Numéro d'identification

ACADEMIE DE MONTPELLIER

UNIVERSITE MONTPELLIER II

SCIENCES ET TECHNIQUES DU LANGUEDOC

THESE

présentée à l'Université de Montpellier II Sciences et Techniques du Languedoc
pour obtenir le diplôme de DOCTORAT

Spécialité : INFORMATIQUE

Formation Doctorale : Informatique

Ecole Doctorale : Sciences pour l'Ingénieur

ANALYSE ET REPRÉSENTATION DE SCÈNES COMPLEXES PAR GROUPEMENT PERCEPTUEL

Application à la perception de structures curvilignes.

par

Laurent ALQUIER

Soutenue le 30 septembre 1998 devant le Jury composé de :

M. HABIB Michel	Professeur, LIRMM Montpellier	Président
Mme ZERUBIA Josiane	Directeur de Recherche, INRIA Sophia Antipolis	Rapporteur
M. JOLION Jean-Michel	Professeur, INSA Lyon	Rapporteur
M. FIORIO Christophe	Maître de conférence, LIRMM Montpellier	Examinateur
Mme. PHILIPP Sylvie	Professeur, ENSEA Paris	Examinateur
M. OUSSALAH Chabane	Professeur, LGI2P Nîmes	Directeur de Thèse
M. MONTESINOS Philippe	Maître assistant, LGI2P Nîmes	Responsable de Thèse

A Lori, pour son infinie patience.
A Linfa, pour avoir quitté l'une de ses neuf vies si tôt.

Je tiens à exprimer ma reconnaissance aux personnes qui ont contribué, de près ou de loin , à l'accomplissement de ce travail. En particulier, je voudrais remercier tout d'abord ceux qui m'ont fait l'honneur de participer au jury de cette thèse.

Monsieur Mourad Oussalah, Professeur à l'Ecole des Mines d'Alès, pour la confiance qu'il m'a accordée en acceptant de diriger ce travail, et Madame Janine Magnier, Directeur du Laboratoire de Génie Industriel et d'Ingénierie de Production pour m'avoir accueilli dans son laboratoire.

Monsieur Philippe Montesinos, Maître assistant au LGI2P, pour son accueil au sein de l'équipe Vision du laboratoire. Il a su par ses conseils, sa disponibilité et les fructueuses discussions que nous avons eues ensemble, me communiquer sa passion pour la vision par ordinateur.

Madame Josiane Zerubia, Directeur de Recherche à l'INRIA et Monsieur Jean-Michel Jolion, Professeur à l'INSA de Lyon, pour le temps accordé à juger ce travail et les conseils précieux qu'ils m'ont donnés pour l'améliorer.

Monsieur Michel Habib, Professeur à l'Université Montpellier II, Monsieur Christophe Fiorio, Maître de Conférence au LIRMM, et Madame Sylvie Philipp, Professeur à l'ENSEA de Paris, pour l'intérêt qu'ils ont porté à mon travail en acceptant de faire partie du jury.

Toute ma sympathie pour les membres du laboratoire, enseignants-chercheurs et doctorants, ainsi que l'ensemble du personnel du site EERIE pour la qualité de leur accueil. Merci, notamment, à Sylvie, Gérard, Mireille, Michel, Vincent, et les autres...

Les membres de l'équipe Vision, les nouveaux comme les anciens, pour tout ce que j'ai pu apprendre à leur contact. En particulier, Nasser Armande pour ses contributions indirectes à ce travail, Houda Chabbi pour ses précieux conseils de rédaction et Pierre Soille pour de nombreuses discussions stimulantes.

Mes deux "cyber-mamans" : Françoise, pour ses encouragements en période de stress, pour avoir subi la relecture des premières versions de ce mémoire et surtout, pour les mercredi matin de la bib'. Annie pour sa confiance, sa bonne humeur contagieuse et son soutien pendant les longs séjours au Centre de Calcul. Vive le F.A.A.L!

Les piliers du Centre de Calcul : François, pour son aide technique inestimable, ses conseils de programmation et sa disponibilité, devant une machine comme autour d'une bonne table. Nicolas et Fabien pour les longues soirées de travail à la lueur des néons.

Sylvain et Christelle pour leur profonde amitié, les nombreuses pauses café et les longues discussions à refaire le monde (en particulier celui des D'ni). Sans oublier Zoé, voyageuse nocturne, grâce à qui j'ai pu rédiger cette thèse sur mon propre ordinateur.

Mes parents et mes proches pour leurs encouragements, en particulier Fabien pour avoir pris le temps de pousser son grand frère à aller voir ailleurs que ses bouquins et l'écran de sa machine.

Enfin Lori, pour m'avoir attendu depuis si longtemps.

Sommaire

Introduction

"On aura compris qu'il n'y a pas d'un côté l'Image, matériau unique, inerte et stable, et de l'autre le Regard, comme un rayon de soleil mobile qui viendrait animer la page d'un livre grand ouvert. Regarder n'est pas recevoir mais ordonner le visible, organiser l'expérience. L'image tire son sens du regard, comme l'écrit de la lecture, et ce sens n'est pas spéculatif mais pratique."

"Vie et mort de l'image" - Régis Debray [Debray, 1992]

Figure 0.1 - *Marché d'esclaves avec buste invisible de Voltaire - Salvador Dali*
Marché d'esclaves avec buste invisible de Voltaire - 1940 - Salvador Dali Museum, San Petersburg, Floride

La perception des objets d'une scène, en particulier leur forme, est l'un des problèmes fondamentaux de la vision par ordinateur. L'acquisition d'images monoculaires reste la méthode la plus économique pour percevoir une scène. Mais elle est également source de nombreux problèmes mal posés, liés aux ambiguïtés créées par la projection de la scène sur l'image. En d'autres termes, une image seule ne contient pas assez d'informations pour une interprétation complète et sans ambiguïté de la scène représentée.

Pourtant, la vision humaine démontre l'existence de mécanismes robustes pour la perception de structures en trois dimensions, même lorsque la scène est perçue à partir d'une simple image. Le système visuel humain présente en particulier de nombreux mécanismes dont le rôle est de guider en permanence la perception dans un flot continu d'informations visuelles. Un certain nombre de ces mécanismes ont été appliqués à la vision par ordinateur. Inspirés de la vision naturelle, l'utilisation de vues multiples, de focus d'attention ou encore de rétine artificielle à résolution polaire ont montré leur intérêt pour résoudre certaines ambiguïtés de la recherche visuelle. D'autre part, des théories issues de la psychologie de la perception, fournissent des contributions importantes. C'est le cas, en particulier, du groupement perceptuel.

Le groupement perceptuel désigne la faculté de la vision humaine à organiser certains éléments visuels en groupes significatifs. Ce phénomène se produit de façon spontanée, avant toute interprétation du contenu de la scène observée. Des expériences psycho-visuelles simples montrent clairement l'importance de certaines relations, telles que la proximité, la fermeture, la continuité ou la symétrie, lors de ces groupements.

L'organisation perceptuelle a, depuis une vingtaine d'années, fait l'objet de nombreuses recherches en vision par ordinateur. Ce problème combinatoire complexe est le plus souvent exprimé comme le groupement d'attributs détectés dans l'image selon des structures régulières plus complexes. Contrairement aux méthodes classiques d'analyse d'images, ce type d'approche met l'accent sur des critères qualitatifs et génériques, au lieu de rechercher des modèles mathématiques précis. Finalement, ces groupements sont utilisés pour initialiser plus facilement des traitements de haut niveau tels que la reconnaissance de formes ou bien la mise en correspondance d'images.

Au cours de notre travail, nous nous sommes intéressés au rôle que peut jouer le groupement perceptuel dans la perception des contours des objets d'une scène. Plus précisément, l'objectif de cette thèse est d'extraire, à partir d'une détection de contours, les éléments caractéristiques des principales structures curvilignes de la scène. Il s'agit donc, dans un premier temps, de définir des critères de qualité permettant de mettre en valeur les contours les plus saillants de l'image. Ces contours dominants servent de point de départ à l'extraction des éléments caractéristiques de la scène.

Nous proposons une approche hiérarchique pour extraire les éléments visuels les plus importants et fournir un ensemble d'hypothèses fortes pour des processus de haut niveau d'interprétation.

Cette méthode procède en trois niveaux d'organisation. Un niveau de *groupements élémentaires* organise d'abord les éléments de contours en chaînes perceptuellement importantes. Le rôle de cette étape est de réduire la complexité des niveaux suivants en estimant de manière robuste l'importance structurelle des éléments de contours. Le niveau suivant recherche des *groupements intermédiaires* tels que des segments, des courbes simples ou des points d'intérêt. Ces éléments de représentation sont extraits à partir des chaînes groupées précédemment. Finalement, ceux-ci sont organisés selon des *groupements de haut niveau* dans une application de mise en correspondance structurelle de jonctions entre deux images.

Nos principales contributions sont l'apport d'un nouveau formalisme générique pour la construction et l'optimisation de réseaux d'éléments visuels localement connectés, l'utilisation de ce type de réseau pour extraire les contours les plus réguliers, une méthodologie de groupement hiérarchique de ces éléments visuels en hypothèses de structures plus complexes, et enfin, l'utilisation de ces structures pour la mise en correspondance de jonctions entre deux images.

Ce mémoire de thèse est divisé en deux parties. La première, théorique, est consacrée à la perception visuelle et au rôle que joue le groupement perceptuel en vision par ordinateur. Cette partie couvre l'ensemble des problèmes, théoriques et pratiques, que posent l'utilisation des contours pour l'analyse d'image.

- *Chapitre 1 : Vision naturelle et artificielle*

  Nous posons dans ce chapitre la problématique de la vision par ordinateur. Afin de situer le contexte de notre travail, nous présentons les principales méthodologies appliquées en vision artificielle, en commençant par les différentes théories de la perception visuelle dont elles sont issues.

- *Chapitre 2 : Interprétation des scènes de contours*

  Après avoir défini le contexte de notre travail, nous abordons dans ce chapitre la question des différentes représentations de l'environnement visuel. Le choix des contours comme élément de base de représentation est ensuite justifié. La problématique de la détection et la modélisation des contours, ainsi que leur rôle en vision par ordinateur sont abordés en détail dans ce chapitre.

- *Chapitre 3 : Groupement perceptuel en vision par ordinateur*

  Ce chapitre est consacré à la place du groupement perceptuel en vision par ordinateur. En particulier, ce chapitre passe en revue les principes d'organisation perceptuelle issus de la psychologie de la vision et la manière dont ces principes ont été appliqués à la vision par ordinateur. Enfin, en guise de conclusion à cette première partie, nous exposons les principes de notre approche en regard de ces problèmes et des travaux antérieurs.

La seconde partie de ce mémoire concerne le travail effectivement réalisé au cours de cette thèse. Elle est divisée en trois chapitres, correspondant aux trois niveaux d'organisation perceptuelle de notre méthode.

– *Chapitre 4 : Saillance structurelle et groupements élémentaires*

  Dans ce chapitre, nous abordons le premier niveau de groupement de notre approche. Il s'applique dès la détection de contours. Sa fonction est de faire ressortir les structures linéaires importantes présentes dans l'image de contours et de produire un ensemble de chaînes, ou groupements, correspondant à ces structures. Ces groupements servent de point de départ à l'extraction d'éléments de représentation réalisée aux niveaux suivants, en réduisant ainsi la complexité de la recherche visuelle aux seules structures d'intérêt de l'image.

– *Chapitre 5 : Eléments de représentation et groupements intermédiaires*

  Nous montrons dans ce chapitre comment extraire de ces groupements élémentaires, un ensemble d'hypothèses géométriques utiles aux niveaux supérieurs de traitement. Dans un premier temps, nous présentons les principes de notre méthode de structuration hiérarchique, issus de l'analyse des groupements élémentaires. Cette méthode est ensuite détaillée pour chaque type d'hypothèse géométrique envisagée et illustrée par une application à différentes scènes réelles.

– *Chapitre 6 : Mise en correspondance structurelle*

  Les éléments visuels extraits à partir des groupements précédents peuvent être comparés à un croquis sommaire des formes saillantes de la scène. Nous montrons dans ce chapitre comment établir des relations structurelles plus complexes à partir de ces éléments visuels afin de mettre en correspondance des structures issues de deux scènes. Nous illustrons enfin ce dernier niveau de groupement par une méthode de mise en correspondance de jonctions.

En conclusion, nous récapitulons enfin les principales contributions de notre approche, les améliorations à apporter et les perspectives ouvertes par cette thèse.

Première partie

Perception de structures curvilignes en vision par ordinateur

Chapitre 1

Vision naturelle et artificielle

Nous posons dans ce chapitre la problématique de la vision par ordinateur. Afin de situer le contexte de notre travail, nous présentons les principales méthodologies appliquées en vision artificielle, en commençant par les différentes théories de la perception visuelle dont elles sont issues.

1.1 Problématique de la vision par ordinateur

La vision par ordinateur concerne l'aspect algorithmique de la perception visuelle, depuis l'acquisition à l'interprétation d'images. Elle partage le domaine de la "vision artificielle" (par opposition à "vision biologique") avec la vision industrielle, plus préoccupée par la conception matérielle de systèmes visuels. C'est, dans le domaine de la perception visuelle, la discipline la plus récente du fait de l'évolution de la puissance de calcul des ordinateurs.

Ses buts fondamentaux sont de comprendre les mécanismes de la vision afin de construire des systèmes de vision artificielle dont les performances sont au moins semblables à celles de la vision humaine. Les domaines de compétence qu'elle couvre vont du traitement du signal à l'intelligence artificielle en passant par l'imagerie (traitement des images).

Pourtant, on est encore très loin de comprendre la vision animale, encore moins la vision humaine. L'idée d'une définition de vision "générale" reste encore inaccessible. En l'absence de définition valable, on pourrait dire que la vision générale est celle dont nous faisons l'expérience, la nôtre. En général, c'est la vision humaine qui sert de référence, de modèle pour évaluer les performances d'un système visuel. Mais alors, elle constitue une référence nécessairement limitée de par ses capacités. La vision humaine est adaptative et rapide, mais elle n'est pas pour autant complètement générale. Contrairement à certains animaux, elle est peu efficace dans l'obscurité ou sous l'eau. Elle ne perçoit qu'un nombre limité de fréquences lumineuses, et ne permet pas de donner une mesure précise des volumes ou des couleurs. Elle ne permet même pas de distinguer une image de son reflet dans un miroir. Il faut garder ces remarques à l'esprit pour éviter des querelles inutiles pour de simples

questions de définitions. On pourrait préférer au terme de "vision générale" le terme de "vision naturelle" [Tarr et Black, 1994b] . Fabriquer un système visuel artificiel pose un certain nombre de questions fondamentales dont la plupart n'ont pas encore été tranchées.

Nous venons de le rappeler, la vision humaine, malgré son efficacité, est tout de même limitée. Un système visuel artificiel doit il alors se comporter comme la vision humaine ? Ne doit il pas plutôt apporter des facultés différentes de la vision humaine ? La vision humaine sert de référence dans la plupart des situations car nous en faisons l'expérience en permanence. On peut toujours faire l'objection qu'il est inutile d'essayer de comprendre d'autres systèmes visuels tant que nous ne maîtrisons pas complètement le fonctionnement du notre. Pourtant, des systèmes visuels radicalement différents tels que celui des insectes ou des animaux pourraient non seulement apporter des éléments de solutions à certains de nos problèmes, mais ils pourraient aussi être riches en enseignements sur notre propre système. En effet, malgré leurs différences apparentes, existe-t-il des mécanismes communs à la vision stéréoscopique humaine, la vision en "facettes" des abeilles ou celle, globale, des caméléons ?

Cette question en appelle d'autres, plus précises, sur les mécanismes de la vision naturelle. Peut-on définir des modules distincts ? (vision, raisonnement, apprentissage, mémoire, interprétation, décision, action) ou bien ces modules font-ils partie d'un système complexe ? La vision est-elle dissociée de l'action ? ou en d'autres termes, "voir" a-t-il un sens indépendant de toute autre tâche ? ou encore plus généralement, la vision est-elle un processus actif ? ou bien passif ? Comme nous le verrons plus loin, la recherche d'une réponse à ces questions a donné lieu à des approches différentes du phénomène de la vision [Aloimonos, 1994] .

Pour en revenir au problème de la vision par ordinateur, nous devons considérer la vision d'un point de vue de traitement de l'information. Les questions qui se posent alors concernent la nature de ces informations et leur représentation vis à vis de la machine. Quelle sorte d'information extraire depuis les images ? Après tout, l'expérience visuelle est faite de textures, couleurs, formes, mouvements, perçus tous à la fois par le même moyen. Une modélisation de la vision doit tenir compte de tous ces aspects.

Pour citer Y. Aloimonos :

"Le problème fondamental d'un système visuel est de déterminer quelle information doit être utilisée dans l'image, et quelle doit être la représentation la plus adaptée pour que la relation entre le système et son environnement soit la plus efficace possible."

Est-ce que cette information doit être décrite sous la forme d'un langage ? Doit-elle être assez générique pour être manipulée par un grand nombre de tâches ? ou doit-elle être spécialisée et adaptée à chaque type de tâche ? Quelle sorte de description produire à partir d'une scène ? un modèle 3D ? une description symbolique ? [Ramesh, 1994] [Brown, 1994] .

Toutes ces questions nécessitent d'élargir le sujet en apportant des précisions sur les différentes façons d'aborder la Vision ainsi que les principales théories qui ont permis d'éclairer certains aspects de la perception visuelle utiles en vision par ordinateur.

1.2 Théories de la perception visuelle

L'étude du problème de la vision a donné lieu à de nombreuses approches, séparées pour des raisons historiques et pour différentes interprétations du fonctionnement de la vision naturelle, qu'elle soit humaine ou bien animale. Le domaine de la "vision" concerne des chercheurs aussi variés que des psychologues, des biologistes, des neuro-biologistes, des ingénieurs, des informaticiens ou mathématiciens.

En laissant de côté l'aspect philosophique, nous pouvons dégager trois familles d'approches [Trivedi et Rosenfeld, 1989] :

- *L'approche psycho-visuelle*, la plus ancienne car attachée aux aspects psychologiques de la perception visuelle.
- *L'approche analytique*, qui cherche à comprendre comment fonctionnent les mécanismes sensoriels et neuronaux de la vision à un niveau biologique. C'est le cas, en particulier, des théories neuro-physiologiques.
- *L'approche calculatoire*, qui traite des problèmes algorithmiques de l'acquisition, du traitement et de l'interprétation des informations visuelles. Il s'agit des théories engendrées par la vision par ordinateur.

Les différentes approches de la vision par ordinateur doivent beaucoup aux théories développées depuis plus d'un siècle par les neuro-physiologistes et les psychologues. Elles subissent également les influences millénaires de courants de pensées scientifiques et philosophiques. Donner une vue d'ensemble de ces différentes théories est donc nécessaire afin de replacer la vision par ordinateur dans son contexte.

1.2.1 Comment définir la vision ?

Etant notre sens le plus développé, la Vision a entretenu l'intérêt de générations de philosophes, en particulier sur le rapport qu'elle tisse entre la perception et le réel. Voir a été longtemps synonyme de Connaissance. Pour Aristote, "voir" signifie connaître l'emplacement des choses par l'intermédiaire de la vue. Pourtant, dans son "Discours de la Méthode", Descartes accorde du crédit *"aux objets visibles mais à condition de les construire avec ordre et mesure, et de bien poser ses équations."* C'est la séparation entre l'esprit et la matière. Dans ce contexte, si notre esprit est essentiellement différent de la matière, alors nous ne pouvons pas connaître le monde matériel directement. Nous ne percevons le monde qu'au travers de sensations qui

nous servent de représentations. Bien avant lui, Platon tenait un discours encore plus extrême. Dans "l'Allégorie de la caverne", il fait la distinction entre les Idées, objets d'intelligence, et leur matérialisation, les objets terrestres que nous percevons, et qui n'en sont que les ombres. Cette distance entre ce que nous percevons et ce qui "est" suggère une existence indépendante de toute perception.

Enfin, sans aller trop loin dans les débats philosophiques sur les rapports entre la perception visuelle et la réalité, l'importance de la vision a aussi son revers en se changeant en dépendance. Interpréter ce que nous voyons nécessite de plus en plus de précautions tant il est facile de manipuler ou de trop dépendre des images. C'est particulièrement vrai lorsqu'il s'agit de scènes inaccessibles à nos sens. Depuis les images de la jeunesse de l'univers transmises par le télescope spatial Hubble, jusqu'à la visualisation de structures à l'échelle atomique, les exemples ne manquent pas.

Ainsi se dessinent deux conceptions de la perception visuelle. L'une suppose une perception du monde directement tel qu'il est alors que pour l'autre, le monde n'est perçu qu'au travers de reconstructions mentales. On retrouvera ces deux aspects jusque dans les différentes approches de la vision par ordinateur.

D'un point de vue évolutionniste, la vision est notre sens le plus utile. Elle est nécessaire à l'accomplissement de tâches essentielles à notre survie : reconnaître des partenaires, amis ou ennemis, identifier de la nourriture, s'orienter, éviter le danger [Aloimonos et Rosenfeld, 1992] . Comme les autres sens, elle permet à l'espèce de survivre et d'évoluer. Mais ce n'est pas sa seule fonction. L'intérêt de la vision dans le domaine artistique n'a pas grand chose à voir avec la survie et pourtant, elle nous permet de développer des capacités d'interprétation et d'association aigües. Le problème étant de trouver le dénominateur commun à toutes ces tâches, s'il existe.

En considérant la vision d'un point de vue purement calculatoire, on pourrait être tenté de la présenter de la manière suivante :

"La vision est un processus qui, à partir d'images d'un environnement extérieur à l'observateur, produit une description utile et dépouillée d'informations superflues." [Marr, 1982]

ou bien :

" La vision, en tant que processus d'intelligence artificielle, peut être considérée comme un problème indépendant, qui fournit un ensemble de données symboliques à des niveaux supérieurs d'interprétation." [Brooks, 1987]

Selon cette conception, la vision fournirait alors une description seule, indépendamment de toute interprétation. L'idée est séduisante car elle permettrait de diviser le problème de la perception visuelle en tâches indépendantes, en une hiérarchie de sous problèmes. Mais c'est oublier bien vite le côté dynamique de la vision. L'oeil est bien plus qu'une caméra figée. Et l'environnement qui nous entoure bien plus qu'une image fixe.

