une approximation du laplacien de l'image à un niveau de résolution. L'opération utilisée est un produit de convolution entre l'image initiale I et un opérateur passe-bande b_k .

$$B_k(i,j) = I(i,j) * b_k(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(m,n) \cdot b_k(i-m,j-n)$$

où $M \times N$ définit le domaine de convolution.

Selon l'importance de taille de l'opérateur, la passe-bande sera d'autant plus étroite et donc la résolution de l'image résultante plus faible. Pour une image initiale de $N = P \times P$ points, nous obtenons au maximum $\text{LOG}_S(N/X_0)$ passe-bandes d'images (cf. annexe1 pour les détails du calcul).



Une implémentation logicielle du principe énoncé ci dessus est coûteux en temps de calcul et en place mémoire [Crowley 84b]. Le coût serait $O(N^2)$ multiplications et $O(N \log N)$ place mémoire (avec N nombre de pixels dans l'image). La réalisation de la transformée de DOLP diffère du principe, tout en permettant l'obtention de résultats similaires avec des coûts inférieurs en nombre de multiplications et en place mémoire. En plus de ces gains en temps et en place, l'algorithme se prête facilement à une réalisation matérielle comme nous le montrons dans le paragraphe suivant.

IV.2.2 Réalisation de DOLP

La transformée de DOLP est un moyen performant d'obtenir à la fois une pyramide de passe-bas d'images et une pyramide de passe-bandes d'images. Les optimisations apportées par rapport à l'algorithme présenté précédemment concerne la construction de cascades de convolutions et du rééchantillonnage.

Cascade de convolutions

La variation d'écart-type S entre deux filtres successifs est choisie égale à $\sqrt{2}$ au lieu de 2 traditionnellement pour conserver un important recouvrement en fréquence entre deux filtres successifs.

Une des bases de la méthode est la propriété de la distribution gaussienne:

$$G(x,y,\sigma_1) * G(x,y,\sigma_2) = G(x,y,\sqrt{\sigma_1^2 + \sigma_2^2})$$

D'où il en découle que

$$G(x,y,\sigma) * G(x,y,\sigma) = G(x,y,\sqrt{2}\sigma)$$

et

$$\underbrace{G(x,y,\sigma) * \dots * G(x,y,\sigma)}_{n \text{ fois}} = G(x,y,\sqrt{n}\sigma)$$

Ainsi, pour faire croître successivement l'écart type d'un filtre initial g_1 par un facteur de $\sqrt{2}$, il suffit d'autoconvoluer n fois le filtre (avec n égal successivement à 1, 2, 4, 8, ...). Soit L_k une image de passe-bas au niveau k de résolution. $L_k = g_k * I$. L'obtention d'un passe-bas au niveau k coûte donc 2^{k-1} autoconvolutions du filtre initial g_1 plus une convolution entre le filtre et l'image. Cette technique est très coûteuse. Il est possible de réduire ce coût en calculant différemment les images de passe-bas.

g_k est un filtre d'écart type $\sqrt{2^{k-1}} \sigma$. Donc d'après la propriété des distributions gaussiennes, $g_k = g_{k-1} * g_{k-1}$.

D'où $L_k = g_{k-1} * (g_{k-1} * I) = g_{k-1} * L_{k-1}$. Ainsi, l'utilisation de l'image de passe-bas précédente permet de diviser le coût en convolution par deux.

Le schéma de la méthode est donc une cascade de convolutions:

$$g_1 * I \rightarrow L_1$$

$$g_1 * L_1 \rightarrow L_2$$

$$g_1 * g_1 * L_2 \rightarrow L_3$$

$$g_1 * g_1 * g_1 * g_1 * L_3 \rightarrow L_4$$

...

Dans un tel schéma, les passe-bandes s'obtiennent facilement et rapidement. En effet dans l'algorithme précédent, $B_k = b_k * I$

$$\text{Or } b_k = g_{k-1} - g_k$$

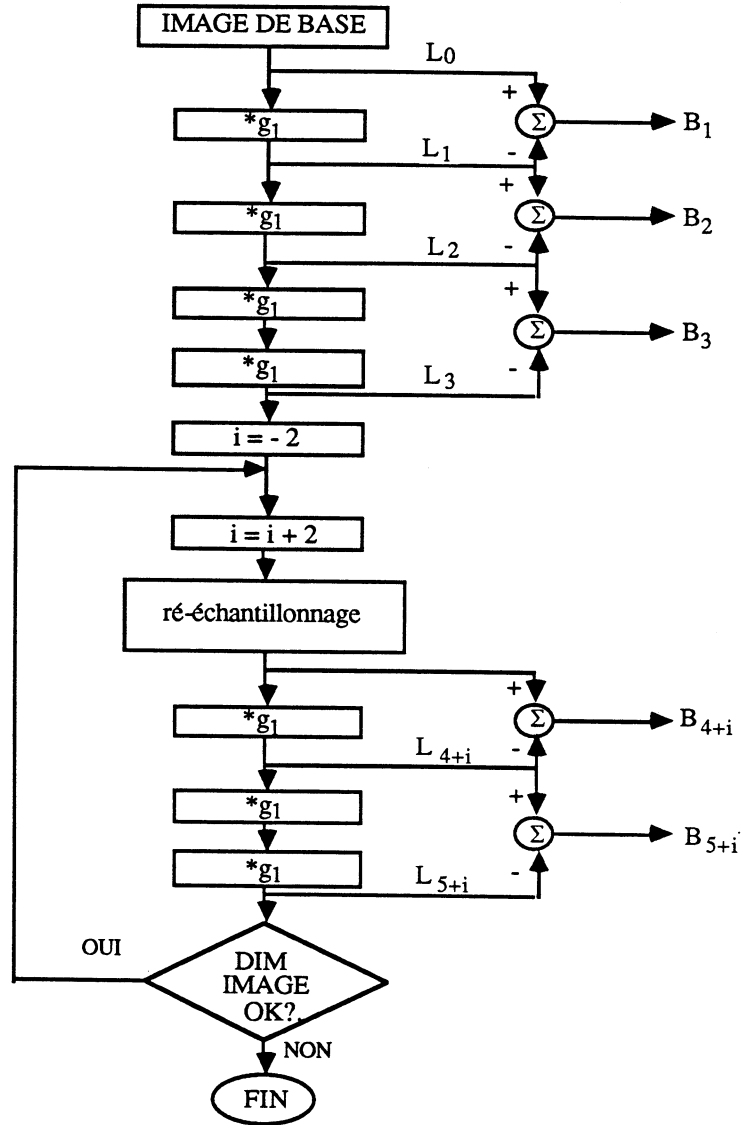
$$\text{Donc } B_k = (g_{k-1} - g_k) * I = g_{k-1} * I - g_k * I = L_{k-1} - L_k.$$

Cependant, bien que l'implémentation de la cascade de convolutions soit aisée, le nombre de convolutions s'accroît rapidement.

Rééchantillonnage

En se rappelant que la convolution avec un filtre passe-bas élimine les hautes fréquences d'un signal, il est possible, après un certain nombre de convolutions avec ce filtre, d'effectuer un rééchantillonnage sans risque de perte d'informations (Théorème de Whittaker-Shannon). Ce rééchantillonnage, en divisant la taille de l'image par 2, revient à multiplier l'écart-type du filtre par 2. D'où au $n^{\text{ième}}$ rééchantillonnage, un filtre g_1 d'écart type σ a virtuellement lorsqu'il est appliqué à l'image, un écart-type de $2^n \sigma$. Ainsi, si L_{k-1} est obtenu par convolution avec un filtre d'écart-type de $2^n \sigma$, L_k rééchantillonné n fois est égal à $g_1 * L_{k-1}$. Opération nécessitant une seule convolution.

Un prototype de DOLP câblé a été réalisé au LTIRF (Laboratoire de Traitement d'Image et de Reconnaissance de Forme) par des membres de l'équipe du professeur Chehikian. Nous en donnons l'organigramme, extrait de la thèse de [Maffres 86].



où L_i désigne l'image de passe-bas de niveau i de résolution.

B_i désigne l'image laplacien de niveau i de résolution.

$g_1 = g * g * g$ avec g la réponse impulsionnelle du filtre séparable gaussien 3x3 utilisé.

$$g = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \frac{1}{16} \left(\begin{bmatrix} 1 & 2 & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \right)$$

La complexité d'un tel algorithme est en $O(N)$ (cf. Annexe2)

La validité du rééchantillonnage dans l'approche DOLP par rapport au théorème de Whittaker-Shannon est démontrée par la figure IV.3.

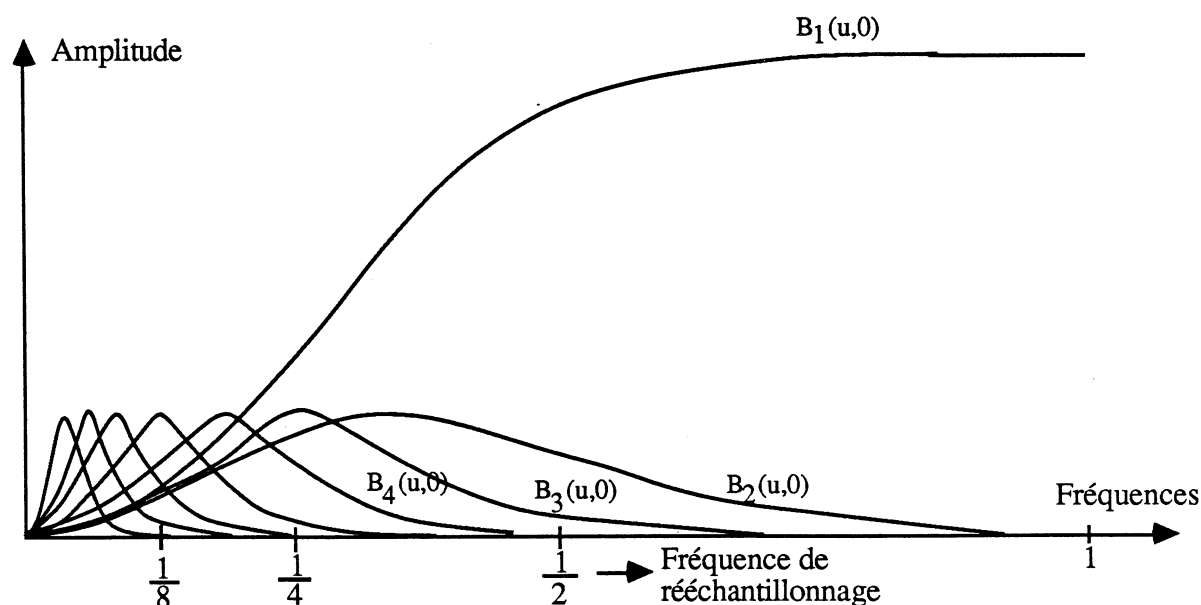


Figure IV.3: Spectre de fréquences des passe-bandes obtenues. Le rééchantillonnage se fait lors de l'obtention de $B_4(u,0)$, $B_6(u,0)$, ..., $B_{2i}(u,0)$, d'où pas de repliement de spectre.

L'intérêt d'utiliser le filtre g défini précédemment, outre le fait qu'il soit séparable, réside dans la possibilité d'obtenir par autoconvolution des filtres qui ont tous une fonction de transfert monotonique en fréquence. En effet:

Soit le filtre $f_1(x) = [1 \ 2 \ 1]$. Sa fonction de transfert est $F_1(w) = 2\cos(w) + 2$. Cette fonction décroît monotoniquement entre les valeurs 4 à $w = 0$ et 0 à $w = \pi$.

Soit un filtre $f_n(x)$ obtenu par n autoconvolutions de f . Sa fonction de transfert correspond à n produits de la fonction $F_1(w)$. Or le produit de fonctions monotoniques donne une fonction monotonique. Par conséquent, quel que soit n , la fonction de transfert $F_n(w)$ est monotonique.

IV.2.3 Résultats

Présentations des pyramides multi-résolutions d'images de passe-bas et de passe-bande obtenues par la transformée de DOLP sur une image composée de pièces diverses (manufacturées, outils, ...).

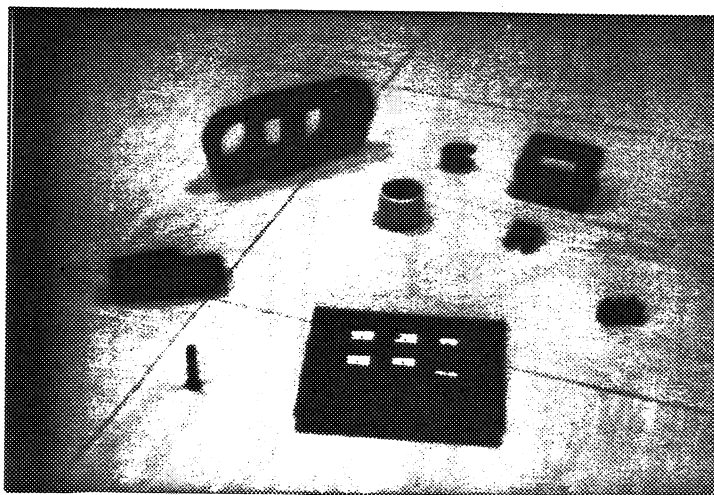
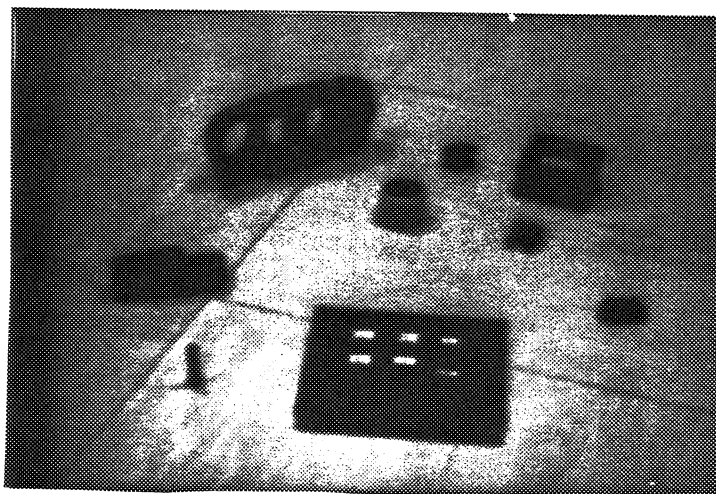
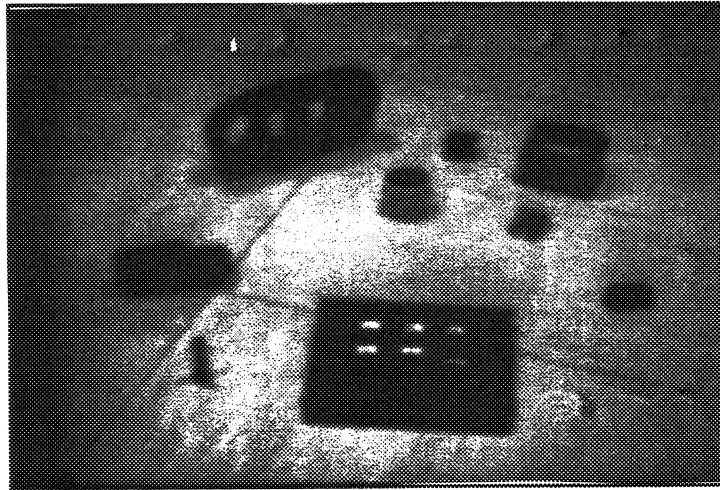


Image d'origine

Pyramide de passe-bas (images de 1 à 10)



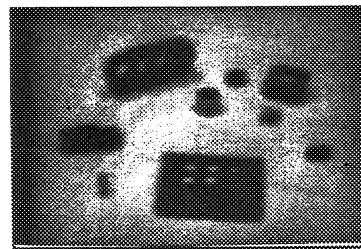
L_1



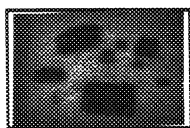
L₂



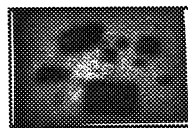
L₃



L₄



L₅



L₆



L₇



L₈

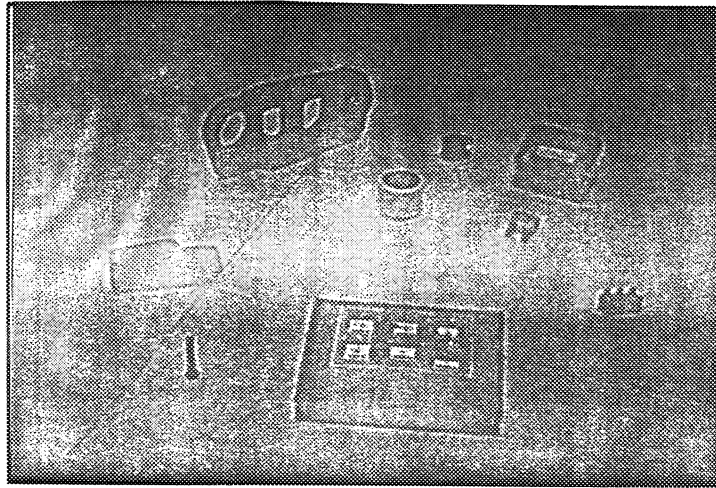


L₉



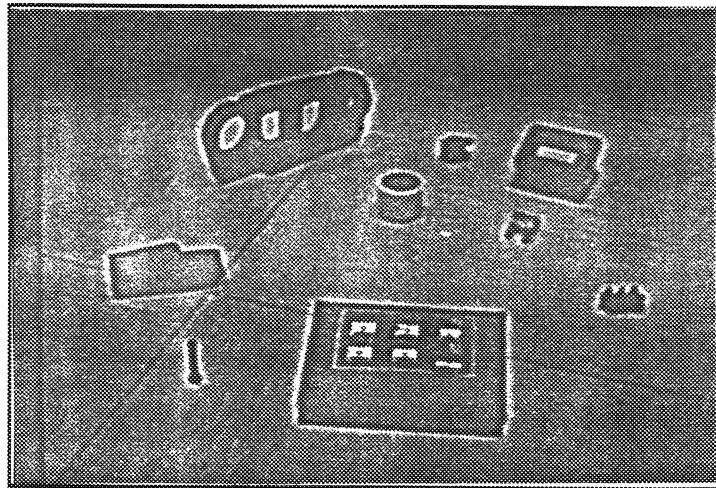
L₁₀

Pyramide de passe-bande (images de 1 à 10)



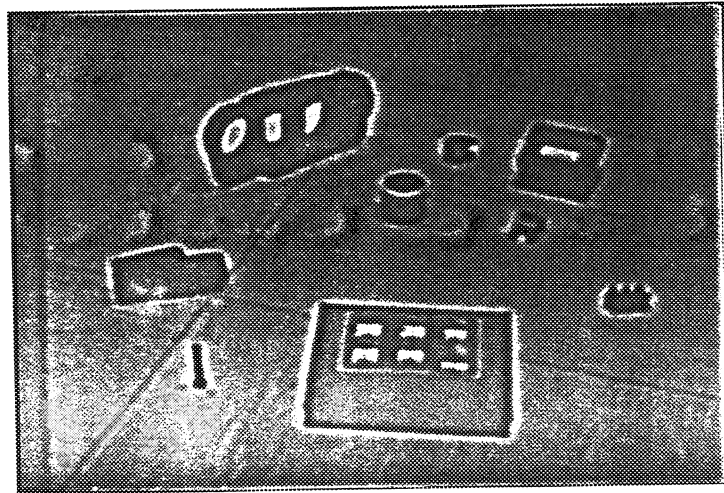
B_1

(Cette image correspond en réalité à un passe-haut de l'image initiale (cf. figure IV.3); elle ne sera donc pas utilisée par la suite)

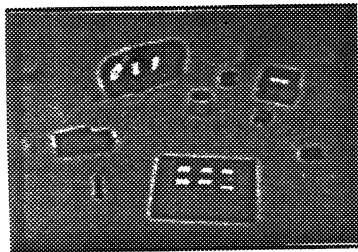


B_2

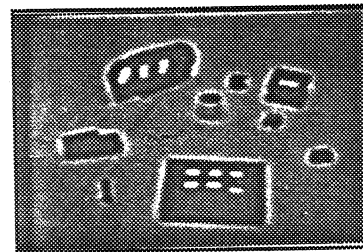
(Remarque: Les crêtes du laplacien se détachent dans l'image. Les contours blanc sont les crêtes positives du laplacien, les contours noirs les crêtes négatives. Ces crêtes encadrent le contour réel de la forme.)



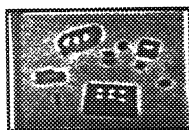
B₃



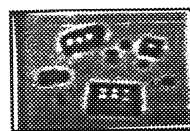
B₄



B₅



B₆



B₇



B₈



B₉



B₁₀

IV.3 Extraction des indices

La transformation de l'image initiale en une pyramide multi-résolutions d'images a permis une redistribution pertinente des informations sans perte d'informations à l'intérieur de la pyramide [Crowley 81]. Il convient maintenant d'extraire ces informations dans un format adapté au problème de la mise en correspondance. Ce sont les indices images. Ci-dessous nous présentons le type d'indice avec lequel nous avons travaillé. Ce sont les "pics". Nous y rajoutons un second indice, les "crêtes", car bien que n'ayant pas été utilisé directement en tant qu'indice lors de la phase d'appariement, il contribue à l'amélioration de la détection des "pics".

IV.3.1 Les pics

Les pics sont des indices de type "point caractéristique" et par conséquent véhiculent peu d'informations sur l'image. Néanmoins, ils possèdent de nombreux avantages à savoir: une manipulation aisée et des possibilités étendues de traitements.

IV.3.1.1 La détection

L'extraction des indices images s'effectue dans chaque image de la pyramide de laplacien. Les indices choisis, appelés pics, correspondent aux extrema de la dérivée seconde (cf. figure IV.4).

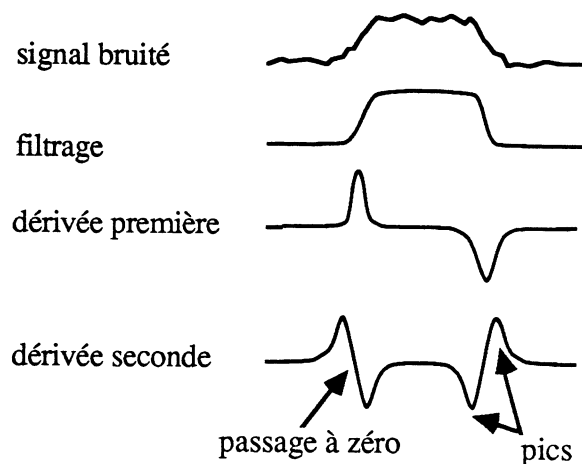


Figure IV.4: Pics dans la dérivée seconde d'un signal

Ces extrema correspondent aux maximum positifs ou aux minimum négatifs dans le laplacien.

Dans la réalisation pratique, les pics dans une image de passe-bande sont les pixels minimum négatifs ou maximum positifs par rapport au 8-voisinage [Crowley 84].

Soit le voisinage du pixel $p(i,j,k)$:

V_2	V_1	V_0
V_3	P	V_7
V_4	V_5	V_6

Un pic est défini de la manière suivante:

$\text{pic} = M_0 \text{ et } M_1 \text{ et } M_2 \text{ et } M_3$ (M_n : Maximum dans la direction n)

avec $M_n = ((|P| \geq |V_n|) \text{ ou } (\text{Sign}(P) \neq \text{Sign}(V_n)))$

et

$((|P| \geq |V_{n+4}|) \text{ ou } (\text{Sign}(P) \neq \text{Sign}(V_{n+4})))$

Un pic représente un maximum de corrélations entre le filtre et la forme perçue. Ainsi, les caractéristiques des formes tels que les coins, les extrémités et la forme globale sont représentées par des pics (cf. figure IV.5).

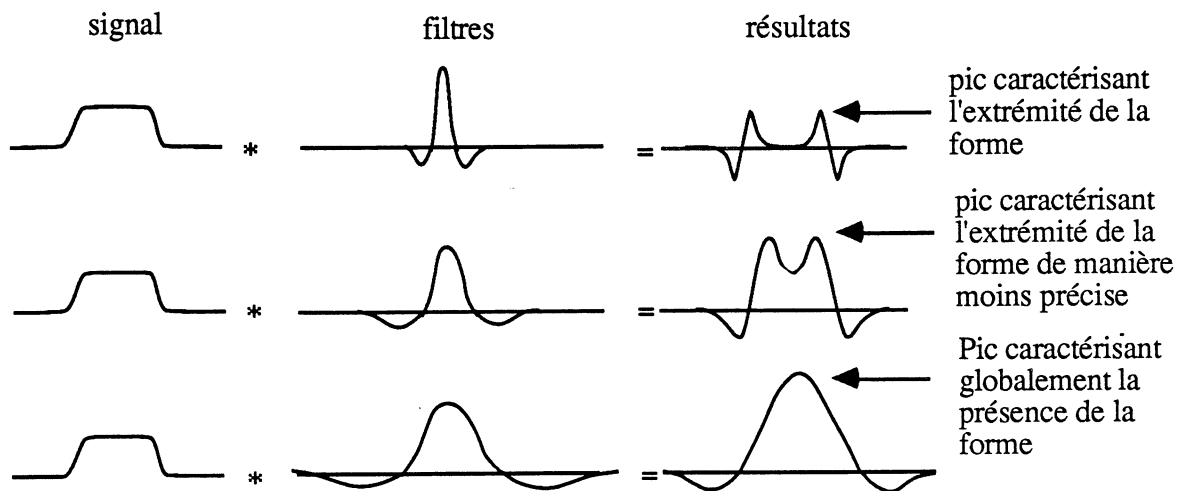


Figure IV.5: Convolution d'un signal 1D avec des opérateurs de différence gaussienne de différentes tailles. Les pics extraits dans les plus hautes résolutions (résultat du haut) caractérisent de manière précise la présence de l'extrémité de la forme. Les pics extraits dans les plus faibles résolutions (résultat du bas) caractérisent la présence de la forme.

Il faut remarquer que seul les coins localement convexes sur une forme sont repérés par des pics. Or toute forme possède son "négatif" qui correspond au complément de la forme dans l'image. A toute partie concave de la forme est associée une partie convexe dans son négatif. Ainsi, il existe les pics représentant la présence de coins dans les parties concaves des formes à l'intérieur du négatif de la forme (cf figure IV.6)

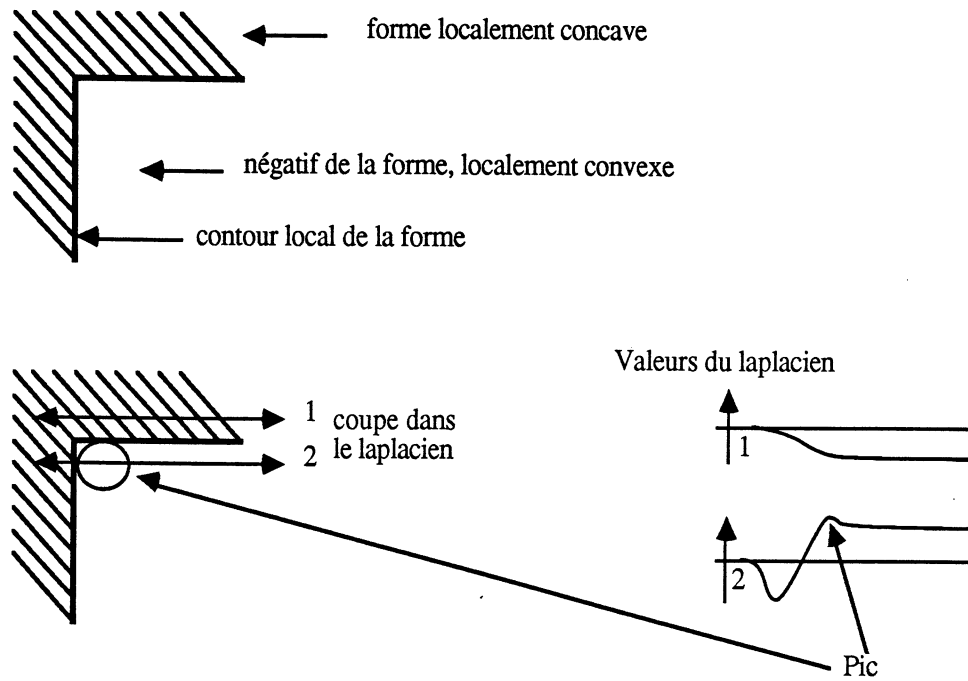


Figure IV.6: Présence d'un pic dans le négatif d'une partie de forme localement concave

IV.3.1.2 Limitation de la détection

Certains pics détectés ne représentent pas les caractéristiques des formes énoncées précédemment.

La première raison en est **le bruit sur les crêtes du Laplacien**.

Les crêtes dans le laplacien sont des lignes maximales en valeur positive ou minimales en valeur négative dans au moins une direction par rapport au voisinage. Elles se situent de part et d'autre du contour des formes. Ainsi, le bruit occasionne des variations de la valeur du laplacien de certains pixels le long de ces crêtes qui peuvent engendrer des pixels maximum positifs locaux ou minimum négatifs locaux dans les quatre directions. Ils représenteront donc des "faux" pics. Nous détaillerons par la suite ce que représentent les crêtes.

La seconde raison en est **l'échantillonnage du signal**.

En théorie, chaque ligne parallèle au contour d'une forme d'intensité constante possède, excepté à proximité des variations de courbure, une valeur constante du laplacien. Parmi ces lignes, la crête du laplacien a une caractéristique particulière. Elle est maximale positive ou minimale négative dans au moins une direction. Les autres lignes parallèles à la crête ont des valeurs du laplacien décroissantes avec la distance les séparant de la crête.

Or, par échantillonnage les pixels représentant la crête se situent à $[-1,+1]$ en distance pixel par rapport à la position théorique de la crête. Ainsi, la variation en distance des pixels approximant la crête par rapport à la position théorique leur confère des valeurs du laplacien différentes. D'où le risque de voir apparaître des "faux" pics le long des crêtes (cf figure IV.7).