Des expériences très simples dans lesquelles des sujets sont placés dans un environnement visuel uniforme révèlent des comportements intéressants. En l'absence de

tout point de repère, l'observateur ne se contente pas de fixer un point indéterminé, mais au contraire parcours inlassablement le champ visuel à la recherche d'un point de référence. Il en résulte un effet de désorientation et une perception de nuances de tons ou de couleurs là où il n'y en a pas [Zakia, 1997] . Le système visuel aurait donc besoin de la présence d'objets dans le champ visuel pour fonctionner. Ce qui amène à la remarque suivante :

"(...) il devrait être acquis que la perception n'est pas passive mais active. L'activité perceptuelle est exploratoire, elle parcours, elle recherche ; ce qui est perçu ne tombe pas seulement sur les capteurs comme la pluie sur le sol. Nous ne voyons pas, nous regardons." [Bajcsy, 1988]

ou encore, d'une façon plus élaborée :

" La Vision est un processus de reconnaissance : Elle est Associative (association de vues ou de propriétés avec des concepts et des représentations), Interpretative (cherche à répondre à des questions spécifiques à propos de l'environnement), Dirigée (chaque comportement oriente vers un certain type de calculs), et Sélective (les informations inappropriées pour la tâche en cours sont rejetées)." [Aloimonos, 1994]

La vision est définie ainsi comme une démarche active afin d'accomplir un certain but. Nous devons alors nous poser la question de savoir si la vision par ordinateur doit se contenter de "voir" ou bien si elle doit "regarder".

Ces différentes définitions suggèrent deux conceptions de la vision. L'une, reconstructive, considère que le rôle de la vision est de fournir une représentation de l'environnement à d'autres niveaux cognitifs. L'autre considère la vision comme partie prenante d'un système complexe, inextricablement liée à l'intention et à l'action. Cette dichotomie se retrouve sous des formes diverses dans les théories de la vision et surtout, en ce qui nous concerne, dans les différentes approches de la vision par ordinateur. Cependant, on peut voir apparaître depuis quelques années des approches nouvelles, liant la robustesse des méthodes reconstructives à la souplesse des approches dynamiques.

1.2.2 Psychologie de la perception visuelle

L'étude de la vision d'un point de vue psychologique est la façon la plus naturelle d'aborder le problème. Nous en faisons continuellement l'expérience, d'une façon si familière qu'il est très facile de sous-estimer la complexité de cette tâche. Il est donc nécessaire de comprendre comment fonctionne la vision à un niveau psychologique pour ne pas être leurrés par notre propre expérience de la vision.

1.2.2.1 Approche psycho-physique et concept de seuil

Ce premier aspect de l'étude de la vision sur un plan psychologique n'est pas exactement une théorie mais plutôt un ensemble de techniques permettant de me-

surer la réponse d'observateurs à des stimuli visuels. C'est la première approche a avoir étudié la vision de façon rigoureuse, en proposant des données fiables et des méthodes d'expérimentation contrôlées [Gordon, 1989] .

Les méthodes psycho-physiques sont à l'origine des théories de Seuils sensoriels. En résumé, mesurer et expliquer un seuil sensoriel consiste à étudier les relations entre un stimulus et la réponse, ou sensation, à ce stimulus. En d'autres termes, il s'agit de déterminer à partir de quelle intensité un stimulus est perçu comme tel.

La démarche psycho-physique peut être globalement définie comme l'étude des relations entre la force d'un stimulus et la force de la sensation perçue en conséquence de ce stimulus. Une procédure psycho-physique répond en particulier aux problèmes de détection (comment mesurer si un sujet est sensible ou non à un stimulus ?), d'identification (comment mesurer l'aptitude du sujet à exprimer une réponse ?), de discrimination (comment évaluer son aptitude à faire la différence entre plusieurs stimuli ?), et enfin, des problèmes d'échelle (comment mesurer la proportion de stimulus réellement perçue par le sujet ?).

En montrant que l'expérimentation est un moyen approprié pour étudier la perception et que l'utilisation de stimuli bien définis dans un environnement contrôlé peut amener à des découvertes remarquables sur le fonctionnement de la perception, les méthodes psycho-physiques ont apporté une contribution essentielle à l'étude de la perception visuelle.

Le principal reproche fait à cette approche est de n'étudier qu'un aspect particulier de la vision à la fois. Les stimuli utilisés sont nécessairement simples et ne prennent donc pas en compte la complexité de l'environnement visuel tel qu'il est couramment perçu. En effet, la vision naturelle doit rarement répondre à un seul stimulus à la fois.

1.2.2.2 Théorie du Gestalt

Parmi les plus anciennes théories sur la psychologie de la vision, la théorie du Gestalt est l'une des plus célèbres de part l'ampleur des phénomènes qu'elle a mis en évidence, dont un grand nombre constituent encore aujourd'hui des problèmes non résolus. Sa contribution est telle que son influence et sa popularité sont toujours d'actualité, malgré les lacunes de sa forme d'origine.

La théorie du Gestalt met l'accent sur le besoin d'étudier le comportement du cerveau dans ses relations avec le monde au travers d'expériences de tous les jours. Son domaine d'application n'est pas limité à la perception visuelle. Les "Gestaltistes" ont expérimenté sur les processus de la pensée (Wertheimer, 1920), la résolution de problèmes (Duncker, 1945), la mémoire et l'apprentissage (Katona, 1940). Elle représente plutôt un système général d'étude de phénomènes psychologiques.

La théorie du Gestalt liée à la perception repose sur un certain nombre de principes mis en évidence par des phénomènes simples. Le premier de ces principes a donné son nom à la théorie, et part de la constatation que dans de nombreux cas, un groupe de stimuli acquiert une qualité supérieure à la somme des qualités des

parties. Par exemple, une mélodie est quelque chose de plus qu'une simple succession de notes, ou bien un carré a quelque chose de plus qu'un simple arrangement de lignes. Cette qualité propre au "tout" a été baptisée en allemand *Gestaltqualität* - qualité de la forme.

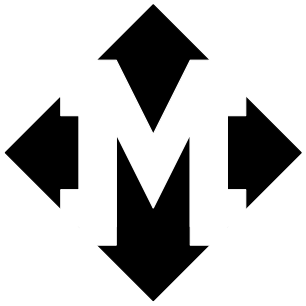

Figure 1.1 - *Illustration de la* Gestaltqualität - *cet arrangement de flèches noires représente quelque chose de plus que la somme de quatre flèches.*

Les autres principes, sont autant d'illustrations de la perception en tant que processus dynamique et organisé : séparation entre objets et fond, groupements perceptuels selon des règles simples (continuité, proximité, orientation, symétrie), principe de simplicité et de cohérence (résumé par le terme allemand d'origine - *Prägnanz*), invariants des formes et des couleurs. En montrant que la perception visuelle obéit dans une certaine mesure à ces principes, la théorie du Gestalt était en faveur d'une origine innée de ces mécanismes (approche "nativiste").

Le chapitre 3 abordera chacun de ces principes d'une façon plus détaillée. Nous nous limiterons pour l'instant à une présentation sommaire de la théorie du Gestalt pour la placer dans le contexte des théories de la perception visuelle.

La théorie du Gestalt a dû faire face à de nombreuses critiques. Leur principal outil d'explication, le modèle de cerveau de Köhler, rendait compte des phénomènes observés sous la forme d'hypothétiques processus de résonance dans le cerveau. Ces processus n'ont encore jamais été démontrés. Leurs arguments ont été critiqués comme étant souvent circulaires : les descriptions des phénomènes qu'ils ont mis en évidence servent souvent d'explications de ces mêmes phénomènes. C'est une chose de montrer que la vision réagit selon des lois de groupements, expliquer pourquoi elle réagit ainsi en est une autre.

En particulier, la théorie du Gestalt n'a pu répondre à des questions telles que : Pourquoi la vision est-elle dynamique ? Quelle est l'origine de ce degré d'organisation que nous pouvons observer ? Comment en prévoir le comportement en présence d'une situation inattendue ? En d'autres termes, il s'agit plus d'une théorie descriptive que prédictive.

Pourtant, son influence sur les autres théories psychologique de la perception est considérable. L'une de ses grandes forces a été d'insister sur le côté phénomènolo-

gique de la vision à partir d'expériences psycho-visuelles et de rester claire sur une question simple : qu'est-ce qu'une théorie de la perception doit essayer d'expliquer ?

1.2.2.3 Fonctionnalisme probabiliste de Brunswik

Le Fonctionnalisme probabiliste est essentiellement le travail d'Egon Brunswik (1903-1955). Cette théorie n'a pas eu le même succès que la théorie du Gestalt, et pourtant, elle a anticipé de nombreux développements contemporains de l'étude de la perception. C'est la première théorie a avoir souligné la nature probabilisite de la perception, l'importance de l'environnement de l'observateur et le rôle de l'évolution dans le développement de la perception visuelle [Gordon, 1989] .

Brunswik fut l'un des premiers à envisager que la recherche en perception devrait refléter la complexité des phénomènes observés. Jusque là, la tendance dans le domaine était de simplifier les mécanismes, de faire des parallèles avec d'autres disciplines, d'utiliser des sujets entraînés à certaines réponses sur un ensemble de stimuli extrêmement artificiel. Ce rejet catégorique des méthodes classiques de psychophysique, méthodes pourtant à l'origine de nombreux résultats en perception, est l'une des principales faiblesses de cette théorie.

Préoccupé par l'influence de l'évolution sur les mécanismes du système visuel, Brunswick propose de prendre en compte les millions d'années d'évolution nécessaires à l'élaboration de tels systèmes. Cette approche de la perception suggère un système visuel guidé par des impératifs de survie. Il ne peut être compris hors de sa relation avec l'environnement de l'observateur. Les mécanismes d'un tel système n'existent alors que parce qu'ils ont une fonction liée à l'évolution.

Enfin, il accorde une importance particulière à la nature incertaine du monde perçu par l'observateur. Non seulement les signaux émis par l'environnement sont incomplets ou brouillés, mais ils ont une nature statistique ou probabiliste. De ce point de vue, ce qu'on perçoit du monde n'est pas seulement incomplet mais plutôt incertain. Percevoir revient à faire la meilleure évaluation sur des signaux pondérés en fonction de précédents échecs ou succès.

Ses travaux sur les groupements par proximité et sur la perception des visages, apportent une nouvelle signification aux phénomènes révélés par Wertheimer : les groupements ont une valeur fonctionnelle. L'action de grouper est utile car elle permet en général de délimiter le contour d'objets. C'est l'une des principales contributions de cette théorie.

Malgré des travaux souvent jugés trop confus, peu convaincants et qui ont contribué à son insuccès, les idées suggérées par Brunswick réapparaissent depuis peu grâce aux avancées en calculs statistiques, qui proposent des méthodes plus efficaces que celles dont il disposait, et surtout, au retour de l'idée selon laquelle l'observateur participe de manière active à la vision. Cette idée d'interaction entre l'observateur et son environnement est particulièrement présente dans les travaux de J.J.Gibson sur la perception immédiate (cf. page 18) et plus récemment, en vision par ordinateur (vision intentionnelle [Aloimonos, 1990]).

1.2.2.4 Empirisme ou paradigme constructioniste

S'il fallait désigner une théorie dominante en perception visuelle ce serait certainement celle de l'Empirisme. Ses principes clairs et la force des expériences menées pour y apporter des arguments en ont fait la théorie la plus populaire et celle qui a eu le plus de succès en perception. Parmi les contributeurs importants à cette théorie, on peut trouver Helmholtz (1821-1894), Ames (1949), Bruner (1951) et plus récemment R. L. Gregory (1974).

La théorie de l'Empirisme part de l'idée que la perception visuelle est quelque chose de plus qu'une simple analyse de stimuli visuels. Cette idée suggère qu'un phénomène intermédiaire de construction intervient entre la stimulation et l'expérience. Elle décrit la perception en tant que processus de construction, capable de déductions qui vont au delà de ce qui est seulement perçu par les sens. Cette approche de la perception visuelle souligne l'importance de l'expérience et des associations d'idées et s'oppose en cela au "nativisme" des Gestaltistes, qui supposent les mécanismes de groupements innés. La perception n'est plus un simple signal d'entrée mais un processus de sélection qui intervient entre l'image rétinienne et l'interprétation.

Des expériences psychologiques relativement simples sur l'attention visuelle ont montré que ce qui est perçu par un observateur subit l'influence du contexte, des idées reçues, des stéréotypes. Les sources d'influence vont même jusqu'à l'état physique ou la faim du sujet. Au cours de ces expériences, des sujets à qui sont montrés des images brèves, ne retiennent qu'une fraction de ces images d'une façon sélective.

Dans les années 40, d'autres expériences sur la résolution de problèmes et l'attention menées par J. S. Bruner conduisirent à la théorie suivante : l'observateur perçoit le monde avec une série d'hypothèses, d'attentes qu'il confronte à ce que ses sens lui fournissent effectivement. Une hypothèse forte nécessite un faisceau de preuves important pour être contredite et autorise à l'inverse une certaine tolérance. En d'autres termes, l'observateur est acteur dans la perception, faisant des hypothèses sur le monde et les vérifiant.

La forme moderne de cette théorie est représentée par les travaux du psychologue anglais R.L.Gregory [Gregory, 1974] . Selon cette théorie, les signaux reçus par les récepteurs sensoriels activent des évènements neuronaux. Ces évènements interagissent avec la connaissance et la mémoire pour fabriquer un ensemble de données à partir desquelles des hypothèses sont faites sur l'environnement. C'est cette chaîne d'évènements que nous appelons "perception".

Un certain nombre d'arguments semblent confirmer cette théorie:

- La perception peut, dans certains cas familiers, anticiper sur nos actions. Lors d'expériences sur le suivi de cible à l'aide d'un pointeur manuel, les sujets opèrent remarquablement bien lorsque les mouvements du pointeur sont réguliers et prévisibles.

- La perception est ambigüe. L'un des exemples les plus classiques est celui du cube de Necker. C'est une figure instable, pour laquelle deux interprétations

coexistent. Cet exemple est un indice fort en faveur d'une sorte de déduction "visuelle". Si la perception était exclusivement liée aux stimuli, un même signal ne pourrait produire deux interprétations.

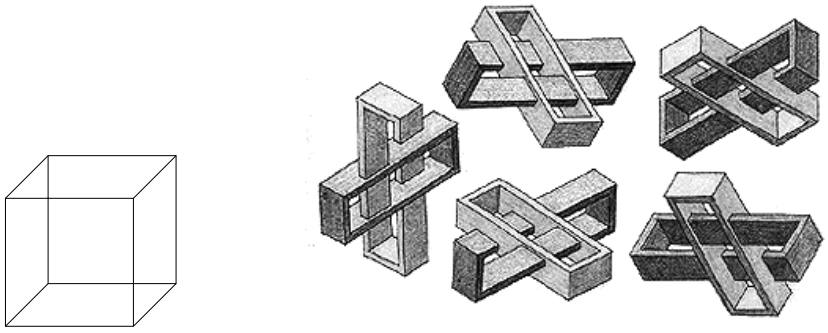

Figure 1.2 - *Cube de Necker et figures impossibles de Oscar Reutersvärd*

– La perception peut être paradoxale. Les objets impossibles de la figure 1.2 sont perçues comme un tout cohérent . Ce phénomène suggère une perception séquentielle des objets avec déduction d'une forme plus globale. En effet, ces figures impossibles sont constituées de parties cohérentes assemblées d'une manière incohérente.

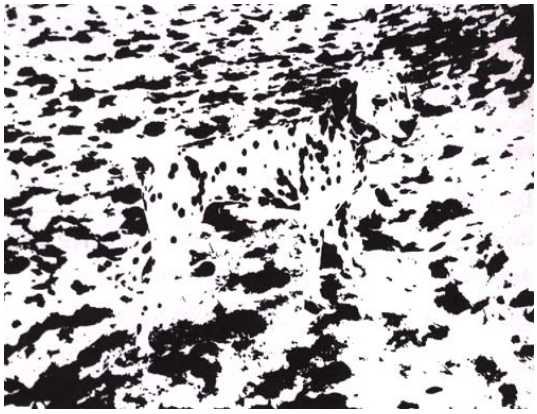

Figure 1.3 - *Dalmatien - Exemple de séparation d'un objet familier avec un arrière plan complexe.*

– La perception montre une capacité particulière à séparer des objets familiers d'un arrière plan complexe. C'est valable pour la perception visuelle comme la perception auditive. Il a été démontré comment un sujet peut séparer une voix familière du bruit de fond d'une foule. La connaissance joue un rôle actif dans la sélection des signaux perçus.

– Des objets improbables tendent à être pris pour des objets probables. L'une des démonstrations de Gregory les plus connues utilise un visage sculpté en creux. Correctement illuminé, le visage est interprété comme sculpté en bosses.

Figure 1.4 - *Influence de stimuli inconscients - Photo de Richard Gicewicz*

– La perception peut être influencée par des stimuli inconscients. Par exemple, lorsqu'on montre à des sujets deux photos apparemment identiques d'un même visage, la plupart choisissent celle pour laquelle les pupilles sont les plus dilatées. Une majorité se montrent incapables de différencier les deux photos, et lorsqu'on demande aux autres s'ils ont remarqué le changement de pupilles, la réponse est négative.

– Enfin, la perception autorise des défauts selon le contexte. L'un des exemple les plus frappants est celui d'un visage pour lequel les yeux et la bouche ont été localement inversés. Confrontés avec l'image d'origine, une majorité de sujets les déclarent identiques.

En application, cette théorie permet d'éclairer certaines illusions comme celles reproduites sur la figure 1.5. Ces quatre illusions ont toutes comme résultat de faire apparaître des objets identiques avec des tailles différentes. L'illusion est induite dans chaque cas par l'activation de l'hypothèse de profondeur alors que les lignes sont coplanaires. Ainsi, dans l'illusion de Ponzo, les lignes obliques suggèrent un effet de profondeur qui conduit à considérer le segment A plus long que B car apparemment plus éloigné de l'observateur. Ce type d'illusion est un exemple frappant de la flexibilité des constructions du système visuel, ainsi que de sa vulnérabilité.

Cependant, malgré ses succès, cette théorie laisse encore dans l'ombre un certain nombre de questions importantes.

Par exemple, rien n'est encore dit sur la nature des hypothèses utilisées par le système visuel pour la perception. De même pour les preuves utilisées par la perception pour confirmer ou infirmer une hypothèse. Leur origine, tout comme leur évolution sont encore indéterminées.

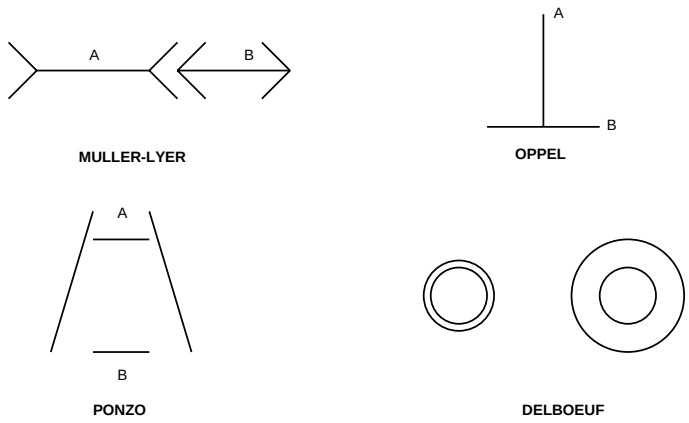

Figure 1.5 - *Illusions géométriques - Dans chaque cas, les segments A et B ont la même longueur. De même, les cercles intérieurs ont le même rayon.*

Les sensations visuelles sont-elles si pauvres en information qu'elles doivent être recomposées en permanence par la mémoire et le raisonnement ? Les images perçues par la rétine contiennent rarement des informations sur un seul objet à la fois, mais plutôt des projections d'objets multiples, des occlusions ou des points de fuite. Autant d'informations riches et immédiatement disponibles. De la même manière, le mouvement et une vision stéréoscopique apportent des indices importants sur l'environnement.

La vision est-elle constamment un processus de construction ? Procède-t-elle hiérarchiquement, à la manière d'un ordinateur ? Si c'est le cas, comment les nouveaux nés acquierent-ils une construction du monde ? A quoi confrontent-ils leurs hypothèses ? Il est probable que nous ayons aussi des mécanismes de perception directe. Dans ce cas, à partir de quel point la perception cesse-t-elle d'être directe pour devenir constructive ? Les processus de construction pourraient n'intervenir que lorsque nous sortons des conditions usuelles de perception [1] .

Ces questions montrent combien l'approche empiriste de la vision est encore activement débattue. Avec la neuro-physiologie, c'est la théorie qui a le plus influencé la vision par ordinateur, en particulier avec le paradigme de Marr (cf. section 1.3.1). L'utilisation de l'ordinateur comme modèle de cerveau est une évolution naturelle dans l'histoire de cette théorie.

1.2.2.5 Perception immédiate

La théorie de la Perception Immédiate, aussi appelée Optique Environnementale (en anglais, *Ecological Optics*), est l'une des théories les plus récentes dans le domaine de la perception visuelle. Elle a vu le jour en grande partie grâce aux travaux

1. Cf. le sous chapitre 1.2.2.5 consacré à la vision immédiate de J. J. Gibson.

du psychologue américain J.J.Gibson (1904-1979) [Gibson, 1950] [Gibson, 1979] . Rejetant à l'origine toute démarche exclusivement empiriste, cette approche est toujours en évolution depuis.

Les question qu'elle pose sont les suivantes : la perception du monde est-elle toujours indirecte ? Est-ce que nous ne "voyons" qu'une représentation du monde ou bien le monde en lui-même ? Le but de cette remise en question est de déterminer comment les organismes vivants sont conscients d'un monde essentiellement neutre.

A partir d'observations sur la nature de la lumière, la relation entre l'observateur et son environnement et le rôle des invariants en perception, Gibson et ses successeurs aboutissent à une conception nouvelle de la perception, en rupture avec la conception constructionniste classique. Leur théorie repose sur trois concepts importants:

- L'environnement contient tous les éléments nécessaires à l'action. Les informations contenues dans le flux optique sous forme d'invariants sont suffisamment élaborées pour permettre des décisions. La lumière elle-même est extrêmement riche et structurée (comme le montrent les hologrammes, il est possible de percevoir un objet en relief à partir d'une image de cet objet sous un "éclairage" laser). L'information existe donc déjà à l'extérieur de l'observateur et celui-ci n'a pas besoin de représentation interne pour l'utiliser.

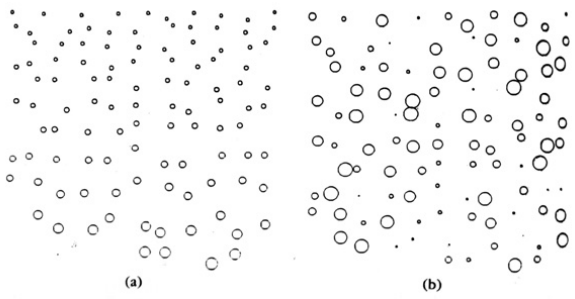

Figure 1.6 - *La présence d'un gradient de texture est un indice immédiatement utilisable concernant l'orientation de la surface. L'exemple (a) semble s'éloigner alors que (b) semble parallèle à l'observateur.*

- L'un des concepts les plus importants de cette théorie est celui des "invariants". Nous ne percevons pas le monde d'une façon aléatoire ou chaotique mais par un flot continu d'images en corrélation permanente. C'est ce qui nous permet de dire qu'un objet ne rétrécit pas réellement lorsqu'il s'éloigne de nous, ou que deux objets plus ou moins éloignés ont en fait la même taille (alors qu'ils ont une taille différente sur l'image).