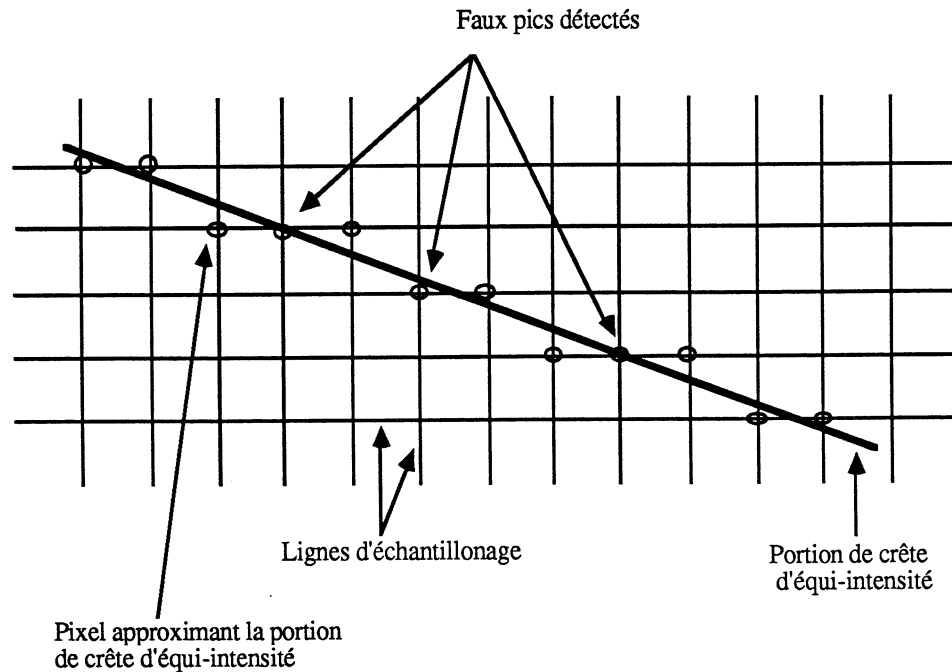


Figure IV.7: "Faux" pics dûs à l'échantillonnage.

La représentation de l'image sera d'autant plus instable que la quantité de "faux" pics augmentera. Ainsi, lors de traitements ultérieurs tels que la mise en correspondance de deux images, ces "faux" pics pourront perturber fortement la qualité des résultats. Ces considérations nous amènent à rechercher les moyens d'éliminer un maximum de "faux" pics.

IV.3.1.3 Amélioration de la détection

Les "vrais" pics sont soit des pics isolés (ils localisent dans ce cas une forme de manière globale), soit des pics liés à une seule crête (ils localisent une extrémité de forme), soit des pics liés à plusieurs crêtes (ils localisent des coins). Les "faux" pics, comme nous l'avons vu précédemment sont eux aussi reliés à deux crêtes (une portion de crête avant le pic et une autre après). Par contre, ils ne localisent pas de variation de direction entre les deux crêtes. Ainsi, il est possible de différencier les "vrais" pics des "faux" en calculant l'angle que forme les crêtes qui se rattachent à ce pic.

Cependant certains "faux" pics répondent à ces propriétés. Ce sont des pics sur des parties de forme proches d'une concavité (cf. figure IV.8).

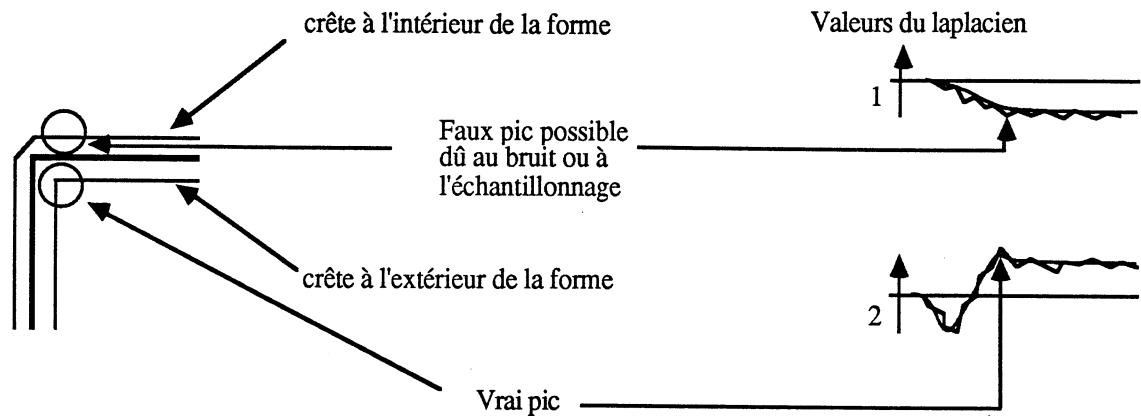


Figure IV.8: "Faux" pics vérifiant le critère d'angle.

Le critère d'angle ne convenant pas à ces faux pics, il faut déterminer, afin de les éliminer, s'ils appartiennent à une partie de forme concave. La détermination de la position du passage à zéro du laplacien par rapport à l'angle intérieur entre deux portions de crête fournit le moyen de détecter des concavités ou des convexités locales.

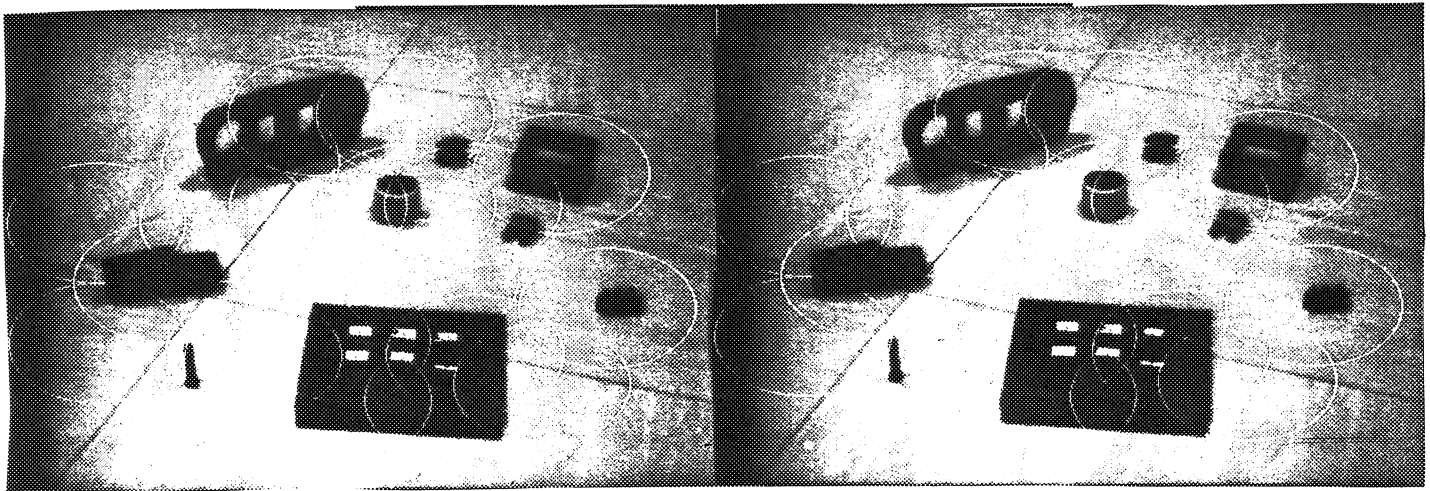
Si le passage à zéro est à l'intérieur de l'angle et qu'il n'existe pas de passage à zéro suffisamment proche à l'extérieur de l'angle, alors le pic se situe sur une portion concave de la forme. Il est donc éliminé.

Si le passage à zéro est à l'extérieur de l'angle, alors le pic se situe sur une portion convexe de la forme. Ce pic est conservé.

IV.3.1.4 Résultats

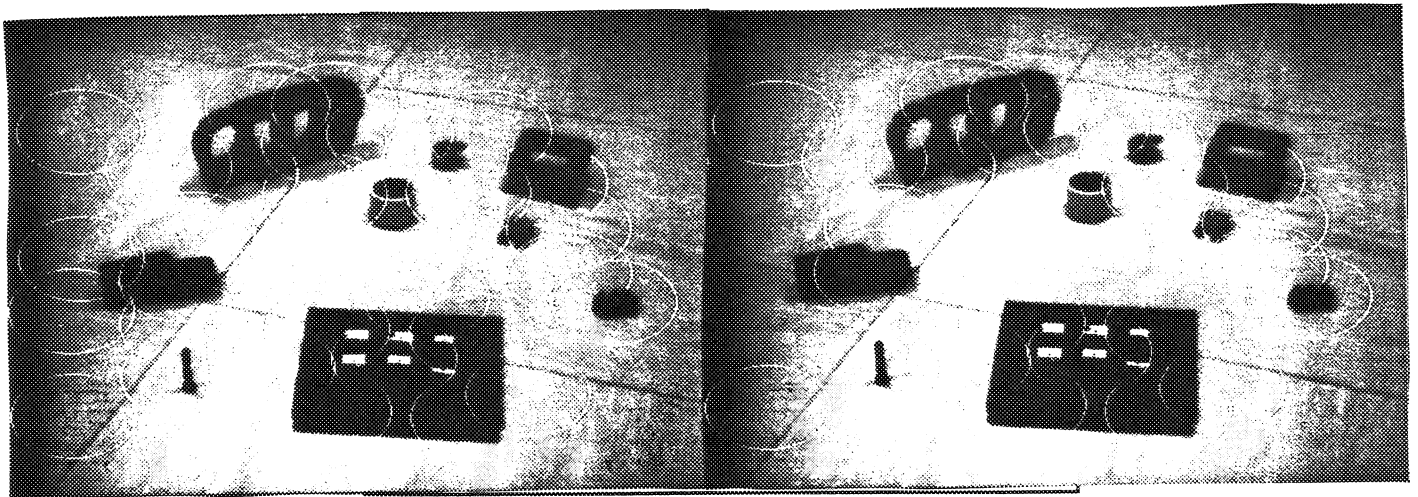
Nous présentons en parallèle les deux techniques d'extraction de pics en vue de simplifier la comparaison visuelle des résultats. La première repose sur l'utilisation du seul critère de maximalité local, tandis que la seconde incorpore les critères supplémentaires d'angle et de détection de concavité locale. Les pics extraits sont visualisés par des incrustations de cercles sur l'image initiale. Le centre du cercle donne la position du pic détecté, et le diamètre donne la taille du lobe positif du filtre de différence gaussienne utilisé. Ainsi, les forts diamètres correspondent aux faibles résolutions et réciproquement.

Après la visualisation des images et des résultats obtenus, nous donnons un tableau comparatif explicite des résultats des deux méthodes sur l'image analysée. Ainsi, il est facile de constater que 80% des "faux" pics sont éliminés dans les hautes résolutions par l'adjonction des deux critères supplémentaires. Le même taux d'élimination a été obtenu sur les scènes de bureaux présentées dans le chapitre suivant (par exemple sur l'une de ces images, on passe de 1179 pics à 331 au niveau 3 de résolution).



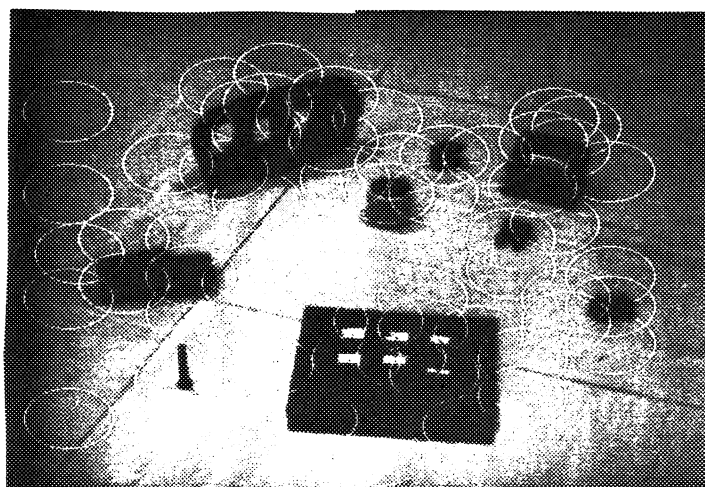
K = 10 Ancienne méthode

Nouvelle méthode

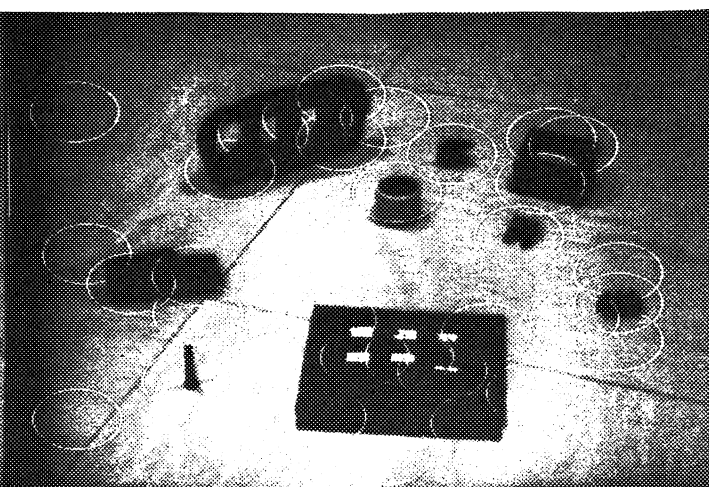


K = 9 Ancienne méthode

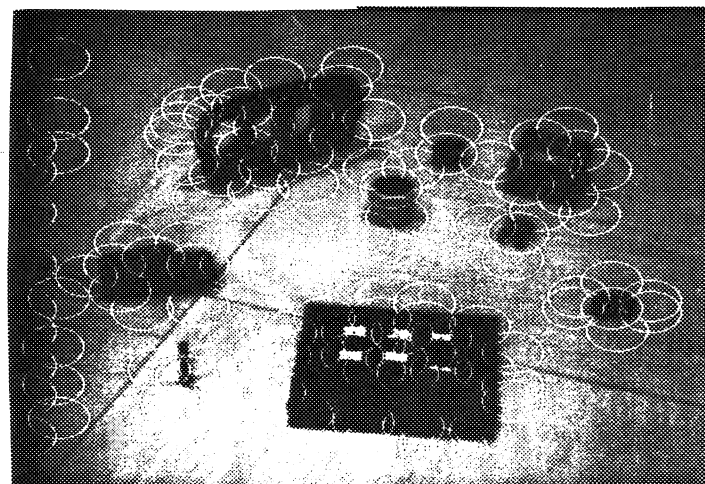
Nouvelle méthode



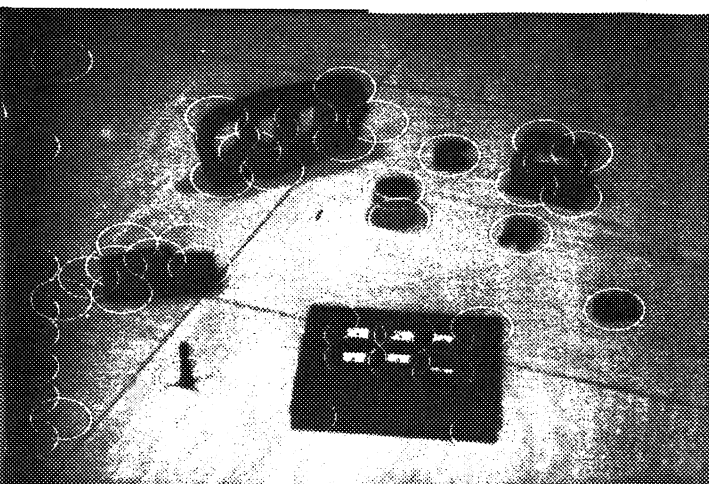
K = 8 Ancienne méthode



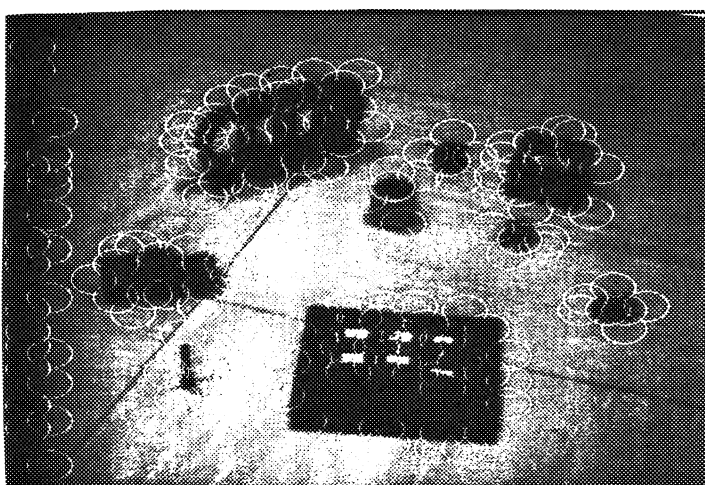
Nouvelle méthode



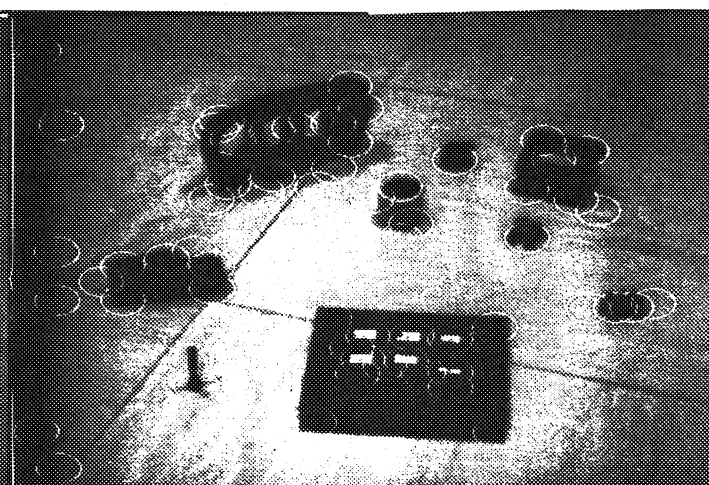
K = 7 Ancienne méthode



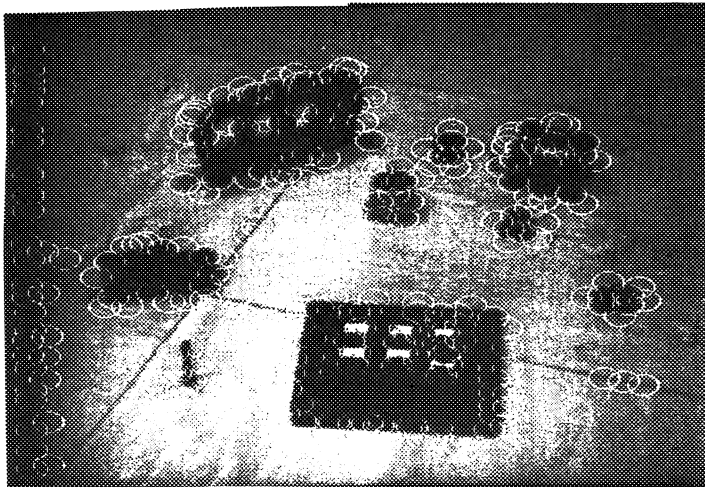
Nouvelle méthode



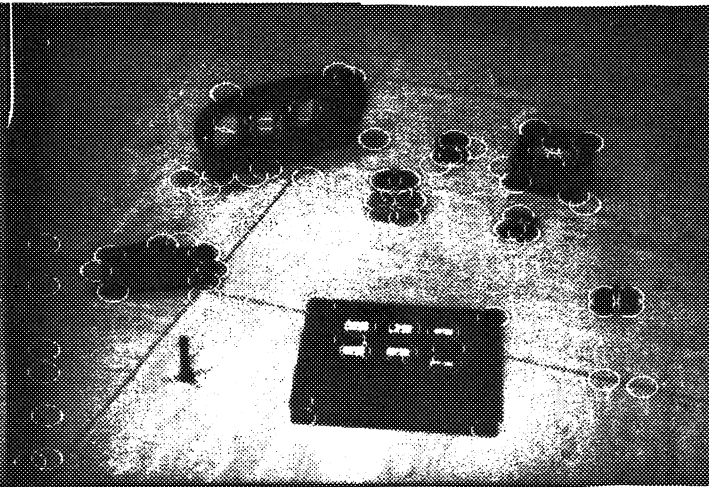
K = 6 Ancienne méthode



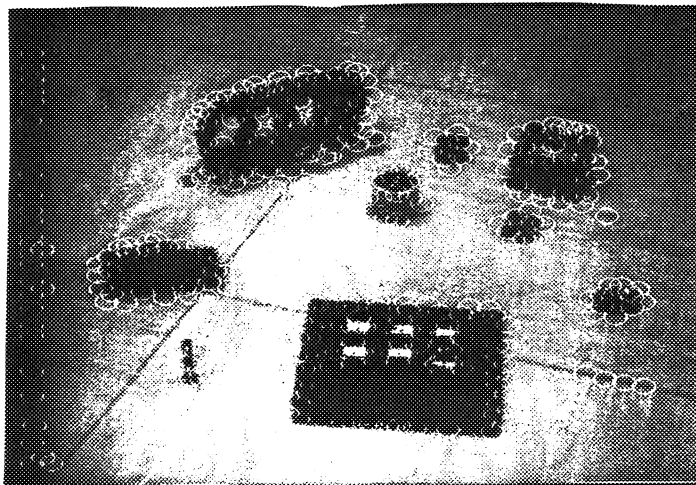
Nouvelle méthode



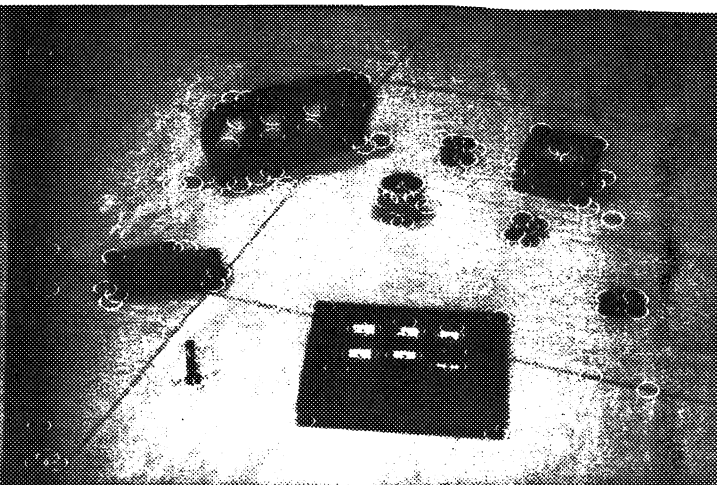
K = 5 Ancienne méthode



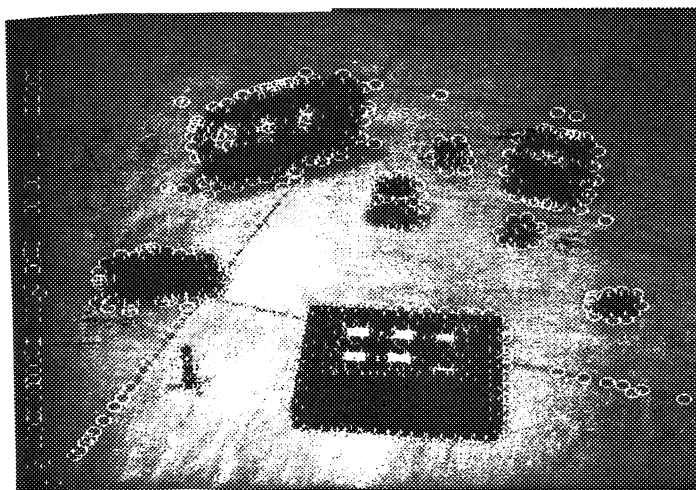
Nouvelle méthode



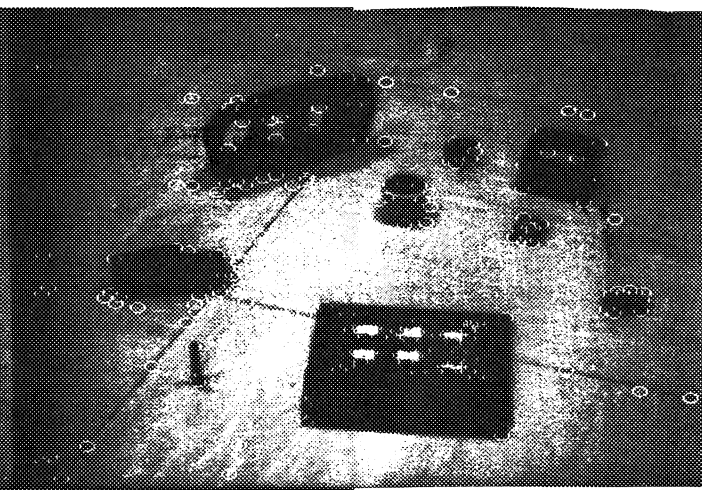
K = 4 Ancienne méthode



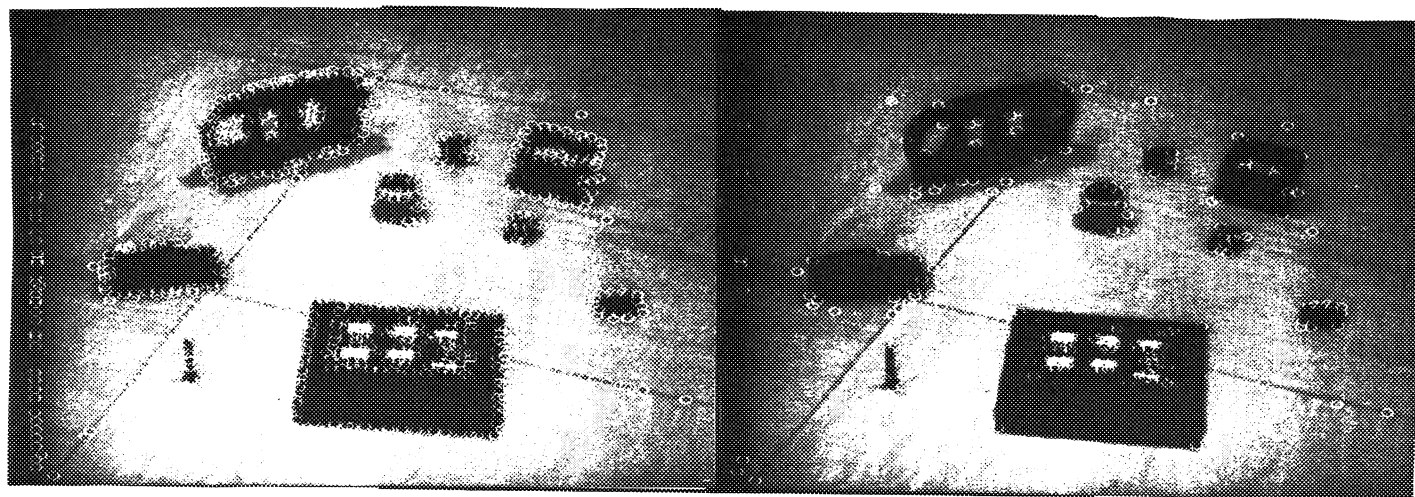
Nouvelle méthode



K = 3 Ancienne méthode



Nouvelle méthode



$K = 2$ Ancienne méthode

Nouvelle méthode

K	Ancienne méthode			Nouvelle méthode		
	Total pics	Vrais pics	Faux pics	Total pics	Vrais pics	Faux pics
10	19	16	3	16	16	0
9	32	23	9	26	23	3
8	58	28	30	32	26	8
7	109	36	73	48	35	13
6	152	47	105	62	46	16
5	236	64	172	84	60	24
4	323	82	241	116	75	41
3	528	110	418	168	104	64
2	699	123	576	195	117	78

Malgré ces bon résultats, on peut constater qu'il subsiste encore environ 20% des "faux" pics initiaux, pour les raisons suivantes :

L'élimination des "faux" pics passe par une phase de calcul d'angle entre une portion de crête de part et d'autre du pic. Ainsi, de petits sauts locaux d'un pixel le long des crêtes peuvent engendrer un angle trop grand et une détection de convexité locale. Cet inconvénient apparaît lorsqu'une portion de la crête théorique se situe entre deux portions de lignes de même valeur laplacien. Dans ce cas, il y a des risques d'oscillation des pixels de la crête extraite d'entre ces deux lignes.

De même, les "faux" pixels sur une portion de crête située le long d'une partie de forme arrondie, ne peuvent être éliminés par un simple calcul d'angle. Ce problème nous incite actuellement à restreindre le champ d'application de la détection des pics. Dans les scènes choisies, les objets possèdent peu ou pas de portions arrondies. Les scènes de bureau respectent bien cette contrainte.

Comme pour la détection des passages à zéro, la détection des crêtes est sensible au contraste. Un faible contraste peut entraîner une rupture de la crête. Ainsi, le pic le plus proche de cette rupture est considéré comme un pic sur une extrémité. Par conséquent, il ne peut être éliminé par un calcul d'angle ou de concavité.

Il en est de même pour le dernier pic sur une crête atteignant le bord de l'image, comme nous pouvons le constater sur les photos présentées.

IV.3.1.5 Caractéristiques sur les pics

Il est possible d'affecter à chaque indice de type pic certains attributs qui permettent de le caractériser. Nous avons retenu comme caractéristiques du pic, sans toutefois toutes les utiliser systématiquement dans nos réalisations : la position dans l'image, le niveau de résolution, la valeur du laplacien, le signe, l'étiquetage, la valeur du voisinage. Nous les détaillons ci-dessous.

La position: Cette caractéristique est très discriminante dans l'image, car à une position donnée ne peut correspondre qu'un seul pic. Par contre, pour un couple de pics correspondant entre deux images, elle n'est pas du tout robuste avec le mouvement de la caméra ou le mouvement dans la scène.

Le niveau de résolution: Cette caractéristique permet uniquement de discriminer les pics entre divers niveaux de résolution, mais pas de faire une distinction entre les pics au sein d'une image.

La valeur du laplacien: Cette caractéristique s'obtient par une lecture directe sur l'image de passe-bande. Cette valeur est bornée, elle permet par conséquent de classer les pics au sein d'une image en M ensembles (où M est le nombre de valeurs possibles du laplacien). D'une image à l'autre, la variation de position relative des formes ou les variations de luminance, entraînent des variations sur valeurs du laplacien.

Le signe: Cette caractéristique permet seulement de discriminer la moitié des pics d'une image. Par contre entre deux images successives, elle est très robuste. Un maximum de courbure sur une partie de forme plus foncée que son négatif aura un pic négatif. Un maximum de courbure sur une partie de forme plus claire que son négatif aura un pic positif. Entre deux images, un pic peut changer de signe si par mouvement et projection perspective le contraste entre la partie de forme et son négatif s'inverse. Expérimentalement, nous n'avons jamais constaté un tel phénomène. Ce cas semble fort rare, et ce d'autant plus que le déplacement entre la prise de vue de deux images successives est faible.

L'étiquetage [Crowley 87]: Le but est de différencier les pics en fonction des pics voisins qui lui sont directement reliés, signifiant qu'il n'existe aucun autre pic le long de la crête reliant deux pics.

Pics taches: Ils ne relient aucun pic voisin par une crête. Ils symbolisent la présence d'une forme, ou d'une partie de forme, et par conséquent sont surtout détectés en très faible résolution, excepté lorsque qu'ils repèrent une petite forme apparaissant en haute résolution.

Pics fins de barre: Ces pics ont un seul voisin. Ils représentent l'extrémité d'une forme repérée à un niveau donné. Ils sont surtout présents en faible et en moyenne résolution.

Pics coins: Ils ont deux voisins. De par la technique d'élimination des pics, ils représentent forcément une courbure maximale de la forme. Ils sont surtout présents en moyenne et en haute résolution.

Pics divers: Ils ont au moins trois voisins qu'ils relient par des crêtes. La figure IV.9 représente un cas de figure où un tel pic est détecté.

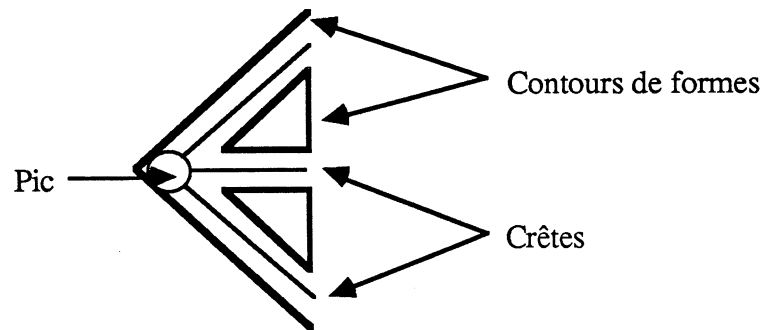


Figure IV.9: Exemple de pic relié à trois crêtes.

La valeur du voisinage: C'est une caractéristique contextuelle, mais locale. En ce qui concerne la robustesse, les occlusions et les variations de luminance font varier la valeur des pixels. La valeur du voisinage de deux pics correspondants entre deux images successives peut présenter des variations importantes. Ainsi, la discrimination et la robustesse sont deux critères qui s'affrontent dans la définition de la taille du voisinage. Plus la taille du voisinage est importante, plus grande est la discrimination; par contre, la robustesse est dans ce cas affectée car les variations en occlusion sont amplifiées.

IV.3.1.6 Conclusion

Nous venons de présenter une réalisation originale permettant d'extraire des points significatifs de l'image. Cette réalisation allie les avantages des deux types de techniques permettant l'extraction des points caractéristiques qui sont respectivement: Les opérateurs d'extraction travaillant directement sur la luminance de l'image et les opérateurs travaillant sur les contours. Ainsi, la plupart des "faux" pics dûs au bruit et à l'échantillonnage obtenus par une méthode appartenant au premier type de technique (extraction des pics du laplacien) est ensuite éliminée par l'utilisation d'une autre méthode appartenant au second type de technique (calcul d'angle entre les crêtes de par et d'autre des pics). Les problèmes inhérents au second type de techniques, à savoir les "faux" points dus aux erreurs de segmentation, sont supprimés par l'utilisation en premier de l'extraction des pics du laplacien.

Il s'avère que la détection des vrais pics nécessite d'exploiter l'information fournie par les crêtes. Pour cette raison nous présentons les "crêtes" du laplacien dans le paragraphe suivant. Il est à noter que ces indices furent utilisés dans de nombreux travaux [Crowley 84][Giraudon 86b][Long 86].

IV.3.2 Les crêtes

IV.3.2.1 La détection

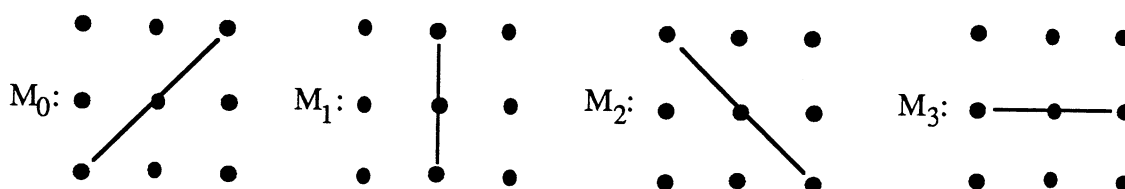
Dans le domaine réel, une crête est une ligne du laplacien de l'image qui est maximale positive ou minimale négative dans au moins une direction. Ces lignes se placent de part et d'autre du contour de la forme [Giraudon 85]. Giraudon a utilisé ces lignes comme primitive stéréo. Les avantages qu'apporte une crête par rapport au passage à zéro du laplacien sont les suivants:

Chaque ligne appartenant à une région; on peut lui attacher des attributs texturaux propres à la région. Ainsi, on obtient une plus grande discrimination entre les primitives lors de la phase d'appariement.

Il est possible de coupler chaque crête positive avec la crête négative appartenant au même passage à zéro en vue de disposer d'une contrainte de cohérence lors de la phase d'appariement.

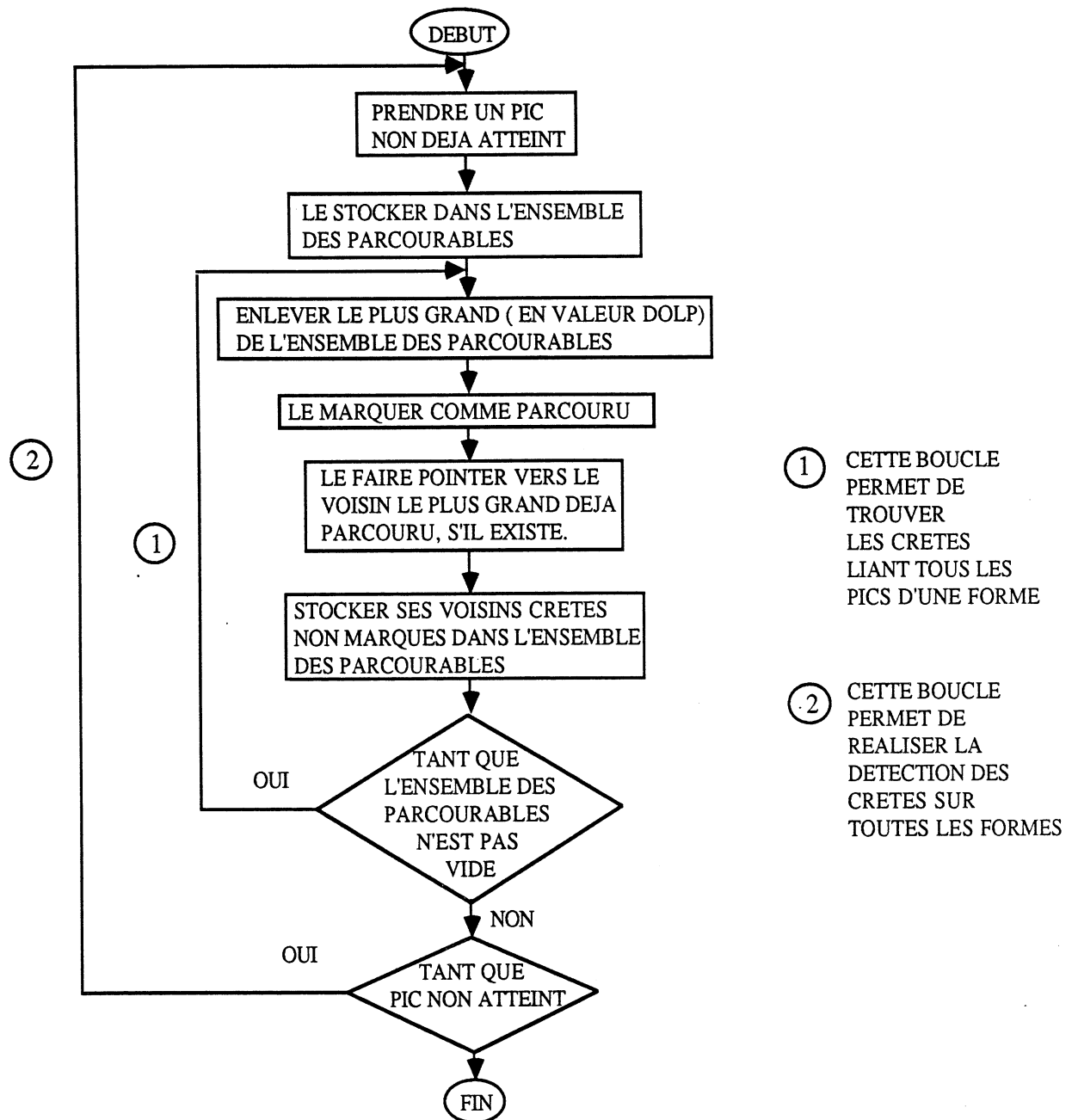
Dans le domaine discret, un pixel est étiqueté comme appartenant à une crête s'il a une valeur maximale dans au moins une des quatre directions. Le pic est un cas particulier de pixel crête.

Pixel crête = M_0 ou M_1 ou M_2 ou M_3 (M_n : Maximum dans la direction n)



Cette méthode fournit un amas de pixels "crête" le long de la crête réelle. Il convient alors d'en extraire les crêtes d'épaisseur égale au pixel.

Pour cela nous avons utilisé une stratégie de recherche en "profondeur d'abord" basée sur la valeur DOLP (algorithme GraphSearch [Nilsson 80]).



Les crêtes sont alors extraites en remontant le chemin de pointeur en pointeur en partant d'un pic. Chaque pixel atteint lors du parcours est un pixel de la crête.

Toutefois, nous ne repérons pas toutes les crêtes. Pour une forme fermée caractérisée par N courbures maximales localement convexes le long de son contour, nous devrions repérer N pics et N crêtes le long de son contour. D'après l'algorithme présenté, chaque pixel crête aura

un seul pointeur, qui par suivi des pointeurs successifs conduira à un pic. Il en est évidemment de même pour les pics qui représentent des pixels crêtes particuliers. Or le pic d'une forme avec lequel est initialisée la recherche des crêtes ne pointe vers aucun voisin car il représente le premier pixel "crête" parcouru pour la forme. Chaque pic de la forme conduit donc vers un autre pic, excepté le premier pris, ce qui implique que nous détectons seulement $N-1$ crête. Comme conséquence de l'algorithme utilisé, chaque crête manquante possèdera deux pixels voisins qui pointent vers deux pics différents (cf. figure IV.10).

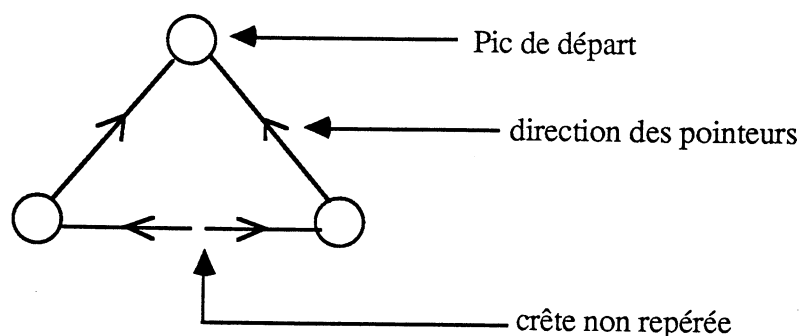
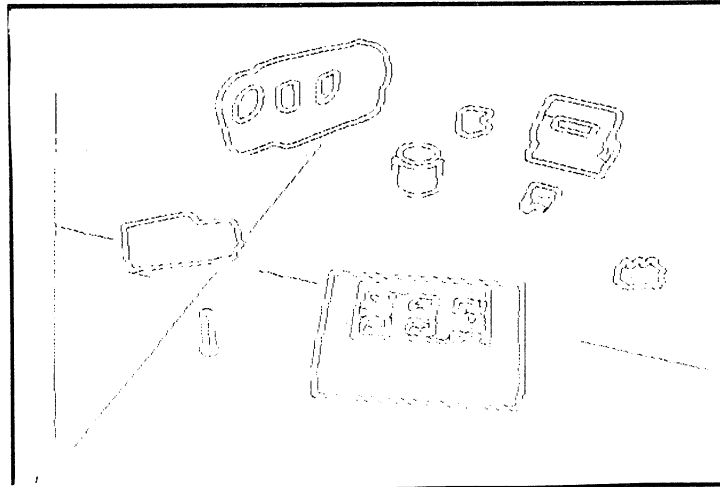


Figure IV.10: Illustration d'un non repérage d'une crête.

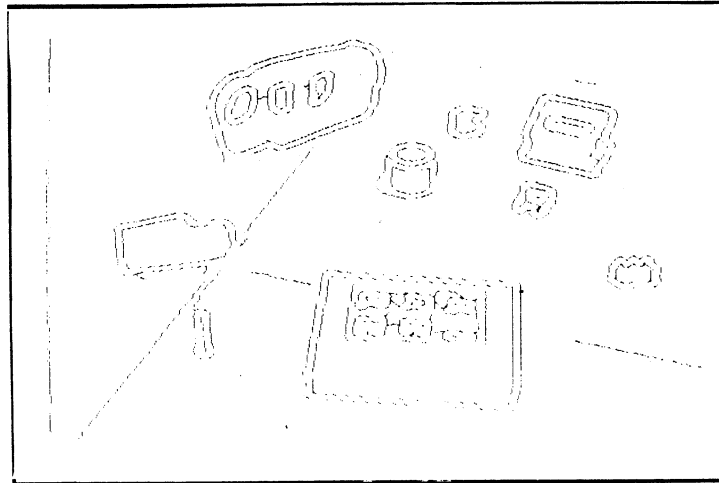
Pour résoudre ce problème, on affecte à chaque pic un numéro d'identification. Durant le déroulement de l'algorithme, lorsque qu'on fait pointer un pixel crête vers le plus grand de ses voisins (en valeur du laplacien), le pixel pointeur prend (sauf si c'est un pixel pic) le numéro du pixel pointé. Ainsi, chaque pixel crête aura le numéro du pic vers lequel il conduit par parcours des pointeurs. Repérer une cassure revient alors à rechercher les pixels voisins n'ayant pas le même numéro d'identification. La crête cassée relie les deux pics dont le numéro d'identification est fourni par les deux pixels voisins. En remontant les pointeurs à partir des deux pixels voisins, on trouve la crête manquante.

IV.3.2.2 Résultats

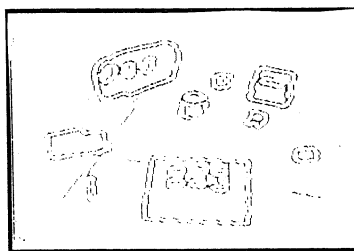
A partir de l'image de pièces diverses, on fournit une représentation sous forme d'une pyramide d'images multi-résolutions. Les crêtes positives et négatives du laplacien des images de niveau de résolution 2 à 10 sont présentées.



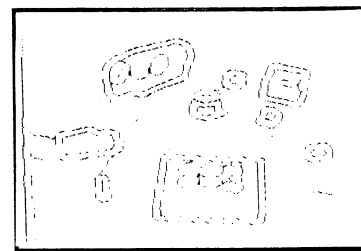
Crêtes au niveau $K = 2$



Crêtes au niveau $K = 3$



Crêtes au niveau $k = 4$



Crêtes au niveau $k = 5$



C_6



C_7



C_8



C_9



C_{10}

IV.4 Les arbres de pics

Nous décrivons ici une autre représentation des informations visuelles que nous avons aussi utilisées. Il s'agit des structures d'indices "arbres de pics" définies par Crowley [Crowley 81].

IV.4.1 La détection

Etant donné qu'il existe une continuité entre les indices de deux passe-bandes d'images successives, chaque pic dans un niveau K de résolution possède au niveau inférieur de plus grande résolution ($K-1$), une région de descendance [Crowley 84], dans laquelle chaque pic est considéré comme un fils. Après de multiples expérimentations, Crowley a choisi $2\sigma + 1$ comme rayon de la région de descendance (où σ est l'écart type du filtre initial). Cette région englobe dans une large majorité de cas les pics fils. Chaque père est alors relié à ses descendants à chaque niveau de résolution.

Nous obtenons par conséquent un arbre de pics représentatif de la forme analysée (ou d'une partie de la forme) [Crowley 84]. Les pics en faibles résolutions donnent un aperçu de la forme globale. Une descente dans l'arbre précise les détails de la forme. La figure IV.11 illustre cet arbre de pics pour la forme trapézoïdale dessinée.

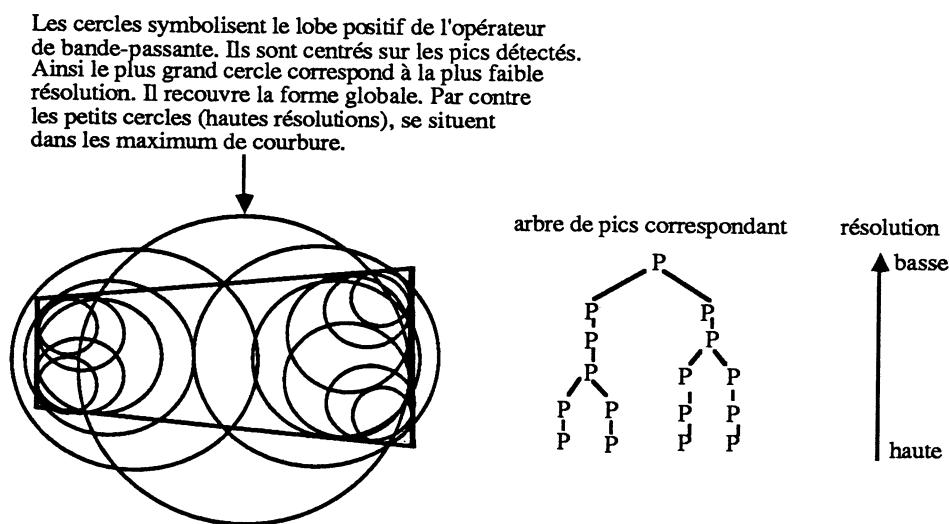


Figure IV.11: A chaque cercle représenté sur la forme trapézoïdale correspond un pic dans l'arbre. Le diamètre du cercle correspond à un niveau de résolution du pic. Le centre du cercle donne la position sur la forme du pic détecté. Dans l'arbre, les pics sont représentés par la lettre P, et les branches qui unissent ces lettres symbolisent les liens entre les niveaux de résolution.

IV.4.2 Limitations

Nous avons vu précédemment que de "faux" indices pics pouvaient être détectés. Par contre, à un "vrai" indice de type pic correspond au plus un indice pic dans l'image suivante. Avec les indices "arbres de pics", il en va tout autrement.

Effectivement, il paraît peu probable d'extraire des arbres de pics à plusieurs niveaux de résolution composés exclusivement de "faux" pics. Par conséquent, de "faux" indices arbres de pics ont peu de chance d'exister. Par contre, le mouvement de la caméra ou des objets dans la scène et les variations de luminance engendrent au sein de l'image, des variations dans les formes et dans leur proximité.

Ainsi entre deux images consécutives on peut observer des regroupements ou des cassures de formes qui interviennent dans les faibles résolutions. La conséquence directe est un regroupement dans l'image courante entre des arbres de pics détectés séparément dans l'image précédente ou une cassure d'un des arbres de l'image précédente en plusieurs arbres de pics.

De même, le mouvement peut engendrer des déformations des coins sur certaines formes. Ainsi, un coin sur une partie d'une forme peut devenir très aigu. Or il peut arriver que pour de tels coins, un pic fils à un certain niveau de résolution ne se situe pas dans la zone d'héritage. D'où une cause possible supplémentaire de cassure de l'arbre de pics. Par conséquent, à un indice "arbre de pic" dans une image peuvent correspondre zéro, un ou plusieurs arbres de pics dans l'image suivante.

IV.4.3 Caractéristiques

La structure étant un regroupement d'indices, toutes les caractéristiques des indices constituent de ce fait des caractéristiques pour la structure. Les liens entre les indices dans la structure lui confère des caractérisations supplémentaires par rapport aux indices de type simple. La robustesse des caractéristiques a été présentée dans la partie "indices pics". Ce paragraphe souligne l'influence des perturbations sur les liens de la structure.

Au sein de la structure, il existe des liens avec les "faux" pics. Ainsi, d'une image à la suivante, les "faux" pics affectent la forme de la structure.

Une autre cause de perturbation, causée par le mouvement des formes dans l'image, a pour effet de déformer les structures. De ce fait, certains niveaux d'embranchement de l'arbre peuvent être modifiés(cf figure IV.12).

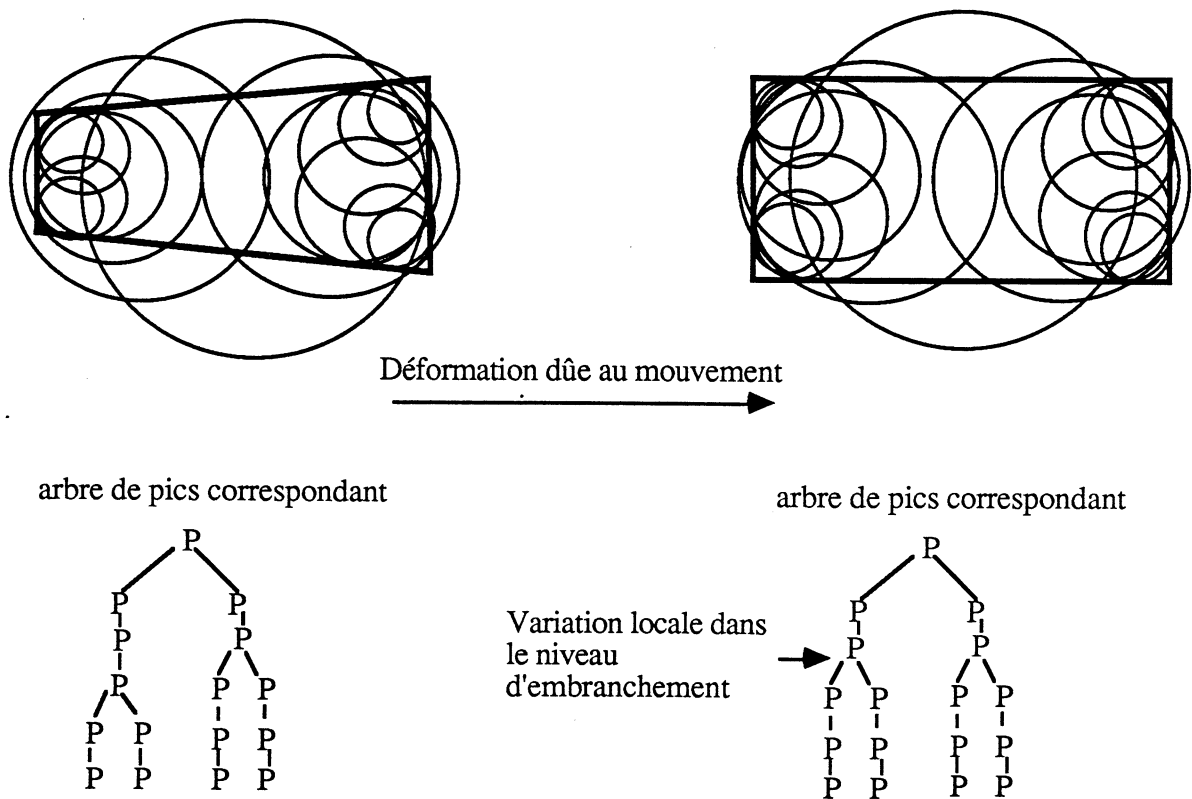
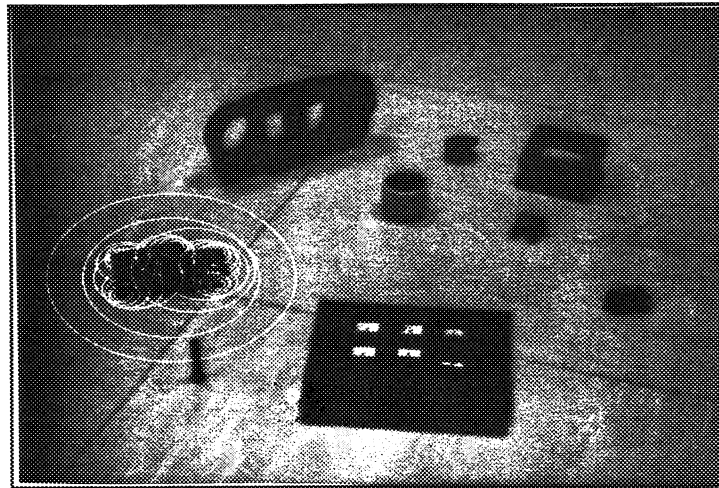


Figure IV.12: Déformations des arbres de pics dues au mouvement

IV.4.4 Résultats

Voici ci-dessous la structure caractérisant la tête du marteau sous forme d'incrustation sur l'image d'origine de cercles de diamètres différents. Suit une représentation explicite de "l'arbre de pics" obtenu (cf. figure IV.13).



Visualisation de la structure caractérisant la tête du marteau

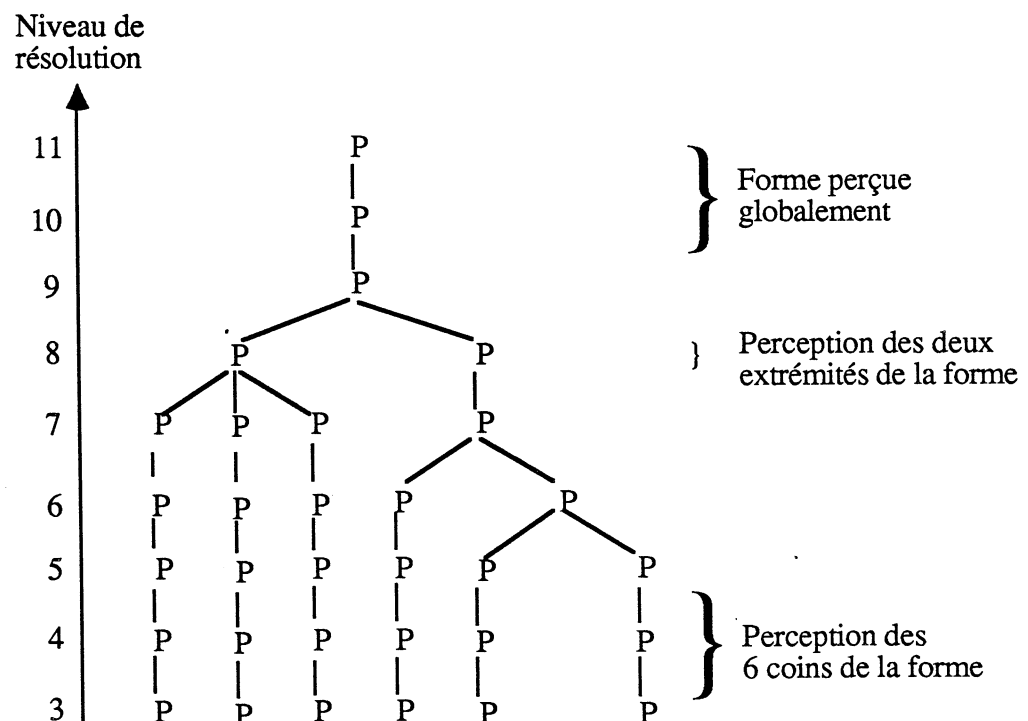


Figure IV.13: Représentation de l'arbre de pics correspondant à la structure visualisée sur l'image précédente.

IV.4.5 Conclusion

Les arbres de pics constituent des indices plus discriminants que les points caractéristiques car plus denses en informations. En effet, la structure d'indices est un moyen adéquat pour intégrer les indices contenus dans les différents niveaux de résolution (les informations locales sur les formes sont situées dans les plus hautes résolutions, les informations plus globales l'étant dans les plus faibles résolutions). De plus, le nombre de structures est faible par rapport au nombre d'indices de type "point caractéristique", car chaque structure correspond à un regroupement de tels indices. Cette approche permet de limiter la combinatoire dans la phase d'appariement. Elle semble donc au premier abord plus adéquate.

IV.5 Relations

Nous avons défini des relations uniquement entre les indices "pics", une détermination de relations entre les indices "arbres de pics" semblant difficile à définir. Pour limiter les ambiguïtés d'appariement lors de la phase de mise en correspondance, nous avons choisi d'utiliser un algorithme de recherche de cliques maximales. Cette démarche nous contraint à définir totalement la compatibilité ou l'incompatibilité entre tous couples d'hypothèses. En se reportant à la définition de la compatibilité introduite au chapitre précédent, ceci implique que le graphe des relations entre les indices soit complet. Afin d'obtenir un tel graphe, nous avons choisi des relations de type spatial. Chaque indice est lié à tous les autres par une des relations suivantes:

- à droite de (D)
- à gauche de (G)
- en haut de (H)
- en bas de (B)
- à droite en haut de (DH)
- à droite en bas de (DB)
- à gauche en haut de (GH)
- à gauche en bas de (GB)
- proche de (P).

Pour chaque indice, on considère un quadrillage à 9 régions afin de définir ces relations (cf. figure IV.14):

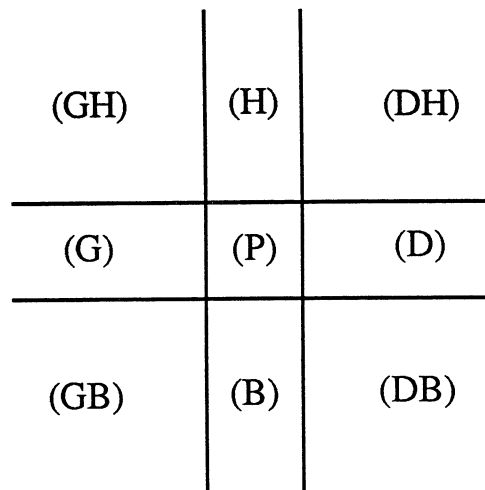
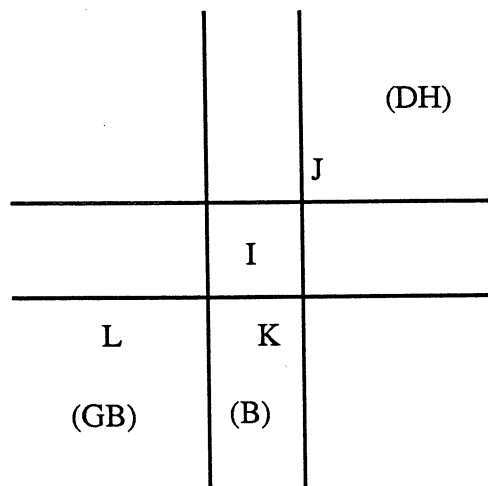


Figure IV.14: quadrillage définissant les relations topologiques

Ce quadrillage est centré sur l'indice courant, ce qui entraîne que tous les autres pics de l'image vont appartenir à l'une des 9 régions de ce quadrillage. L'index de la région du quadrillage à laquelle appartient un pic définit la relation qui le lie à l'indice courant (cf. figure IV.15).

Soient les indices : I,J,K,L



Donc J est a DH de I
K est a B de I
L est a GB de I

Figure IV.15: Exemple de détermination des relations.