- Le dernier concept important est plus difficile à appréhender. Dans une démarche qui fait écho aux arguments de Brunswick, Gibson accorde une certaine importance à l'influence de la fonction des objets que nous percevons sur notre

propre perception. D'où l'idée d'un potentiel, d'une capacité, que représente chaque partie de notre environnement (concept désigné en anglais par *affordance*, littéralement "capable de"). Les objets qui nous entourent guident notre perception sur ce qui est possible ou non de faire.

L'idée directrice de cette théorie est celle de la perception immédiate. Les informations utiles à l'action sont extraites directement de l'image. Par comparaison, l'approche empiriste classique suppose une reconstruction interne de la perception visuelle sur laquelle interviennent les processus d'analyse et de reconnaissance. Selon Gibson, le système perceptuel ne répond pas à des stimuli mais extrait en permanence des invariants à partir d'un flot continu. L'analyse et la reconnaissance devraient donc pouvoir agir directement à partir des informations visuelles, sans faire intervenir de représentation intermédiaire.

L'un des mérites de cette approche récente est d'avoir souligné l'importance de l'étude de l'environnement et la richesse des signaux reçus par un observateur actif. Elle représente une tentative de lutter contre la distinction entre l'organisme et son environnement, entre ce qui se passe à l'extérieur et à l'intérieur de l'observateur.

Mais en tant que théorie récente et nouvelle, elle a aussi ses faiblesses. L'une des plus importantes est une tendance à sous-estimer la difficulté pour le système visuel que pose le problème de la détection d'invariants (difficulté qui n'a pas cessé d'être démontrée avec la vision par ordinateur). Les concepts centraux d'invariants et de "potentiel" semblent vagues et difficiles à appréhender. Comment détecter ces invariants ? comment les prédire ? Cette théorie n'y apporte pas de réponse précise pour l'instant.

1.2.3 Théories neuro-physiologistes

L'aspect psychologique de la vision n'est pas la seule manière d'aborder le problème de la perception visuelle. Les théories neuro-physiologiques constituent l'approche biologique de ce problème.

En partant du principe que le cerveau fonctionne à partir d'échanges entre neurones, cette approche s'attache à comprendre le fonctionnement biologique des mécanismes de la perception visuelle. Le but est d'arriver à expliquer des comportements par des mécanismes neuronaux. Au lieu de tenter d'expliquer la richesse de la perception dans son ensemble, les neuro-physiologues se contentent d'essayer d'expliquer comment des informations simples sont perçues, codées et sous quelle formes elles sont manipulées par le cerveau.

La question posée est celle-ci : est-il possible de faire abstraction du côté psychologique de la perception pour aller directement aux phénomènes physiques et anatomiques ?

Il existe un lien étroit entre la psychologie et la neuro-physiologie. Les découvertes sur le système nerveux influencent les travaux de psychologues. A l'inverse, les travaux des neuro-physiologues tentent d'expliquer les découvertes des psychologues.

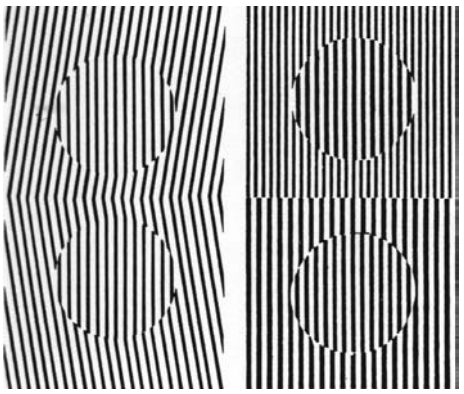

Figure 1.7 - *Exemple de modification apparente de la fréquence spatiale par le système visuel. Dans l'exemple de gauche, les disques ont la même orientation verticale - l'orientation de l'arrière plan semble dévier les disques dans le sens opposé. Dans le cas de droite, les disques ont la même fréquence spatiale - celui du haut parait pourtant avoir une fréquence spatiale plus élevée.*

Ces travaux dépendent d'une façon toute aussi étroite des avancées en anatomie du système nerveux.

Ces théories ont apporté des contributions importantes au domaine de la perception visuelle. Elles ont révélé la présence de détecteurs de caractéristiques spécifiques à un niveau neuronal[2]. Elles ont montré l'existence de capacités d'analyse et de synthèse du système visuel à un niveau neuronal[3]. Enfin, elles ont mis en évidence des correspondances directes entre les mécanismes psychologiques et physiologiques (pour la perception de la couleur par exemple). Ces contributions ont montré les avantages de mettre en commun les résultats de plusieurs disciplines comme la Psychologie, la Biologie, et Neurologie.

Mais ce parti pris de donner un aspect matériel, biologique, à la recherche sur la vision, loin de termes psychologiques abstraits a tout de même ses limites. Les interactions entre les différentes parties du système nerveux sont d'une complexité telle qu'on ne pourra en rendre compte encore longtemps qu'en utilisant des termes psychologiques. Comment exprimer en termes neuro-physiologiques des phénomènes tels que ceux mis en évidence par les Gestaltistes ? Il est tout aussi difficile de tenir compte du côté subjectif de la vision en termes uniquement biologiques. De plus, la neuro-physiologie suppose une relation directe de cause à effet entre une stimulation de la rétine et une réponse perceptuelle, ce qui est mis en doute par l'aspect probabiliste de la vision avancé par Brunswick, mais aussi par les expériences

2. Voir à ce sujet le concept des Champs Réceptifs *Receptive fields*, des cartes rétinotropiques et des détecteurs de configurations (taille, couleur, orientation) [Hubel et Weisel, 1962]

3. Un exemple de ces capacités est donné par la réponse du système visuel à des variations de fréquences spatiales [Sekuler et Blake, 1985] .

d'attention visuelle de l'école Empiriste.

"La Neuro-physiologie a permis d'établir de nombreuses cartes rétinotropiques dans le cortex visuel et dans diverses zones corticales. Ces zones et leurs connexions sont en permanence ramifiées vers d'autres zones mais suggèrent néanmoins l'idée de parties du cerveau dédiées à certaines tâches, ce qui pourrait être un indice d'un comportement de reconstruction de la part du cerveau.

Pourtant, les théories psycho-physiques démontrent des comportements inattendus de la part du cerveau. Même si les conditions expérimentales sont par définition exceptionnelles, ces démonstrations jètent un doute sur une approche reconstructive de la vision." [Brown, 1994]

Enfin la neuro-physiologie n'est pas une théorie générale de la perception visuelle. Les travaux de cette discipline apportent une solution satisfaisante à des problèmes bien délimités.

1.3 Méthodologies de la vision par ordinateur

Dans le contexte des théories de la perception visuelle, la vision par ordinateur est la dernière discipline en date. A l'image des théories psychologiques, la vision par ordinateur a progressivement évolué en deux approches principales. La plus ancienne, l'approche de "vision reconstructive", correspond à la conception empiriste de la vision et dérive essentiellement des travaux de David Marr. C'est la première véritable approche d'un point de vue calculatoire de la vision, jusque là consacrée essentiellement à l'analyse d'images. La plus récente est dite de "vision intentionnelle". Elle hérite d'une conception dynamique et active de la vision liée aux travaux de Gibson.

Le but de ce sous-chapitre est de présenter les caractéristiques de ces deux principaux paradigmes afin d'en tirer les points forts et les limites. Nous verrons en fin de ce chapitre dans quelle mesure ces deux approches de la vision participent de plus en plus à une conception commune, plus ouverte aux différents aspects de l'expérience visuelle. On pourra se reporter aux articles de [Tarr et Black, 1994a] et de [Jolion, 1994] pour une analyse comparative détaillée de chaque approche.

1.3.1 Paradigme reconstructif

A l'image de l'approche Empiriste pour les théories de la perception, le paradigme reconstructif, en anglais *Recovery Paradigm*, est l'approche classique de la vision par ordinateur. En quelques mots, elle consiste à tirer du monde visible une représentation symbolique de ses propriétés, à la fois géométriques et physiques, et à exploiter cette représentation pour un certain nombre de tâches de haut niveau : reconnaissance, localisation, déplacement.

Héritière de l'école Empiriste de la perception visuelle, cette approche part du principe que le monde possède une structure, et par conséquent un certain nombre de régularités qui doivent être utiles à sa représentation.

1.3.1.1 Processus descriptif et paradigme de Marr

Au confluent de recherches en intelligence artificielle, théorie de l'information, cybernétique et informatique, David C.Marr propose le premier cadre véritable d'une approche de la vision d'un point de vue calculatoire. Son travail n'est pas seulement fondamental pour la vision par ordinateur, c'est aussi une tentative de clarifier la conception de systèmes de traitement de l'information en général. Selon lui, le but de la vision en tant que système de traitement de l'information est de décrire l'environnement extérieur.

De l'image imprimée sur notre rétine à la conscience que nous avons du monde, l'activité de notre cerveau s'applique à une représentation du monde qui nous est intérieure. Les neurones qui le constituent ne manipulent pas des images mais une représentation symbolique d'une scène élaborée à partir d'images. La vision est rapportée au problème de la construction d'une telle représentation.

Pour accomplir cette tâche, Marr propose une méthodologie pour l'analyse de tout processus, y compris celui de la vision. Trois niveaux distincts sont ainsi définis, chacun apportant ses propres interrogations quant à la conception du système.

- *Niveau calculatoire*

  Les questions posées à ce niveau concernent le but du système. Quel est l'objectif de la méthode ? Pourquoi est-elle appropriée ? De quelle façon est-elle accomplie ? Quels sont les tâches accomplies et leur rôles dans le processus ?

  En prenant comme exemple la perception de contours, ces questions pourraient être les suivantes. En quoi est-il important de percevoir des contours ? Si le système visuel peut percevoir des contours, quelle est l'utilité de cette information pour l'observateur ? Pourquoi le système visuel devrait-il les rendre explicites ?

- *Représentation et algorithme*

  Ce niveau s'adresse à la modélisation du système proprement dit. Quelles sont les représentations nécessaires pour les entrées et les sorties ? Par quel algorithme passer des unes aux autres ? Quelle est la structure de la représentation ?

  Le point de départ dans le cas de contours est bien sûr l'image rétinienne de l'environnement, c'est à dire, une distribution d'intensités lumineuses. La sortie du système devrait être une représentation symbolique des contours présents dans la scène. Il reste encore à déterminer comment passer de l'un à l'autre - c'est la fonction de ce niveau.

– *Implémentation physique*

  C'est le niveau matériel du système. Comment implémenter chaque niveau de description ? Quel type de capteurs ? Quel langage pour les algorithmes ? Quel type de support pour la représentation finale ?

Cette méthodologie apporte des contraintes au problème de la vision par ordinateur. Marr suggère que ces contraintes devraient être extraites des propriétés du monde visible : *"décrire la géométrie des surfaces visibles à partir des propriétés observables de l'illumination, la texture, les contours de ces surfaces."* [Marr, 1982]

Dans ce contexte, la vision fonctionne comme un système de traitement de l'information, et en tant que tel, il est nécessaire de la diviser en niveaux de traitements. Marr adopte une approche modulaire pour simplifier ce problème complexe en une hiérarchie de problèmes plus simples. En effet, des expériences telles que stéréogrammes de Bela Julesz [Julesz, 1960] démontrent qu'il est possible de percevoir certains aspects d'une scène, en l'occurence, la différence de profondeur (stéréodisparité), indépendamment de toute autre information visuelle.

Cette découverte ouvre la voie à une fragmentation du problème de la vision en différents modules de perception. Marr propose alors une décomposition de la perception visuelle en quatre niveaux de traitement :

– *L'image*

  C'est le niveau le plus bas : l'image rétinienne. Sa fonction est de représenter la distribution de l'intensité lumineuse sur la rétine.

– *Ebauche primaire - "Primal sketch"*

  La fonction de ce niveau est de rendre explicite, à partir des intensités lumineuses, des informations géométriques sur la façon dont elles sont organisées. La possibilité de détecter des surfaces commence à ce niveau.

– *Ebauche intermédiaire - "*$2\frac{1}{2} - D$ *sketch"*

  Ce niveau rend explicite l'orientation de ces surface et fournit une estimation de leur profondeur. Cette représentation intermédiaire est encore liée à l'observateur et n'a pas encore de caractère global.

– *Représentation 3-D*

  Ce niveau donne une représentation tridimensionnelle des objets indépendante de l'observateur. Cette représentation est un modèle symbolique de la scène telle qu'elle est perçue.

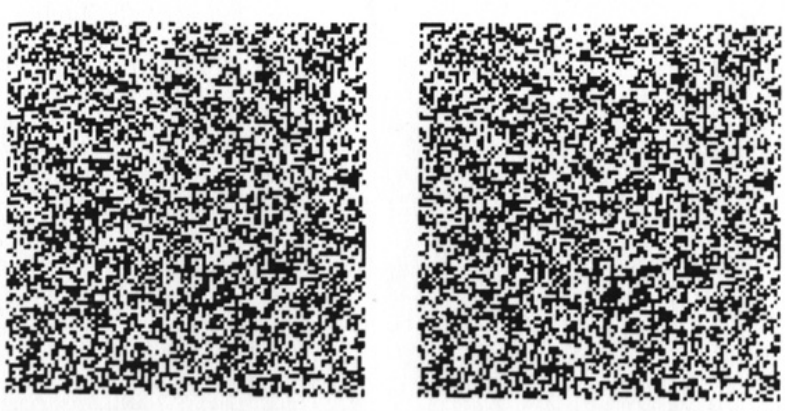

Figure 1.8 - *Deux stéréo-grammes de points aléatoires. La superposition des deux images, à l'aide d'un stéréoscope par exemple, fait apparaitre la forme d'un carré dont la profondeur est différente du reste de l'image.*

1.3.1.2 Représentation et reconstruction

"Reconstruction" et "représentation" sont des concepts clés de cette théorie, mais sont souvent confondus. D'après la propre définition de Marr :

" Une représentation est un système formel qui rend explicite certaines entités ou types d'information, tout en spécifiant comment y parvenir. Et j'appellerais le résultat de l'usage d'une représentation pour une entité donnée, une description de cette entité selon cette représentation."

La reconstruction d'une scène désigne en général l'élaboration une réplique géométrique de la scène observée. C'est un problème particulièrement difficile lorsqu'il est appliqué à des données qu'on ne contrôle pas (scènes d'extérieur, éclairage variable, robotique mobile). La réplique est une abstraction qui doit forcément éliminer des informations pour rendre certaines connaissances explicites. Ce genre de modèle est donc forcément simplifié - l'environnement visuel est trop complexe pour être reconstitué dans ses moindres détails. De plus, le choix des simplifications à apporter ne prend son sens qu'en connaissant à l'avance à quelle fin sera utilisée la représentation [Fischler, 1994] .

Ce sont donc deux concepts liés mais pas interchangeables. Reconstruire une scène signifie produire une représentation 3D de la scène. Représenter une scène n'est pas forcément synonyme de la reconstruire - la représentation peut (et même doit dans certaines situations) être symbolique et n'avoir rien à voir avec la représentation d'origine.

Une représentation de scène suffisamment générale pour convenir à tous les usages relève encore du domaine de la théorie (quand ça ne touche pas à la philosophie). Dans la pratique, le modèle dépend étroitement de l'usage auquel il est

destiné. Par exemple, les tâches de reconnaissance et d'évitement d'un obstacle ne nécessitent pas le même modèle (l'une a besoin de caractéristiques précises alors que l'autre peut se contenter de l'information de volume).

"Plus généralement, est-ce qu'un système visuel peut se permettre de construire des modèles génériques (par conséquent, très complexes) et n'en utiliser qu'une partie ?" [Sandini et Grosso, 1994]

Si l'intégralité des informations nécessaires aux tâches visuelles (parcours, évitement, reconnaissance) est bien contenue dans la représentation, reste encore à exploiter ces informations. La représentation tridimensionnelle peut se révéler à ce titre une représentation bien peu efficace. De la même manière, un dictionnaire est une très bonne représentation pour chercher un mot, mais elle n'est pas du tout adaptée à la recherche d'un mot associé à une définition donnée.

La représentation d'une idée ne doit pas forcément ressembler à cette idée. Il lui suffit d'être associée à cette idée de façon fiable, tout comme la lettre 'A' peut-être représentée par '65' dans une table ASCII. Adaptée à la vision par ordinateur, cette notion permet de considérer des représentations directement issues de divers traitements, sans pour cela avoir besoin de reconstituer la représentation d'origine. C'est le cas notamment avec des représentations hiérarchiques (espaces échelles par exemple).

1.3.1.3 Limites et perspectives

Depuis des années, le paradigme reconstructif s'est montré viable pour une majeure partie de la communauté de vision par ordinateur. Pourtant, les développements récents d'autres types d'approches remettent en cause la vision en tant que processus descriptif (voir les sections suivantes). Malgré ces critiques, l'approche reconstructive constitue un cadre de travail solide pour la compréhension de la vision, artificielle et humaine [Tarr et Black, 1994a] .

Lorsqu'on dresse la liste des problèmes soulevés par ces nouvelles approches, il n'est pas surprenant de retrouver les principales objections faites par Gibson contre les théories empiristes de la vision.

- Est-il possible de produire une représentation objective d'une scène?

  Une représentation objective de l'environnement implique qu'elle soit assez générique pour être utilisable par de multiples tâches visuelles (si ce n'est toutes). Nous venons de voir que c'est pour l'instant une idée impossible en pratique. Les possibilités sont si nombreuses qu'une telle représentation serait tout simplement trop complexe pour être modélisée.

  Mais la réponse à cette question n'est pas incompatible avec la méthodologie de Marr. Lors de la conception d'un système, la représentation dépend des réponses apportées au niveau théorique. Une représentation assez générique

pour convenir à tous les usages de la vision devrait être définie en réponse à tous ces usages. Le fait que notre vision soit adaptée à de multiples usages suggère que cette représentation devrait exister, mais elle reste encore hors de portée.

– La scène n'est pas la seule source de contraintes.

  En donnant à l'observateur une représentation interne de son environnement, le paradigme de Marr suit la conception empiriste de la vision. Si l'approche reconstructive sépare, par définition, l'observateur de son environnement, elle n'exclue pas pour autant l'intervention de mécanismes de hauts niveaux pour guider la perception. Marr suggère que la principale source de contraintes provienne des propriétés physiques des surfaces, mais rien dans son approche ne rejette l'influence possible de l'expérience, l'intention ou la connaissance du contexte.

  Comme le soulignent [Ballard et Brown, 1992] dans leur argumentation pour une vision "dirigée" :

  " La complexité des tâches liées à la reconstruction de scènes est exponentielle - complexité qui peut être réduite par l'exploitation de connaissances contextuelles ou d'informations fournies par des tâches dédiées. La réduction de complexité peut être immense lorsque des contrôles de haut niveau sont injectés dans des opérations de niveaux inférieurs. "

  Cette conception de la vision n'est pas incompatible avec une approche reconstructive. Par contre, la situation est moins claire pour des contraintes liées à des tâches spécifiques ou de temps réel. Comment rester efficace avec une représentation générale lorsqu'il s'agit d'accomplir des tâches spécialisées? ne vaudrait-il pas mieux utiliser des représentations différentes et adaptées à chacune des tâches?

– L'approche reconstructive est trop figée.

  En effet, l'approche de Marr laisse peu de place au côté dynamique de la vision. Comment définir à l'avance une représentation qui puisse faire face à l'imprévisibilité du monde? De nombreux aspects de la vision humaine ne sont pas pris en compte : structure polaire de la rétine, exploration visuelle par saccades, apprentissage [Ballard et Brown, 1992] [Brunnström *et al.*, 1996] .

La plupart des reproches faits à l'approche reconstructive ne sont pas purement théoriques. Ils viennent de considérations pratiques sur l'efficacité des systèmes visuels artificiels. Comme nous venons de le souligner, une représentation assez générique pour permettre un traitement efficace à tous les niveaux reste encore hors de portée. En l'absence d'une telle représentation, des travaux se sont tournés vers d'autres pistes. L'approche de "vision intentionnelle" (en anglais *Purposive Vision*) en est une.

1.3.2 Approche "intentionnelle"

Dans un premier temps, des solutions partielles ont été apportées à l'approche reconstructive avec l'utilisation "d'observateurs actifs" (*Active Vision* en anglais) [Aloimonos *et al.*, 1988] . L'activité de l'observateur se limite ici au contrôle de points de vues pour apporter de nouvelles contraintes au problème. Ce terme de "vision active" ne doit pas être confondu avec celui de "vision intentionnelle" développée dans cette section. L'observateur est toujours séparé de l'environnement observé - il se contente de changer de point de vue pour lever des ambiguïtés. A l'inverse, la "vision intentionnelle" considère l'observateur et son environnement comme un seul et même système. Pour résumer, on pourrait dire que la vision active est un premier pas en direction d'une vision intentionnelle.

Cette approche est motivée par un soucis de performance, surtout en temps de réponse. Abandonnant l'idée d'une représentation générale de la vision, elle se concentre sur la résolution de problèmes précis d'une façon la plus efficace et la plus rapide possible. Depuis quelques années, c'est une véritable méthodologie qui a vu le jour, issue de la vision robotique pour une grande partie, mais aussi de la théorie de perception immédiate de Gibson, comme le souligne [Edelman, 1994] :

"Des développements récents en vision par ordinateur et en neuro-sciences ont montré qu'un certain nombre de caractéristiques utiles à des tâches visuelles telles que la localisation d'objets ou la reconnaissance de formes peuvent être extraites de l'image directement, un peu à la manière de la perception immédiate de Gibson."

1.3.2.1 Processus dynamique et interactif

Reprenant les arguments de Gibson sur la perception directe, les partisans de cette approche soulignent le caractère actif de la perception visuelle. La vision est un processus dynamique qui ne calcule que ce qui est nécessaire, par opposition à maintenir une représentation interne complète [Aloimonos, 1990] .

La vision est considérée ici comme un processus flexible, adaptatif. Elle doit avoir une part d'apprentissage, de changements internes pour s'adapter à de nouvelles situations de façon autonome. Elle est enfin considérée comme une démarche volontaire et interactive afin d'accomplir un certain but. C'est le principe majeur de l'approche : prendre en compte dans un même système visuel le but de ce système et l'environnement dans lequel il va évoluer.

Le but n'est plus seulement la représentation, mais un ensemble de tâches simples liées à la vision pour lesquelles sont utilisés des méthodes et représentations appropriées. La vision intentionnelle nécessite des reconstructions partielles, adaptées au problème recherché et par conséquent flexibles au lieu de chercher une reconstruction symbolique et générique. L'introduction de comportements et de buts bien déterminés permet de rendre le problème de vision réalisable dans de nombreuses situations en temps constant (par opposition aux méthodes itératives des approches par reconstruction).

Paradigme "reconstructif"	**Paradigme "intentionnel"**
Reconstruction du monde et de ses propriétés	Reconnaitre des situations utiles à l'accomplissement d'une certaine tâche
Sujets de recherche : Méthodes d'extraction des aspects d'une scène à partir d'images	*Sujets de recherche :* Connaissant une tâche, décomposition du problème en sous-tâches à résoudre séparément. Agencer ensuite ces sous-tâches
Outils : Analyse quantitative, théorie de la régularisation	*Outils :* Analyse qualitative

Table 1.1 - *Comparaison entre les paradigmes "reconstructifs" et "intentionnels", d'après Y. Aloimonos (1990)*

L'approche intentionnelle va encore plus loin dans la prise en compte de l'environnement au sein du système visuel. Dans ce contexte, l'étude de la vision n'est pas concevable sans une connaissance de la façon dont les capteurs fournissent des informations sur la scène observée. Par opposition, Marr place les capteurs dans le niveau d'implémentation physique, indépendamment du reste du système. Le processus de la vision n'est pas considéré indépendamment du reste, mais placé au sein d'un système global entre perception et connaissance. On retrouve ici l'idée d'une interaction complexe entre le monde et l'observateur.

Le but de la vision "intentionnelle" est donc de construire un certain nombre de fonctions visuelles (comportements) en vue d'accomplir certaines tâches visuelles considérées comme essentielles par un agent. La perception et l'action sont liées en boucle. Ces comportements de haut niveau sont ensuite utilisés pour déduire des comportements de bas niveau plus complexe.

1.3.2.2 Vision "dirigée"

L'approche intentionnelle s'inspire des travaux de Gibson dans son opposition à l'idée de représentation purement interne. A l'inverse, elle ne va pas non plus jusqu'à supposer l'existence d'une représentation purement externe du monde, prête à être utilisée directement par l'observateur. La tendance actuelle se tourne plutôt vers une approche interactive entre observateur et monde.

Un cas extrême de la perception active, la vision dirigée (ou *animate vision* en anglais), proposée par Ballard et Brown illustre bien cette démarche de conception d'un système de vision intentionnelle [Ballard et Brown, 1992] . A partir d'observations du fonctionnement de la vision humaine au niveau de son récepteur principal, l'oeil, Ballard et Brown se posent la question suivante : Comment arrive-t-on à une

conception globale et stable du monde avec une exploration du champ visuel si dynamique? En effet, l'oeil parcours son champ de vision par à-coups (saccades). De plus, il montre des particularités qui pourraient servir de modèles pour un système visuel artificiel : par exemple, la rétine présente une meilleure résolution autour de l'axe optique. L'importance des mouvements de l'oeil a enfin été confirmée par des travaux en neuro-physiologie qui ont montré l'existence de structures du cerveau dont l'état d'activation est directement lié aux mouvements de l'oeil.

Ces observations permettent d'avancer une explication possible : une appréhension du monde à partir d'un système aussi dynamique n'est rendu possible que par la capacité du système visuel à accomplir rapidement des tâches particulières. Un système de vision dirigée tel qu'avancé par Ballard et Brown ne peut être efficace que dans la relation qu'il établit entre l'observateur et son environnement. Cette relation est basée avant tout sur une certaine régularité. D'un côté, le monde extérieur obéit à des lois physiques. Il possède une structure qui peut être prédite dans une certaine mesure. Le monde peut être aussi utilisé comme une mémoire externe, interprétée par des mouvements oculaires et modifié par des actionneurs, ce qui permet de se passer de représentation intermédiaire et de gagner en efficacité.

D'un autre côté, l'observateur et son système visuel obéissent aussi à certaines règles. Le système visuel n'est pas complètement général. Il n'a pas non plus un ensemble arbitraire de capacités mais plutôt un jeu réduit de perceptions et d'actions possibles (appelées "instructions"). Pour suivre le modèle de la vision humaine, et garantir un fonctionnement optimal, ces instructions s'appliquent au centre du champ visuel, là où la résolution optique est la meilleure. Cette capacité nécessite un système de contrôle du centre d'attention. Enfin, pour faire face à l'imprévu, le système visuel doit posséder une part d'apprentissage. D'où un ensemble de mécanismes d'évaluation des situations pour déterminer quelle action appliquer à des situations similaires à celles rencontrées lors de l'apprentissage.

De cette démarche, ils tirent trois grands principes de la vision dirigée. Ces principes représentent les facultés les plus importantes que doit remplir le système.

- Les tâches visuelles doivent être simplifiées par une approche séquentielle. On entend par "approche séquentielle" le parcours du champ visuel à la manière de la lecture d'une bande. L'absence de représentation complexe du monde est compensée par l'utilisation d'un jeu de tâches visuelles réduit.

- Le contrôle du point de vue est nécessaire pour placer le centre d'attention aux points d'application de ces tâches.

- Le système doit être capable d'apprendre pour compenser l'aspect imprévisible du monde (au sens de "situations imprévues"). En particulier, il doit pouvoir évaluer une situation rencontrée, la comparer avec des classes de situations déjà rencontrées et autoriser des modifications autonomes des tâches élémentaires pour apporter une action si la situation est jugée similaire.

Ballard et Brown rappellent en particulier que la première tâche d'un système de vision est qu'il doit fonctionner correctement. Un système de vision intentionnelle ne cherche pas la meilleure réponse possible de chacun de ses composants. La qualité de la réponse du système dans son ensemble est plus importante que la qualité de la réponse de chacune de ses parties. La perception est ici convertie immédiatement en actions, et les conflits sont pris en charge par l'architecture du système. Les avantages de cette approche sont certains. En n'utilisant que les représentations les plus adaptées pour chaque comportement, le système est assuré d'une réponse optimale [Sandini et Grosso, 1994] .

Pour faire un parallèle avec l'approche de Marr, cette conception de la vision ne cherche pas l'analyse des images mais leur "compréhension". Les niveaux de représentation ne sont pas étudiés séparément mais d'un point de vue global. Dans ce contexte, les traitements de bas niveaux sur l'image sont remplacés par des mécanismes d'exploration de l'image. C'est la notion de détection "sélective" (*smart sensing*) que P.J. Burt définit comme un *"rassemblement d'informations sélectif, dicté par une tâche, à partir du monde extérieur. Un processus actif dans lequel l'observateur, homme ou bien machine, sonde et explore son environnement visuel à la recherche d'informations."* [Burt, 1988]

Cette exploration de l'environnement visuel se veut un début de réponse pour une segmentation indépendante de toute tâche de haut niveau. Elle suggère de partir d'abord des contraintes matérielles des capteurs (ce qu'il est possible de faire) pour définir ensuite ce qu'il est préférable d'en faire. L'idée est de n'utiliser qu'un nombre limité de techniques génériques. Par exemple : détection d'une zone d'intérêt, réduction de la quantité d'informations, traitement particulier de cette zone.

1.3.2.3 Intérêt et limites

Comme nous venons de le voir, la vision intentionnelle s'oppose à la vision classique sur de nombreux points. La vision humaine est considérée comme une source d'inspiration pour atteindre un haut niveau d'éfficacité, et non comme un modèle. L'approche intentionnelle place l'accent sur la recherche d'un ensemble de techniques génériques, communes à toutes les situations rencontrées. Par opposition, l'approche reconstructive se consacre à la recherche d'une représentation commune aux tâches de haut niveau d'interprétation.

Depuis quelques années, elle a su prouver son intérêt en soulevant des questions importantes apparemment laissées de côté par l'approche classique. Mais en tant qu'approche récente, elle reste encore sujette à controverses [4].

L'un des principaux reproches qui pourraient être faits porte sur l'absence de définition formelle de cette méthodologie. L'une des grandes forces du paradigme de

4. En particulier, on pourra se reporter au débat lancé par [Tarr et Black, 1994a] et aux réponses [Aloimonos, 1994] [Ramesh, 1994] [Brown, 1994] [Edelman, 1994] [Tsotsos, 1994] [Fischler, 1994] [Aggarwal et Martin, 1994] [Christensen et Madsen, 1994] [Sandini et Grosso, 1994] et finalement [Tarr et Black, 1994b]

Marr est la clarté de son approche. Les différentes variantes de la vision intentionnelle portent les noms de vision "active", "dynamique", "comportementale" (*behavorial vision*), "dirigée" (*animate vision*), "intentionnelle" (*purposive vision*). Autant de définitions différentes, aussi difficiles à définir clairement les unes que les autres. Pourtant, elles partent toutes de l'idée de remplacer la recherche d'une représentation unique par la recherche de méthodes génériques [Tsotsos, 1994] .

La définition de ces méthodes est tout aussi difficile. Certains parlent de "tâches visuelles", d'autres de "comportements" ou encore "d'instructions". Comment définir les tâches visuelles? Combien de tâches sont nécessaires à la conception d'un système? comment les choisir? L'agencement de ces tâches pose également des problèmes importants. Est-il préférable de déduire des tâches de bas niveau de perception à partir d'un ensemble de comportements de haut niveau comme c'est le cas en vision comportementale? ou bien doit-on faire émerger des comportements de haut niveau à partir d'instructions élémentaires comme le suggère la vision dirigée? Dans les deux cas, la manière de s'assurer de l'émergence de comportements efficaces reste encore floue.

De nombreuses questions fondamentales, comme par exemple l'architecture de ces systèmes et les relations des différentes tâches entre elles, restent sans réponse. Comment faire face à un conflit entre tâches concurrentes? C'est le cas aussi des mécanismes d'apprentissages. Si les systèmes créés à partir de l'approche intentionnelle ont montré des performances particulièrement efficaces, cela n'a été le cas pour l'instant que pour des problèmes précis. La question de l'apprentissage et de la réaction à des situations totalement imprévues reste encore entière.

1.3.3 Approche "système"

Tout comme il semble impossible de trouver une représentation symbolique suffisamment générique pour tenir compte de la complexité du monde, il semble tout aussi impossible de trouver des comportements suffisamment génériques pour faire face à cette même complexité. Comme on peut le constater, chacune des deux approches, reconstructive et intentionnelle, prise de façon exclusive, pose encore quantités de questions non résolues. Cette constatation conduit de plus en plus à une conception plus consensuelle de la vision par ordinateur, avec l'idée générale de mettre en commun les qualités des deux approches.

Dans la pratique, la vision par ordinateur est un problème mal posé exprimé en général avec insuffisamment de contraintes. La plupart des tâches visuelles sont constituées de problèmes NP Complets tels que la recherche visuelle, l'étiquetage de Waltz, l'interprétation de scène avec occlusions. Chaque approche tente de rendre ce problème soluble par ajout de contraintes.

En vision "reconstructive", les contraintes proviennent de l'environnement, en particulier des propriétés géométriques et physiques des surfaces observées (lissage, continuité, rigidité, persistance temporelle). C'est une approche généralement "ascendante" (*bottom-up*) pour parvenir à une représentation générique de la scène.

A l'inverse, l'approche "intentionnelle" est moins unanime sur la démarche à suivre. Elle est en général "descendante" (*top-down*), guidée par des tâches de haut niveau spécifiques à résoudre (localisation, évitement, déplacement, fixation d'un centre d'intérêt). C'est le cas de la vision comportementale. D'autres variantes suggèrent l'inverse et privilégient les contraintes issues d'une perception directe de l'environnement et de comportements élémentaires. On retrouve cette démarche en vision dynamique et en vision dirigée. Dans les deux cas, la différence avec l'approche reconstructive est l'absence d'une représentation symbolique du monde commune aux hauts niveaux d'interprétation.

Malgré leurs différences, ces deux approches de la vision par ordinateur sont intimement liées. La vision intentionnelle est-elle possible sans reconstruction ? Si elle ne cherche pas à fabriquer une représentation générique, elle a tout de même besoin de représentations flexibles et adaptées à chaque tâche visuelle qui la compose. D'autre part, peut-on décider d'une représentation de scène sans définir au préalable le but du système ? L'approche reconstructive sous entend forcément des tâches à accomplir, ne serait-ce que celle de fabriquer une représentation symbolique de l'environnement.

En résumé, tout système de vision doit procéder à une représentation ou une reconstruction à un certain niveau et tout système de vision doit avoir un but à atteindre [Ramesh, 1994] . La position maintenue par Marr tout comme Aloimonos suppose une division des tâches visuelles en modules indépendants. C'est une idée intéressante d'un point de vue calculatoire - il suffit de décomposer le problème de la Vision en problèmes élémentaires et d'assembler le tout en un système plus global [Tsotsos, 1994] . Cette position, héritière de travaux en neuro-biologie de la fin des années 70 tend à devenir obsolète devant des découvertes plus récentes dans ce domaine. Bien qu'il ait été démontré que certaines zones du cerveau sont dédiées à des tâches spécifiques (mouvements, couleur), [Allman et Kaas, 1971] [Zeki, 1977] des travaux plus récents ont mis en évidence une intense communication entre le cortex visuel et les autres zones du cerveau, ainsi qu'un va et vient constant entre approches ascendantes et descendantes [Felleman et Van Essen, 1991] . Tout porte à penser que chaque approche ne résoud qu'une partie d'un problème plus global. Chacune apporte des éléments de solutions, mais on est encore bien loin d'une solution générale. De plus en plus émergent des tentatives de coopération entre les deux approches, par exemple en effectuant des allers et retours entre différents niveaux de représentation ou par injection de contraintes liées à des tâches précises dans des méthodes de reconstruction.

1.3.3.1 Reconstruction intentionnelle

L'une de ces tentatives récente de concilier les deux approches a été résumée par [Christensen et Madsen, 1994] sous le terme de "reconstruction intentionnelle". En partant du fait que les méthodes reconstructives et intentionnelles sont insuffisantes prises individuellement, Christensen souligne leur caractère complémentaire

et suggère un cadre de travail permettant de les faire collaborer au sein d'un même système.

La quantité considérable d'informations visuelles à traiter et les ressources de calcul tout de même réduites dont nous disposons imposent des modèles de représentation et de traitement de taille limitée. La maintenance de ces modèles de façon continue nécessite de réévaluer en permanence les représentations internes du système en fonction du contexte et des buts à atteindre. L'approche intentionnelle se montre très utile pour atteindre ce résultat et constitue un bon point de départ, mais on ne doit pas négliger les apports de techniques issues de la reconstruction. Celles-ci apportent une grande robustesse aux représentations internes du système. Il faudrait donc adapter les méthodes de reconstruction aux besoins des comportements de vision intentionnelle. Chaque approche peut bénéficier de méthodes actives à un niveau algorithmique pour évaluer des méthodes de reconstruction.

Le côté intentionnel du système doit être amené à choisir parmi plusieurs stratégies de reconstruction, et imposer des contraintes (ou hypothèses) sur ces méthodes pour les rendre plus efficaces.

Le côté reconstruction fournit au système des méthodes robustes et fiables. Ces méthodes peuvent être utilisées à la manière de la perception directe de Gibson. Le système pourrait ainsi prendre des décisions à partir de représentations locales, adaptées à des tâches spécifiques.

Cette conception de la vision semble confirmée par des travaux en psychologie visuelle selon lesquels la reconnaissance visuelle se base uniquement sur des informations 2D alors que l'information 3D n'est utilisée que pour interagir avec l'environnement. Les reconstructions sont alors locales et uniquement liées aux centres d'intérêts [Biederman, 1987] . Dans ce contexte, les recherches en vision reconstructive sont considérées en tant que développement d'outils fiables à utiliser au sein de systèmes dont la stratégie est intentionnelle. Elles doivent fournir non seulement des techniques mais aussi des spécifications sur les circonstances d'application et les contraintes nécessaires. L'approche intentionnelle doit quant à elle servir de "lien" pour intégrer des méthodes développées par une approche de reconstruction, en systèmes opérationnels.

Pour mieux cerner quelles tâches visuelles appliquer et selon quelles contraintes, le système visuel doit répondre à la question : Comment la vision peut-elle aider une certaine action ou tâche ? Ce genre d'approche ne résout pas pour autant le problème de la vision d'un coup de baguette magique. Un certain nombre de questions liées à chaque approche restent toujours à résoudre. L'idée d'utiliser les qualités d'un aspect du système pour apporter des solutions ou des contraintes à l'autre aspect devrait permettre d'en résoudre quelques unes. Christiensen souligne à ce propos le rôle important que devraient jouer deux types de primitives encore peu exploitées en vision : primitives fonctionnelles et primitives contextuelles. Du fait des difficultés de modélisation qu'elles soulèvent, ces primitives nécessitent encore l'intervention humaine.

La fonction des objets devrait pouvoir jouer un grand rôle pour la sélection et

la reconnaissance. La vision humaine montre en effet une remarquable aptitude à classer des objets par catégorie. Connaître la fonction d'un objet à rechercher dans une image permettrait d'utiliser son contexte pour réduire l'espace de recherche et de définir quel type de contrainte appliquer. Cette notion est semblable à celle de potentiel de l'environnement (*affordance*) de J. J. Gibson. Le principal problème étant de modéliser ces fonctions.

Un objet peut avoir plusieurs fonctions selon le contexte, l'action ou même l'intention envisagés. Un autre problème est de reconnaître le contexte d'une scène. Par exemple, les méthodes à appliquer ne seront pas les mêmes entre une scène d'extérieur et un bureau. L'utilisation de ce genre de primitives suppose des recherches d'invariants propres à chaque contexte et des méthodes pour les détecter : distribution des textures ou des couleurs (scènes d'extérieur), détection de structures de référence (des structures verticales et horizontales sont autant d'indices d'une scène d'intérieur). Le problème se complique lorsqu'on remarque que ces invariants sont propres à un domaine spécifique d'application. Des applications en imagerie satellite, médicale, de robotique mobile, ou d'astronomie n'ont pas les mêmes exigences ni les mêmes besoins.

1.3.3.2 Approche "système" de la vision par ordinateur

Durant les vingt dernières années, l'application de la vision artificielle à des problèmes industriels est passée de l'idée utopique d'un système de vision général et adaptatif, à une conception plus pragmatique du problème. Aujourd'hui, la "vision industrielle" permet des solutions extrêmement performantes pour l'exploitation d'informations visuelles sous forme de systèmes dédiés à des tâches précises [Batchelor et Whelan, 1997] [Freeman, 1988] [Freeman, 1989] . Après avoir apporté de nombreuses méthodes de traitement et d'interprétation d'images à la vision industrielle, la vision par ordinateur pourrait désormais bénéficier en retour des leçons tirées par la vision industrielle sur les techniques de conception de systèmes visuels.

L'idée principale est de considérer la vision en tant que système, et d'analyser ses différentes composantes selon ce point de vue. La vision ne fait plus partie d'un système auquel elle fournit des données selon une certaine représentation, mais devient elle même un système à part entière [Jolion, 1994] .

Dans ce contexte, on appelle un système une entité organisée, composée d'éléments interdépendants. Ces éléments doivent être compris selon leurs relations mutuelles au sein de l'entité globale. On retrouve un principe de globalité cher aux Gestaltistes : le tout représente plus que la somme de ses parties. Comme suggéré par les approches intentionnelles, le système visuel serait constitué de l'observateur (humain ou machine) et de la scène. L'idée est alors d'exprimer les caractéristiques du système et de ses éléments dans toute leur complexité.

L'étude du système doit d'abord prendre en compte les caractéristiques de l'environnement étudié. Celles-ci vont des limites physiques du système (est-ce que la scène observée est en 2D, en 3D ? est-elle statique ? dynamique ?) jusqu'aux limites

conceptuelles (de quoi le système doit-il être capable ? détection ? reconnaissance ? mobilité ? manipulation ? le système est-il figé ou doit-il être capable d'évoluer ?).

Chacun des composants du système doit être complètement défini. Ces composants peuvent être, par exemple, les détecteurs d'indices visuels, des modules de décision, ou d'action, ou encore, des composants de la scène observée (une surface peut transformer la lumière en la réfléchissant par exemple). Les composants du système sont définis par leurs propriétés intrinsèques (propriétés physiques pour des éléments de la scène ou bien propriétés d'implémentation pour des algorithmes ou des capteurs). C'est à ce niveau que devraient être décrites leurs propriétés fonctionnelles et les conditions dans lesquelles elles s'appliquent.

Les composants du système peuvent être passifs, et agissent alors comme des mémoires. Leur seul but est alors de stocker des informations et de les restituer à la demande. Ils peuvent aussi être actifs, avec un but spécifique. Leur fonction est alors d'interpréter des informations d'entrée et de fournir des informations de sortie en fonction de leur propre but. On définit comme "informations" l'ensemble des éléments manipulés par le système ou bien échangés par ses composants. Une information peut très bien être de nature physique (lumière, texture), symbolique (primitive géométrique, représentation locale) ou encore événementielle (pour activer, désactiver ou synchroniser des composants par exemple).

Enfin, l'étude d'un système visuel dans ce cadre de travail suppose une attention particulière portée à son architecture. La façon dont les différents composants sont liés entre eux, et surtout la nature de leurs échanges sont indispensables à l'équilibre du système. De ce point de vue, le système définit un réseau de communication composé de tous les supports utiles à l'acheminement d'informations entre ses différentes parties.

L'étude des différentes parties du système devrait apporter des contraintes mais aussi une meilleure compréhension des problèmes liés à chaque composant. La complexité de cette approche doit être considérée comme une source de richesse pour l'introduction de contraintes. Elle peut être réduite dans une certaine mesure en définissant une hiérarchie de priorités entre composants. Une autre possibilité pour profiter de cette complexité est l'idée de "consensus". Plutôt que de chercher "la" méthode optimale pour résoudre une certaine tâche, pourquoi ne pas permettre l'utilisation de diverses sources d'information et d'évaluer les plus appropriées ? Par exemple, extraire les contours selon différentes techniques et définir lesquelles fournissent les résultats les plus utiles pour la tâche courante.

Cette méthodologie souligne enfin l'importance d'une étude des relations entre composants très tôt dans la conception du système. Il ne s'agit pas de définir chaque composant indépendamment des autres et d'essayer d'assembler le tout ensuite, mais plutôt de partir d'une vue d'ensemble du système et de définir chaque composant en fonction des autres. C'est le rôle de chaque partie du système de contribuer à l'équilibre du tout et éviter ainsi au maximum l'intervention extérieure d'un opérateur.

L'application réussie de cette méthodologie sur des problèmes spécifiques de vision industrielle est de bonne augure pour le développement de systèmes visuels

artificiels de plus en plus génériques. Elle offre un cadre de travail intéressant pour intégrer les principes issus des approches reconstructives et intentionnelles au sein d'une même conception de la vision artificielle.

1.4 Conclusion

En résumé, pour tenter de résoudre le difficile problème de la perception visuelle par ordinateur, deux principales approches ont été abordées. La plus ancienne, dite *reconstructive*, vise à extraire d'une scène une représentation générique sur laquelle appliquer des raisonnements et prendre des décisions. L'autre, plus récente, est dite *intentionnelle* et cherche à explorer une scène à l'aide d'un système de méthodes génériques.

Chacune de ces approches hérite de conceptions différentes de la vision. Celles-ci remontent à des théories de la perception visuelle qui s'opposent depuis des années pour apporter une explication au fonctionnement biologique et psychologique de la vision. L'une, qualifiée d'*empiriste*, fait la distinction entre l'observateur et son environnement. Celui-ci ne perçoit pas l'environnement directement mais par l'intermédiaire de représentations symboliques interprétées par le cerveau. L'autre approche, dite *environnementale*, place l'observateur au sein de l'environnement qu'il perçoit directement. Cette séparation peut être suivie jusqu'à un niveau philosophique sur la perception de la réalité.