Les 5 zones H,B,G,D,P sont petites et étroites. Nous les considérons comme des zones de transition, nécessaires lors de la définition de la compatibilité entre les relations. La

signification du terme de transition et la dimension de ces cinq zones seront explicitées dans le chapitre suivant (lors de la phase d'appariement globale), car c'est à ce moment que se justifie leur importance.

IV.6 Conclusion

La phase de représentation des informations de l'image est une étape primordiale qui détermine la validité, la robustesse, et la rapidité de la phase ultérieure d'appariement. Afin de répondre à l'énoncé de ces critères, nous avons choisi d'utiliser une représentation multi-résolutions de l'image. Dans chaque image de bande passante d'une pyramide obtenue par convolutions successives de filtres de différence de gaussiennes, nous avons extrait des indices s'accordant bien avec un schéma multi-résolutions. Ce sont les "pics" avec des relations topologiques les liant au sein de l'image ou des indices composés "arbres de pics".

Chacun d'eux nous amènera, dans le chapitre suivant, à concevoir la phase d'appariement des indices d'une manière différente. Les résultats obtenus sur des images réelles permettront de dégager l'indice, à savoir le "pic" ou "l'arbre de pics", en vue d'obtenir le meilleur compromis entre la validité, la robustesse et la rapidité.

V APPARIEMENT PAR SIMILITUDE D'ARBRES DE PICS

Ce chapitre présente notre première réalisation sur l'appariement.

V.1 Introduction

Le travail décrit traite de la mise en correspondance de deux images successives d'une scène statique d'intérieur. Les indices images sur lesquels s'effectue l'appariement sont les arbres de pics. L'algorithme de mise en correspondance est basé uniquement sur un critère de similitude. Ce critère étant local, les indices doivent être suffisamment discriminants et la zone de recherche assez réduite pour restreindre les ambiguïtés potentielles. La présentation de cette méthode est découpée en trois parties:

- l'appariement des indices
- l'évaluation des résultats obtenus
- la détermination du domaine d'application.

V.2 Mise en correspondance

Notre approche, qui vise l'appariement des structures de deux images successives, est basée sur le calcul de la dissemblance de ces structures. Celle-ci est perçue par le biais d'une comparaison des caractéristiques des arbres de pics. Une valeur nulle de la dissemblance laisse supposer un appariement plausible, une valeur élevée le rejette. L'algorithme que nous proposons est basé sur les deux étapes suivantes:

- La détermination de la zone de recherche.
- L'appariement par similitude.

V.2.1 Zone de recherche

Le mouvement de la caméra étant connu, l'information la plus pertinente permettant de restreindre l'espace de recherche réside dans l'exploitation des épipolaires. Le calcul de la zone de recherche et des indices y appartenant, peut se décomposer selon le schéma suivant:

- sélection d'un ou de plusieurs pixels appartenant à l'indice considéré,
- calcul pour ce (ou ces) pixel(s) de la (ou des) épipolaire(s) dans la seconde image,
- définition de l'appartenance à la zone de recherche.

Les techniques d'appariement présentées dans la littérature font état principalement de l'utilisation de segments de droites comme indices images [Saint Vincent 86] [Ayache 85] [Heraud 87]. La première étape pour déterminer la zone de recherche consiste donc à sélectionner le ou les pixels à partir desquels sont calculées les épipolaires. Les deux pixels extrémités des segments sont couramment utilisés. Dans la seconde étape on effectue le calcul des deux épipolaires correspondant aux points extrémités. Les segments ayant alors une intersection avec la région située entre les deux épipolaires sont considérés comme appartenant à la zone de recherche. Une autre approche, similaire dans son principe, utilise le point milieu des segments de droite.

Notre approche repose, quant à elle, sur l'association de structures "arbres de pics". Les pixels sélectionnés correspondent aux pics de la structure. Ainsi, en traçant les lignes épipolaires issues de tous les pics d'une structure de la première image, on définit un faisceau d'épipolaires multi-niveaux dans la seconde image. La structure associée aura, en théorie, chacun de ses pics lié à l'une des lignes du faisceau. Cependant, à cause des incertitudes sur la position des pics et des épipolaires dûes au processus de calibration des caméras, les pics de la structure associée ne se situent pas obligatoirement sur les épipolaires, mais dans leurs voisinages. A cause des déformations qu'engendre le mouvement, des occlusions, des variations de luminance, du bruit sans oublier les erreurs dûes au processus d'extraction des indices, la structure associée peut de plus ne pas avoir tous ses pics dans le voisinage des épipolaires. Ainsi, dans nos applications, une structure appartient à la zone de recherche si au moins la moitié de ses pics sont situés au voisinage des épipolaires du faisceau.

V.2.2 Appariement par similarité

L'appariement peut être décomposé en trois phases:

- La recherche des caractéristiques permettant de discriminer les structures.
- Le calcul d'une fonction de coût pour quantifier la similitude entre deux structures.
- La recherche des appariements.

V.2.2.1 Caractéristiques discriminantes

Il convient, pour restreindre les ambiguïtés d'appariement, de dégager les caractéristiques des structures apparaissant et les plus robustes tout en minimisant les pertes d'informations. La connaissance du type de scène analysée et des contraintes sur notre système permettent de trouver celles qui conviennent à nos réalisations.

Notre démarche s'applique à des scènes d'intérieur. Celles-ci sont, en général, composées principalement d'objets polyédriques. On peut donc admettre que la plupart des formes obtenues par projection sont des polygones.

Dans le cas de la stéréovision, tous les objets de la scène sont fixes les uns par rapport aux autres. La cause des occlusions est donc la différence de point de vue des deux caméras.

Nous contraignons encore notre système en supposant un faible déplacement entre deux prises de vue, ce qui implique, dans la majorité des cas, des occlusions faibles, voire inexistantes. Cette hypothèse limite aussi les déformations des formes présentes dans les images.

Moyennant ces remarques, nous en déduisons que les maxima de courbure dans l'image sont, en général, des points anguleux. Ainsi, chaque pic détecté aux niveaux de résolution les plus élevés correspondent aux points anguleux dans l'image. Les occultations et les déformations étant faibles d'une image à l'autre, on conserve la plupart des points anguleux entre les deux images, ainsi que les points représentant la présence d'une forme globale. Ainsi, le nombre de pics à chaque niveau de la structure nous semble une caractéristique robuste et discriminante moyennant les contraintes de départ que nous imposons à notre système.

V.2.2.2 Fonction de dissemblance

Le critère permettant de quantifier l'appariement entre deux structures sera défini au moyen d'une mesure de leurs dissemblance (elle doit être nulle lorsque les deux structures sont identiques et d'autant plus élevée que les deux structures sont différentes). De plus, pour une variation donnée du nombre de pics, la dissemblance est d'autant plus importante que le nombre de pics est faible. Donc nous décidons de pénaliser davantage une variation en nombre de pics lorsque ces derniers sont peu nombreux. Ces constatations nous conduisent à utiliser une fonction de dissemblance non invariante en translation.

Cette fonction, similaire à celle utilisée dans les travaux relatifs au suivi de cellules sanguines dans une séquence d'images [Levine 83] , peut s'écrire de la manière suivante :

$$\frac{1}{m} \cdot \left(\sum_{k=1}^m \frac{|P_k(i) - P_k(j)|}{\text{Max}(P_k(i), P_k(j))} \cdot W_k \right)$$

Dans laquelle m représente le nombre de niveaux de la structure,

$P_k(i)$ le nombre de pics de la structure i au niveau de résolution k ,

et W_k un poids affecté à chacun des niveaux de résolution.

La présence du facteur W_k est ainsi justifiée:

La présence du bruit aux niveaux de résolutions les plus élevés peut modifier le nombre de pics entre deux structures identiques. Cette variation ne doit pas pénaliser dans de grandes proportions la fonction de coût. Par conséquent, une variation de même ordre apparaissant dans les niveaux de basse résolution est davantage significative.

Donc le coefficient W_k est un facteur pondérateur qui diminue la réponse de la fonction si la différence du nombre de pics entre deux structures intervient aux niveaux de résolution les plus élevés.

V.2.2.3 Génération des appariements

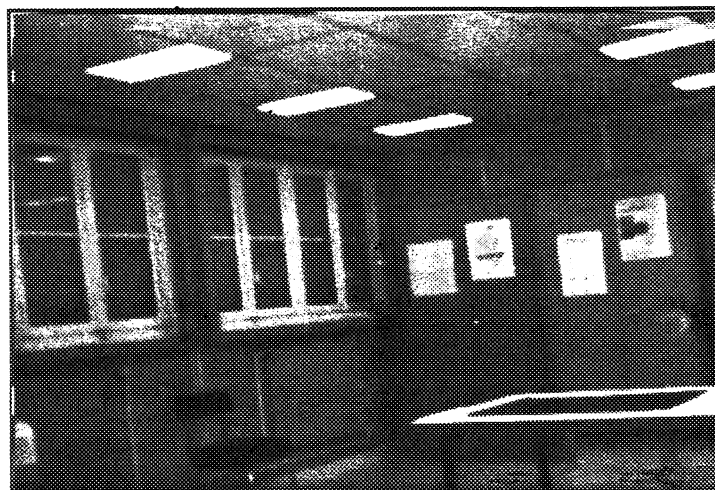
Pour une structure S de la première image, on définit une zone de recherche dans la seconde. Pour chaque structure appartenant à cette zone, on calcule la valeur de dissemblance. Si cette

valeur est représentative (très faible valeur), l'appariement est plausible. Une hypothèse d'appariement est alors créée. Ainsi, pour chaque structure dans la première image, on lui associe un ensemble de structures dans la seconde qui peuvent lui être appariées.

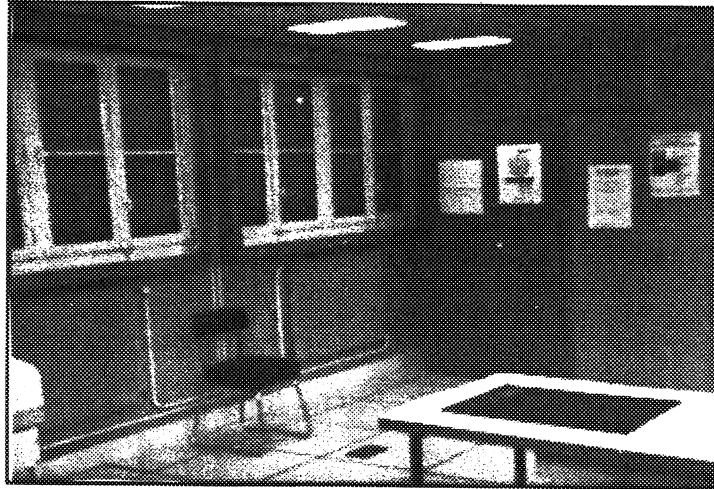
Il convient maintenant d'extraire l'ensemble des appariements "valides". Pour chaque structure de la première image, on retient parmi toutes celles pouvant lui être associés, les deux qui lui ressemblent le plus. Les valeurs de dissemblance associées à chacun de ces deux appariements potentiels sont alors comparées. Si l'écart est suffisamment grand, l'hypothèse de plus faible coût (qui est dans ce cas un minimum local), est retenue. Dans le cas contraire, il y a ambiguïté d'appariement.

V.3 Applications et évaluations des résultats

L'application décrite concerne la mise en correspondance d'images stéréoscopique acquises sur des scènes de bureaux. Les photos V.P.1 et V.P.2 représentent les deux images que l'on désire appairer.

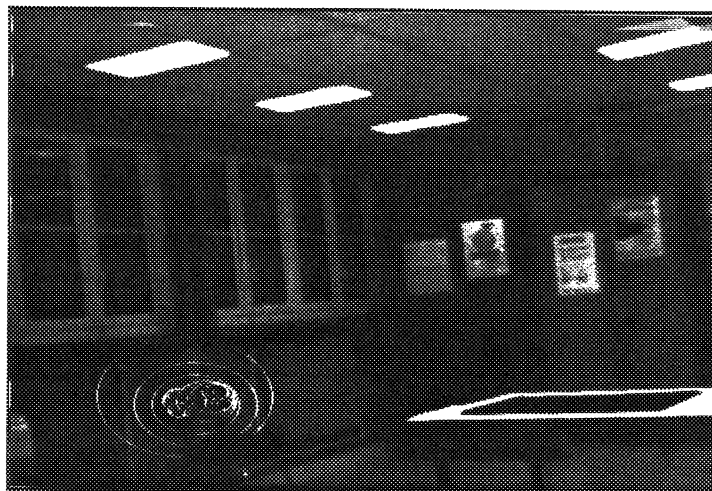


V.P.1



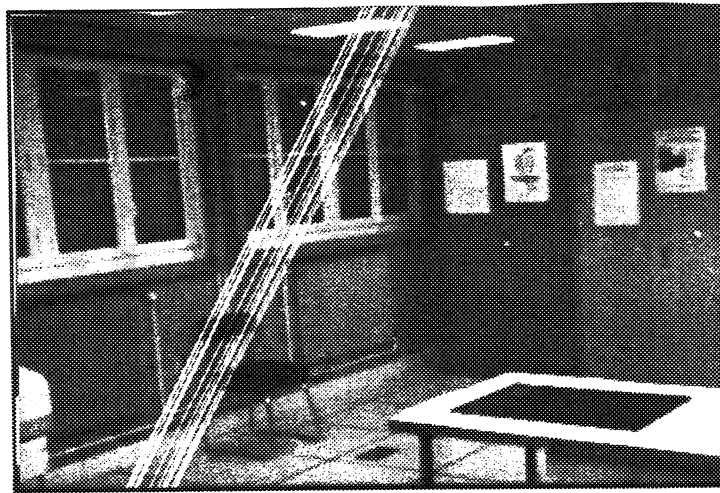
V.P.2

Afin de présenter des résultats compréhensibles, nous limitons la visualisation du processus de mise en correspondance à l'appariement d'une seule structure choisie arbitrairement. La structure choisie dans la première image, à savoir un dossier de chaise, est représentée sur la photo V.P.3.



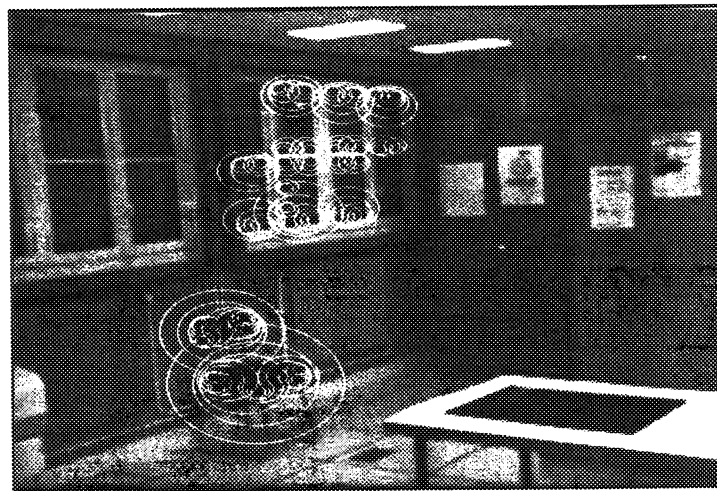
V.P.3

La photo V.P.4 visualise les épipolaires dans la seconde image correspondant à tous les pics de la structure représentant le dossier de chaise.



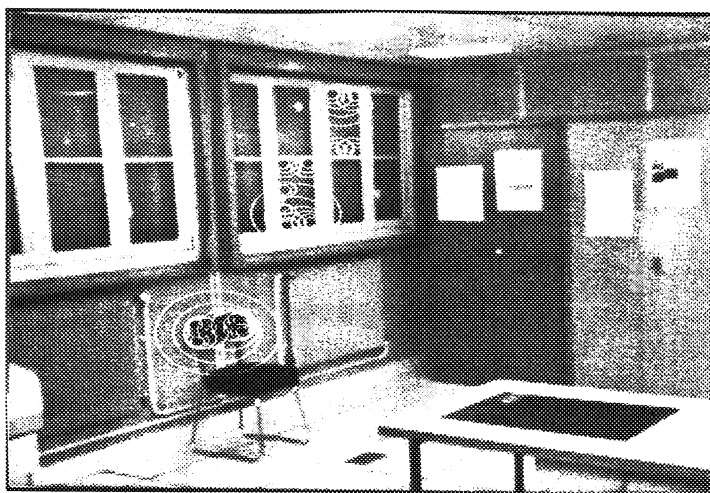
V.P.4

Toutes les structures qui possèdent au moins un pic placé sur ce faisceau d'épipolaires sont représentées sur la photo V.P.5.



V.P.5

En sélectionnant uniquement les structures qui vérifient que le nombre de pics liés au faisceau sur le nombre total de pics est élevé, on réduit considérablement le nombre de candidats à l'appariement. Le rapport a été fixé arbitrairement à 0.5, les expérimentations ayant montré que son ajustement n'était pas critique. La photo V.P.6 représente les trois structures restantes dans la zone de recherche.



V.P.6

Le calcul de la dissemblance entre la structure de la première image et chacune des trois structures candidates à l'appariement est alors exécuté. Les résultats sont fournis sur le tableau V.T.1. On vérifie que la dissemblance est minimale entre les structures correspondant respectivement au dossier de chaise dans la première et la seconde image. La valeur est de plus très faible (0.041) et se présente comme un minimum local (la seconde plus petite valeur de dissemblance étant égale à 0.35). L'appariement de ces deux structures peut alors être réalisé sans ambiguïté.

Structure choisie dans la première image	Structures de la seconde image situées dans la zone de recherche	Dissemblance avec la structure de la première image
Dossier de chaise	Dossier de chaise	0.041
	Fenêtre 1	0.35
	Fenêtre 2	0.39

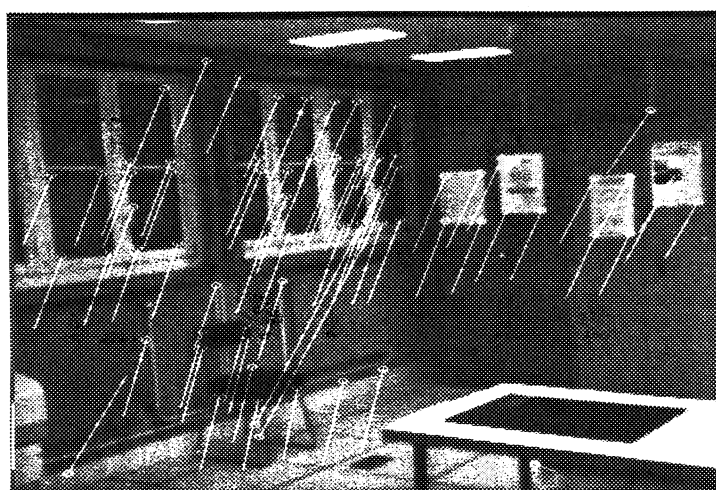
Tableau V.T.1

Le processus décrit ci-dessus est poursuivi pour toutes les structures de la première image. Chaque appariement obtenu étant formé de deux structures légèrement différentes, nous recherchons la sous-structure commune afin de ne conserver que les éléments communs. Cependant, il faut remarquer que seuls les pics en haute résolution sont localisés avec précision. De ce fait, lors d'une phase ultérieure (telle que la reconstruction 3D), une information fiable pourra être acquise en considérant les pics de haute résolution. Pour cette raison, les photos V.P.7 et V.P.8 ne représentent que les appariements présents au niveau 3 dans lesquelles:

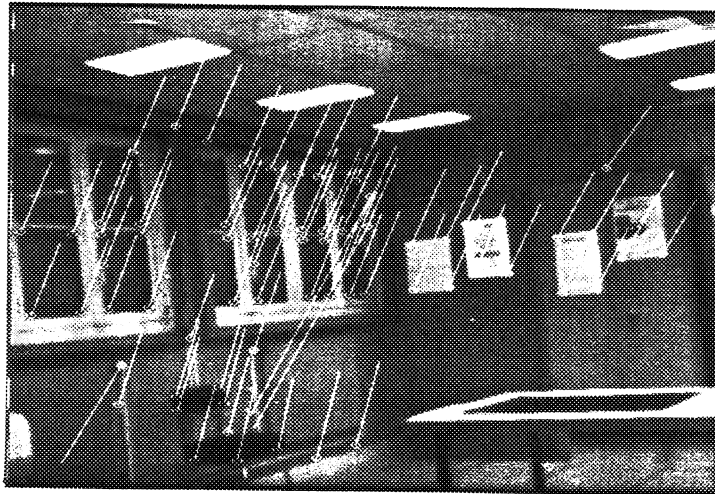
Les cercles représentent la position des pics appariés dans l'image.

Les segments de droite fournissent le déplacement des pics associés dans l'image suivante.

Les résultats en nombre de pics appariés et de faux appariements sont présentés dans le tableau V.T.2.



V.P.7



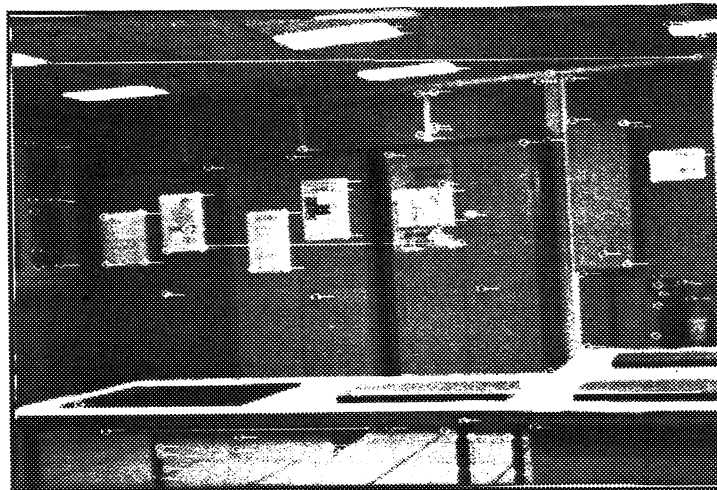
V.P.8

Nombre total de pics dans la première image	330
Nombre total de "vrais" pics dans la première image	164
Nombre total de pics dans la seconde image	363
Nombre total de "vrais" pics dans la seconde image	171
Nombre total d'appariements	72
Nombre d'appariements justes	68
Nombre de "vrais" pics parmi les appariements justes	62

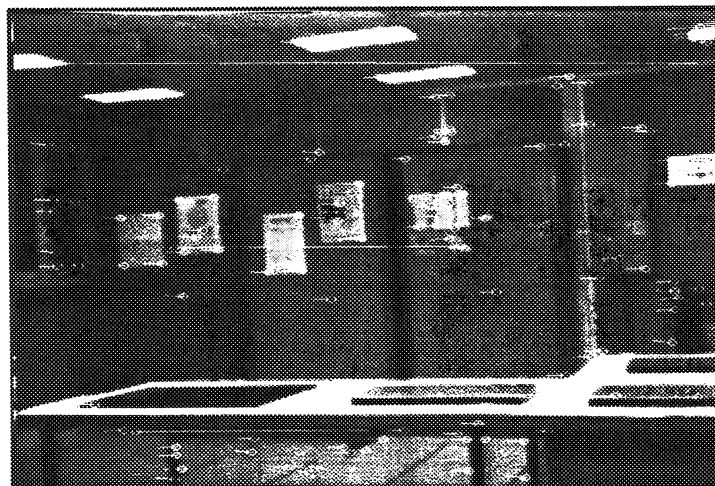
Tableau V.T.2

Le faible pourcentage de "vrais" pics appariés (41%) est dû à un déplacement trop important de la caméra entre les deux images, entraînant des déformations trop grandes des structures. Ainsi, la table et les lampes n'ont pu être mises en correspondance. Cependant, le même algorithme appliqué à une paire d'images acquises à des emplacements spatialement proches apparie alors correctement la table et les lampes.

Les photos V.P.9 et V.P.10 illustrent les résultats obtenus. Les résultats numériques sont ensuite fournis dans le tableau V.T.3. On peut constater que le pourcentage d'appariements corrects est bien meilleur (60%) mais reste néanmoins encore faible.



V.P.9



V.P.10

Nombre total de pics dans la première image	331
Nombre total de "vrais" pics dans la première image	168
Nombre total de pics dans la seconde image	332
Nombre total de "vrais" pics dans la seconde image	180
Nombre total d'appariements	119
Nombre d'appariements justes	112
Nombre de "vrais" pics parmi les appariements justes	101

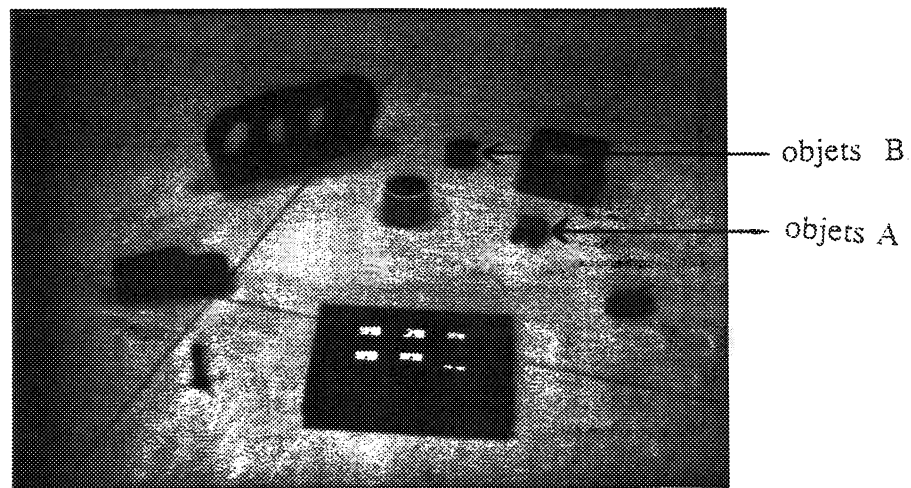
Tableau V.T.3

V.4 Détermination du domaine d'application

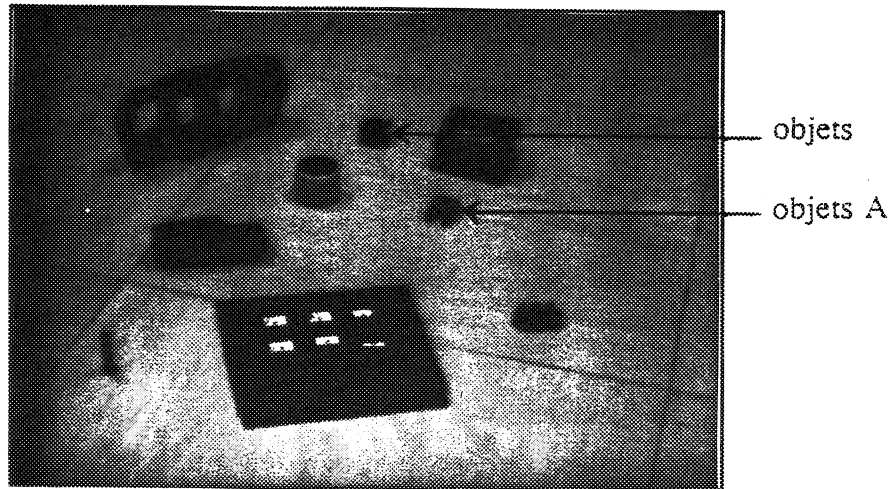
Les expérimentations montrent que les arbres de pics par le biais de la caractéristique "nombre de pics à chaque niveau" permettent de bien discriminer les formes dans les images, à condition toutefois que le mouvement entre l'acquisition de deux images ne soit pas "trop important". Néanmoins, le critère utilisé, qui est uniquement local, ne permet pas de lever les ambiguïtés d'appariement. Ces dernières apparaissent dans le cas où il existe des formes semblables dans la zone de recherche. De plus, nous avons précisé dans le chapitre précédent que les arbres de pics peuvent se casser ou fusionner d'une image à la suivante. Dans ces conditions, les appariements deviennent impossibles.

Les remarques précédentes permettent de constater que "les arbres de pics" ne sont pas bien adaptés à une mise en correspondance d'images sur une scène complexe. Par contre, cette représentation semble convenir à la modélisation de formes 2D en vue d'une correspondance modèle-forme dans l'image (la reconnaissance de caractères par exemple). Néanmoins, aucune forme perçue ne doit être occultée pour éviter toute amputation de la structure correspondante. A cause de la non-invariance des structures "arbres de pics" par homothétie, la taille des formes doit être connue lors de la modélisation. Cela permet aussi de déterminer, en fonction des formes à percevoir, le niveau de plus faible résolution en-dessous duquel il ne faut pas descendre sous peine de fusion de deux "arbres de pics" correspondant à deux formes différentes. Ainsi en respectant ces conditions, à chaque structure est associée une forme perçue.

Ce type d'application est illustré par une mise en correspondance entre deux images acquises avec une caméra mobile. Les objets perçus (outils, pièces mécaniques, etc ...) sont également mobiles (cf. photos V.P.11 et V.P.12).



V.P.11



V.P.12

Afin de respecter les hypothèses nécessaires à l'utilisation de l'algorithme:

Le mouvement de la caméra s'effectue à la même profondeur par rapport au plan de travail. Il n'y a donc pas de variation significative en taille des formes perçues.

Le mouvement des objets peut être inconnu à condition qu'il n'engendre pas de variations significatives des formes. Cela revient à percevoir des formes 2D dans des positions inconnues en translation et en rotation.

Le mouvement des objets peut être important et différent d'un objet à l'autre. D'où absence de zone de recherche privilégiée et surtout de contrainte de relations entre les objets dans la scène. Ceci empêche l'utilisation de relations contextuelles entre les formes en vue de limiter les ambiguïtés d'appariements.

Les occlusions sont interdites.

Les structures acquises dans la première image servent de modèle à retrouver dans l'image suivante.

Les formes présentes dans les images ont été correctement appariées, à l'exception du corps du cylindre et des objets A et B. Le tableau V.T.4 indique pour chaque forme de la première

image, les deux formes de la seconde qui lui ressemblent le plus (c'est à dire les deux valeurs de dissemblance les plus réduites). L'appariement est correct quand la valeur obtenue est faible et présente un minimum local dans l'ensemble des hypothèses. On vérifie sur le tableau que l'écart entre les deux valeurs est significatif (compris entre 0.1 et 0.26) excepté pour le corps du cylindre et les pièces A et B. En ce qui concerne les formes A et B, elles ne peuvent être appariées sans ambiguïté car les deux objets dont elles sont la projection sont identiques et orientés de la même manière. On peut vérifier que malgré une faible valeur de dissemblance, la différence entre les deux plus faibles valeurs est insuffisante pour valider l'appariement (0.02 d'écart). Le corps du cylindre, quant à lui, n'a pu être apparié en raison d'un très faible contraste par rapport au fond, ce qui engendre la perte de "vrais" pics d'une image à l'autre. De plus, sa base arrondie possède de "faux" pics qui n'ont pu être éliminés. Il en résulte des dissemblances trop importantes entre les deux structures correspondant au corps du cylindre dans les deux images respectives.