Notre approche se place dans la perspective de méthodes compatibles pour les deux paradigmes de la vision par ordinateur. En fournissant une représentation hiérarchique de la scène, elle suit les étapes classiques de segmentation, structuration et interprétation propres aux méthodes reconstructives.

Pourtant, en faisant appel à des principes de groupements perceptuels tout au long de la chaîne de traitements, cette approche reste ouverte à l'intervention de processus de vision intentionnelle. En effet, elle ne produit pas de représentation particulière de la scène, mais plutôt un ensemble d'hypothèses sur les éléments visuels les plus importants.

Ces éléments de représentation extraits par les différents niveaux de groupements perceptuels forment une ébauche qualitative des structures visuelles importantes de l'image. Les chapitres 4 à 6 illustrent l'utilité des méthodes issues de la théorie Gestaltiste afin de guider la formulation de ces hypothèses. Celles-ci peuvent être ensuite utilisées comme centres d'attention par une tâche de reconstruction ou bien directement en tant que représentations partielles dans un système de vision intentionnelle.

L'idée principale est de remplacer autant que possible toute connaissance à priori des objets de la scène par une connaissance générique du type de scène observée.

Chapitre 2

Analyse et interprétation des scènes de contours

Après avoir défini le contexte de notre travail, nous abordons dans ce chapitre la question des différentes représentations de l'environnement visuel. Le choix des contours comme éléments de base de représentation est ensuite justifié. La problématique de la détection et de la modélisation des contours, ainsi que leur rôle en vision par ordinateur sont détaillés à partir de le sous-chapitre 2.2. Enfin, en guise de conclusion, nous exposons les points clés de notre travail en regard de ces problèmes.

2.1 Représentation de scènes complexes

Que ce soit pour construire une représentation complète, dans le cadre du paradigme de Marr, ou bien pour maintenir des représentations partielles dans un système visuel dynamique, les traitements sont fondamentalement semblables. Il s'agit dans un premier temps de détecter des éléments visuels intéressants, puis de les organiser en structures intermédiaires pour leur donner une forme adaptée à la tâche visuelle.

Par exemple, l'image d'un bureau peut être décomposée selon des plans verticaux, un plan horizontal et des barres verticales, ou bien selon des murs et une table. Ces éléments peuvent donc être éventuellement composites, créés à partir de représentations de plus en plus complexes. Par ordre croissant de complexité et d'organisation, ces représentations peuvent être rapprochées des différents niveaux de traitement suggérés par Marr. D'une manière plus générale, ils sont désignés respectivement par "acquisition", "segmentation", "structuration" et "interprétation". Bien que les traitements de chaque niveau soient en principe indépendants, chacun de ces niveaux utilise les représentations des niveaux inférieurs pour fabriquer sa propre représentation.

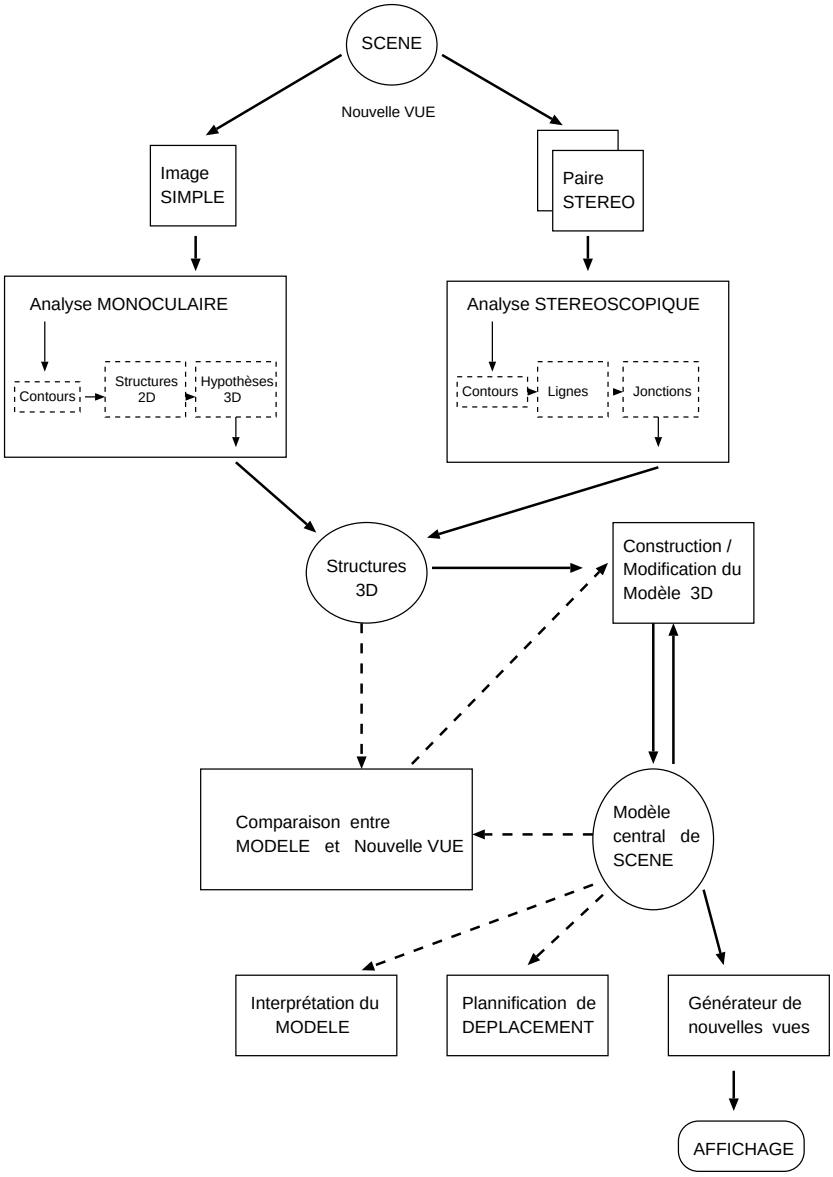

Figure 2.1 - *Exemple de système visuel - tiré de "3D Mosaic Scene Understanding System" - Herman M. et Kanade, T. - 1986*

2.1.1 Acquisition des images

On entend par "image" le niveau de représentation le plus bas. Il s'agit de l'information brute, telle que perçue directement par les capteurs. Celle-ci correspond à la projection sur un plan en deux dimensions d'informations en trois dimensions émises par l'environnement observé. La nature de ces informations dépend du type de capteur utilisé. Ces capteurs peuvent être de deux types : actifs ou passifs.

Un capteur est dit *actif* lorsqu'il mesure la différence entre un signal émis préalablement et son "écho". A l'aide de laser ou d'ultrasons , ils donnent la possibilité d'explorer activement l'environnement. Ces capteurs sont constitués d'un émetteur et d'un récepteur. Ils permettent de générer une carte de profondeur (*range image*) par triangulation (si on connaît déjà la position de l'émetteur par rapport au récepteur) ou par décalage de phase (temps de vol) dans le cas d'ultrasons. L'intensité lumineuse des images obtenues par ces moyens donne directement une information de profondeur. La richesse de cette information permet de constituer une ébauche intermédiaire des surfaces directement à partir des images de profondeur. Par contre, les informations d'intensité lumineuse des surfaces ou de leur texture ne sont pas accessibles par ce moyen.

A l'inverse, les capteurs *passifs* se contentent de recevoir les radiations lumineuses émises par l'environnement. Ces radiations sont projetées sur un plan à l'aide d'un système optique afin de former une image, qui devient alors une modélisation de l'image rétinienne de notre système visuel. On parle de systèmes "monoculaires" dans le cas d'une seule caméra, et de systèmes de "stéréo-vision" dans le cas de caméras multiples. Ces systèmes sont les plus fréquemment utilisés car ils permettent une acquisition plus économique et plus facile à mettre en place que les systèmes actifs. En projetant l'information lumineuse sur un plan image, la perte de l'information de profondeur peut être compensée par la perception des textures, de l'illumination des surfaces et l'acquisition de plusieurs vues.

Dans tous les cas, ces images sont des reconstructions élémentaires de la réalité. Les images obtenues sont modélisées à partir d'un maillage du récepteur. Le maillage le plus utilisé est un quadrillage rectangulaire, plus facile à manipuler sous la forme d'une matrice de points. D'autres maillages sont possibles selon les applications et les propriétés topologiques désirées : polaires, triangulaires, hexagonaux. Les éléments qui constituent ces maillages sont désignés par des *pixels*. Ils représentent chacun un échantillon de l'information reçue par le capteur. On parle alors d'images généralisées, ou de représentations *iconiques* (représentations semblables à la scène observée) [Ballard et Brown, 1982] .

2.1.2 Segmentation d'indices visuels

La segmentation correspond à la détection d'éléments de l'image susceptibles de provenir d'une cause commune. Ces éléments sont autant d'indices sur les propriétés des surfaces des objets présents dans la scène observée. On appelle "pré-traitement"

(*early processing*) les opérations d'extraction de ces indices, aussi appelés "indices visuels". L'étude de l'intensité lumineuse sur l'image permet de détecter des indices différents en fonction des propriétés observées.

Ainsi, les discontinuités de l'intensité lumineuse de l'image reflètent, en général, des phénomènes physiques tels que les contours des objets présents dans la scène. De nombreux travaux ont été menés sur les différentes façons de définir, détecter et utiliser les contours [1]. Les contours servent de représentation de base aux descriptions géométriques utilisant des segments (scènes polyhédriques) ou des courbes.

L'approche complémentaire pour l'étude des images porte sur la détection de zones dont les propriétés sont similaires. Les indices visuels recherchés sont alors définis par des régions, et représentent les zones de l'image où l'intensité lumineuse est relativement uniforme. En généralisant l'idée d'homogénéité à la définition de motifs répétitifs, les régions ainsi définies se rapportent à des textures particulières [Ballard et Brown, 1982] . Qu'elles soient obtenues à partir des intensités lumineuses, de la couleur ou de textures, les régions forment des représentations particulièrement utiles pour délimiter des surfaces susceptibles d'appartenir à un même objet dans la scène.

Quel que soit l'indice visuel considéré, sa détection doit tenir compte des défauts des capteurs (bruit, discrétisation de l'image). Une certaine connaissance liée à la scène commence à devenir utile à partir de ce niveau. Des hypothèses sur le type de scènes, de surfaces, d'éclairage peuvent orienter la décision sur les indices visuels à extraire et les techniques d'extraction. Par exemple, un environnement artificiel (une scène d'intérieur) se prête plus à une approche par contours alors qu'une approche par textures sera toute désignée pour une scène de forêt.

2.1.3 Structuration des indices visuels

Pour l'instant, la fonction de ces représentations élémentaires est d'extraire de l'image des informations susceptibles de correspondre à des centres d'intérêt. Le rôle de la segmentation est essentiellement de réduire le volume d'information contenue dans l'image pour rendre son analyse plus accessible. Le but de cette étape est donc de structurer les indices visuels détectés par la recherche d'indices plus évolués concernant la forme des objets observés. En particulier, il s'agit de déterminer l'orientation spatiale des surfaces pour enfin percevoir les objets indépendamment de l'observateur.

Ce problème a fait l'objet de nombreuses approches qui ont toutes en commun l'utilisation d'hypothèses particulières. Les trois familles d'approches que nous citons ici sont parmi les plus représentatives de ces hypothèses et des contraintes qu'elles introduisent.

1. La suite de ce chapitre discute plus en détail des représentations de scènes à partir de contours.

– **"Shape from X"**

Cette catégorie de techniques permet de reconstituer la forme des surfaces par l'utilisation d'hypothèses fortes sur la nature des objets ou les conditions d'observations : illumination, textures et contours.

L'étude de l'illumination de la scène et de l'ombrage provoqué sur les objets est connue sous le nom de *shape from shading*. Dans des conditions bien déterminées, l'intensité lumineuse à la surface d'un objet est directement fonction du type de surface, et de la position relative entre l'éclairage et la caméra. Des travaux de B. K. P. Horn [Horn, 1975] à ceux d'Alex P. Pentland [Pentland et Bichsel, 1994] de nombreuses recherches ont été menées pour étendre les limites d'application de cette technique [Zhang *et al.*, 1994] .

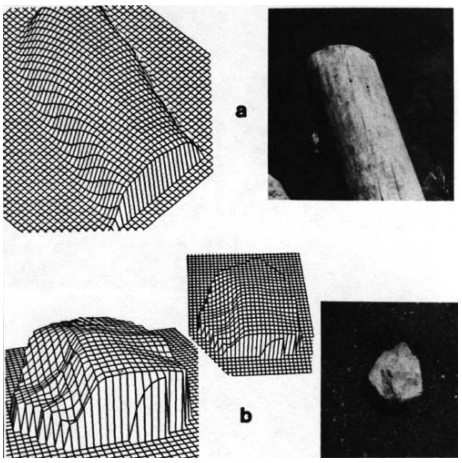

Figure 2.2 - *Deux exemples de surfaces extraites à partir de l'illumination de la scène (shape from shading).*

J.J.Gibson [Gibson, 1950] fut l'un des premiers à souligner l'importance de la texture comme indicateur de forme et de profondeur. Ses observations ont été à l'origine de techniques exploitant des gradients de textures pour évaluer la courbure et l'orientation des surfaces [Bajcsy et Lieberman, 1976] [Kender, 1978] . Ces techniques reposent en particulier sur l'étude des changements de densité et de taille des motifs de la texture sur l'image. Cependant, comme elles nécessitent des textures relativement régulières, les difficultés de segmentation en textures rendent ces méthodes peu utilisables dans des cas généraux.

Les hypothèses peuvent enfin porter sur les contours des objets observés. Le principe est de reconstituer les normales aux surfaces visibles à partir de leurs contours. Il est alors possible de reproduire la surface à partir d'un champ de normales. Cette approche donne d'assez bons résultats dans le cas d'objets

dont la structure est connue à l'avance, comme par exemple des objets de révolution ou bien à base de cylindres généralisés [Zerroug et Nevatia, 1996b] [Malik et Maydan, 1989] .

Même si elles donnent d'assez bons résultats dans des situations bien précises, les hypothèses demandées par ces méthodes sont en général trop sévères pour une utilisation générale.

– **Stéréo-vision**

Les méthodes relevant de la stéréo-vision consistent à utiliser plusieurs images d'une même scène, selon des points de vues différents, pour mettre en correspondance des objets communs à chaque image. Cette mise en correspondance permet de mesurer la disparité entre les deux images et de reconstituer la profondeur par triangulation [Zhang, 1993] [Faugeras et Robert, 1994] [Eric et Grimson, 1993] [Hartley et Sturm, 1997] . Les paramètres de la caméra (angles de vues, focale) peuvent être déterminés à l'avance grâce à un banc stéréoscopique calibré. S'ils sont inconnus, on parle alors de stéréo-vision non calibrée. De nombreux travaux portent sur cet aspect pour s'affranchir autant que possible d'hypothèses trop contraignantes sur les paramètres de la caméra [Luong et Faugeras, 1993] [Ayache et Lustman, 1991] .

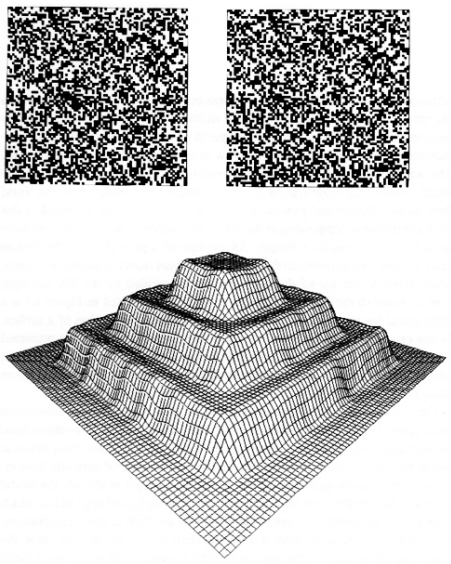

Figure 2.3 - *Déduction de la profondeur à partir de la disparité entre une paire stéréoscopique - ici, un stéréogramme.*

La stéréo-vision permet de déterminer l'orientation spatiale des contours et des surfaces, et par là, la construction de facettes 3D. Contrairement à l'approche précédente, l'approche par stéréo-vision tente de conserver des contraintes aussi génériques que possibles. Celles-ci reposent par exemple, sur une

recherche d'invariants ou sur une hypothèse d'un faible changement de point de vue d'une image à l'autre.

– **Méthodes actives**

Ce dernier type d'approche adopte une démarche active pour la recherche de contraintes sur la perception des surfaces. Ces méthodes sont actives au sens où elles supposent une modification appropriée de certains paramètres de la scène ou du système visuel afin d'étudier l'influence de ces variations sur l'image.

A titre d'exemple sur ces paramètres, on peut citer l'analyse du mouvement de la caméra. On définit alors un flot optique par l'ensemble des vecteurs de déplacements d'une image à l'autre. L'étude de ce champ de vecteurs permet, entre autres, de définir les parties rigides en mouvement et d'en déduire une estimation des objets présents dans la scène [Gibson, 1979] .

L'utilisation active d'autres paramètres de la caméra a été proposée. A titre d'exemple, [Grossmann, 1987] et [Pentland, 1987] ont suggéré de reconstituer la profondeur par l'analyse de l'image d'une seule caméra à focale variable. Cette technique mesure la quantité de "flou" introduite autour des contours par un changement de focale pour en déduire la distance de ce contour à la caméra. Malgré des côtés attractifs, tels que l'utilisation d'une seule caméra, cette méthode demande une connaissance approfondie du système optique de la caméra et de l'éclairage de la scène.

Une alternative à la modification des paramètres de la caméra consiste à agir directement sur l'environnement observé. On peut trouver dans cette catégorie des méthodes basées sur l'étude des variations lumineuses d'une mire projetée sur la scène. La déformation d'un faisceau laser étalonné ou la projection d'une grille sur une surface permet d'évaluer le "profil" de cette surface et d'en déduire sa forme et son orientation dans l'espace [Shrikhande et Stockman, 1989] [Strauss, 1992] . A l'instar des méthodes de *shape from shading*, cette dernière catégorie reste confinée à des cas bien précis où l'environnement est contrôlé.

A l'aide d'hypothèses plus ou moins contraignantes, ces différentes approches ont toutes en commun la reconstitution de la forme des parties visibles des objets. Cette reconstitution de la géométrie des surfaces reste partielle et limitée à ce que perçoit l'observateur. Les représentations utilisées à ce niveau sont plus abstraites, faisant appel à des champs de vecteurs et des modèles géométriques de surfaces.

2.1.4 Représentation de haut niveau

Alors que les niveaux inférieurs se contentent de produire une représentation partielle de certaines propriétés de la scène, depuis l'intensité lumineuse aux surfaces visibles, le but de ce dernier niveau de traitement est d'obtenir une représentation globale de la scène et des relations entre ses composants. Ce niveau fait le lien

entre les représentations partielles obtenues par les niveaux inférieurs et une représentation de la forme des composants de la scène indépendante du point de vue de l'observateur. C'est donc un niveau d'intégration entre les modules visuels des autres niveaux dans une même représentation.

Marr et Hishima donnent une série de critères pour une représentation de formes appropriée.

- *Accessibilité :* la représentation doit pouvoir être construite au prix d'un temps et de ressources de calcul raisonnables.

- *Usage :* la représentation doit être adaptée à l'application voulue. Une représentation à base de plans est peu adaptée pour décrire des sphères. De la même manière, l'utilisation de primitives géométriques est peu adaptée à une scène naturelle.

- *Unicité :* Ce critère est indispensable pour pouvoir localiser, identifier ou reconnaître un objet. Il suppose une représentation peu dépendante du point de vue, centrée sur chaque objet et non sur l'observateur.

- *Stabilité :* La représentation doit être suffisamment générique pour résister à des perturbations dues aux différentes étapes de la reconstruction. Elle doit permettre de définir une mesure de similarité globale entre objets.

- *Sensibilité :* A l'inverse du critère précédent, la représentation doit aussi autoriser des différences faibles entre objets.

A partir de ces critères, plusieurs approches deviennent possibles pour représenter une scène. L'approche la plus naturelle consiste à adopter une structure isomorphe à la scène observée. Marr suggère cette solution pour représenter la scène selon une hiérarchie de représentations à base de volumes géométriques simples tels que les cylindres généralisés. D'une manière semblable, Pentland souligne l'importance de choisir une représentation fidèle à la façon dont on perçoit naturellement une scène. Il propose pour cela de rechercher directement dans l'image les indices de la présence d'un ensemble de modèles simples et de représenter la scène à une échelle semblable à ce que nous percevons. Les modèles proposés utilisent un ensemble de superquadriques et de représentations fractales pour tenir compte de formes naturelles complexes [Pentland, 1986] . L'idée est de rapprocher la représentation de l'objet reconstruit de celle de modèles géométriques pour en faciliter la reconnaissance.

A l'opposé de ces représentations explicites de la scène, d'autres travaux ont développé d'idée de représentations plus abstraites, plus adaptées aux tâches complexes relevant de l'intelligence artificielle : décision, déduction, apprentissage, raisonnement. C'est le cas par exemple de l'indexation automatique de modèle à partir de caractéristiques de la scène observée [Jacobs, 1992] [Beis et Lowe, 1994] .

Dans tous les cas, les représentations de haut niveau doivent être adaptées à des tâches complexes allant de la reconnaissance à l'apprentissage. Leur conception doit

donc tenir compte des besoins des méthodes impliquées dans ces tâches (bases de connaissances, réseaux sémantiques, graphes d'inférence) [Ballard et Brown, 1982] [Wechsler, 1990] .

Notre objectif est l'extraction de primitives 2D structurées à partir d'images de scènes complexes. Les contours des objets présents dans la scène forment les éléments de base de représentation dans notre travail. Avant de développer cet objectif en fin de chapitre, nous commençons par rappeler l'importance des contours comme principaux éléments de représentation.

2.2 Scènes de contours

La notion de contours est d'abord une représentation cognitive des limites d'une forme, d'une surface ou d'un objet. La silhouette des objets, et par conséquent, leurs contours, est généralement perçue avant leur interprétation [Attneave, 1954] . Marr cite en exemple les travaux en neurologie de Warrington et Taylor (1973) sur des patients atteints de lésions cérébrales. Ces patients, à qui on demande de décrire différents objets, se montrent incapables de les reconnaître et de les nommer, alors qu'ils peuvent les décrire dans leurs moindres détails. D'autres travaux de neurophysiologie démontrent que l'analyse des formes 2D est facilitée par l'existence de zones cérébrales sensibles à la continuité de propriétés visuelles telles que l'intensité, la couleur ou la texture [Perrett et Oram, 1993] . Ces observations suggèrent que les formes des objets ont une représentation indépendante de leur interprétation ou de leur manipulation.

Notre capacité à interpréter des croquis ou des dessins sont autant d'indices psychologiques qui tendent à prouver que la silhouette d'un objet est souvent suffisante pour son identification. Ces arguments ont poussé Marr et ses successeurs à privilégier la forme des objets comme principale source de représentation, les autres indices de texture, couleur ou mouvement étant considérés comme secondaires.

D'une manière plus générale, les contours peuvent être définis comme des frontières permettant de séparer une forme d'un arrière plan, ou d'autres formes. Ces frontières sont le résultat de changements plus ou moins brusques de certaines caractéristiques visuelles, telles que la fréquence spatiale, la phase, l'intensité lumineuse, la direction, la vitesse ou encore, la densité de texture. En général, l'intensité lumineuse est la caractéristique privilégiée pour la détection des contours.

D'un point de vue plus pratique enfin, la détection de contours demande relativement peu d'hypothèses préalables, par opposition aux techniques basées sur l'illumination par exemple. De plus, ils offrent une grande simplicité de modélisation et d'exploitation par rapport à la manipulation de régions par exemple. Enfin, une représentation de scène à partir de lignes (droites ou courbes) est bien adaptée à l'usage de modèles géométriques explicites d'objets (constitués de sommets, arêtes et courbes).

Nous nous attacherons dans un premier temps à définir les différentes contraintes

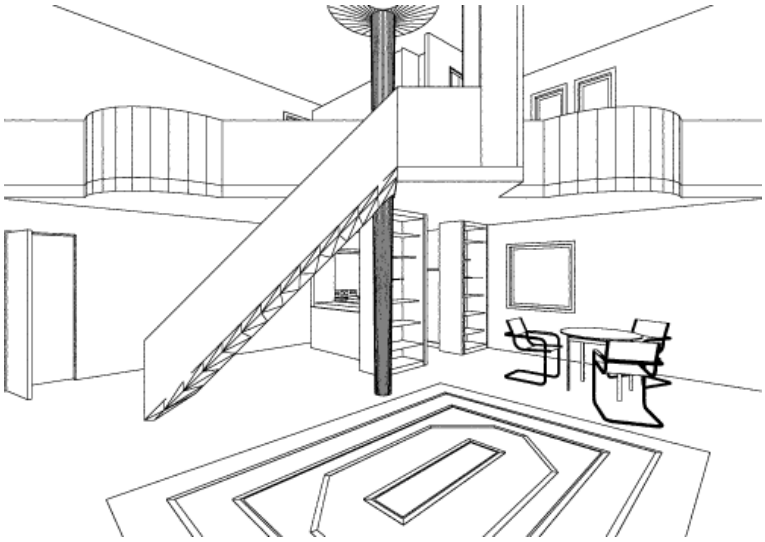

Figure 2.4 - *Les contours permettent souvent d'interpréter les objets d'une scène et sa structure 3D.*

rencontrées pour la représentation de scènes à partir de contours. Du type de scènes observées à l'acquisition d'images, les sources d'ambiguités sont nombreuses. Ces contraintes une fois définies, nous aborderons les principales méthodes disponibles pour chaque étape de la construction d'une représentation de scène, depuis la détection des contours jusqu'à la perception des structures. En guise de conclusion à ce chapitre, nous exposerons une vue d'ensemble de notre approche en regard de ces méthodes.

2.2.1 Scènes de contours et représentations

Le type de scène observée apporte nécessairement des simplifications sur la représentation des objets à partir de leurs contours. La complexité de ces scènes suit l'ordre chronologique dans lequel elles ont été étudiées ainsi que l'évolution de la puissance de calcul des ordinateurs.

Les premiers travaux sur la représentation de scènes à partir de contours se sont placés dans l'hypothèse d'une scène polyhédrique simple, l'observation portant essentiellement sur des prismes, de surface homogène, posés sur un arrière plan uniforme. En l'absence d'ombres et de réflections, les contours rectilignes sont dans ce cas directement liés à la silhouette de l'objet. Entamés dès les années 60 par L. G. Roberts [Roberts, 1968] ces travaux ont été étendus par la suite à l'interprétation de scènes polyhédrique complexes, constituées d'objets multiples et d'occlusions entre ces objets. Au début des années 70, Huffman et Clowes proposent une interprétation de telles scènes à l'aide d'un étiquetage cohérent des arêtes des objets observés. Dans ce contexte, une ligne joignant deux intersections sur l'image peut provenir

d'une arête convexe, d'une arête concave, ou bien d'un bord de l'objet occultant de la matière en arrière plan. Une ligne ne pouvant avoir plusieurs étiquettes à la fois, seules quelques configurations sont utilisables parmi toutes les combinaisons possibles. Cet étiquetage conduit à classer les jonctions entre lignes selon un catalogue réduit [Regier, 1991] .

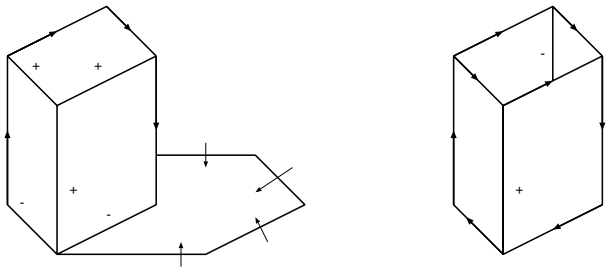

Figure 2.5 - *Exemples simples d'étiquetage dans une scène de blocs et d'origami. Les (+) désignent des arêtes convexes les (-) des arêtes concaves. Les arêtes étiquetées par un → signalent la présence de matière derrière la facette. Les flèches transversales désignent le cas particulier d'ombres.*

Des contribution importantes à ces travaux ont été apportées par Waltz (1975), avec l'introduction d'ombres et l'étiquetage de polyhèdres complexes, et par Kanade (1978) avec une interprétation de scènes à partir de facettes planes, ce qui autorise des objets éventuellement creux. De nombreux travaux ont été menés depuis sur ces deux types de scènes complexes, classiquement désignées par des *scènes de blocs* pour les travaux dérivés de ceux de Waltz, et par *scènes d'origami* pour ceux issus de ceux de Kanade [Ballard et Brown, 1982] [Parodi, 1996] .

Aussi robustes soient-elles, ces méthodes d'interprétation supposent une segmentation parfaite des contours des objets observés selon des segments de droites reliés entre eux par des jonctions. Si une segmentation suffisamment claire peut être obtenue pour des scènes d'objets polyhédriques, il en est autrement pour des objets quelconques, présentant des parties courbes. L'ajout d'arètes courbes dans la représentation de la scène augmente d'autant plus la complexité d'un étiquetage cohérent des contours. De plus, les surfaces ombrées et texturées des scènes réelles sont autant de sources de perturbations quand à la détection des contours.

Depuis Marr, les travaux sur l'interprétation de scènes de contours (*line drawings*) couvrent tous les niveaux de représentation, depuis la détection précise de contours, à l'interprétation de la nature des arêtes détectées (segments ou arcs) et à la reconstruction tridimensionnelle de la scène [Nalwa, 1988] [Straforini *et al.*, 1992] [Cooper, 1993] [Cowie et Perrott, 1993] .

Ces méthodes restent adaptées à des images d'objets réguliers, à partir desquels

une représentation, éventuellement schématique, à base de lignes et de courbes est encore possible. Il en est autrement de scènes naturelles, dont les formes présentent des échelles de régularité difficilement modèlisables à partir de modèles géométriques explicites [Pentland, 1986] .

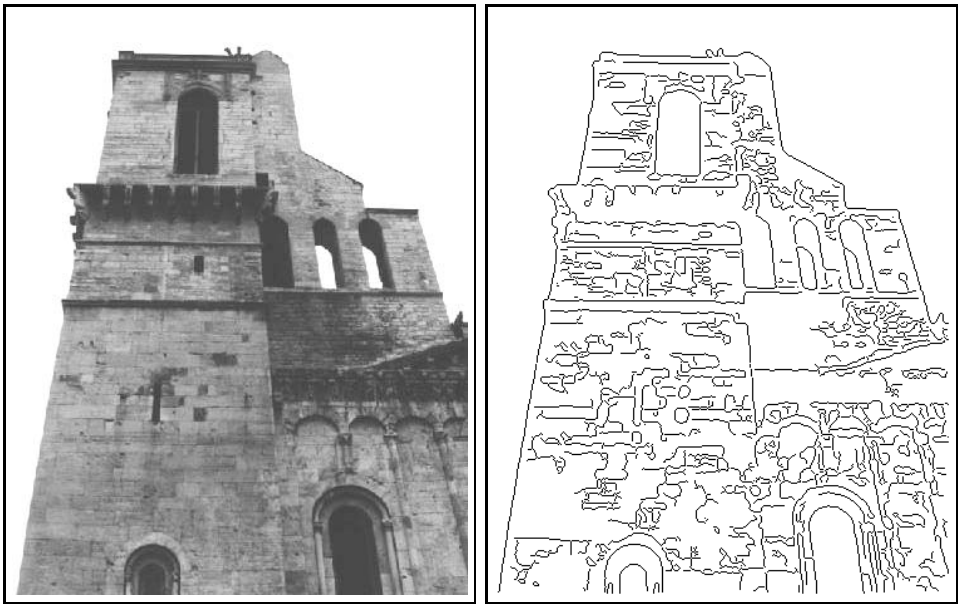

Figure 2.6 - *Exemple de détection de contours*

2.2.2 Sources d'ambiguités et contraintes

Dans le cas de scènes quelconques, l'exploitation des contours pour représenter la forme des objets est à l'origine de nombreuses sources d'ambiguités. Nous nous plaçons désormais dans l'hypothèse d'images représentant l'intensité lumineuse perçue par une caméra. Comme le montre l'exemple de la figure 2.6 cette intensité lumineuse est déclinée en nuances de gris. On peut aussi constater sur cet exemple le type de problème posé par une représentation à base de contours.

Ces problèmes peuvent être tout d'abord liés à l'acquisition même des images. Selon les conditions de prises de vues, le système optique peut introduire des perturbations dans l'image. En particulier en cas d'éclairages particulièrement faibles, ou de contrastes trop peu marqués, la résolution des capteurs peut ne pas être suffisante pour tenir compte des différences d'intensité lumineuse. Cela peut être le cas de surfaces de couleurs différentes mais d'intensités proches. Les niveaux de gris correspondant à chacune de ces couleurs (vert clair et jaune vif par exemple) peuvent être semblables et finir par atténuer un contour pourtant bien visible. A l'inverse, des éclats lumineux trop intenses peuvent provoquer des "débordements" de lumière d'un capteur sur ses voisins.

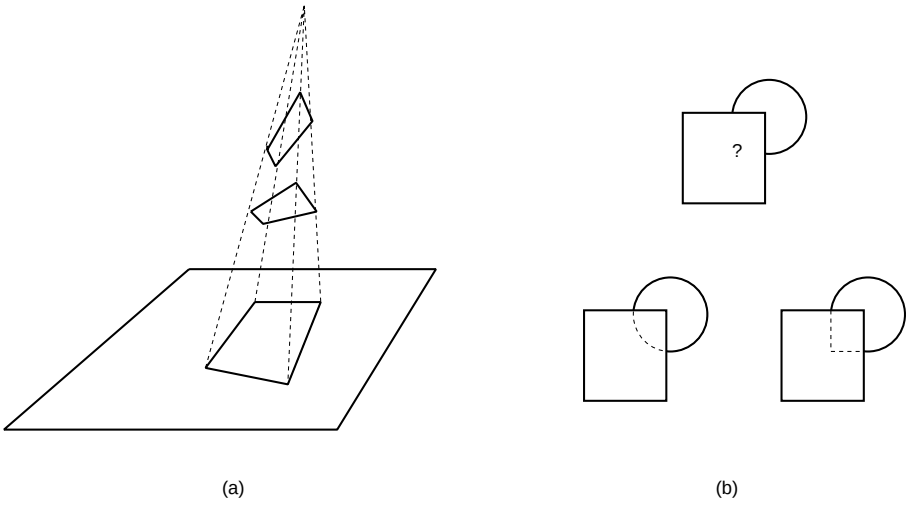

Figure 2.7 - *Ambiguïtés de projection (a) et d'occlusion (b).*

Le codage de l'image en un quadrillage, rectangulaire ou hexagonal par exemple, impose un échantillonnage de l'intensité lumineuse. Des problèmes de précision apparaissent alors pour des objets trop éloignés ou avec des arêtes trop fines. Leurs contours deviennent trop irréguliers à l'échelle de l'image. Le même problème se pose avec des surfaces texturées dont les variations sont trop fines pour être échantillonnées correctement. On assiste alors à l'apparition de phénomènes de moirées selon la loi de Shannon sur la fréquence d'échantillonnage. Enfin, la projection de la scène en perspective au travers du système optique provoque une déformation des objets proches de la caméra ainsi qu'une perte de netteté pour leur image.

Ces ambiguïtés peuvent être réduites par une connaissance du système optique utilisé (pour tenir compte des distorsions par exemple), ou par des méthodes actives de changement de focale ou de point de vue. En pratique, le système optique est représenté par le modèle de sténopé (*pin-hole*) dans lequel l'image est projetée au travers d'un point.

Le principe même de la projection de la scène sur un plan image pose une ambiguïté fondamentale due à la perte de la profondeur. Une infinité d'objets, des plus simples aux plus improbables, peuvent être disposés de manière à avoir la même projection selon un point de vue déterminé. Et pourtant, malgré cette infinité de possibilités, notre propre système visuel nous permet d'interpréter des dessins de façon cohérente, en ignorant délibérément cette infinité de solutions. Cette aptitude à interpréter des dessins suggère la possibilité d'aboutir à une méthode d'interprétation artificielle.

Ce problème de la projection est aussi à l'origine des occlusions, une source ambiguïté plus difficile à lever sans l'aide d'une connaissance préalable des objets observés. Qu'elles soient provoquées par la superposition de plusieurs objets dans l'image ou bien par la présence d'un objet en limite du champ visuel, les occlusions

interdisent simplement une interprétation complète en rendant invisibles des parties de la scène. Ici encore, il est possible de lever ces ambiguïtés dans une certaine mesure par des vues multiples.

Enfin, les contours observés sur une image ne sont pas forcément synonymes de frontières entre objets. Selon la définition adoptée pour les détecter, les contours peuvent représenter des discontinuités de distance par rapport à l'observateur (occlusions, alignements accidentels), des discontinuités d'orientation de surface, des changements dans les propriétés de la surface (reflets, textures) ou encore, des effets d'éclairages (éclats lumineux, ombres). Il peut être alors judicieux de comparer les résultats de détecteurs de contours selon plusieurs définitions.

2.3 Définition et détection des contours

En règle générale, un contour est associé à un changement brusque de propriétés physiques ou géométriques dans l'image. La détection de contours consiste donc à extraire de l'image une information spécifique les concernant. Mis à part quelques exceptions (contours déformables et contours fictifs, cf. pages 57 et 62), la détection de contours à partir d'une image produit en général une autre image. C'est à partir de cette image de contours que les représentations de plus haut niveau pourront être plus facilement construites.

Le premier problème que pose la représentation de scènes réelles à partir de contours est la définition même de ces contours. Cette définition conditionne les méthodes de détection et le type de scène "observable" par celles-ci. Ces méthodes peuvent être classées selon le type de discontinuité observée.

2.3.1 Discontinuité d'intensité

L'intensité lumineuse est l'information primordiale issue de l'image rétinienne. Elle est donc directement utilisable sans analyses supplémentaires (comme c'est le cas pour les frontières entre régions). Idéalement, les différents types de discontinuités peuvent être modélisés selon leur aspect mono-dimensionnel : marche d'escalier (*step-edge*), crête ou pic, porte (fonction de Heavyside), rampe. En réalité, les nombreuses discontinuités évoquées précédemment viennent perturber le signal. Ce qui ramène la détection de contours au problème de différenciation d'un signal bruité.

Le signal bruité est, dans ce cas, la fonction d'intensité lumineuse de l'image, notée $I(x, y)$. Les différentes méthodes de différenciation de cette fonction ont en commun l'application d'un opérateur de détection. On parle alors de 'filtrage' de l'image, l'opérateur étant la réponse impulsionnelle du filtre [Monga et Horaud, 1993] .

On peut distinguer deux types d'approches selon l'ordre de différenciation du filtre utilisé. L'approche "gradient" consiste à détecter les maxima locaux après application d'un filtre de différenciation du premier ordre. Les contours correspondent en effet à une forte différence d'intensité, donc à un maximum du gradient. Les

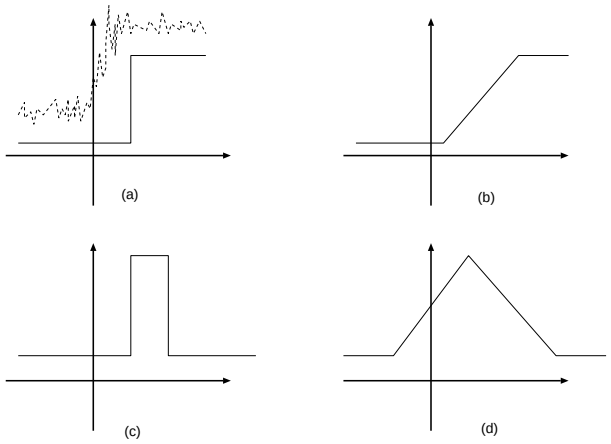

Figure 2.8 - *Exemples classiques de contours d'intensité 1D - (a) marche, (b) rampe, (c) porte, (d) crête. La courbe en pointillés représente l'allure d'un contour réel, bruité.*

maxima sont détectés dans la direction du gradient pour tenir compte de l'aspect bi-dimensionnel du signal. D'une manière similaire, l'approche "laplacien" détecte les passages par zéro de la réponse à un filtre du second ordre, c'est à dire, les endroits de l'image où la variation du gradient est nulle. En théorie, la détection des valeurs nulles du laplacien devrait suffire mais le calcul du laplacien n'étant qu'une approximation dépendante d'un certain échantillonnage de l'intensité, on recherche les passages par zéro par ses changements de signe.

Les différentes variantes de ces deux approches peuvent être classées selon le type d'opérateur appliqué [Deriche, 1990] .

2.3.1.1 Opérateurs locaux de dérivation.

La première manière de filtrer l'image consiste à discrétiser les directions selon lesquelles le gradient où le laplacien sera calculé. Réalisée à l'aide de masques directionnels, la différenciation est effectuée par convolution de l'image par ces masques. De nombreux masques de convolution ont été proposés pour le calcul du gradient ou du laplacien.

Parmi les masques du premier ordre les plus connus, citons les masques de Roberts (masque 2×2 selon des axes orientés à $45°$), ou de Prewitt et Sobel (masques 3×3, suivant les axes Ox et Oy). On peut noter que ces derniers sont le résultat de l'application successive d'un masque de lissage dans une direction puis d'une dérivation dans la direction orthogonale. Des masques plus complexes tiennent compte d'un plus grand nombre de directions, comme les masques de Kirsch (8 masques 3×3 orientés selon les multiples de $\frac{\pi}{8}$). L'orientation retenue pour le gradient est celle du masque donnant la plus forte réponse.

$$Masques\ de\ Prewitt\ (c=1)\ et\ Sobel\ (c=2)\ :\ \begin{bmatrix} 1 & 0 & -1 \\ c & 0 & -c \\ 1 & 0 & -1 \end{bmatrix}\ et\ \begin{bmatrix} -1 & -c & -1 \\ 0 & 0 & 0 \\ 1 & c & 1 \end{bmatrix}$$

L'estimation du laplacien peut être obtenue de la même manière par un masque de convolution 3×3 de la forme suivante :

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}\ ou\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

En pratique, ces filtres locaux donnent un meilleur résultat avec un lissage préalable de l'image pour atténuer l'influence du bruit de l'image. Ce lissage peut être appliqué, par exemple, par moyennage local des intensités lumineuses ou bien par le choix de la valeur médiane des intensités sur une fenêtre centrée autour de chaque pixel (filtre médian). Pour leur opérateur du second ordre, Marr et Hildreth appliquèrent un lissage préalable par un opérateur isotrope gaussien [Marr, 1982] .

Malgré un lissage préalable, ces méthodes locales sont trop sensibles au bruit. La modélisation du bruit n'intervient pas, en effet, dans leur définition. D'autres méthodes ont donc été introduites pour tenir compte du bruit et de la géométrie des contours à détecter.

2.3.1.2 Optimisation d'opérateurs géométriques

Hueckel (1973) fut le premier à rechercher un opérateur optimal de détection de contour. Le principe de cet opérateur est de partir d'un modèle géométrique idéal de contour, et d'ajuster ses paramètres de façon à ce qu'ils correspondent au mieux aux données d'une portion de l'image. Le modèle est défini localement par une droite délimitant un disque de taille fixée en deux régions homogènes. En utilisant une base de fonctions orthogonales définies dans le domaine de Fourier, il propose une optimisation de l'écart quadratique entre modèle et image [Monga et Horaud, 1993] . Cette méthode étant assez lourde à mettre en oeuvre, elle est précédée par une estimation de la présence d'un contour possible par opérateur gradient. Mais ici aussi, de nombreux problèmes apparaissent en raison de l'absence de définition du bruit dans la méthode.

L'opérateur de détection SUSAN proposé par [Smith et Brady, 1995] part d'une idée similaire. En délimitant un disque de recherche autour d'un pixel central, cet opérateur compare les intensités des pixels du disque à celle du pixel central (noyau). La proportion de pixels semblables au noyau, permet de définir un opérateur non linéaire réagissant à la présence de coins et contours. L'absence de calculs de dérivées le rend robuste au bruit et permet une bonne localisation.

2.3.1.3 Optimisation d'opérateurs de convolution

Au lieu d'optimiser les paramètres d'un modèle géométrique, [Canny, 1983] propose de rechercher une fonction anti-symétrique donnant une réponse optimale pour les contours par convolution avec la fonction intensité. Il définit alors trois critères pour optimiser l'opérateur recherché :

- Détection au voisinage des contours (maximisation du rapport signal sur bruit au point de contour pour rendre faible la probabilité de détecter de faux contours).
- Localisation précise des points de contours (maximisation de l'écart type de la position des contours).
- Réponse unique à un contour (limitation du nombre de maxima locaux détectés en réponse à un seul contour).

De ces trois critères, Canny obtient une équation différentielle dont une solution est de la forme suivante en 1D :

$$f_0(x) = c.e^{\alpha|x|}.sinwx$$

$\alpha = m.w$ représente la largeur du filtre, c'est un compromis entre localisation et détection. L'opérateur de Canny correspond à un filtre à réponse impulsionnelle finie, déterminé sur un intervalle $[-W, W]$ et présentant une pente S à l'origine[2].