IMAGE 1

IMAGE 2

formes	Première forme plus semblable	Dissemblance	Seconde forme plus semblable	Dissemblance
Vis	Vis	0	Dessus cylindre	0.187
Disquette	Disquette	0.027	Gros objet	0.152
Sucre	Sucre	0.025	Dessus cylindre	0.289
Marteau	Marteau	0.022	Perforatrice	0.148
Pièce A	Pièce A	0.028	Pièce B	0.048
Pièce B	Pièce A	0.048	Pièce B	0.028
Corps cylindre	Corps cylindre	0.2	Vis	0.357
Dessus cylindre	Dessus cylindre	0	Vis	0.187
Perforatrice	Perforatrice	0.037	Marteau	0.137
Gros objet	Gros objet	0.040	Disquette	0.152

Tableau V.T.4

V.5 Conclusion

Le calcul de la transformée de DOLP fournit une pyramide multi-résolutions d'images passe-bandes. De chaque image constituant cette pyramide, on extrait les extrema locaux. Les pics sont alors liés à travers les résolutions afin de créer des arbres de pics. La mise en correspondance est basée sur ce type d'indices. Une fonction de dissemblance, qui quantifie analytiquement le degré d'appariement de deux structures, est alors créée. Le calcul de la dissemblance exploite le nombre de pics à un niveau donné et ce, pour tous les niveaux de résolution. Si la réponse en dissemblance est faible, on crée une hypothèse d'appariement qui peut être validée à condition que ce minimum corresponde à un extrema local. Dans le cas où le mouvement de la caméra est connu, on utilise des faisceaux d'épipolaires pour limiter l'espace de recherche des appariements. Ce type de technique est adapté à l'appariement entre des modèles bidimensionnels et des formes observées.

Dans le cas d'appariement d'images acquises sur une scène réelle, on envisage de travailler directement sur les pics. Ces derniers se prêtent aisément à la définition de relations topologiques. De plus, ils se montrent robustes face aux déformations dues au mouvement de la caméra et à la position relative des formes entre elles au sein des images. Le chapitre suivant présente un nouvel algorithme de mise en correspondance qui repose sur l'utilisation de ces pics.

VI APPARIEMENT DE PICS AVEC HERITAGE ET CLIQUE MAXIMALE

Ce chapitre présente notre seconde réalisation sur l'appariement.

VI.1 Introduction

Nous développons ici une technique d'appariement basée sur l'utilisation des pic aux divers niveaux de résolution [Stelmaszyk 88c] à l'opposé des arbres de pics utilisés dans le chapitre précédent. Bien que l'approche multi-résolution soit conservée, les liens de descendance au sein même de l'indice ne sont pas exploités. Ceci permet de limiter les influences de variation des positions relatives entre les formes. De plus, il apparaît que ces indices pics se prêtent aisément à la définition d'un critère d'appariement global.

VI.2 Principes de la méthode

VI.2.1 Introduction

Lors de la phase d'appariement d'indices, les ambiguïtés font ressortir l'intérêt d'une approche de mise en correspondance globale. Cependant, l'explosion de la combinatoire inhérente à ces méthodes réduit considérablement leurs attraits dans le cas où le nombre d'indices est important. Afin de limiter la combinatoire, il convient de restreindre l'espace de travail en s'appuyant par exemple sur un sous-ensemble de l'image. [Skordas 88] propose de partitionner son image en fenêtres de taille 64x64. Le découpage étant arbitraire, l'inconvénient majeur de cette approche est le non-centrage de cette fenêtre sur une zone d'intérêt. Aussi, il nous a semblé intéressant de considérer une approche multi-résolutions dans laquelle le partitionnement est fourni par les informations acquises dans les niveaux de résolution inférieurs.

L'idée intuitive de ce partitionnement peut être illustrée par un exemple :

Supposons qu'il faille trouver l'appariement d'un mot écrit sur une affiche, elle-même dans un bureau contenant d'autres affiches. Le partitionnement consiste alors à apparier l'affiche par un critère global prenant en compte les relations contextuelles dans toute l'image. Ensuite, le mot est apparié par un critère global, mais basé cette fois-ci uniquement sur les relations contextuelles au sein de l'affiche.

Notre algorithme est basé sur cette idée intuitive. La mise en correspondance des indices entre deux images est d'abord réalisée dans la plus faible résolution. On apparie des formes perçues globalement. Puis ces appariements permettent de répartir dans la résolution supérieure les indices en zones associées. L'appariement est alors effectué à ce niveau de résolution uniquement entre les indices des zones associées, d'où une limitation des coûts. Le processus s'arrête lorsque les indices sont appariés au niveau de résolution souhaité. Il faut souligner que pour des raisons explicitées dans la suite de ce chapitre, le déplacement de la caméra entre deux acquisitions d'image est faible.

Notre méthode s'apparente aux techniques de "coarse to fine".

VI.2.2 Coarse to fine

Le principe d'une approche "coarse to fine" consiste à générer des appariements dans les niveaux de faible résolution et à les propager dans la pyramide d'images multi-résolutions.

Avant de décrire notre algorithme, il nous paraît opportun de présenter en détail un algorithme de "coarse to fine" présent dans la littérature afin de souligner ses propriétés intéressantes, et de dégager les points faibles pour lesquels on proposera une solution. Notre choix s'est porté sur les travaux de [Moravec 77] qui nous semblent les plus représentatifs de la méthode.

Moravec traite une séquence d'images, en vue de suivre des "points d'intérêts". Chaque image I de la séquence est découpée en carré $n \times n$. La moyenne en intensité à l'intérieur de chaque carré est alors calculée en vue de constituer une nouvelle image. Il utilise cette démarche successivement avec $n = 2$, $n = 4$, $n = 8$, $n = 16$. Chaque image ainsi obtenue a une plus faible résolution que l'image qui la précède, ce qui donne une représentation multi-résolutions de l'image initiale I . Sur la première image de la séquence, il extrait des "points d'intérêts". Ceux-ci appartiennent à l'un des carrés du découpage de l'image. Le pixel correspondant à la

moyenne sur ce carré représente le point d'intérêt en plus faible résolution. Ainsi, pour chacun d'entre eux, il recherche le pixel correspondant dans l'image de plus faible résolution et définit une fenêtre autour de ce pixel. Il recherche alors par corrélation dans l'image suivante de la séquence, et ce à la même résolution, l'emplacement de la fenêtre correspondante. Le maximum en corrélation lui fournit cet emplacement. Cet appariement étant effectué à la plus faible résolution, il est imprécis. Il affine donc la correspondance précédente en recommençant le même processus au niveau supérieur en résolution. La nouvelle fenêtre choisie est centrée sur le pixel correspondant au point d'intérêt. L'appariement précédent définit, dans la seconde image de la séquence au niveau supérieur en résolution, une zone de descendance. La recherche de la fenêtre associée est alors effectuée exclusivement dans la zone de descendance (qui correspond à la zone de recherche) ce qui limite le coût en corrélation. Le maximum de corrélation dans cette zone de recherche met en évidence un appariement qui fournit une nouvelle zone de descendance. Ce processus est réitéré pour tous les niveaux de la pyramide d'images.

Les avantages qu'apporte cette approche multi-résolutions par rapport aux techniques classiques de corrélation sont évidents :

Le coût en temps de calcul des appariements est considérablement réduit grâce à la limitation de la zone de recherche. Le coût serait énorme s'il devait effectuer le calcul de corrélation sur toute l'image initiale.

Le bruit est négligeable dans les faibles résolutions. L'erreur de corrélation est ainsi limitée en ciblant la région d'appariement dans les faibles résolutions.

Par contre, les inconvénients de cette approche résident dans les erreurs d'appariement dues à l'utilisation d'une technique locale de corrélation, et à la propagation systématique de ces erreurs aux plus hautes résolutions. Ces constatations nous amènent à introduire un critère global d'appariement et à vérifier les héritages d'appariements précédents.

VI.2.3 Présentation de la méthode

Le principe de la méthode développée peut être représenté par la figure VI.1 :

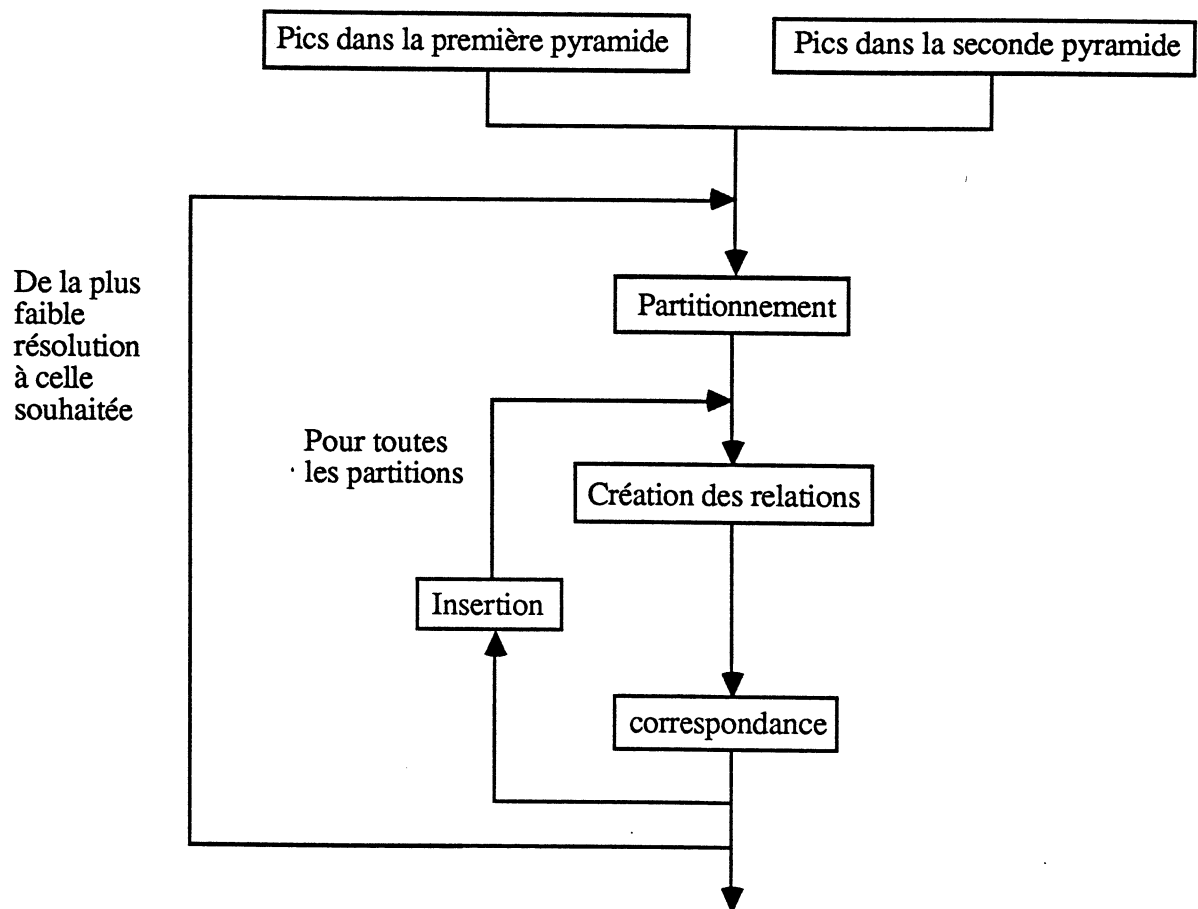


Figure VI.1 : Schéma de la méthode.

Pour chaque image, on constitue une pyramide de Laplacien dans laquelle on extrait les pics (le principe de cette technique a été présenté dans le chapitre IV).

Dans la plus faible résolution, les pics sont appariés en prenant en compte les contraintes contextuelles sur l'image entière. Ils représentent, dans le cas général, des formes perçues globalement. Or ils possèdent une région de descendance dans le niveau de résolution supérieur. Ainsi, un appariement entre deux pics induit une association de leur région de descendance (cf. figure VI.2).

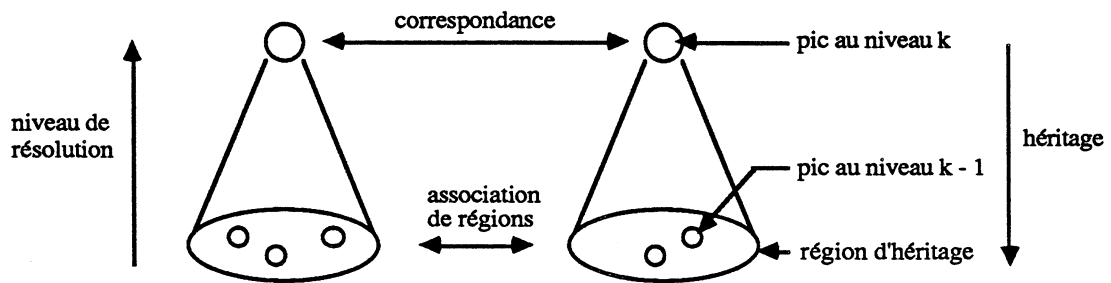


Figure VI.2 : Association de régions par héritage d'appariement de pics.

Les régions de descendance des pics appariés constituent notre **partitionnement** de l'image. Ainsi, chaque pic dont le père a été apparié précédemment appartient à l'une de ces régions. Le complémentaire dans l'image de ces régions constitue aussi une région. Elle englobe les pics n'ayant pas de père, les pics ayant un père non apparié et les pics non contenus dans la zone de descendance. Cette région est associée à la région complémentaire de l'autre image. Tous les pics dans la nouvelle résolution sont donc répartis dans ces diverses régions (cf. figure VI.3).

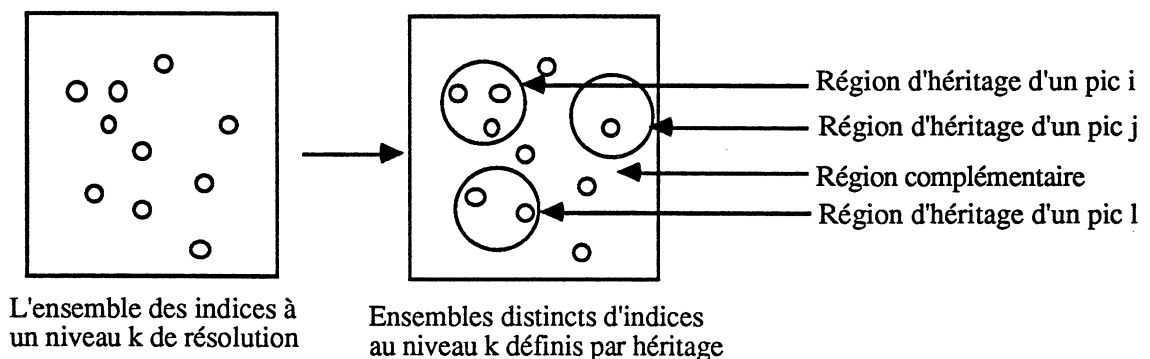


Figure VI.3 : Répartition de l'ensemble des pics dans les régions d'héritage et la région complémentaire.

Par conséquent, au niveau de résolution supérieur, on cherche l'**appariement** d'un pic uniquement parmi les pics de la région associée. Cette mise en correspondance est réalisée de manière globale et ce, uniquement au sein de la région. Ainsi, ne prenant pas en compte le contexte de toute l'image, on réduit fortement les coûts. Il est à noter que le problème de mauvais héritage peut entraîner des non-appariements de pics alors que le correspondant existe effectivement. Pour prévenir ce problème, les pics non appariés lors de la mise en correspondance des pics appartenant aux régions de descendance, sont obligatoirement remis en jeu par **insertion** dans la région complémentaire. L'appariement des pics appartenant à cet ensemble est donc effectué en dernier.

L'information contextuelle étant mise en compte uniquement au sein de la région, les **relations** sont calculées uniquement entre les indices d'une même région, d'où un nouveau gain en temps de calcul (cf. figure VI.4).

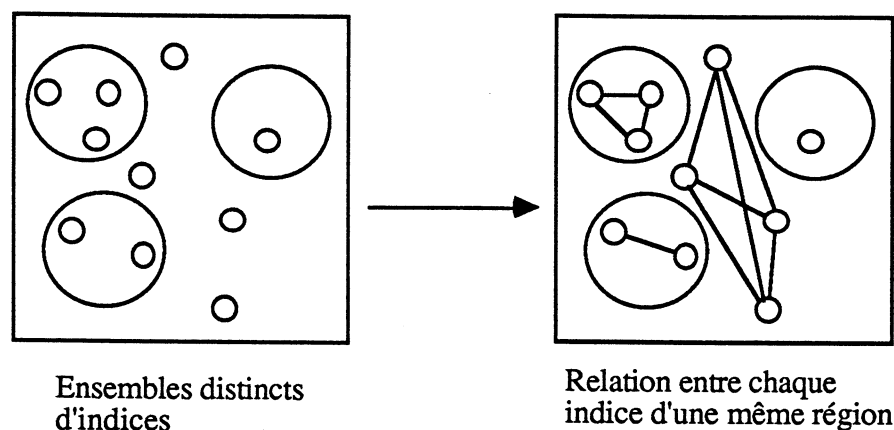


Figure VI.4 : Création d'un graphe complet pour chaque ensemble de pics.

L'appariement étant réalisé à ce niveau, on réitère le processus jusqu'à l'appariement des pics au niveau de résolution souhaité.

Dans l'organigramme présenté précédemment, il reste à détailler le module de mise en correspondance entre les pics de deux régions associées. Ceci fait l'objet de la section suivante.

VI.3 Mise en correspondance

L'appariement est effectué à chaque niveau successif de résolution à partir des plus faibles. La mise en correspondance des indices entre deux images est répartie en appariements d'indices entre régions associées. La présentation de l'appariement est découpée suivant le schéma général introduit dans le chapitre III. Une première partie présente la détermination de la zone de recherche, une seconde traite de la création d'hypothèses d'appariement, la troisième met l'accent sur la phase d'appariement global.

VI.3.1 Zone de recherche

On combine la contrainte épipolaire et la disparité héritée du niveau de résolution inférieur pour restreindre la zone de recherche des correspondants possibles.

VL3.1.1 Contrainte épipolaire

On travaille dans des scènes de type MCSO dans lesquelles le mouvement de la caméra est connu; de ce fait on peut utiliser la contrainte épipolaire (cf. chapitre III sur les zones de recherche). Ainsi pour chaque indice d'une image, nous sommes en mesure de déterminer une zone épipolaire dans l'autre image où se situe, s'il existe, l'indice associé.

VL3.1.2 Zone définie par disparité

La différence de position au sein des images entre deux pics appariés (p_1, p_2) au niveau k de résolution fournit une disparité. De ce fait, la zone de recherche pour chaque héritier du pic p_1 (p_{fils1}) est définie par un cercle dans la seconde image de centre égal à la somme entre la valeur des coordonnées de p_{fils1} et la disparité entre (p_1, p_2), et de rayon égal à la variation maximale possible en disparité. Cette variation prend en compte l'erreur sur la disparité des parents et la variation maximale de disparité entre deux niveaux de résolution.

Erreur sur la disparité

Du fait de la digitalisation, de l'influence du bruit, du niveau de résolution et de la proximité d'un autre pic, la position d'un indice dans l'image est imprécise. L'incertitude associée à la position d'un pic est de 1 pixel [Crowley 81].

Variation de disparité

Le mouvement de la caméra entraîne par projection perspective une déformation des formes visualisées, ce qui se traduit pour les objets polyédriques par une variation en angle des coins. Par conséquent, le pic inscrit dans le coin varie en disparité d'un niveau au suivant (cf.figure VI.5).

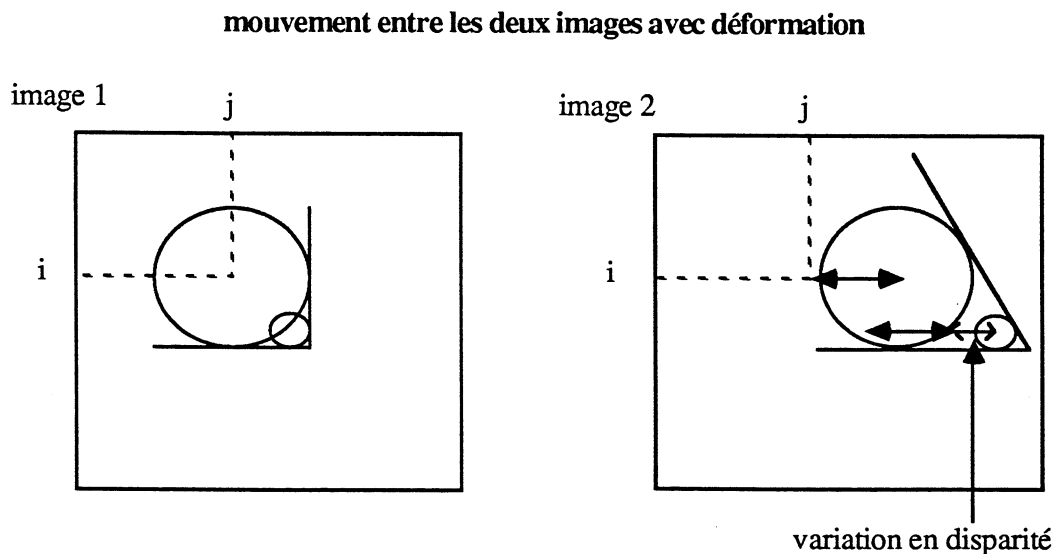
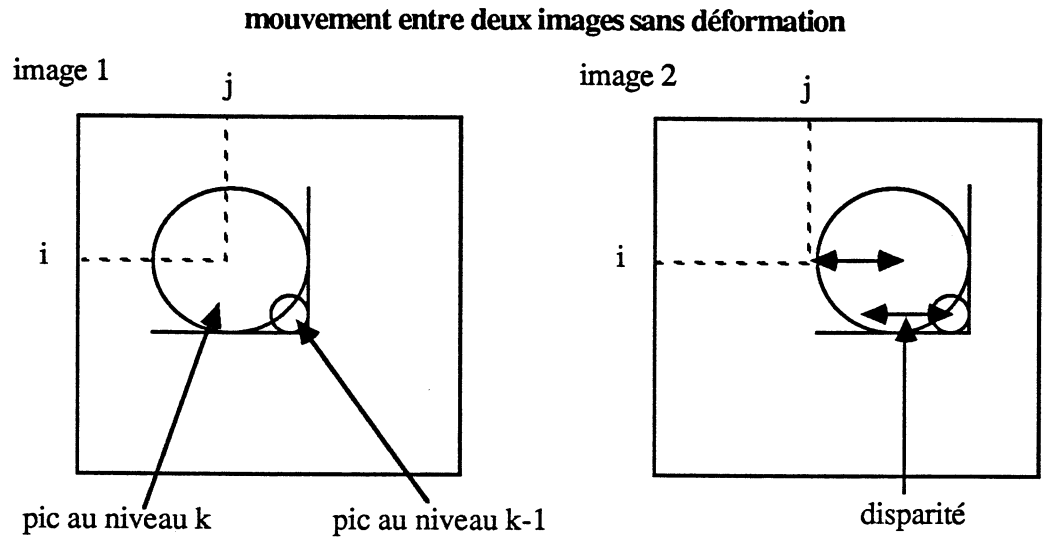


Figure VI.5 : Variation de disparité entre deux niveaux successifs de résolution. La variation est due aux déformations dans l'image engendrées par le mouvement de la caméra. Il est à noter que pour des raisons de meilleure lisibilité de l'exemple, les proportions de taille des pics entre les deux niveaux successifs de résolution ne sont pas respectées.

Cette variation est d'autant plus petite que le mouvement de la caméra est faible. C'est une des raisons pour lesquelles on utilise dans notre système une faible amplitude du mouvement de la caméra.

Du fait de la contrainte sur le déplacement de la caméra, la variation de disparité est très faible (de l'ordre de quelques pixels). L'erreur sur la position des pics étant d'un seul pixel, on prend une variation de disparité égale à quelques pixels (de 3 à 5, les expérimentations ayant montré que l'ajustement n'était pas critique).

VL3.2 Création des hypothèses d'appariement

Les zones de recherche étant très restreintes, il y a peu d'indices à différencier. Nous avons retenu une seule caractéristique pour nos indices pics, à savoir leur signe. D'autres caractéristiques, telles la valeur du Laplacien, seront utilisées dans la phase d'appariement global.

La fonction de dissemblance choisie est la suivante :

$$|s_1 - s_2| \quad \text{où } s_1 \text{ et } s_2 \text{ caractérisent le signe du pic respectivement dans la première et la seconde image.}$$

Nous créons une hypothèse d'appariement entre deux indices si leurs signes sont identiques, c'est à dire si la valeur de dissemblance est nulle.

Le calcul de cette dissemblance est réalisé pour chaque indice de la première image avec les indices de la seconde situés dans la zone de recherche. Nous obtenons un ensemble H d'hypothèses d'appariement. Celles-ci sont ensuite confirmées ou rejetées dans la phase de compatibilité entre les hypothèses.

VL3.3 Compatibilité entre les hypothèses

soit deux hypothèses H_{ia} et H_{jb} :

$$H_{ia} = (G_i, D_a) \quad \begin{array}{l} \text{où } G_i \text{ est un indice de la première image.} \\ D_a \text{ est un indice de la seconde image.} \end{array}$$

$$H_{jb} = (G_j, D_b) \quad \begin{array}{l} \text{où } G_j \text{ est un indice de la première image.} \\ D_b \text{ est un indice de la seconde image.} \end{array}$$

Dans notre application, les deux hypothèses H_{ia} et H_{jb} sont compatibles, si la relation liant les indices G_i et G_j de la première image est compatible avec la relation liant les indices D_a et D_b de la seconde image. C'est à dire si :

$((G_i \neq G_j) \text{ et } (D_a \neq D_b))$ (car unicité de l'appariement) et

$R(G_i, G_j)$ est compatible avec $R(D_a, D_b)$

où R est la relation spatiale définie dans la phase de représentation de l'image.

Cette définition pose le problème de la caractérisation du critère de compatibilité entre nos relations.

VL3.3.1 Compatibilité relationnelle

Notre volonté d'obtenir un graphe complet représentant chaque image nous amène à créer des relations définissables entre tout indice. Celles-ci, présentées dans le chapitre IV, sont les suivantes : "en haut de" (H), "à droite en haut de" (DH), "à droite de" (D), "à droite en bas de" (DB), "en bas de" (B), "à gauche en bas de" (GB), "à gauche de" (G), "à gauche en haut de" (GH) et "proche de" (P). Cependant, il faut souligner qu'un type de relation identifié dans une image n'est pas toujours conservé dans la suivante (cf. figure VI.6).

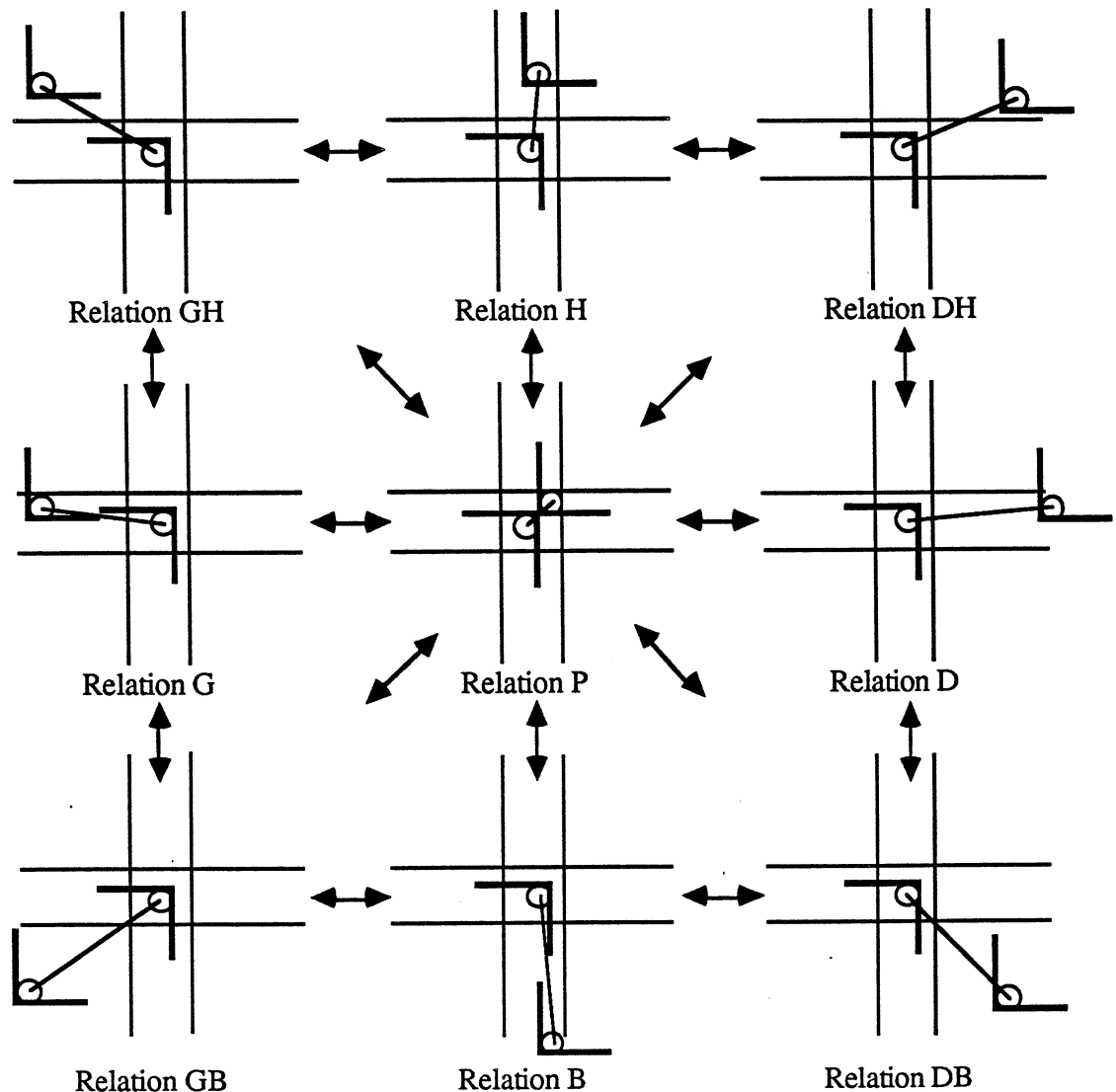


Figure VI.6 : D'une position relative de deux indices (cercles sur la figure), on peut passer à la plupart des autres positions par un mouvement adéquat de la caméra. Les doubles flèches indiquent les changements unitaires possibles. Pour atteindre une position quelconque, on emprunte un chemin dont le point de départ est la position courante et les étapes des changements unitaires successifs. Dans chacune des positions présentées, le quadrillage définissant les relations est centré sur l'indice courant.