Les critères d'optimisation donnent une fonction de régularisation dont la dérivée est l'opérateur recherché. Pour des raisons de coût de calculs, le filtre de dérivation est approché en pratique par la dérivée première du filtre gaussien défini par :

$$f_0(x) = \frac{1}{\sqrt{2\Pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$$

[Deriche, 1987] apporte une solution à l'équation de Canny étendue aux filtres à réponse impulsionnelle infinie et démontre ensuite que cette solution présente un indice de performance optimal lorsque w tend vers 0.

$$f_0(x) = \frac{S}{w}.e^{\alpha|x|}.sinwx$$

Dans le cas où $w \to 0$, on obtient le filtre de dérivation suivant :

$$f_1(x) = S.x.e^{-\alpha|x|}$$

2. L'intervalle est défini par : $W = \frac{1}{w}$

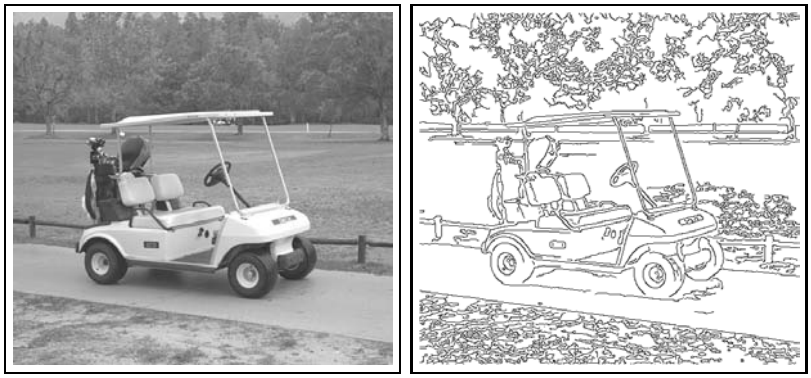

Figure 2.9 - *Détection de contours par application du filtre de Canny*

Figure 2.10 - *Détection de contours par application du filtre de Deriche.*

Dans les deux cas, la détection des contours commence par une étape de lissage (en général, un lissage gaussien), puis par l'application du filtre de dérivation pour obtenir l'image du gradient. Les contours sont représentés par les extrema du gradient détectés dans la direction de celui-ci et seuillés pour éliminer les fausses détections.

Enfin, dans une démarche semblable à celle de Canny, [Castan *et al.*, 1990] proposent un filtre optimal de lissage pour la détection des passages par zéro du laplacien :

$$f(x) = a.e^{-\alpha|x|}$$

L'image lissée est soustraite à l'image originale pour en estimer le laplacien. L'image résultat est alors binarisée en mettant à 1 les points positifs et 0 les autres. Les contours sont enfin définis par les frontières des régions ainsi formées.

2.3.1.4 Variantes de filtrage

Ces filtres optimaux sont les plus efficaces et les plus couramment utilisés. Ils présentent le meilleur compromis entre une détection relativement précise des contours et un faible coût calculatoire. Le filtre de Deriche, en particulier, admet une écriture récursive des produits de convolutions entre filtre et image. Enfin, ils tiennent compte du bruit dans leur modélisation et permettent une application à différentes échelles. Des travaux récents augmentent encore les performances de localisation et robustesse au bruit de ces filtres par une généralisation en précision inter-pixel [Montesinos et Datteny, 1997] [Devernay, 1995] [Fiorio, 1995] .

Le dernier type d'améliorations envisageables porte sur la préparation de l'image à filtrer. Nous avons déjà vu qu'une étape préalable de lissage permet d'atténuer l'influence du bruit sur la détection des contours. Pour mieux distinguer les véritables contours des fausses détections dues au bruit, il peut être intéressant de détecter les contours à différentes échelles de lissage. Les contours véritables devraient en effet rester stables sur plusieurs échelles. L'utilisation d'espaces échelles permet ainsi d'optimiser les paramètres de détecteurs de contours classiques [Lu et Jain, 1992] . D'autres approches multi-échelles ont été développées pour répondre au problème de la perte de détails lors du lissage. La plupart des lissages appliqués par les filtres classiques, comme le lissage gaussien, sont isotropes. Le principe de ces approches est donc d'utiliser un lissage anisotrope pour éliminer les faibles perturbations de niveaux de gris tout en conservant les contours des structures importantes. Ainsi, [Perona et Malik, 1990] proposent une détection de contours présents à plusieurs échelles à l'aide d'un opérateur de diffusion anisotrope.

2.3.2 Modèles de contours actifs

L'une des principales critiques des méthodes de détection de contours par filtrage est l'importance qu'elles attachent à des calculs locaux et surtout, l'influence que le bruit de l'image exerce sur les résultats. En réponse à ces critiques, des méthodes considérant des modèles globaux de contours ont vu le jour. Parmi ces méthodes, les modèles de contours actifs ont été introduits à la fin des années 80 par Kass, Witkin et Terzopoulos afin d'extraire des contours continus et uniformes à partir d'images trop bruitées pour les méthodes classiques [Kass *et al.*, 1987] .

Un contour actif, ou bien *snake*, est un contour déformable, dont les paramètres sont optimisés de façon à ce qu'il suive au mieux les contours de l'image. En pratique, il s'agit d'une courbe tracée sur l'image et à laquelle est associée une énergie. Cette énergie est composée d'un terme interne, propre à la géométrie de la courbe, et d'un terme externe, imposé par l'image. Ces deux termes étant antagonistes, le tracé de la courbe est déformé itérativement de manière à optimiser son énergie et ainsi, obtenir une solution satisfaisante pour chacun des deux termes.

Si $v(\ s\ )$ est la position le long du snake (décrit de manière paramétrique), l'expression élémentaire de l'énergie d'un contour actif est la suivante :

$$E^0_{snake} = \int_0^1 E_{int}(\, v(\, s\,)\,)ds + \int_0^1 E_{ext}(\, v(\, s\,)\,)ds + \int_0^1 E_{con}(\, v(\, s\,)\,)ds$$

où :

- $E_{int}(\, v(\, s\,)\,)$ représente l'énergie propre du snake, aussi appelé *terme de régularisation*. Il tient compte de la géométrie de la courbe, en particulier, sa longueur, sa forme et sa courbure. La minimisation de cette énergie conduit à lisser la courbe le long du contour et à réduire sa longueur.

- $E_{image}(\, v(\, s\,)\,)$ représente les forces d'attraction de l'image sur le snake. En général liée au gradient de l'intensité lumineuse, la minimisation de cette énergie force la courbe à suivre au plus prés les contours de l'image.

- $E_{con}(\, v(\, s\,)\,)$ regroupe les contraintes externes imposées au contour (le bord de l'image ou bien une zone interdite par exemple).

Ce modèle donne des résultats remarquables dans des situations où les méthodes classiques échouent en raison d'un bruit trop élevé ou de contours peu contrastés. Il est particulièrement adapté à des images bruitées, ou présentant des contrastes trop faibles pour obtenir des contours continus. Il permet également de suivre l'évolution d'un contour dans une séquence d'images, par exemple, une valve cardiaque en imagerie médicale.

Pourtant, cette méthode présente quelques faiblesses. La fonction de l'énergie n'étant pas convexe, il n'existe pas de moyen direct d'obtenir un minimum. Des méthodes itératives telles que l'algorithme GNC (*Gradual Non Convexity*) proposé par [Blake et Zisserman, 1987] et appliqué ensuite par M. O. Berger [Berger, 1991] permettent d'atteindre des solutions mais nécessitent une initialisation du tracé de la courbe à proximité d'une position optimale. D'autres variantes ont été développées depuis pour simplifier l'initialisation du modèle [Neuenschwander *et al.*, 1997] [Lai et Chin, 1993] , augmenter sa stabilité [Gunn et Nixon, 1996] , ou évaluer rapidement la courbure d'un contour [Williams et Shah, 1992] .

D'autres modèles plus élaborés tiennent compte de modèles paramétriques explicites tels que des coins, segments, ellipses ou B-splines, et d'une estimation du taux de lissage local de l'image le long du contour [Blaszka et Deriche, 1994a] . La convergence d'un contour actif pose également problème en cas d'objets multiples, de jonctions entre objets ou bien d'objets non convexes. Un début de solution à ce problème est apporté par les modèles de contours actifs géodésiques, qui permettent des séparations et fusions le long de la courbe afin de suivre les contours d'objets multiples [Caselles *et al.*, 1997] [Sapiro, 1997] [Deriche et Faugeras, 1996] .

Cette approche présente l'avantage de produire une représentation de contours sous forme paramétrique et non plus comme une seule image de points de contours. Le contour détecté est directement utilisable pour une interprétation, sans passer par une étape de structuration.

2.3.3 Coins, sommets

Les coins et sommets forment des configurations particulières de contours en deux ou trois dimensions. Etant localisés à l'intersection de deux ou plusieurs contours, les coins peuvent être déterminés à partir d'une détection de contours classique. Après un chaînage des points de contours, les coins correspondent alors à des points de coupures pour une approximation polygonale ou bien des points de courbure maximale pour une détection de courbes. Ce type de démarche, détaillé plus longuement dans la section suivante, peut manquer de précision à cause du lissage introduit pour la détection de contours. En effet, le lissage de l'image provoque une atténuation des intersections entre contours, d'où un effet de "coin arrondi" et un déplacement de la localisation du coin à l'intérieur de l'angle formé par les contours.

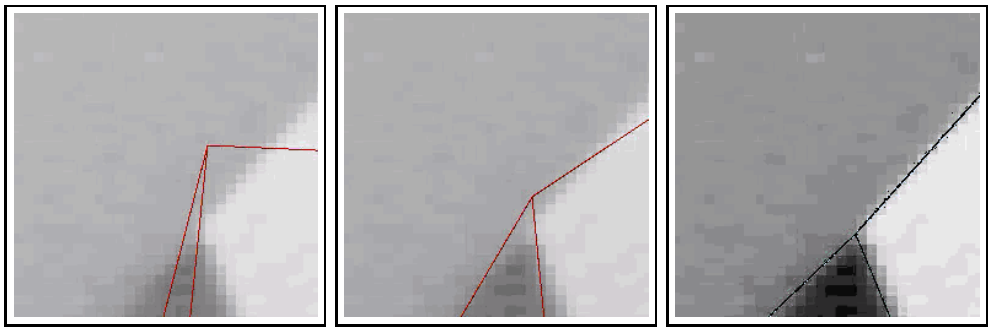

Figure 2.11 - *Détection précise de coins - convergence d'un modèle de coin vers une position optimale (à droite) - Méthode de Blaszka et Deriche.*

Une détection précise des coins nécessite des méthodes particulières, adaptées à ce type de configuration. Parmi les nombreux détecteurs de coins existants, les plus répandus mesurent la présence possible de coins sur l'image par un produit entre l'amplitude du gradient et le taux de variation de la direction du gradient. On pourra citer par exemple les détecteurs de [Kitchen et Rosenfeld, 1982] , [Noble, 1988] ou [Harris et Stephen, 1988] . Plus récemment, de nouvelles approches ont été introduites par [Rohr, 1992] ou [Deriche et Giraudon, 1993] utilisant des propriétés de géométrie différentielle et de modèles de coins. On pourra se reporter à ces derniers pour une étude comparée des méthodes les plus répandues de détection de contours. Cette dernière approche permet de définir des modèles de contours, coins et jonctions triples de manière extrêmement précise [Blaszka et Deriche, 1994b] .

2.3.4 Réseaux fins

Ces réseaux correspondent à des structures linéaires de l'image, comme par exemple des vaisseaux sanguins ou bien des routes. Ils correspondent au cas particulier de contours de type "crête" ou bien "toit".

Une première approximation de ce type de contour peut être obtenue à l'aide de méthodes de classiques de squelettisation, rapides mais peu précises. D'une manière plus analytique, Haralick (1984) propose d'approcher la fonction d'intensité localement par une fonction polynômiale qu'il suffit de dériver analytiquement. Huertas et Médioni (1986) proposent la même démarche pour estimer le laplacien. En pratique, l'approximation est réalisée à l'aide des polynômes de Tchebicheff, et conduit, dans le domaine discret, à un calcul des coefficients par application de masques de convolutions 3×3. Malheureusement, cette méthode souffre de l'influence importante du bruit sur l'approximation polynômiale [Monga et Horaud, 1993] .

En suivant une démarche semblable à celle de Canny, [Ziou, 1991] obtient un filtre optimal à réponse impulsionnelle infinie adapté à ce type de contours. Ce filtre est séparable et récursif, ce qui permet une implémentation efficace. Cependant, l'utilisation de modèles mathématiques pour évaluer l'orientation des contours rend leur détection approximative.

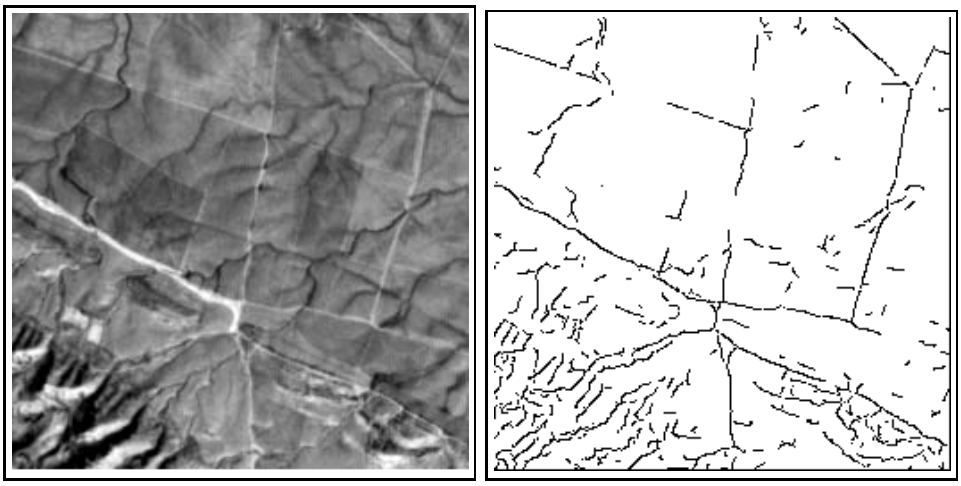

Figure 2.12 - *Détection de réseaux fins sur une image satellitaire. Un paramètre d'échelle σ permet de définir la largeur maximale des structures détectées. Ici, $\sigma = 1$. - Méthode de Armande, Monga et Montesinos.*

En considérant la fonction d'intensité comme une surface, Armande et Monga [Monga *et al.*, 1995], [Armande *et al.*, 1995], en étudient les propriétés différentielles afin d'extraire les lignes de crêtes. Après une extraction des dérivées partielles du premier au troisième ordre à l'aide de filtres gaussiens, les courbures principales de la surface sont calculées. La détection des passages par zéro de la courbure le long des directions principales permet de détecter les lignes de crêtes.

Il existe enfin des détecteurs spécialisés, comme par exemple les détecteurs de lignes de [Tupin *et al.*, 1996] adaptés au cas particulier des images radar à ouverture synthétique, ou encore la méthode adaptée aux images satellitaires SPOT proposée par [Merlet et Zerubia, 1996] .

2.3.5 Frontières de régions homogènes

Une approche complémentaire des méthodes précédentes consiste à segmenter préalablement l'image d'intensité lumineuse en régions homogènes. Les contours sont alors définis par les frontières entre régions. Notre propos n'est pas de passer en revue les différentes méthodes de détection de régions mais de donner un aperçu des méthodes générales utilisant les régions[3].

Figure 2.13 - *Comparaison entre une détection de contours avec filtre de Deriche (image de gauche) et l'extraction des frontières entre régions (image de droite).*

Une première manière de définir des régions consiste à regrouper les pixels selon un critère d'homogénéité de leur intensité. Les régions qui forment alors l'image définissent une partition de celle-ci au sens mathématique du terme. En ce sens, chaque région est maximale (il ne peut pas y avoir d'autre regroupement après la segmentation). Les différentes méthodes d'extraction de régions peuvent se séparer en deux classes. D'un côté, des méthodes procédant par *Division* dans lesquelles l'image est fragmentée en régions indivisibles (à l'aide de pyramides de résolutions par exemple). D'un autre côté, les méthodes procédant par *Fusion* regroupent les pixels en régions maximales[4].

La notion de région homogène peut être étendue à la détection de motifs répétitifs à plus grande échelle. On parle alors de *textures*. La segmentation en textures est difficile du fait de la complexité des modèles statistiques utilisés. Elle est applicable dans les situations où les variations d'intensité au sein des motifs interdisent les méthodes classiques de détection de contours ou régions [Malik et Perona, 1990] [Dunn *et al.*, 1994] [Geman *et al.*, 1990] .

3. Pour plus de détails sur la segmentation en textures et régions, on pourra se reporter à [Ballard et Brown, 1982] (chapitres 5 et 6) et [Monga et Horaud, 1993] (chapitre 4).

4. cf. [Nevatia, 1982] , chapitre 8.

2.3.6 Contours fictifs

Nous percevons enfin une dernière catégorie de contours qui ne peut pas être exactement définie par une discontinuité de propriétés de l'image. Il s'agit des contours subjectifs, ou encore, *contours fictifs*, mis en évidence entre autres, par Kanizsa. Dans la figure 2.14, un triangle "blanc", fictif, apparaît plus intensément que l'arrière plan. Il en est de même pour le disque "blanc" formé par les extrémités des segments disposés en cercle.

Figure 2.14 - *Figures de Kanizsa - les "formes" fictives apparaissent d'une intensité plus grande que le fond.*

Les travaux qui ont été menés pour modéliser la perception de contours fictifs peuvent être divisés en deux familles. L'une, plus théorique, s'attache à modéliser un étiquetage cohérent des éléments de la scène de manière à reproduire le phénomène de la perception d'un contour fictif [Thornber et Williams, 1997] . L'autre, plus algorithmique, ne s'attache qu'à une partie du problème, comme par exemple, la fermeture des contours [Williams et Jacobs, 1994] .

Ces contours apparaissent clairement à la vision naturelle. Ils semblent constitués d'alignements de discontinuités compatibles entre elles. Les Gestaltistes ont montré l'importance des mécanismes de groupements en vision naturelle[5] . La continuité semble être un critère si important que même des discontinuités locales peuvent être regroupées en formes cohérentes. Dans le cas présent, les formes "blanches" fictives sont perçues comme un triangle ou un cercle posés sur une forme en arrière plan (segments ou cercles noirs).

Cette dernière remarque peut être généralisée aux autres définitions de contours. Un contour peut être ainsi perçu comme un alignement régulier de points de discontinuité. En effet, la détection de contours consiste, en général, à localiser les discontinuités locales de propriétés visuelles. Nous avons vu comment cette approche, trop locale, doit intégrer des critères plus globaux pour être robustes au bruit, comme c'est le cas avec les contours actifs. Ce sont les groupements continus de points de contours qui permettent de faire la différence entre les contours véritables et les fausses détections.

5. Ces arguments développés plus précisément au chapitre 3.

2.4 Structuration des contours

Les contours une fois détectés, il est nécessaire de les regrouper en entités cohérentes. Cette étape de structuration permet de passer des pixels de l'image de contours à des structures représentatives de la scène, susceptibles d'être utilisées pour des représentations de plus haut niveau. En pratique, les primitives les plus utilisées sont les segments, les courbes et les jonctions entre ces éléments. Selon les cas, les jonctions peuvent être considérées comme des relations entre segments ou courbes, ou bien comme des entités à part entières.

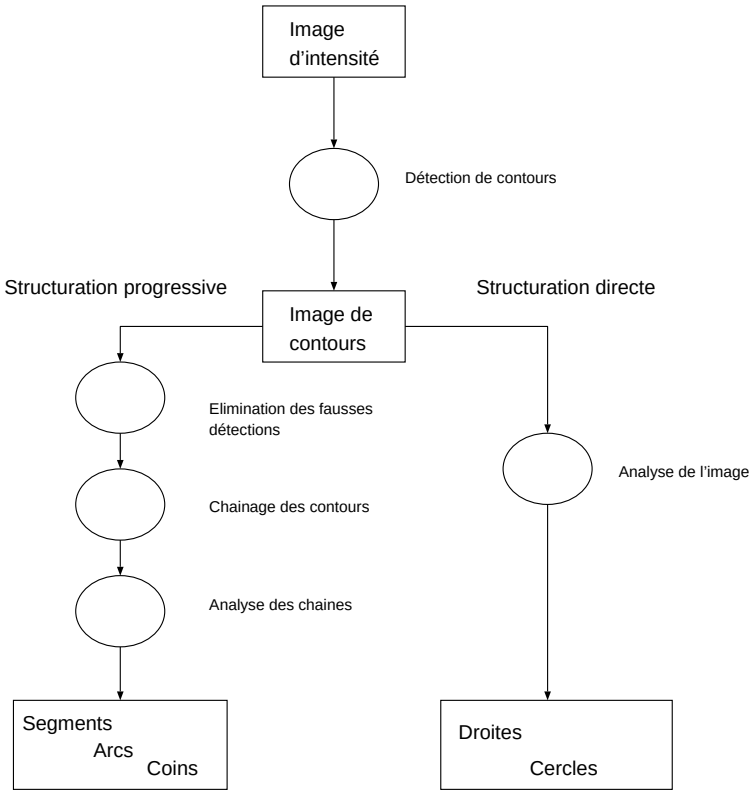

Figure 2.15 - *Structuration de contours après détection - deux niveaux d'application.*

La recherche des meilleurs groupes de points de contours pose un problème de classe N-P complet. Les approches qui existent pour le résoudre ou s'approcher d'une solution se différencient selon leur niveau d'application. La détection de ces primitives peut s'appliquer directement à l'image de contours ou encore, l'image d'intensité. Afin de réduire la complexité d'une telle recherche, d'autres méthodes passent par des représentations intermédiaires pour extraire de l'image de contours une hiérarchie de formes géométriques de plus en plus complexes. Le but de ces techniques reste de réduire le volume de données contenues dans une image, lever

des ambiguïtés issues de la projection de la scène, et rendre l'image manipulable par des systèmes d'interprétation de plus haut niveau.

Notons que les principes d'organisation perceptuelle définis par les Gestaltistes peuvent s'appliquer à chaque niveau de représentation intermédiaire. La frontière entre groupement perceptuel et structuration est suffisamment souple pour permettre de considérer que la plupart des méthodes exposées dans cette partie relèvent de groupements perceptuels. Pour simplifier, nous considérons les méthodes classiques de structuration comme une recherche quantitative de propriétés géométriques (segments de droites, arcs de cercles). Par comparaison, les méthodes relevant du groupement perceptuel recherchent des propriétés visuelles génériques (continuité, proximité, régularité). Le chapitre 3 étant consacré exclusivement à l'application de ces principes en vision par ordinateur, nous nous concentrons pour l'instant sur les approches classiques de structuration des contours.

2.4.1 Structuration directe

Pour faire l'économie de représentations intermédiaires trop nombreuses, une première approche de la structuration consiste à rechercher des primitives géométriques dès les traitements de bas niveau. Nous l'avons vu précédemment, l'utilisation de modèles de contours actifs permet d'extraire de l'image des primitives géométriques directement utilisables sous forme paramétrique. Ces modèles peuvent être appliqués directement à l'image d'intensité [Blaszka et Deriche, 1994a] [Kass *et al.*, 1987] .

L'utilisation d'images issues d'une détection de contours permet naturellement de réduire la complexité du problème en se préoccupant uniquement des points de l'image susceptibles d'appartenir au tracé des primitives recherchées.