Ainsi, d'une relation H entre deux indices, il est possible sur l'image suivante d'avoir une relation DB liant ces deux mêmes indices. Ces forts changements topologiques entre deux images font qu'une relation peut évoluer vers la plupart des autres dans l'image suivante. Ceci entraîne la compatibilité entre toutes les relations (inintéressant car non discriminant lors de l'appariement global). Par conséquent, il est nécessaire de restreindre les variations possibles en relation.

On admet que deux relations sont compatibles si l'évolution entre elles est unitaire. Par exemple, la relation H peut se transformer sur l'image suivante uniquement en GH, DH, P ou rester H. La représentation de la compatibilité peut être visualisée par la figure VI.7.

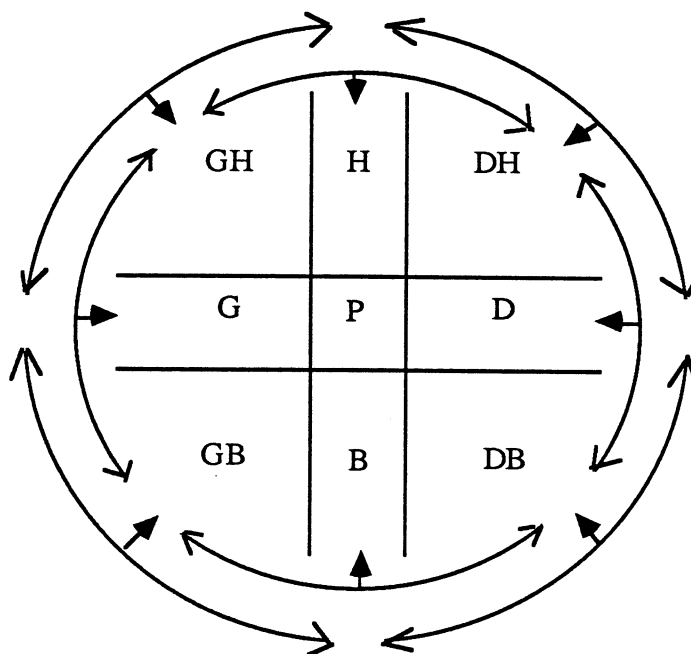


Figure VI.7 : Les arcs de cercles définissent la compatibilité relationnelle par rapport à la zone courante pointée au milieu de l'arc. A noter que la zone P est compatible avec toutes les autres.

La variation unitaire en relation est rendue possible par la contrainte initiale visant à limiter l'amplitude du mouvement de la caméra (cf. section VI.3.1.2) et par la taille des cinq zones de transition (P), (G), (D), (B), (H).

La taille de ces zones est dictée par la région d'héritage. En effet, si les zones de transition sont trop larges, la zone P centrée sur un pic de la région d'héritage couvrira entièrement cette dernière, ce qui entraînera une compatibilité relationnelle entre toutes les hypothèses construites sur la base des pics appartenant à la région. D'où l'inefficacité d'un critère global d'appariement recherchant le plus grand sous-ensemble d'hypothèses mutuellement compatibles. Par contre, si la zone est très fine, les changements topologiques permis sont considérablement limités. La taille des régions d'héritage est de 7×7 pixels (car de rayon $2\sigma + 1$ où σ , correspondant à l'écart type du filtre initial lors de la transformée de DOLP, est fixé à 1). Ainsi, pour limiter les ambiguïtés d'appariement entre régions d'héritage associées,

la largeur des zones doit être inférieure à 7 pixels et avoir un nombre impair de pixels pour assurer la symétrie (la zone P étant centrée sur le pic courant). Notre choix se limitant à une zone de 1, 3 ou 5 pixels, on adopte 3 qui s'avère être un bon compromis.

VI.3.3.2 Evaluation de la compatibilité

L'implémentation logicielle de la compatibilité relationnelle est évidente. En effet, un simple codage nous permet de trouver facilement la (non) compatibilité entre deux index comparés. Le codage est donné dans la figure VI.8:

01100000	11000000	10000001
00110000	11111111	00000011
00011000	00001100	00000110

Figure VI.8 : Codage des relations en vue d'une détermination aisée de la compatibilité.

La compatibilité est testée en réalisant un "et" logique entre les deux codes correspondant aux zones à comparer. Il y aura incompatibilité, dans le seul cas où le résultat de ce test est égal à zéro. Ainsi DH qui a pour code 10000001 est compatible avec H (11000000) car le "et" logique donne un résultat différent de zéro (10000000). Par contre il n'est pas compatible avec DB (00000110) car le "et" logique entre ces deux valeurs donne 0.

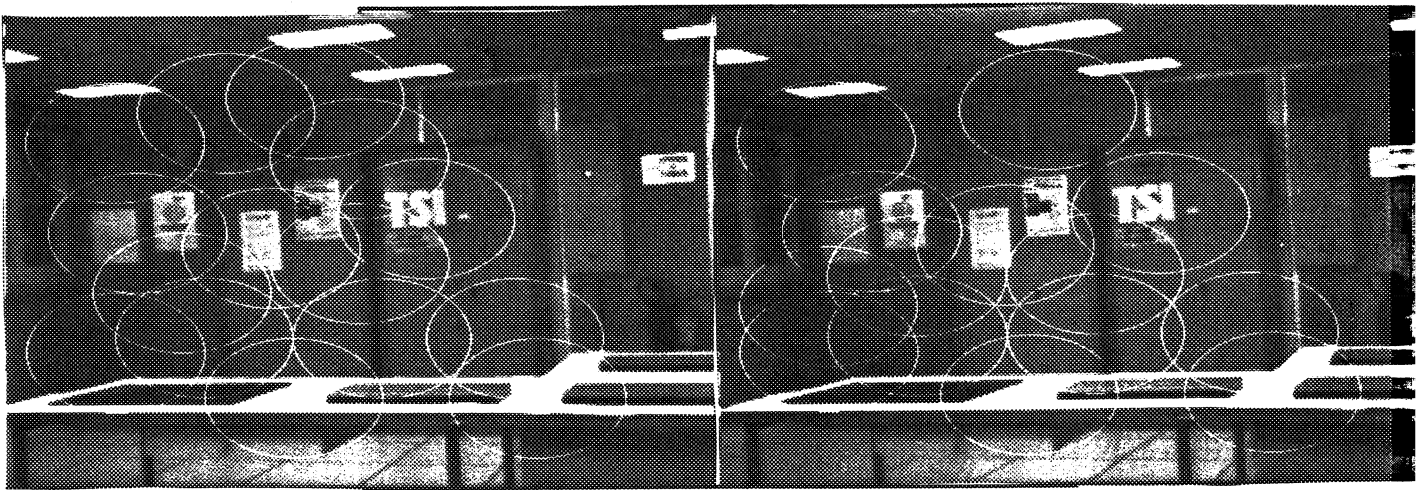
VI.3.3.3 Recherche de la plus grande clique maximale

Le critère de compatibilité étant défini, on détermine celle-ci entre chaque couple d'hypothèses. Ceci permet de créer un graphe associé dans lequel chaque noeud est une hypothèse et chaque arc définit la compatibilité entre les deux noeuds qu'il relie.

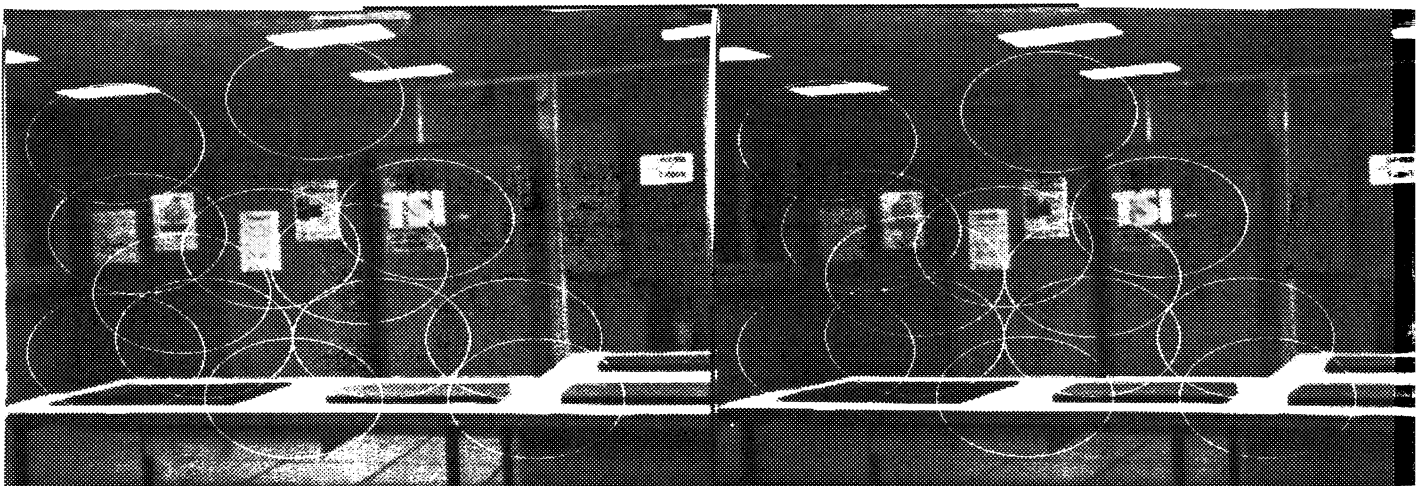
Le plus grand sous-ensemble d'hypothèses d'appariement mutuellement compatibles est obtenu par l'extraction de la plus grande clique maximale du graphe associé. Celle-ci fournit l'ensemble des appariements valides selon notre critère de compatibilité. En cas de litige, c'est à dire plusieurs plus grandes cliques de même taille, une comparaison sur le niveau Laplacien des indices associés permet d'effectuer le choix. L'algorithme utilisé pour rechercher les cliques est celui décrit dans [Bolles 79]. Il est détaillé dans l'annexe 3.

VI.4 Applications et résultats

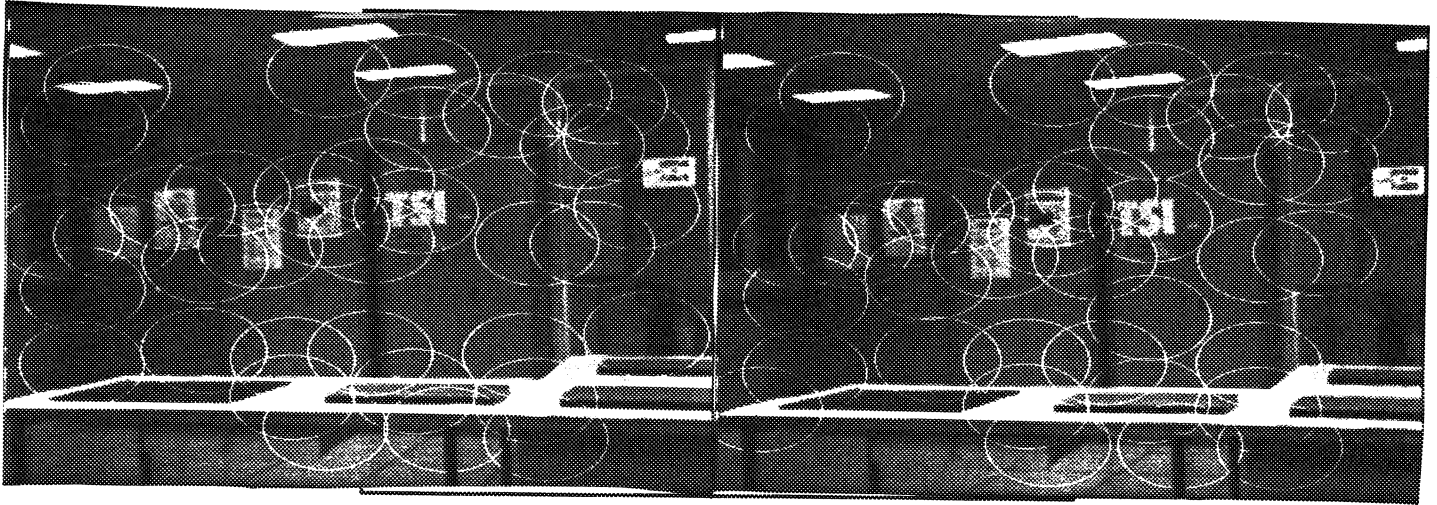
Les photos ci-dessous représentent les résultats des appariements à chaque niveau de résolution. La première image visualise les pics extraits, la seconde les pics appariés dans les régions d'héritage et la troisième l'appariement total après la mise en correspondance des pics des régions complémentaires. Au niveau 10 de résolution, le plus faible dans notre application, l'héritage est inexistant. Suit un tableau présentant les résultats numériques au niveau 3 de résolution.



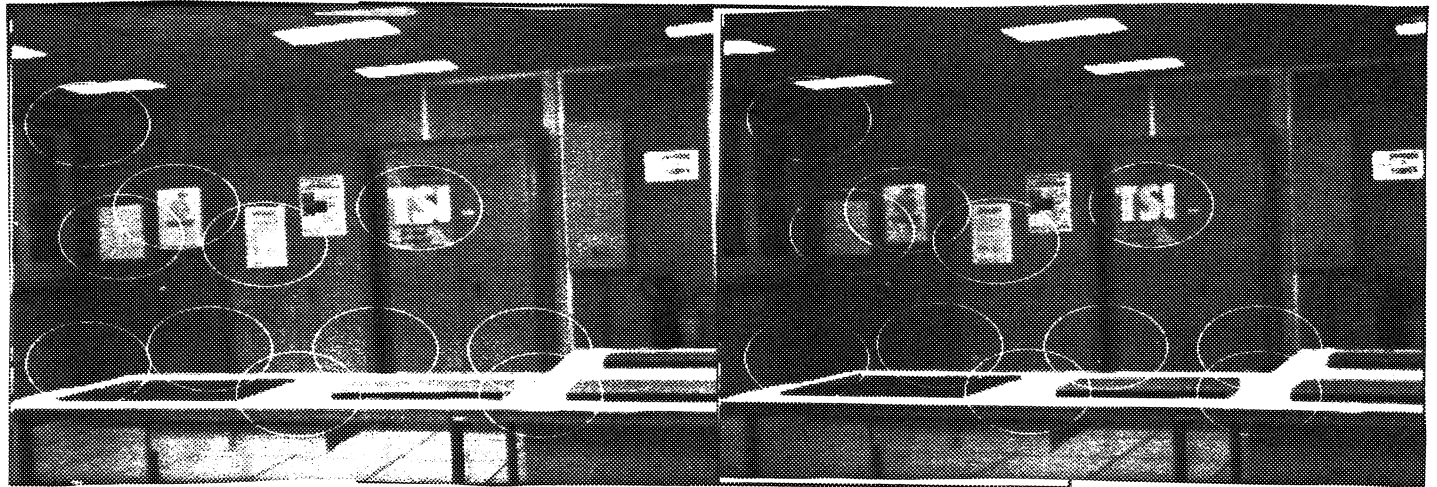
Pics au niveau 10



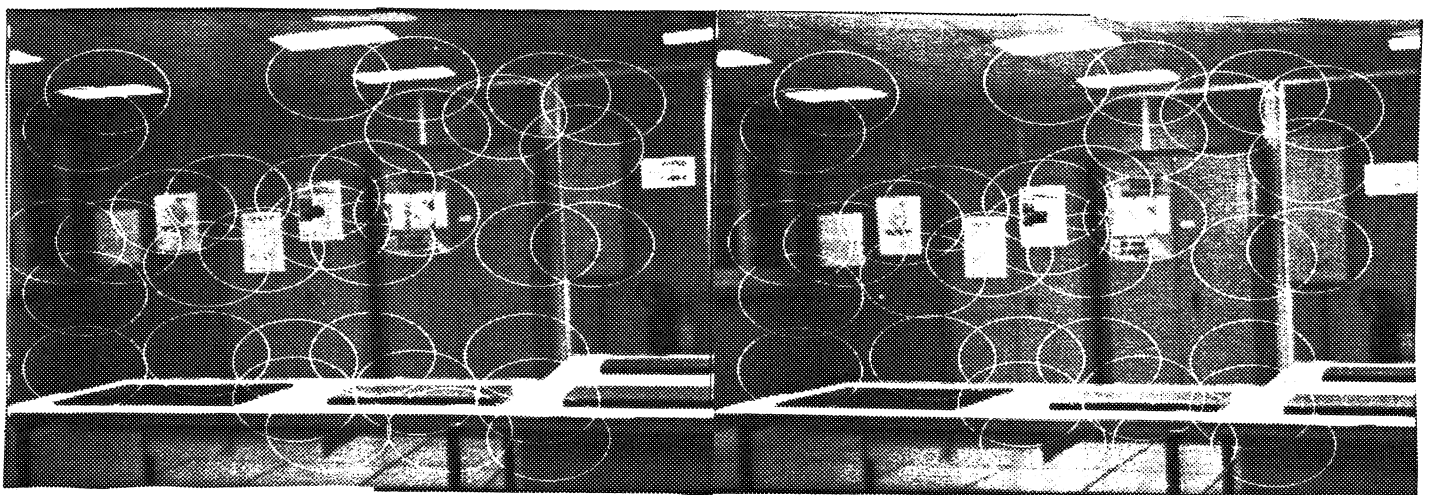
Pics appariés au niveau 10



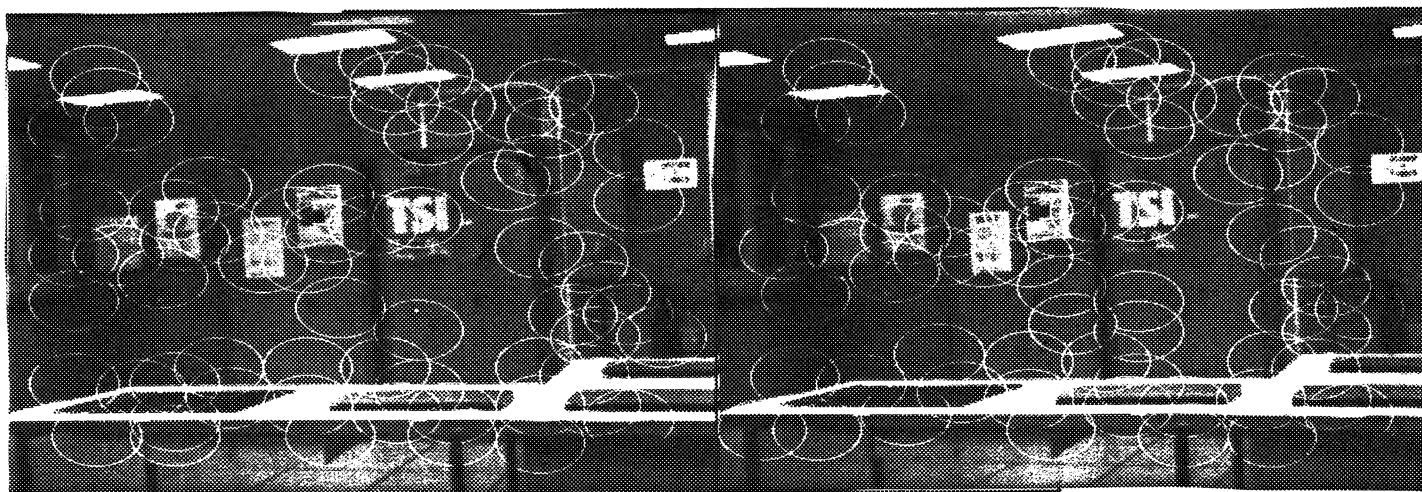
Pics au niveau 9



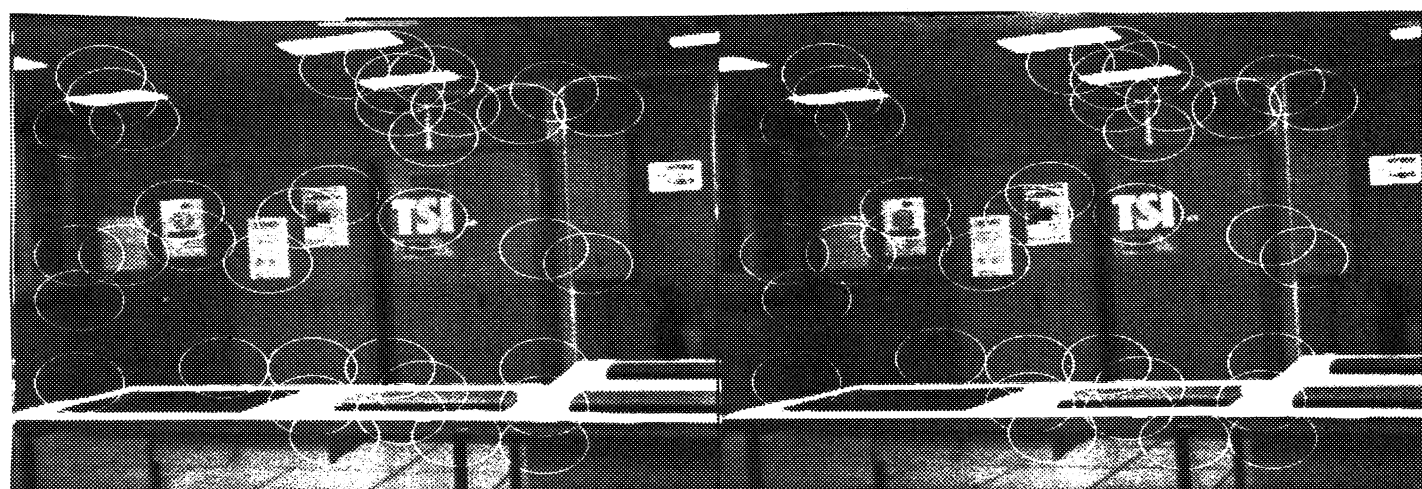
appariements hérités au niveau 9



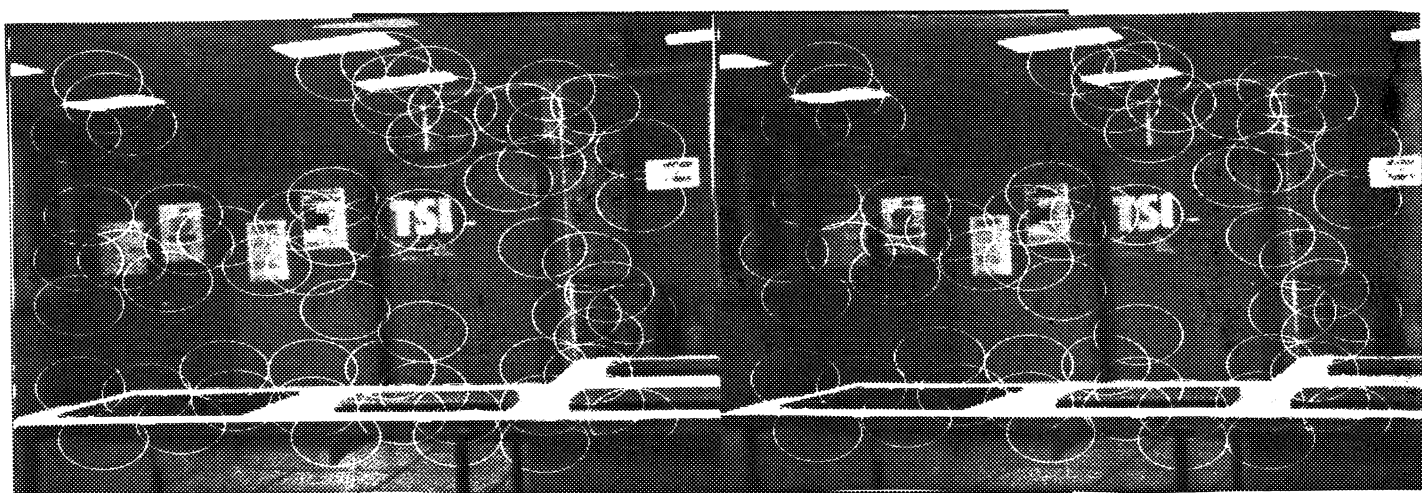
Pics appariés au niveau 9



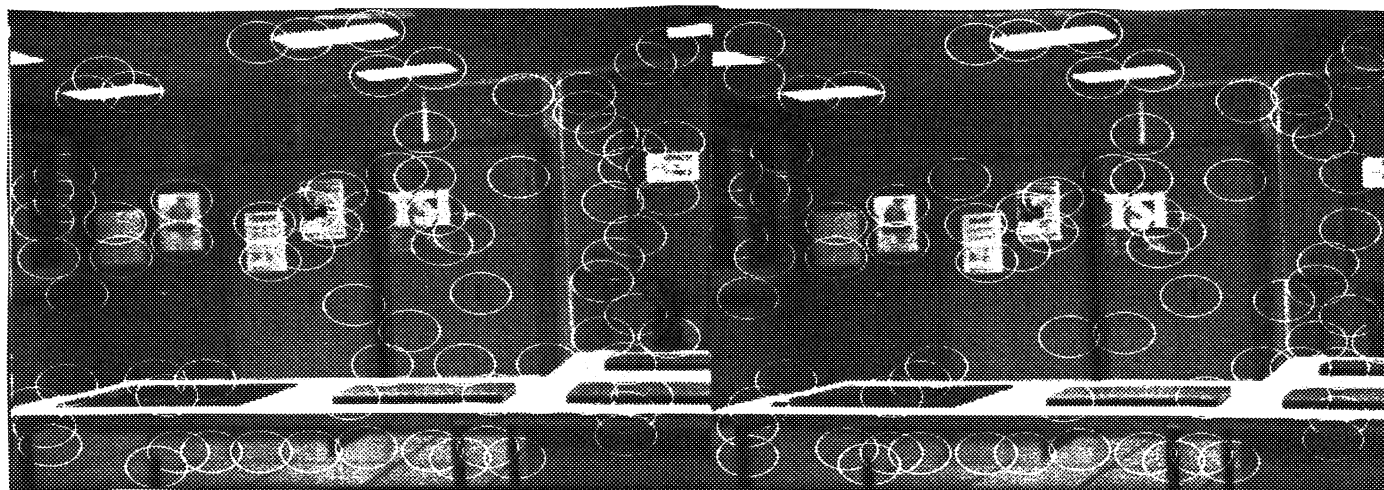
Pics au niveau 8



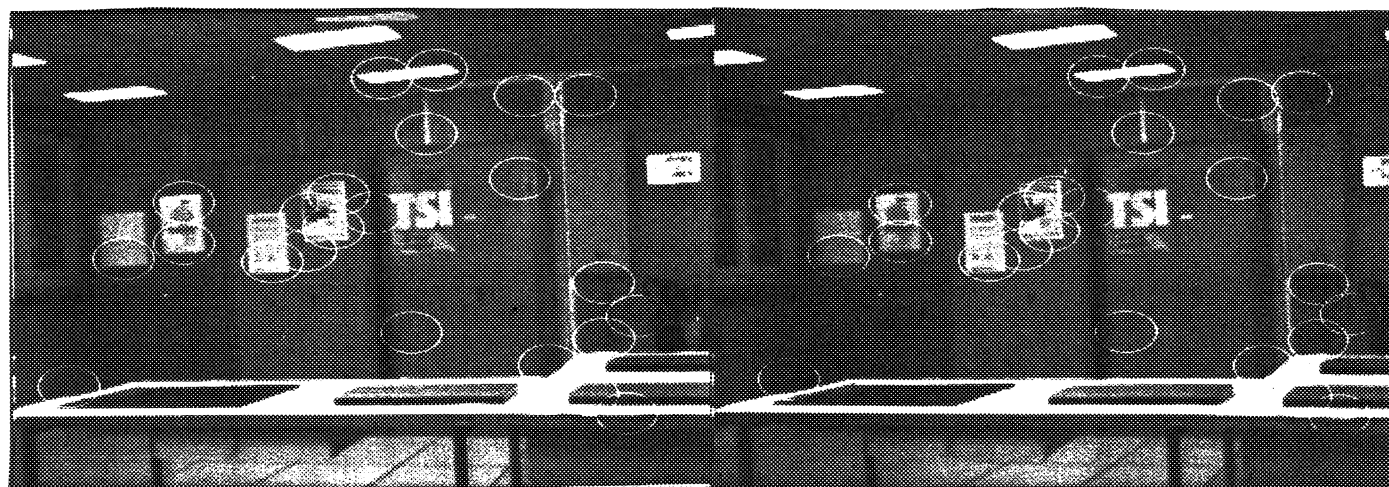
appariements hérités au niveau 8



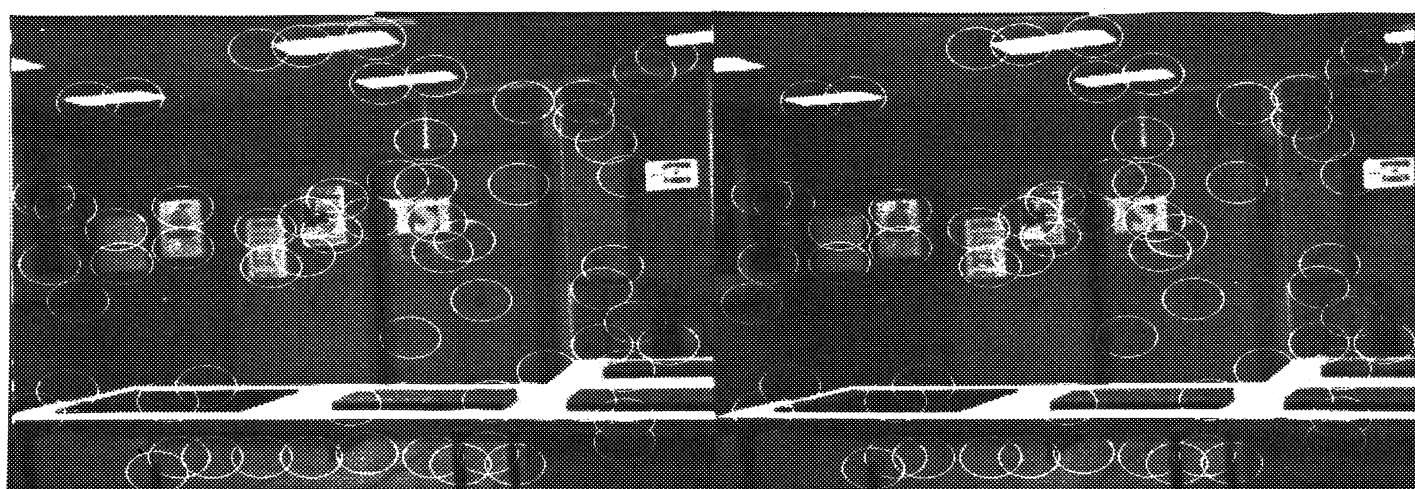
Pics appariés au niveau 8



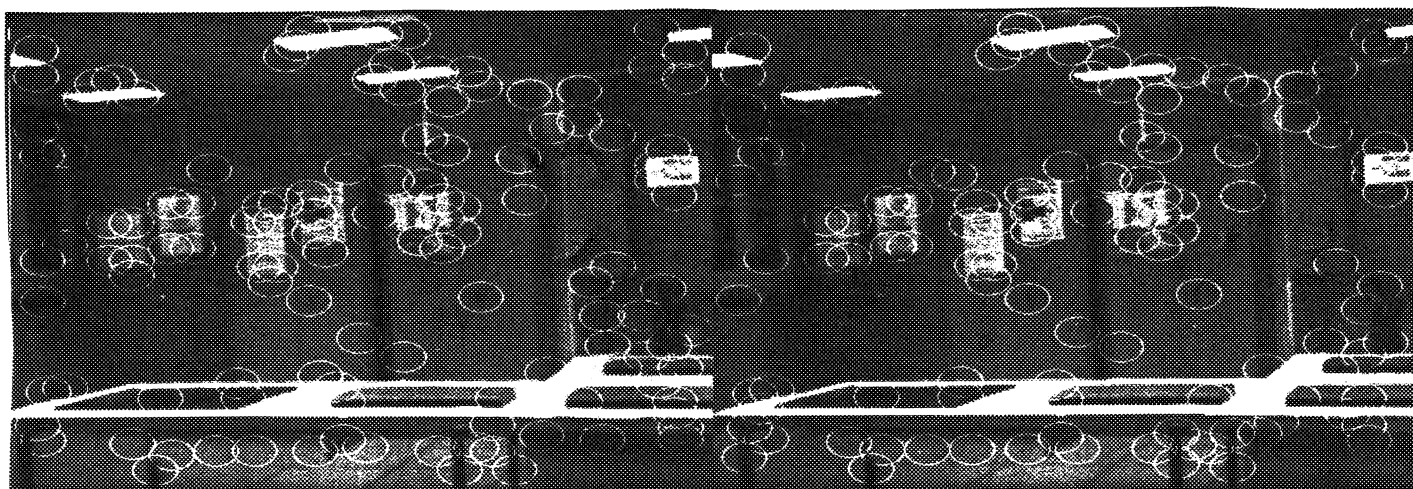
Pics au niveau 7



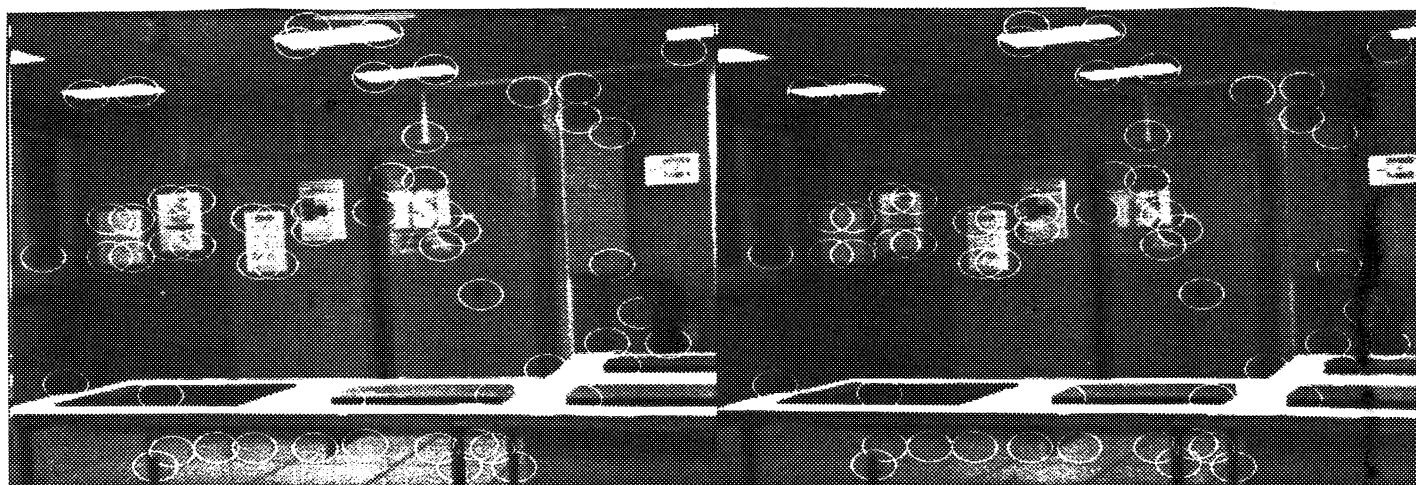
appariements hérités au niveau 7



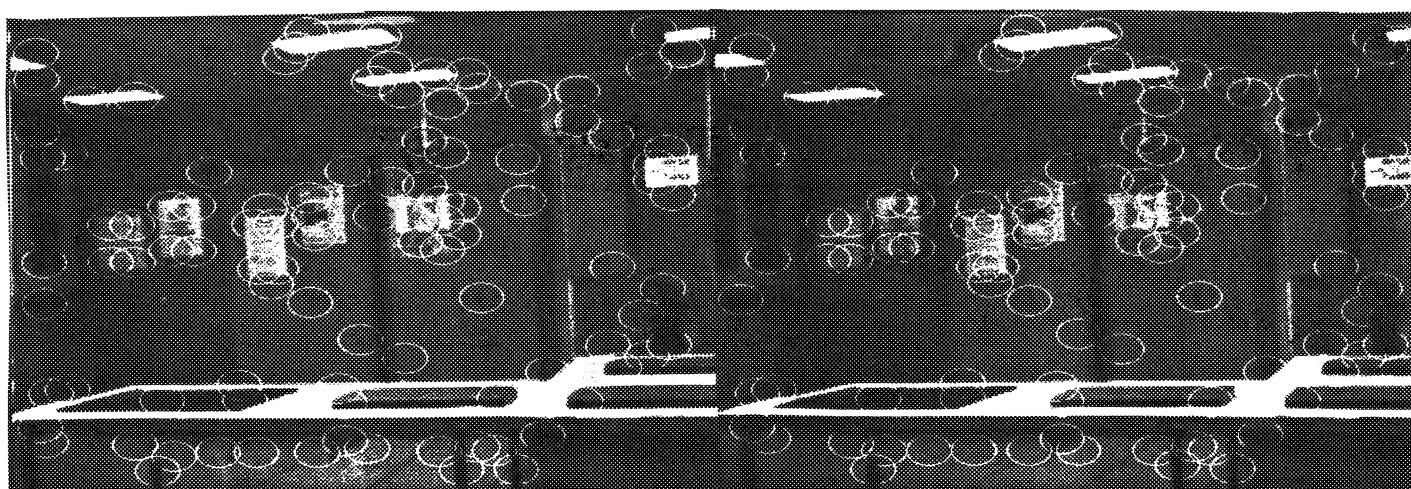
Pics appariés au niveau 7



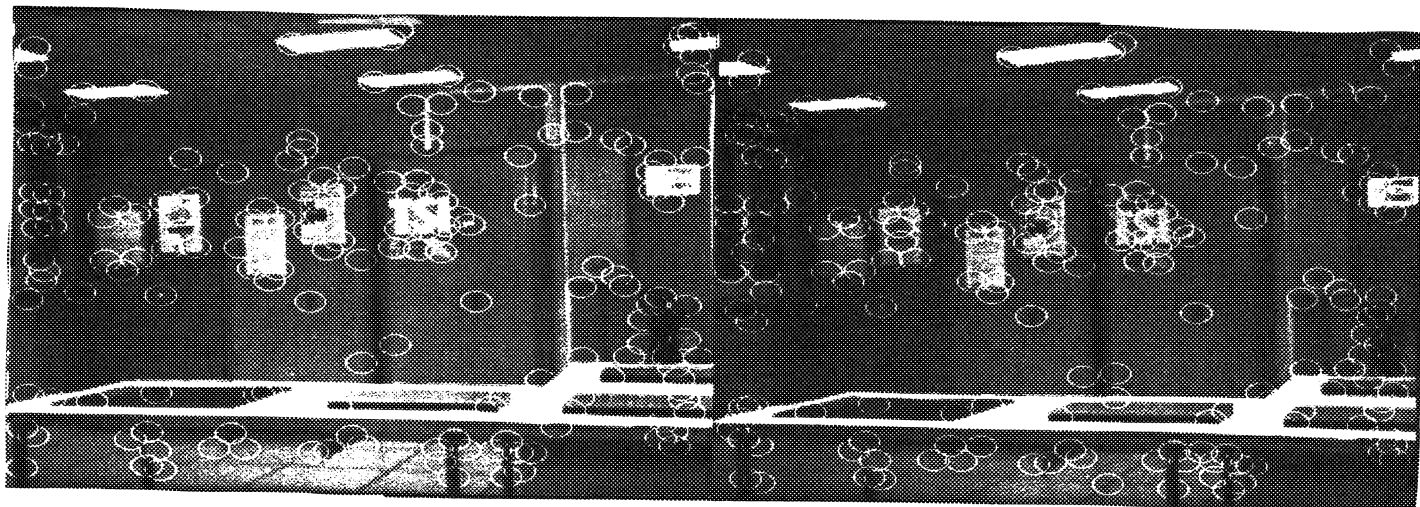
Pics au niveau 6



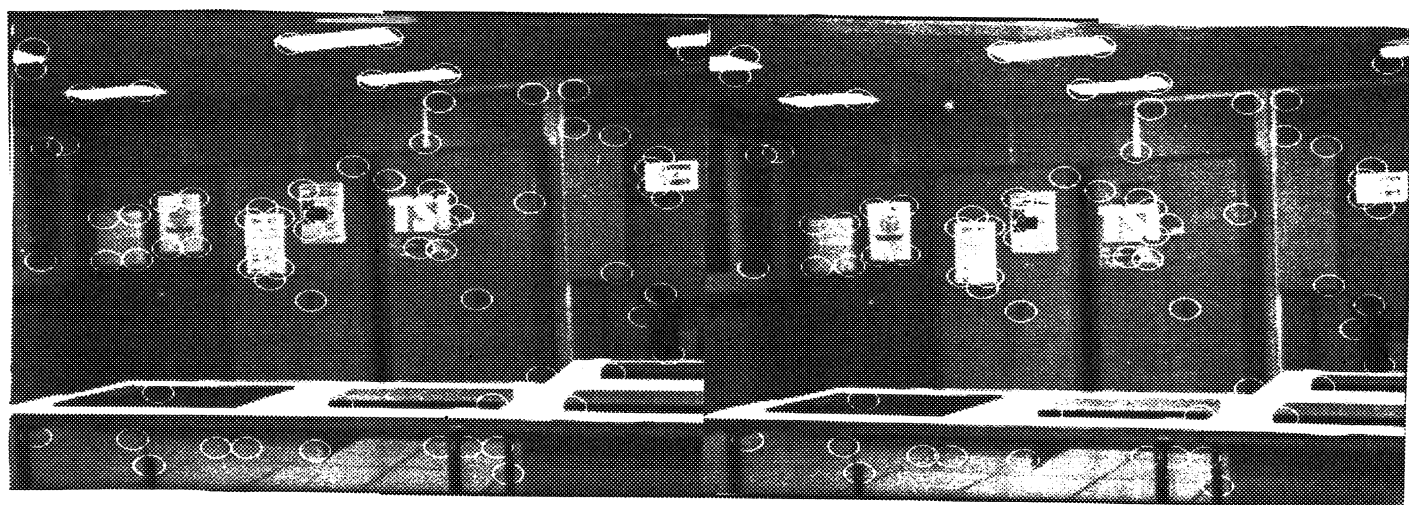
appariements hérités au niveau 6



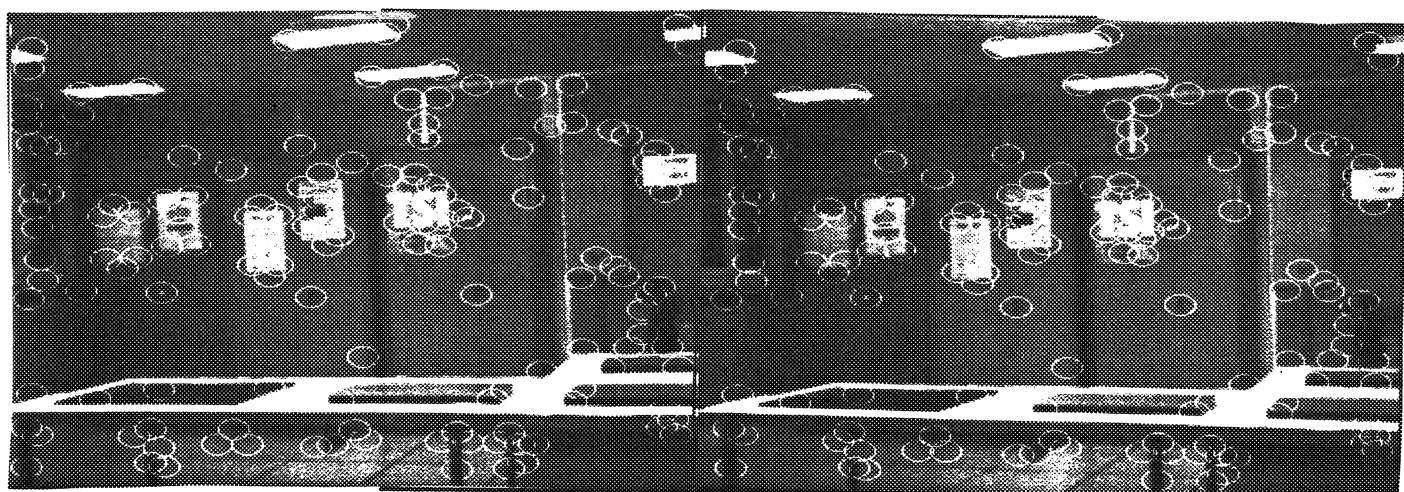
Pics appariés au niveau 6



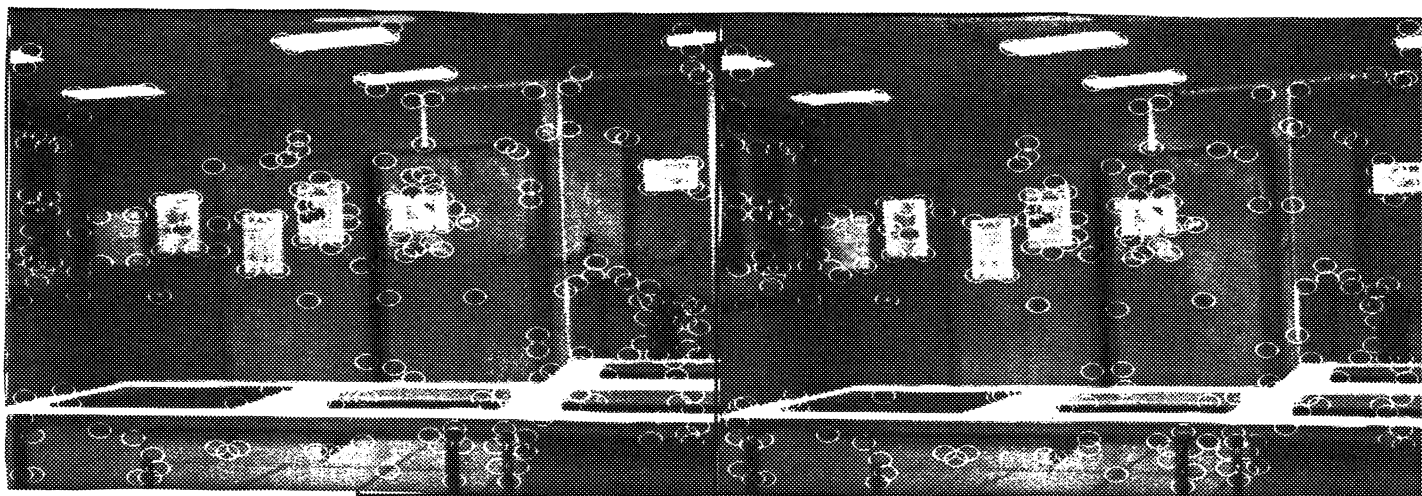
Pics au niveau 5



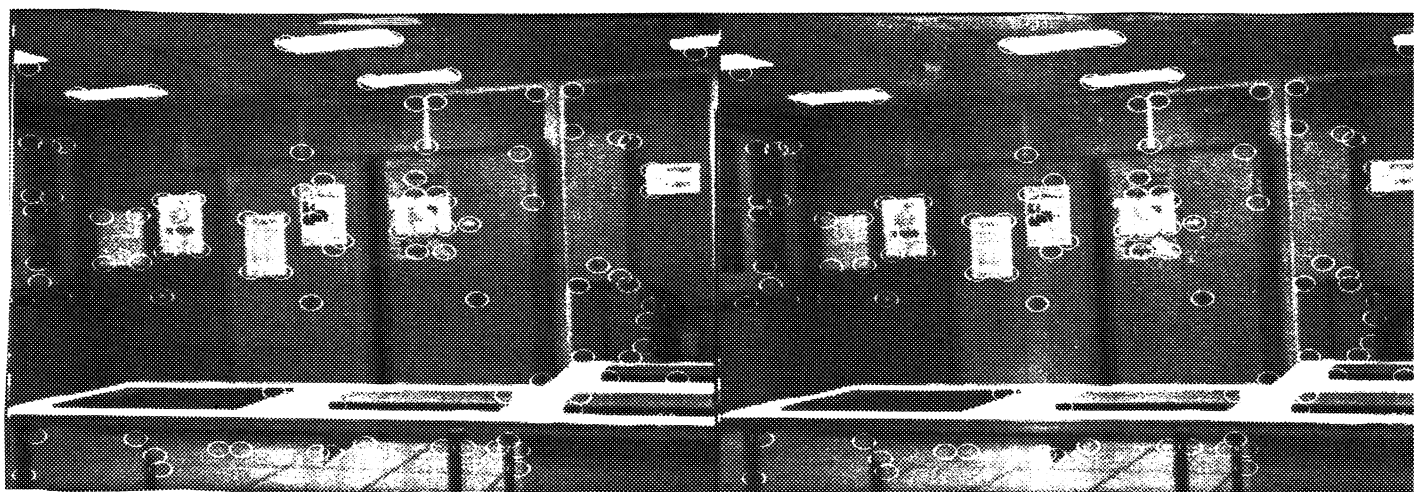
appariements hérités au niveau 5



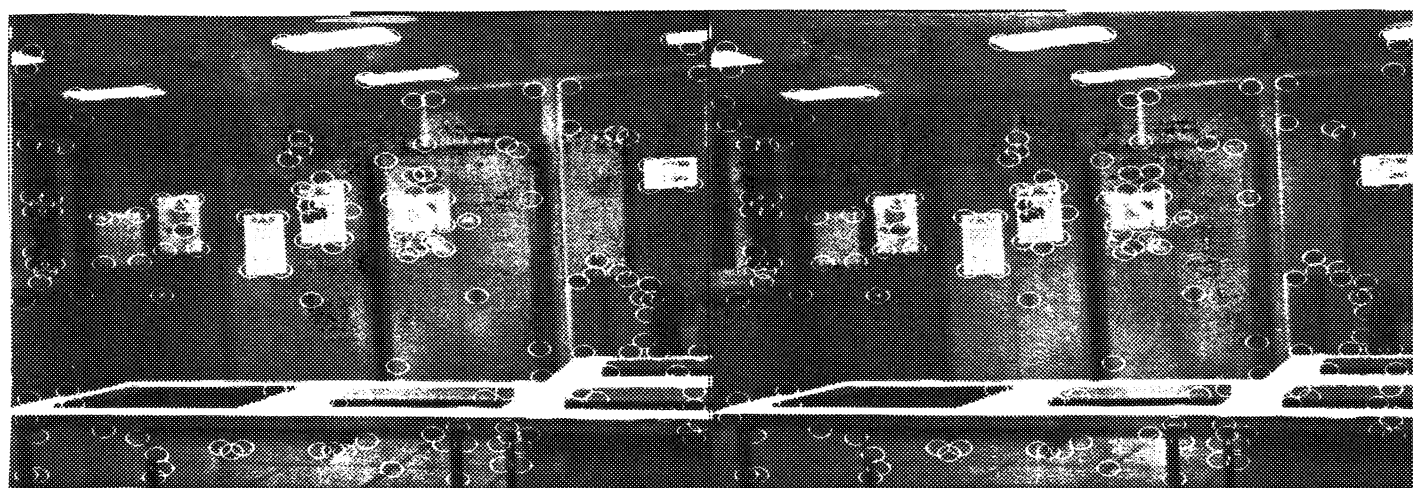
Pics appariés au niveau 5



Pics au niveau 4



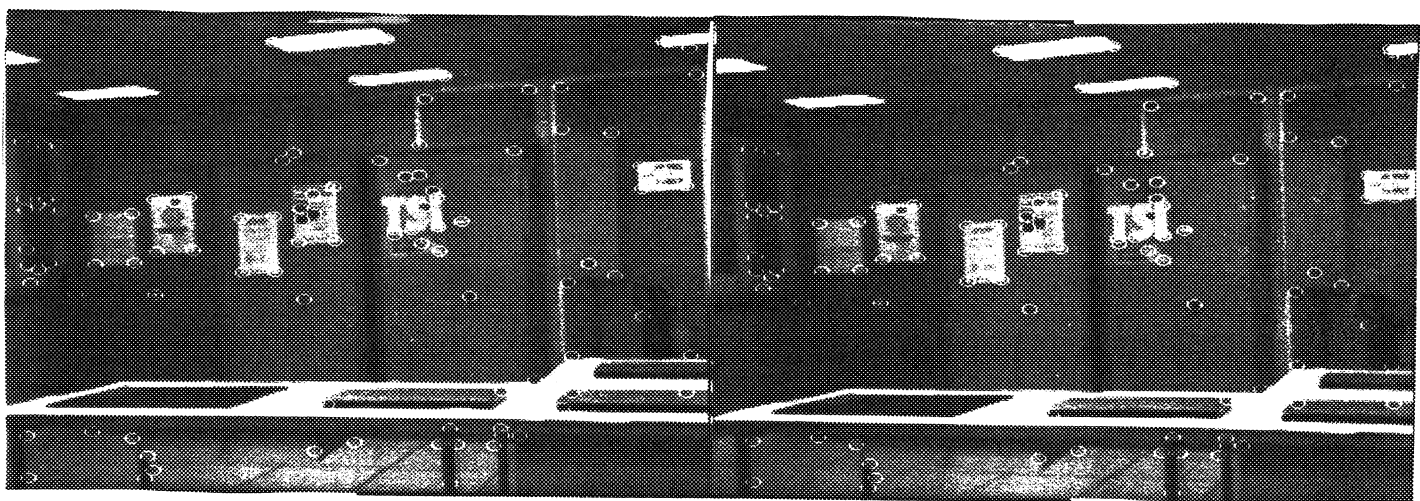
appariements hérités au niveau 4



Pics appariés au niveau 4



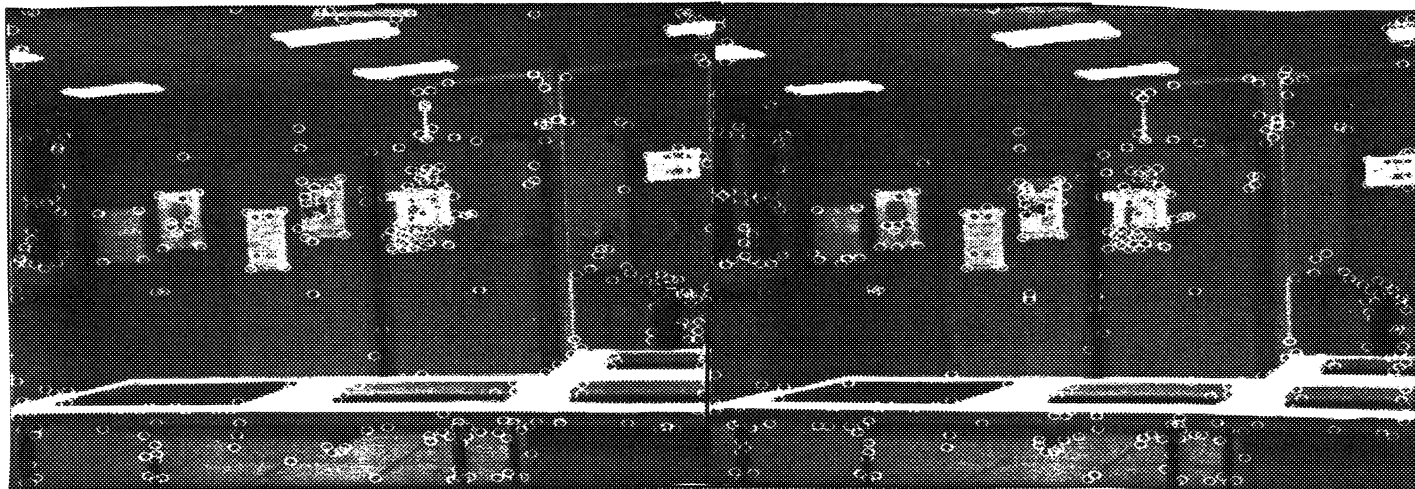
Pics au niveau 3



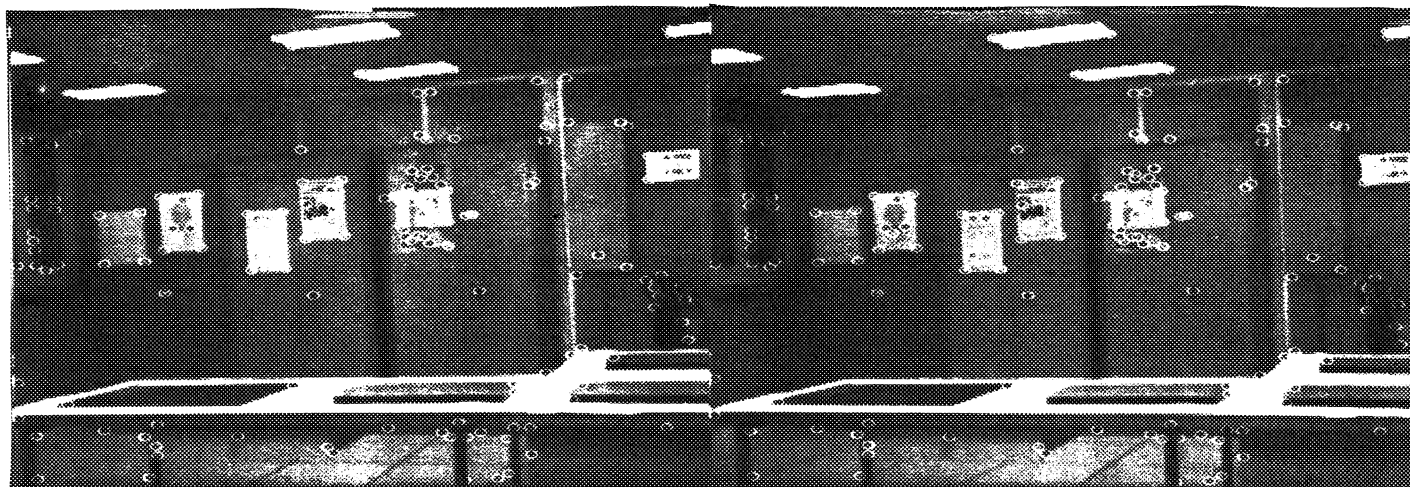
appariements hérités au niveau 3



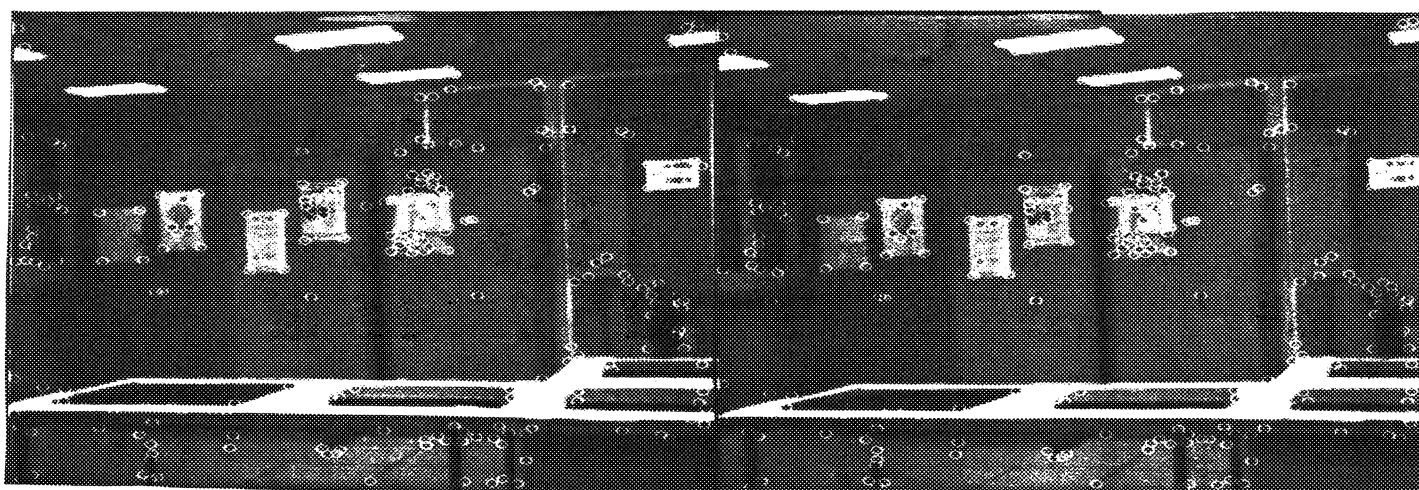
Pics appariés au niveau 3



Pics au niveau 2

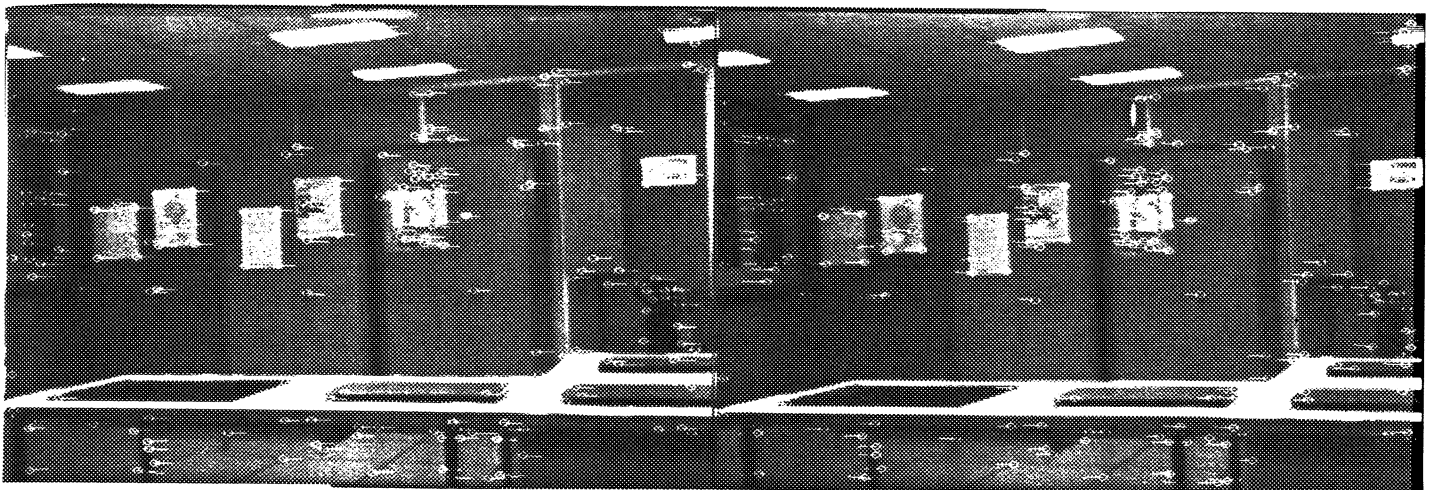


appariements hérités au niveau 2



Pics appariés au niveau 2

Nombre total de pics dans la première image	331
Nombre total de "vrais" pics dans la première image	168
Nombre total de pics dans la seconde image	332
Nombre total de "vrais" pics dans la seconde image	180
Nombre total d'appariements	221
Nombre d'appariements corrects	201
Nombre de "vrais" pics parmi les appariements justes	163



Visualisation des appariements au niveau 3 de résolution. Les cercles représentent la position des pics appariés et les segments le déplacement des pics associés dans l'autre image.

VI.5 Conclusion

Les images sont représentées par une pyramide multi-résolutions de Laplacien. De chaque niveau de résolution sont extrait les pics. Puis, excepté pour la plus basse résolution, l'ensemble des pics est réparti en sous-ensembles grâce aux héritages des appariements du niveau précédent. Ce partitionnement est nécessaire pour limiter les coûts exorbitants d'un appariement global basé sur la recherche des cliques maximales.

Chaque sous-ensemble d'indices est ensuite représenté par un graphe complet dans lequel chaque noeud est un indice pic caractérisé par son attribut, et chaque arc une relation topologique entre deux pics. La phase de mise en correspondance est alors répartie en divers appariements globaux entre sous-ensembles d'indices et ce, en initialisant le processus au niveau de plus faible résolution jusqu'à celle souhaitée.

Dans notre système, l'appariement est basé sur l'emploi d'indices de type "point" (cas des pics). Pourtant, leur quantité, leur faible taux de discrimination et leur pauvreté lors d'une reconstruction tridimensionnelle présentent généralement d'importants inconvénients lors de l'appariement et de phases ultérieures. Notre système prend en compte ces trois problèmes et en limite les influences.

Le nombre d'indices de type "point" dans une image est souvent plus élevé que le nombre d'indices de type "contour" ou "région". La conséquence en est une augmentation des coûts lors de l'appariement. Dans notre système, on peut constater effectivement que le nombre de pics est élevé dans les hautes résolutions. Cependant, on limite le problème du coût en partitionnant l'image en fonction d'informations extraites dans de plus faibles résolutions.

Les indices de type point sont en général peu discriminants. Ceci risque d'engendrer une importante quantité d'hypothèses d'appariement, d'où un coût en temps élevé pour limiter les ambiguïtés. Dans notre système, nous discriminons les pics uniquement sur leur caractéristique signe ce qui permet de différencier seulement la moitié des pics. Cependant, nous limitons la quantité d'hypothèses d'appariement car nos zones de recherche sont extrêmement réduites.

Les systèmes d'appariement sont souvent envisagés en fonction d'une phase ultérieure telle la reconstruction tridimensionnelle. Compte tenu de la pauvreté d'informations

fournies par les indices de type point, on lui préfère généralement des indices de type "contour". Il faut cependant constater que dans notre système, les pics sont reliés par des crêtes. Ainsi, un appariement de pics induit un appariement de crêtes. Moyennant une phase de conversion des résultats de l'appariement, il est possible de faire la reconstruction avec les crêtes (indices de type "contour").

Par contre, les avantages apportés par les indices pics lors de la phase d'appariement sont nombreux. En effet, ils facilitent et augmentent la validité de l'appariement.

Ils permettent une propagation aisée d'héritage d'appariements d'indices à travers les résolutions, ce qui fournit un partitionnement efficace tenant compte d'informations extraites dans le niveau de résolution inférieur.

De plus, ils autorisent la création aisée de graphes complets représentant chaque image. Par conséquent, on est assuré d'obtenir un graphe unique associé entre deux images. Par ce biais, on évite une recombinaison de résultats provenant de traitements appliqués à divers graphes associés, d'où un gain en temps et en validité des résultats.

La connaissance de la compatibilité ou l'incompatibilité de tout couple d'hypothèses est une nécessité lors de la recherche du plus grand sous-ensemble d'hypothèses mutuellement compatibles. Dans notre cas, chaque image étant représentée par un graphe complet, et les relations choisies entre indices étant toutes comparables, on possède cette connaissance. On évite de ce fait l'emploi de propagations transitives des relations de compatibilité et d'incompatibilité, très coûteuses et sources d'erreurs du fait de leur non-transitivité.

Ainsi, l'utilisation des indices "pics" associée à une technique "coarse to fine" d'appariement permet d'obtenir de bons résultats. En effet, la quasi-totalité des vrais pics a été appariée (163 "vrais" pics appariés sur 168 dans la première image et 180 dans la seconde). Ces résultats peuvent être utilisés dans une phase ultérieure telle la reconstruction tridimensionnelle de la scène perçue, ou l'analyse d'une séquence d'images en vue d'estimer le mouvement des objets.

VII CONCLUSION

Dans cette thèse, nous nous sommes intéressés aux problèmes de l'appariement d'indices images. En vue d'une meilleure compréhension des techniques existantes, nous avons proposé une synthèse de l'état de l'art dans les domaines de la représentation des informations images et de la mise en correspondance d'indices. Nous avons ensuite mis l'accent sur l'intérêt d'une représentation multi-résolutions de l'image et le mettons en valeur par deux techniques différentes d'appariement.

Dans un premier temps, en vue de limiter les coûts d'appariement, nous nous orientons vers l'utilisation d'indices très discriminant (structures d'arbres de pics), et de zones restreintes de recherche. Un calcul préalable fournit une pyramide multi-résolutions de bandes passantes de l'image initiale. De chaque image constituant cette pyramide, on extrait des pics qui sont des extrema locaux du laplacien. Des caractéristiques de formes telles les coins, les extrémités ou une vue globale de la forme sont représentées par des pics. Le bruit dans l'image et l'échantillonnage conduisant à la détection d'un nombre important de pics sans signification physique, on en élimine une bonne partie par l'introduction d'un critère d'angle et de concavité locale. Les pics sont ensuite liés à travers les résolutions afin de créer des arbres de pics.

La technique de mise en correspondance de ces indices est basée uniquement sur une fonction de dissemblance qui quantifie analytiquement le degré d'appariement de deux structures. Si la réponse en dissemblance est faible, on crée une hypothèse d'appariement qui peut être validée à condition que ce minimum corresponde à un extremum local.

Les expériences effectuées avec cette technique montrent que les indices sont très discriminants, mais malheureusement trop sensibles aux variations qui surviennent dans les images. Cette démarche convient plutôt à l'appariement entre modèles bidimensionnels et formes observées. Les expériences mettent aussi l'accent sur la nécessité d'un critère global de mise en correspondance, permettant de limiter efficacement les ambiguïtés d'appariement.

Ces considérations nous orientent vers une autre technique d'appariement. Pour restreindre les ambiguïtés, nous prenons en compte les informations contextuelles présentes dans les images. Celles-ci permettent de déterminer la compatibilité entre les hypothèses d'appariement. Ainsi, trouver l'appariement valide consiste à rechercher le plus grand sous-ensemble d'hypothèses mutuellement compatibles selon notre critère contextuel. Pour cela, on recherche la plus grande clique maximale dans un graphe d'hypothèses d'appariement, formulation mathématique élégante de la solution à notre problème.

Cette méthode nous amène à acquérir un graphe complet représentant les images. Dans ce but, nous choisissons d'utiliser les indices "pics", car ils se prêtent aisément à la définition de relations topologiques adéquates. D'autre part, la méthode nous contraint à partitionner de manière adéquate l'image en vue de restreindre la forte combinatoire. Ce partitionnement est fourni par des héritages dans un schéma multi-résolutions.

Comme précédemment, la représentation de l'image passe par un traitement préalable qui fournit une pyramide multi-résolutions de passe-bandes de l'image initiale. De chaque image de cette pyramide, on extrait les pics. Cependant, pour limiter les influences des variations dans les images, les liens de descendance ne sont pas exploités au sein de l'indice.

L'appariement des indices correspond à un schéma "coarse to fine". Des relations topologiques sont créées entre tous les pics dans chacune des deux images de plus faible résolution. Ainsi, chacune de ces images est représentée par un graphe complet. L'appariement des pics s'effectue d'abord dans cette plus faible résolution, car le nombre et la densité des indices est faible. Un critère de similarité permet de créer des hypothèses d'appariement. Des compatibilités entre les hypothèses sont ensuite établies en fonction des relations topologiques. On construit alors un graphe associé dont les noeuds sont les hypothèses et les arcs les compatibilités. Une recherche dans ce graphe de la plus grande clique maximale produit le plus grand sous-ensemble des hypothèses mutuellement compatibles, et donc les appariements dans la plus faible résolution. Ceux-ci permettent de répartir dans la résolution supérieure les indices en régions associées. Les relations topologiques sont créées uniquement entre les indices d'une même région d'héritage. L'appariement est alors réparti en diverses mise en correspondance d'indices entre régions associées. Cette répartition permet de restreindre énormément le coût de la recherche de la plus grande clique maximale. Le processus s'effectue de même jusqu'à ce que les indices pics soient appariés au niveau de résolution souhaité.

Par rapport à d'autres techniques, notre méthode apporte plusieurs contributions :

Elle permet de partitionner de manière adéquate l'image grâce à l'introduction d'héritages au sein d'une pyramide multi-résolutions. Cette démarche s'avère importante pour réduire la forte combinatoire inhérente à l'emploi de techniques globales d'appariement.

En introduisant un appariement global, elle limite les erreurs d'appariement et leur propagation à travers les résolutions (problèmes préoccupants dans les techniques classiques de type "coarse to fine").

Néanmoins, il reste des problèmes à approfondir :

Lors des expérimentations, nous avons constaté que 80% des "faux" pics ont été supprimés dans les hautes résolutions par l'introduction d'un critère d'angle et de concavité locale. Cependant, il subsiste 50% de "faux" pics parmi l'ensemble total des pics à un niveau de haute résolution. Ceci a pour conséquence, une augmentation du nombre d'hypothèses d'appariement, d'où une croissance des coûts et des erreurs potentielles de mise en correspondance.

Nous avons aussi constaté que des pics pouvaient être situés à l'extérieur de la région d'héritage induite par leur père (c'est le cas par exemple lorsqu'ils se situent dans un coin formant un angle trop aigu). Ces pics sont alors insérés lors du partitionnement de l'image dans la région complémentaire aux régions d'héritage. Par conséquent, lors de l'appariement global, si le nombre de pics dans cette région est trop élevé, le coût risque d'être très important.

En vue de résoudre ces problèmes, nous envisageons d'expérimenter l'idée des gènes introduite lors de conversations personnelles avec Alain Chehikian. Cette idée est née de la constatation expérimentale que les pics, à un niveau de résolution donné, sont déjà annoncés dans le niveau inférieur. Le gène est l'indicateur de la position que peut avoir un pic dans la résolution supérieure. Il correspond soit à une extrémité de crête, soit à un pic.

Nous pensons que le problème d'héritage peut être résolu, ou du moins amélioré, en recherchant les héritiers dans un petit voisinage autour de chaque position induite par un gène. Si cette solution est valide, on peut alors envisager d'utiliser les résultats obtenus pour améliorer l'élimination des "faux" pics. En effet, ceux-ci vu leur caractère aléatoire ont peu de chances d'avoir plusieurs ancêtres. Par conséquent, les "vrais" pics peuvent être discriminés grâce à une comparaison avec un seuil minimal portant sur le nombre d'ancêtres.

BIBLIOGRAPHIE

- [Asada 86] H. Asada, M. Brady
The curvature primal sketch,
IEEE Trans. on PAMI-8, No. 1, January 1986.
- [Aubert 86] D. Aubert
Analyse du mouvement portant sur le trafic routier,
DEA, LIFIA, INPG, Grenoble, juillet 1986.
- [Ayache 84] N. Ayache, B. Faverjon, J.D. Boissonnat, B. Bollack,
Manipulation automatique de pièces industrielles en vrac planaire
1er colloque IMAGE, CESTA, Biarritz, mai 1984.
- [Ayache 85] N. Ayache, B. Faverjon
Un algorithme rapide de stéréovision passive utilisant la
prédiction et la vérification récursive d'hypothèses.
Proc. 5eme congrès RFIA, Grenoble, Nov 1985.
- [Ballard 82] H. Ballard, C.M. Brown,
Computer Vision, Prentice Hall Inc., 1982.
- [Bolles 79] R.C. Bolles
Robust feature matching through maximal cliques,
Proc. SPIE Technical symposium on imaging and assembly,
Washington, D.C., April 1979.
- [Bolles 82] R. C. Bolles, R.A. Cain
Recognizing and locating partially visible objects:
The local-Feature-Focus method.
IJRR 1,3, P. 57-82
- [Bolles 86] R.C. Bolles, R. Horaud,
3DPO: A three-dimensional part orientation system,
International journal of robotics research, 5(3) : 3-26, 1986.
- [Burt 81] P.J. Burt
Fast filter transforms for image processing,
Computer Graphics Image Processing, vol. 16, pp. 20-51, 1981.
- [Burt 82] P.J. Burt, C. Yen, X. Xu,
Local correlation measures for motion analysis, a comparative study.
IEEE conf. Publication, Vol. 14, Las Vegas, June 1982.
- [Canny 86] J.F. Canny
A computational approach to edge detection.
IEEE Trans. on PAMI, Vol. PAMI-8, N.6, November, 1986.
- [Castan 84] S. Castan, J. Shen,
A stereo vision algorithm taking into account the perspective distortions,
7th International Conference on Pattern Recognition,
Montreal, Canada, 1984.

- [Crowley 81] J.L. Crowley
A representation for visual information.
PhD thesis, Carnegie-Mellon University, November, 1981.
- [Crowley 84a] J.L. Crowley, A.C. Parker
A representation for shape based on peaks and ridges in the difference of low-pass transform.
IEEE Trans. on PAMI, March, 1984.
- [Crowley 84b] J.L. Crowley, R.M. Stern
Fast computation of the Difference Of Low Pass transform.
IEEE Trans. on PAMI, March, 1984.
- [Crowley 87] J.L. Crowley, D. Aubert
Using Scale-Space Extrema for Depth from Motion and Stereo.
Workshop in Computational Vision Multi-Resolution of Images, DIKU, Copenhagen, November 23-26, 1987
- [Deriche 87] R. Deriche
Optimal edge detection using recursive filtering
Proceedings first international conference on computer vision, London, 8-12 June 1987
- [Dreschler 83] L.S. Dreschler and H.H. Nagel
On the selection of critical points and local curvature extrema of region boundaries for interframe matching.
In image sequence processing and dynamic scene analysis, T.S.HUANG, ed. Berlin, West Germany: spring-verlag, 1983
- [Duda 73] Duda, R.O. and P.E. Hart
Picture processing and scene analysis
Wiley, New York, 1973.
- [Fennema 79] C.L. Fennema, W. Thompson,
Velocity determination in scenes containing several moving objects,
Computer vision graphics and image processing 9, pp 301-315, 1979.
- [Gambotto 85] J.P. Gambotto, O.Monga
A parallel and hierarchical algorithm for region growing.
Conference CVPR, San Francisco, June, 1985.
- [Giraudon 85] G. Giraudon
Edge detection from local negative maximum of second derivative,
in Proc. of Computer Vision and Pattern Recognition,
San Francisco, June 1985
- [Giraudon 86a] G. Giraudon
An efficient edge chaining algorithm
Internal Report INRIA, 1986.
- [Giraudon 86b] G. Giraudon
utilisation des extrema locaux de la dérivée seconde comme primitives stéréo,
2eme colloque IMAGE, Nice, avril 1986.
- [Grimson 81] W.E.L. Grimson,
From images to surfaces: A computational study of human
early vision system,
Cambridge, MA: M.I.T. Press, 1981.

- [Grimson 85] W.E.L. Grimson,
computational experiments with a feature based stereo algorithm,
IEEE Trans. on PAMI, vol. PAMI-7, n°1, january 1985.
- [Hannah 84] M.J. Hannah,
Description of SRI's baseline stereo system.
Technical report 342, SRI International, October 1984.
- [Hebert 83] M. Hebert,
Reconnaissance de formes tridimensionnelles,
Thèse, "université de Paris-sud", Paris, Septembre 1983.
- [Herault 87] L. Herault
Calcul et utilisation de la fonction de courbure pour la détection
de propriétés intrinsèques d'une courbe image en vision par ordinateur,
Rapport de DEA, ENSIEG, INPG.
- [Horaud 87] R. Horaud, T. Skordas
Stereo correspondance through feature grouping and maximal
cliques. Rapport de Recherche LIFIA , Sept 1987.
- [Horn 81] B.K.P. Horn, B.G. Schunck,
Determining optical flow,
Artificial intelligence, Vol. 17, 1981.
- [Horowitz 74] S.Horowitz, T. Pavlidis
picture segmentation by a directed split-and-merge procedure,
proc. IJCPR-II, pp.424-433, 1974
- [Huertas 86] A. Huertas, G. Medioni
detection of intensity changes with subpixel accuracy
using laplacien-gaussian masks.
IEEE Transactions on pattern analysis and machine intelligence,
Vol. PAMI-8, NO. 5, September 1986
- [Jacquot 88] J.L Jacquot, T. Court, D. David, A.Pirson
Une classe d'extracteurs de contours performants adaptés à une implantation
matérielle fonctionnant à cadence video.
Communication personnelle.
- [Jain 81] R. Jain
Dynamic scenes analysis using pixel-based processes,
IEEE, computer 14(8), august 1981.
- [Kass 86] M. Kass,
Computing visual correspondance,
From pixels to predicates, chapter 4, Ed. Alex P. Pentland, 1986.
- [Levine 83] M.D. Levine, P.B. Noble, Y.M. Youssef,
A rule-based system for characterizing blood cell motion,
Image sequence analysis, ed. T.S. Huang, pp 663-709, 1983
- [Long 86] P. Long,
Données 3D obtenues par mise en correspondance stereo
de primitives mixtes lignes-régions,
2eme colloque IMAGE, Nice, Avril 1986.

- [Lux 85] A. Lux
Algorithme et contrôle en vision par ordinateur.
Thèse d'état, LIFIA, Grenoble, 1985.
- [Maffres 86] M. Maffres,
CODEVI: une architecture adaptée à la vision par ordinateur,
Thèse, INPG, Grenoble, Juin 1986.
- [Marr 76a] D. Marr
Analyzing natural images: a computational theory of texture vision.
Cold Spring Harbor Symp. Quant. Biol. 40, 647-662
- [Marr 76b] D. Marr
Early processing of visual information.
Phil. Trans. R. Soc. B (Lond.) 275, 483-524.
- [Marr 79] D. Marr, T. Poggio
A computational theory of human stereo vision.
Roy. Soc. London, vol. B 204, pp. 301-328, 1979
- [Marr 80] D. Marr, E. Hildreth
Theory of edge detection,
In Proc. Roy. Soc. London, Vol. B207, PP. 187-217, 1980.
- [Marr 82] D. Marr
Vision
Freeman, San Francisco, 1982
- [Medioni 84] G. Medioni, R. Nevatia,
Matching images using linear features,
IEEE Trans. on PAMI, Vol. PAMI-6, No. 6, November 1984.
- [Monga 87] O. Mong
A new segmentation method and its application to stereo vision,
Proc. of SPIE's, Technical Symposium on Optics,
Electro-optics and Sensors, 17-22 May, Orlando, 1987.
- [Moravec 77] H.P. Moravec
Towards automatic visual obstacle avoidance,
5th IJCAI, Cambridge, August 1977.
- [Moravec 81] H.P. Moravec,
Rover visual obstacle avoidance
IJCAI 81, PP. 785-790
- [Negre 83] R. Negre, R. Samy, R. Joly,
Traitement d'images spatio-temporelles et son application
à la poursuite par une caméra vidéo.
Neuvième colloque sur le traitement du signal et ses applications,
Nice, Mai 1983.
- [Nilsson 80] N. J. Nilsson,
Principles of artificial intelligence,
Tioga Publishing Company, 1980.

- [Price 82] K. E. Price,
Symbolic matching of images and scene models,
IEEE, Proc. of the workshop on Computer Vision :
Representation and Control, New Hampshire, 1982.
- [Ramparany 88] F. Ramparany,
Un cadre pour la perception multisensorielle en robotique,
Thèse à paraître, LIFIA, INPG, 1988.
- [Rosenfeld 76] A. Rosenfeld
Scene labeling by relaxation operations
IEEE Trans. on Systems, Man, and Cybernetics,
Vol. SMC-6, No. 6, June 1976.
- [Rosenfeld 82] A. Rosenfel, A. C. Kak,
Digital Picture Processing,
Academic Press, Vol. 2, 1982.
- [Rosenfeld 84] A. Rosenfeld
Multiresolution Image Processing and Analysis,
Springer-Verlag, Berlin.
- [Rueb 87] K.D. Rueb, A.K.C. Wong
Analysis of point feature representation of a perspective image.
IEEE, Int. Conf on Robotics and Automation, Raleigh, 1987.
- [Saint Vincent 86] A.R. de Saint Vincent
Perception et modélisation de l'environnement d'un robot mobile:
Une approche par stéréovision.
Thèse, "Université Paul Sabatier", Toulouse, 1986.
- [Skordas 88] T. Skordas, R. Horaud,
Mise en correspondance de structures 2-D dans une paire d'images
stéréoscopique,
Actes du second atelier scientifique TIPI, Aussois, Avril 1988.
- [Souvignier 83] V. Souvignier,
PVV. Un système d'interprétation d'images par prédiction et vérification,
Thèse, "LIFIA-INPG", Grenoble, Juin 1983.
- [Stelmaszyk 85] P. Stelmaszyk, P. Bonnet, J.G. Postaire
Analyse de scènes dynamiques par recherche des contours en mouvement,
5eme Congrès de Reconnaissance de formes et d'Intelligence Artificielle,
AFCET, 27-29 Novembre 1985.
- [Stelmaszyk 88a] P. Stelmaszyk, J.G. Postaire, P. Bonnet
Analyse de séquences d'images par détection des contours en mouvement.
Application à l'assistance de la conduite d'un tramway.
R.A.I.R.O. APII, Vol. 22, No 2, 1988.
- [Stelmaszyk 88b] P. Stelmaszyk, C. Discours, A. Chehikian,
A fast and reliable token tracker,
IAPR Workshop on computer vision, Tokyo, Japan, october 1988.
- [Stelmaszyk 88c] P. Stelmaszyk, D. Aubert, J. Crowley,
Mise en correspondance structurelle en vision multirésolution,
Second atelier scientifique TIPI, Aussois, Savoie, avril 1988

- [Williams 81] T. D. Williams
Computer interpretation of a dynamic image from a moving vehicle.
Ph.D, University of Massachussetts at Amherst, 1981
- [Wrobel 86a] B. Wrobel, O. Monga
segmentation d'images naturelles: coopération entre un détecteur-contour
et un détecteur-région.
11eme colloque GRETSI, Nice, 1987.
- [Wrobel 86b] B. Wrobel
Stéréovision : Mise en correspondance symbolique de régions.
Rapport CRIN 86-R-112, 1986
- [Xu 87] G. Xu, S. Tsuji, M. Asada,
A motion stereo method based on coarse-to-fine control strategy,
IEEE Trans. on PAMI, Vol. PAMI-9, No. 2, March 1987.
- [Yachida 83] M. Yachida,
Determining velocity maps by spatio-temporal neighborhoods
from image sequence,
Computer vision graphics and image processing 21,
pp 262-279, february 1983.

Annexe 1

Calcul du nombre de niveau dans la pyramide

Le dernier niveau de résolution calculable est tel que la taille de l'image est approximativement égale à celle du filtre. Or le filtre à un niveau k de résolution a une taille $S^k.X_0$ (où X_0 est la taille du filtre initiale). Par conséquent, le niveau maximal $k_{\max}$ est tel que $S^{k_{\max}}.X_0 = N$ (où N est le nombre de pixels de l'image).

$$\text{Ainsi } S^{k_{\max}} = N/X_0$$

$$\text{Donc } k_{\max} = \text{Log}(N/X_0)/\text{Log}S = \text{Log}_S(N/X_0).$$

Annexe 2

Coût en nombre de multiplications de l'algorithme de la transformée de DOLP

- Calcul de $b_{\max}$, nombre maximal d'itérations:

Il y a arrêt de l'algorithme lorsque la taille du filtre est approximativement égale à celle de l'image. Soit X_0 la taille du filtre g_1 et N le nombre de pixels de l'image initiale. La taille de l'image à la $b_{\max}$ itérations est $N/2^{b_{\max}}$ (en division entière). Ainsi, le nombre maximal d'itérations ($b_{\max}$) de l'algorithme est tel que :

$$\frac{N}{2^{b_{\max}}} \approx X_0$$

$$\text{Donc } b_{\max} = \frac{\text{Log}\left(\frac{N}{X_0}\right)}{\text{Log}(2)} = \text{Log}_2\left(\frac{N}{X_0}\right)$$

- Coût du calcul d'un couple de passe-bandes (B_{2i}, B_{2i+1})

Soit le filtre g de taille F . Celui-ci étant séparable, la convolution entre g et une image de N pixels se décompose en une convolution ligne et une convolution colonne, chacune coûtant FN multiplications. D'où un coût total de $2FN$. Or le filtre g_1 utilisé se décompose en trois convolutions du filtre g . Par conséquent, le coût est de $6FN$ multiplications.

L'obtention de la passe-bande B_1 coûte donc $6FN = (18FN)/3$.

L'obtention du couple (B_2, B_3) coûte $6FN + 2.6FN = 18FN$ (la multiplication par 2 du coût étant introduite par la double convolution par g_1)

L'obtention d'un couple (B_{2i}, B_{2i+1}) (quelque soit $i = 2 \dots b_{\max}+1$) coûte $6F.N/2^{i-1} + 12F.N/2^{i-1} = 18F.N/2^{i-1}$.

- Coût en nombre de multiplications de l'algorithme de la transformée de DOLP:

L'obtention de toutes les passe-bandes coûte :

$$\begin{aligned} & 18FN\left(\frac{1}{3} + 1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{b_{\max}}}\right) \\ &= 18FN\left(\frac{1}{3} + \sum_{j=0}^{b_{\max}} \frac{1}{2^j}\right) \\ &= 18FN\left(\frac{1}{3} + \frac{1 - \left(\frac{1}{2}\right)^{b_{\max}+1}}{1 - \frac{1}{2}}\right) = 18FN\left(\frac{7}{3} - \left(\frac{1}{2}\right)^{b_{\max}}\right) \end{aligned}$$

Or dans le cas de Dolp $F = 3$,

Donc le nombre total de multiplications est $54N(7/3 - (1/2)^{b_{\max}})$

Annexe 3

L'algorithme de recherche de cliques maximales

L'algorithme présenté est extrait de [Bolles 79]. Il permet d'énumérer toutes les cliques maximales d'un graphe, c'est à dire tous les sous graphes complets.

L'appel de la procédure nécessite trois paramètres qui sont : C, P, PuS. Ces derniers, sont des ensembles de noeuds du graphe.

C représente une clique qu'il s'agit d'étendre en une clique maximale.

P représente l'ensemble des candidats potentiels permettant d'étendre la clique. Ces éléments peuvent ne pas être connectés entre eux, mais ils sont obligatoirement directement connectés à tous les éléments de C.

PuS représente l'union entre l'ensemble P et un ensemble S englobant des éléments permettant d'étendre C mais exclus arbitrairement à un instant donné du déroulement de l'algorithme.

L'algorithme est le suivant :

maxcliques (C, P, PuS)

BEGIN

SetElements x, y;

IF PuS = EmptySet THEN List C as a Maximal Clique

ELSE

BEGIN

x <- Choose (PuS);

ForEach y in $(P \cap \overline{\text{neighbors}(x)})$ DO

BEGIN

P <- P - {y};

maxcliques ($C \cup \{y\}$, $P \cap \text{neighbors}(y)$, $\text{PuS} \cap \text{neighbors}(y)$);

END

END

END

L'appel de `neighbors(y)` retourne l'ensemble des noeuds directement connectés à `y`. Par convention, `y` n'est pas considéré comme son propre voisin. Afin de rechercher toutes les cliques maximales du graphe, l'appel de la procédure doit être : `maxcliques ({}, <l'ensemble de tous les noeuds>, <l'ensemble de tous les noeuds>)`.

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

- . Serge CASTAN , Professeur
- . Jack G. POSTAIRE , Professeur

Monsieur AUBERT Didier

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 12 janvier 1989



Pour le Président de l'IN.P.-G.
et par délégation
le Vice-Président
P. VENNEREAU

Thèse de Monsieur Didier AUBERT

"Mise en correspondance d'indices images en résolutions multiples"

Résumé :

Cette thèse traite de la mise en correspondance d'indices images en vue d'une reconnaissance d'objets ou d'une reconstruction tridimensionnelle ultérieure. Après une synthèse des techniques existantes portant sur la représentation de l'image et sur l'appariement, nous développons nos propres choix. Grâce à des filtres Laplaciens, on obtient une pyramide multirésolution de Laplacien de l'image initiale. Dans chacune des images de cette pyramide, on extrait les pics du Laplacien. Ceux-ci sont alors soit liés à travers les résolutions pour former des "arbres de pics", soit liés au sein de chaque image par des relations topologiques formant ainsi des graphes complets.

Notre première technique d'appariement est basée sur les indices "arbres de pics". Ces indices sont mise en correspondance grâce à un critère de similitude. Les limitations de cette technique nous conduisent à envisager une autre méthode d'appariement basée sur les graphes complets. La recherche de la plus grande clique maximale dans un graphe associé, permet d'obtenir les appariements solutions. Cependant son coût exorbitant, nous conduit à introduire un partitionnement efficace des images de la pyramide basé sur une technique d'héritage dans un schéma de type "coarse to fine". Cette méthode fournit de bons résultats en un coût raisonnable.

Mots clés : Pics du Laplacien, Multi-résolution, relations topologiques, appariement, héritage, cliques maximales.