C'est le cas de la transformée de Hough, méthode désormais classique pour la recherche de formes géométriques paramétriques. La transformée de Hough divise l'espace des paramètres d'une primitive géométrique en intervalles. Chaque point de l'image ajoute un vote dans cet espace de paramètre pour chaque primitive passant par ce point. Au final, les primitives ayant reçu le plus de votes sont retenues [Princen *et al.*, 1994] [Palmer *et al.*, 1997] . Bien que très efficace en temps de recherche, la transformée de Hough est extrêmement gourmande en ressources mémoires. La taille de l'accumulateur de votes dépend directement de la dimension de cet espace. En pratique, son usage reste limité à la détection de droites et de courbes paramétriques simples.

Cette restriction a contribué au développement de méthodes statistiques afin de réduire l'espace de recherche. Ainsi [Roth et Levine, 1993] proposent une extraction de primitives géométriques à partir de tirages aléatoires parmi un ensemble de points. Pour chaque tirage, le plus simple modèle de primitive géométrique passant par ces points est évalué et retenu s'il correspond à une erreur suffisamment faible. L'estimation du nombre de tirages nécessaires en fonction du nombre de points permet de limiter la recherche. Ce type de méthode permet l'extraction de droites comme de coniques, directement sous leur forme implicite.

2.4.2 Structuration progressive

L'approche classique pour passer d'une représentation à une autre lorsque ces représentations sont trop différentes consiste à réduire la complexité du problème en utilisant une série de représentations intermédiaires. L'image issue d'une détection de contours sert alors de point de départ à cette structuration.

Une bonne représentation de scènes doit respecter un certain nombre de critères. Ces critères sont nécessaires pour obtenir une représentation suffisamment stable pour pouvoir comparer deux vues de la même scène par exemple. Comme toute représentation de formes, le découpage des contours en primitives géométriques doit rester invariant par transformation géométrique. Il doit également rester stable devant de faibles perturbations et occlusions. Les primitives résultant du découpage doivent être décrites simplement, tout en tenant compte de différents niveaux de détails. Le dernier critère à considérer est le coût de calcul qui doit rester raisonnable.

L'organisation, de façon hiérarchique, des primitives géométriques en formes de plus en plus complexes, constitue une bonne approche pour une telle représentation. En effet, l'extraction de chaque primitive à partir d'une portion réduite de l'image assure une certaine stabilité en cas de faibles perturbations ou d'occlusions d'une partie de la scène. De plus, l'aspect local de cette extraction permet également une détection plus efficace. Nous abordons à présent les étapes classiques de ce type d'approche, par ordre d'application dans la chaîne de traitements.

1. **Elimination des fausses détections.**

   Les images de détection de contours présentent de nombreuses imperfections qu'il est nécessaire d'atténuer ou de corriger pour permettre une meilleure extraction de primitives. En effet, le choix de l'opérateur de détection de contours introduit un certain nombre de discontinuités, en particulier autour des coins et des jonctions. Une étape préliminaire à cette extraction est donc la fermeture de ces discontinuités et la mise en valeur des structures linéaires dans l'image. Les méthodes utilisées à cette fin vont de la modélisation par champs de Markov pour fermer les discontinuités [Urago *et al.*, 1992] [Urago *et al.*, 1995] à l'utilisation de filtres directionnels pour renforcer les orientations le long des contours, suivi d'une relaxation stochastique de ces orientations pour mettre en valeur les structures linéaires de l'image [Parent et Zucker, 1989] [Duncan et Birkhölzer, 1992] .

2. **Chaînage des contours.**

   Cette étape est nécessaire pour regrouper les pixels restant selon des chaînes de points connexes [Giraudon, 1987] . Elle marque un changement de représentation, entre la matrice de points que constitue l'image et une première forme de structure linéaire. En tant qu'étape transitoire, elle se doit d'être rapide tout en préservant les relations entre les contours.

3. **Segmentation des contours chaînés.**

En général, la segmentation des contours peut être effectuée de deux manières, indifféremment du type de primitives recherchées. Une première approche consiste à fusionner les points de chaque chaîne en parties homogènes, par exemple, en minimisant l'écart entre une portion de chaîne et un modèle de primitive (droite, arc). Les techniques développées pour une détection de primitives directement à partir de l'image peuvent être appliquées aux points de la chaîne. Par exemple, on peut trouver dans [Gupta *et al.*, 1993] une version adaptée de la transformée de Hough pour décomposer un contour en segments de droites.

L'autre type d'approche procède par division de chaque chaîne selon des points de coupure. La mesure de l'importance d'un point sur une chaîne dépend de nombreux facteurs comme l'échelle ou la résolution à laquelle cette chaîne est observée. L'application destinée au partitionnement de la courbe tient une place importante dans ce partitionnement. Par exemple, on pourra accorder une tolérance plus grande aux changements de courbure si on recherche des objets polyhédriques [Fischler et Wolf, 1994] [Wuescher et Boyer, 1991] .

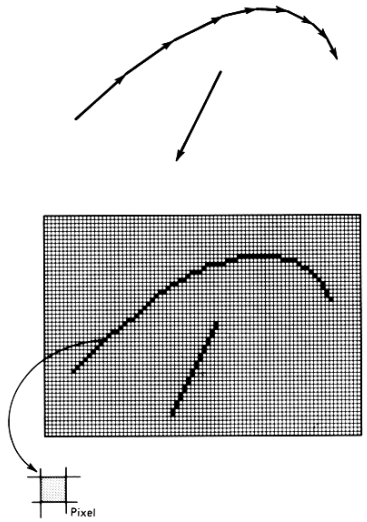

Figure 2.16 - *Ambiguïtés entre segments et arcs en géométrie discrète*

Parmi les nombreuses méthodes de détection de coins ou points de coupure à partir d'une chaîne, la définition la plus utilisée est celle de points de courbure maximale [Wu et Wang, 1993] [Fairney et Fairney, 1994] [Tsang *et al.*, 1994] . Cette détection pose le problème du calcul de la courbure dans un espace discret. Ce calcul est sujet à de nombreuses sources d'erreur dues aux approximations des dérivées et au lissage introduit pour les calculer. On pourra trouver dans [Worring et Smeulders, 1993] une étude complète des différentes définitions possibles pour la courbure, ainsi que les méthodes d'estimation envisageables et leurs erreurs associées. Ce problème étant semblable à ce-

lui de la détection de contours dans un espace mono-dimensionnel, un certain nombre de méthodes proposent d'améliorer le calcul de la courbure dans un espace discret par une approche multi-échelle. La chaîne est interprétée comme une courbe à différentes échelles de lissage, à partir desquelles les points dominants sont détectés. Les points de coupure retenus sont les points les plus stables sur un certain nombre d'échelles [Rattarangsi et Chin, 1992] [Fermüller et Kropatsch, 1992] .

En fonction de la méthode choisie pour détecter les points de coupure, les chaînes peuvent être partitionnées en segments de droites ou en courbes. Nous aborderons plus en détail dans le chapitre 5 les méthodes de segmentation adaptées à chaque type de primitive géométrique, ainsi que les problèmes posés par la différenciation entre segments et arcs.

2.5 Hauts niveaux de représentations

Une fois détectés les éléments de représentation, il est nécessaire de modéliser les relations entre ces éléments pour aboutir à une représentation plus complète de la structure de la scène. A ce niveau là, on suppose avoir obtenu une représentation schématique de la scène, constituée d'arêtes (segments ou courbes) et de jonctions.

Afin de constituer des structures appartenant à des objets propres, les différentes méthodes de représentation dépendent du type d'application recherchée ainsi que du modèle de représentation de haut niveau choisi. Nous proposons dans cette partie un aperçu des différents types de représentations de haut niveau utilisées à partir de scènes de contours.

2.5.1 Représentations bi-dimensionnelles

Le premier type de représentation envisageable consiste à combiner les primitives géométriques en formes 2D plus complexes, suffisamment caractéristiques pour être présentes sur différentes vues de la scène. La scène peut être alors représentée sous forme d'un graphe traduisant les relations hiérarchiques entre ces groupements de primitives [Tomita et Koizumi, 1992] . Un graphe reliant segments et jonctions permet ainsi d'extraire des polygones, convexes ou non, par une recherche de cycles [Wong *et al.*, 1991] . De même, une étude des droites issues des segments permet d'évaluer la position de points de fuites, utiles pour une reconstruction tridimensionnelle ultérieure [Straforini *et al.*, 1993] [Tai *et al.*, 1993] .

L'association de couples de courbes ou de lignes brisées en "vis à vis", constitue une classe de formes fréquemment utilisée. Ces formes peuvent être divisées en deux groupes. D'une part se trouvent les "rubans", obtenus en généralisant la notion de parallélisme à une paire de courbes, moyennant une certaine tolérance. D'autre part se trouvent les symétries, éventuellement penchées. Le terme de "symétrie penchée" en anglais *skewed symmetry* a été introduit par Kanade en 1981. Il désigne une

symétrie entre deux courbes à un angle constant par rapport à un axe incliné. Ce type de symétrie est particulièrement utile pour la recherche d'objets de révolutions dans la scène [Posch, 1992] [Gross et Boult, 1994] . On pourra se reporter à [Ponce, 1988] pour une classification complète de cette classe de formes.

Dans le cas de formes plus générales, des représentations multi-échelles permettent de rendre compte de différents niveaux de détails. Ces représentations peuvent être explicites (à partir d'approximation polygonale successive par exemple) [Bengtsson et Eklundh, 1991] [Chen *et al.*, 1996] ou bien complètement abstraites, à l'aide de descripteurs de formes issus de courbures étudiées à différentes échelles [Mokhtarian et Mackworth, 1992] [Dudek et Tsotsos, 1997] .

2.5.2 Représentations tri-dimensionnelles

La difficulté que pose la perception 3D à partir de simples images a favorisé le développement de méthodes d'interprétations exploitant des représentations 2D. Pourtant, une représentation tri-dimensionnelle restent la façon la plus naturelle pour percevoir les volumes et évaluer les distances. Nous présentons sommairement les nombreuses représentations 3D de scènes utilisées dans des systèmes de vision par ordinateur et ainsi que les principales méthodes d'extraction à partir de représentations 2D.

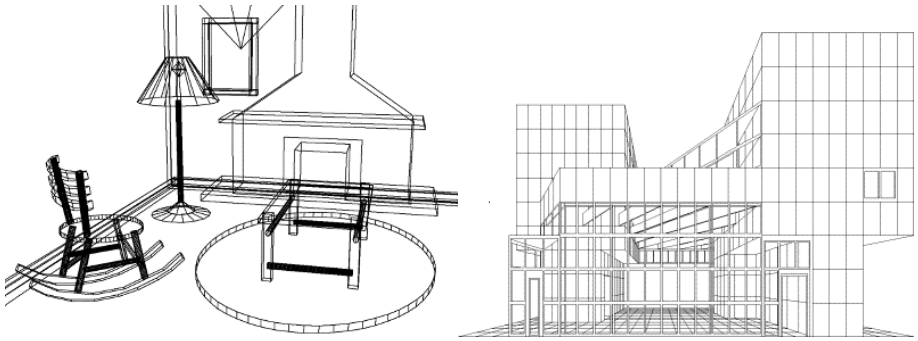

Figure 2.17 - *Modélisation de scène par représentation en fil de fer et par frontières. Le modèle "fil de fer" présente trop d'ambiguïtés pour représenter correctement la profondeur.*

2.5.2.1 Fil de fer

La plus simple des représentations 3D de scènes est la représentation "fil de fer". Elle est constituée d'un graphe de relations dont les noeuds sont des points d'intérêts (sommets, coins, jonctions) et les arcs des primitives les reliant (arêtes rectilignes et courbes). Cette représentation est relativement peu utilisée en vision par ordinateur, en particulier parce-qu'elle ne contient aucune information de volume ni de surface,

et parce-que de nombreuses ambiguïtés d'interprétations la rendent difficile à obtenir (un exemple classique de ces ambiguïtés est donné par le cube de Necker).

C'est pourtant la plus directe à obtenir à partir des primitives des niveaux précédents. Des méthodes de constructions partielles de représentation en fil de fer ont été proposées pour des objets de formes assez génériques, à l'aide de vues multiples de la scène [Pollard *et al.*, 1991] [de Jong et Buurman, 1992] . Dans certains cas, la détection de points de fuites et l'utilisation de connaissances sur la scène sous forme d'arbres permettant d'extraire une représentation 3D de ce type depuis une seule image. Ce type d'approche a été appliqué avec succès à la perception tridimensionnelle de couloirs en vision monoculaire [Brillault, 1992] .

2.5.2.2 Représentation par frontières

Directement issue des travaux sur les mondes de blocs et origami, la représentation par frontières est en général un graphe décrivant la connectivité entre un ensemble de surfaces ou facettes d'une part, et un ensemble de courbes décrivant les intersections entre ces surfaces d'autre part. Ce type de représentation est aussi désigné par le terme de "B-Rep" (*Boundary Representation*) ou "FEG" (*Face Edge Graph*) [Schreiber et Ben-Bassat, 1996] .

La représentation par frontières est une version non ambigüe de la représentation "fil de fer", qui ne tient compte que des surfaces visibles. La représentation "fil de fer" peut servir de point de départ à la construction d'une telle représentation. [Shpitalni et Lipson, 1996] proposent par exemple d'explorer les différents cycles du graphe d'une représentation "fil de fer" pour extraire des facettes. En recherchant des cycles qui ne présentent pas d'intersection avec eux mêmes, l'étiquetage des arêtes en facettes s'en trouve simplifié. Ce type de méthodes est tout de même réservé à des objets polyhédriques.

Dans le cas d'objets plus généraux, la construction de facettes est obtenue par mise en correspondance d'arêtes entre deux ou plusieurs images [Chabbi, 1993] [Tarel, 1996] . Cette construction peut être améliorée si un étiquetage des arêtes délimitant des faces visibles est effectué au préalable [Huynh et Owens, 1994] . Cet étiquetage n'étant pas forcément possible dans le cas d'objets quelconques, il est souvent plus avantageux de n'effectuer qu'un étiquetage partiel. D'un point de vue théorique, cette démarche peut être justifiée par le comportement de la vision humaine en présence de figures impossibles. En particulier, on peut constater des phénomènes de pseudo-stabilité, ou de concurrence entre plusieurs représentations partielles, lorsqu'on observe de telles figures. [Cowie et Perrott, 1993] proposent à ce titre un modèle d'étiquetage partiel qui tient compte de ces phénomènes et tentent de reproduire des mécanismes d'interprétation plus proches de ceux de la vision naturelle.

D'un côté plus pratique, [Malik et Maydan, 1989] ont montré comment utiliser un étiquetage partiel en coopération avec des informations d'illumination (*shape from shading*) afin de déterminer des parcelles de surfaces continues. Ces surfaces

sont représentées à l'aide d'un champ de vecteurs normaux. Il est ensuite possible de passer d'un champ de normales à une représentation 3D des surfaces à l'aide d'une optimisation de surfaces paramétriques, contraintes par ces normales et d'éventuelles discontinuités. Cette méthode a été appliquée avec succès, entre autres, à la reconstruction de surfaces et la modélisation de terrains par [Terzopoulos, 1988] et [Bolle et Vemuri, 1991] .

L'utilisation d'autres contraintes, comme l'exploitation de points de fuites dans le cas de scènes polyhédriques permet aussi de réduire les ambiguïtés d'étiquetage et de construire une représentation 3D [Straforini *et al.*, 1992] [Parodi et Piccioli, 1996] .

2.5.2.3 Représentations par révolutions

Ce type de représentation est généralement défini par un ou plusieurs axes de symétrie et un ensemble de courbes de profils. Le "cylindre généralisé" est un cas particulier de représentation par révolution abondamment utilisé en vision par ordinateur. La représentation d'un cylindre généralisé est simplement constitué d'un axe de révolution et d'une courbe de profil, pas nécessairement liée à l'axe.

La reconstruction de cylindres généralisés à partir de simples images de contours est rendue possible par l'étude des différentes symétries contenues dans l'image, comme l'ont montré [Ulupinar et Nevatia, 1993] [Zerroug et Nevatia, 1996a] . Dans le cas de scènes relativement simples, cette reconstruction peut aussi bénéficier d'hypothèses sur l'illumination des surfaces afin d'augmenter la qualité des résultats.

Pour des objets réels, les contraintes utilisées pour extraire le profil et l'orientation du cylindre dans l'espace sont d'ordre géométriques, comme la présence de points de courbure nulle dans [Richetin *et al.*, 1991] ou fermeture de contours et étude du profil dans [Zerroug et Nevatia, 1996b] . On pourra se reporter à ces derniers pour une présentation détaillée des différentes approches de la reconstruction de cylindres généralisés à partir de contours 2D. Bien que limitée à une certaine catégorie d'objets, ce type de représentation offre l'avantage de décrire simplement une grande classe d'objets tout en conservant des propriétés permettant une segmentation et reconstruction 3D.

2.5.2.4 Représentations par Géons

A l'inverse des précédentes, les représentations par géons rendent compte du côté qualitatif des objets observés. Le principe de "géon" (*geometric ions*) fut introduit en 1987 par Biederman afin de modéliser une reconnaissance de forme par composants. Les géons sont des volumes élémentaires auxquels sont associés un certain nombre d'attributs relatifs aux propriétés de ces composants (symétrie, courbure, profil de coupe). Biederman propose un catalogue d'une trentaine de géons suffisant pour décrire de vastes classes de formes.

Des travaux récents portent sur la génération automatique de telles représentations à partir de l'étude des configurations d'un graphe de relations entre arêtes et

jonctions [Nguyen et Levine, 1996] . Les géons peuvent être utiles pour représenter des objets de manière grossière. Leur côté purement qualitatif empêchent toutefois leur utilisation à des fins de reconnaissance ou de mise en correspondance. En effet, les géons décrivant des classes d'objets, une même représentation peut correspondre à une infinité de variantes. Par exemple, deux tasses différentes auront la même représentation sous forme de géons.

2.5.2.5 Graphes d'aspects et représentations composites

Pour simplifier la reconnaissance de formes à partir de modèles, il peut être utile de décrire un objet à l'aide de différentes vues caractéristiques. C'est le principe des graphes d'aspects. Ce type de représentation est un graphe dont les noeuds sont les vues caractéristiques de l'objet. Les arcs sont représentés par des "évènements visuels", c'est à dire, des changements dans la structure de la projection de l'objet au franchissement de certaines arêtes. Un objet est ainsi représenté selon certaines, ou toutes ses projections possibles.

L'intérêt d'une telle structure est de produire une représentation exhaustive de toutes les projections possibles d'un objet. Il suffit ensuite de comparer les formes caractéristiques extraites de l'image avec les différentes vues pour retrouver l'orientation de l'objet.

Les graphes d'aspects sont exposés à une explosion combinatoire, tant en place mémoire qu'en temps de traitements lorsque les objets décrits sortent de quelques cas très simples. Il existe peu d'algorithmes pour une manipulation d'objets non polyhédriques. Ces graphes nécessitent de plus une qualité de segmentation difficile à obtenir à partir d'images réelles.

La construction automatique de tels graphes se heurte au difficile problème du choix des vues caractéristiques d'un objet. Ce choix peut être en partie automatisé par un échantillonnage d'une sphère de points de vues possibles et une étude de l'évolution de la projection d'un objet selon ces différents points de vue [Weinshall et Werman, 1997] . Devant la difficulté d'une telle représentation, des approches plus pragmatiques se concentrent actuellement sur l'étude des graphes obtenus à partir de certaines classes d'objets, comme les solides de révolutions par exemple [Eggert et Bowyer, 1993] .

2.5.2.6 Représentations volumétriques

Parmi les différentes représentations 3D possibles, les deux dernières variantes sont directement issues de systèmes de CAO. Pour des raisons différentes, chacune de ces représentations est extrêmement difficile à extraire de façon automatique.

La Géométrie Constructive (*CSG - Constructive Solid Geometry*) utilise un ensemble réduit de primitives géométriques 3D (sphères, cylindres, boites), définis par des attributs (échelle, translation, rotation). Un objet est, dans ce contexte, représenté par un arbre d'opérations logiques entre ces primitives, telles que l'ajout

ou la soustraction de volumes. Ce type de représentation se prête bien à des mesures de volumes ou d'intersections. En revanche, à l'exception peut-être de scènes d'objets polyhédriques, il reste extrêmement difficile de construire un arbre d'opérations à partir d'une représentation 2D extraite de l'image [Laurentini, 1997] .

La décomposition en cellules est un autre type de modèle volumétrique. Elle consiste à représenter le volume occupé par un objet à l'aide de volumes élémentaires (ou *voxels*). En général, il s'agit de boites rectangulaires ou bien, dans le cas d'objets courbes, de super-quadriques [Ballard et Brown, 1982] .

En raison de la connaissance qu'elles exigent concernant les objets, ces représentations sont plutôt adaptées à l'utilisation de modèles auxquels on chercherait à comparer des formes caractéristiques extraites de l'image. Par exemple, une représentation par frontières ou en "fil de fer" peut être aisément dérivée d'une représentation volumétrique et faciliter ainsi la comparaison avec une représentation construite à partir de l'image.

2.5.3 Représentations sans reconstruction

Les représentations précédentes ont mis en évidence la difficulté d'obtenir une représentation 3D à partir de projections 2D. L'image correspond seulement aux caractéristiques des parties visibles des objets d'une scène. Même à l'aide de vues multiples, l'écart est souvent trop important à franchir entre images et représentations 3D. C'est particulièrement vrai lorsqu'il s'agit de reconnaître des objets à partir de modèles. Dans ce cas, la reconstruction doit être compatible avec le modèle utilisé. Une représentation de haut niveau n'est pas nécessairement 3D. Elle peut remplir les critères de stabilité exposés au début de ce chapitre sans pour autant être isomorphe à la scène observée. Afin de franchir cet écart, il est souvent préférable d'avoir recours à des représentations intermédiaires associant de façon efficace les caractéristiques d'objets telles qu'elles pourraient être observées sur une image et un modèle de haut niveau qui pourrait être la cause de ces caractéristiques. Ce type de représentation facilite la reconnaissance, même partielle, d'objets dans une scène et permet la construction automatique de modèles à partir d'images.

Dans une première phase d'apprentissage, des relations géométriques entre objets sont décrites de façon paramétrique à partir d'un certain nombre de vues caractéristiques de l'objet. Ces relations forment des points dans un espace de paramètres qu'il suffit de regrouper sous forme d'un index. Lors de la phase de reconnaissance, les relations géométriques entre primitives extraites de l'image sont recherchées dans cet index, afin d'obtenir une correspondance avec un ou plusieurs objets. Une vérification peut être enfin réalisée à partir des propriétés restantes de chaque objet afin de déterminer lequel est le plus probable.

Les techniques d'indexation de modèles ont fait l'objet de nombreuses revues, comme celles de [Gros, 1994] ou bien [Pope, 1994] . Des exemples de construction automatique de modèles peuvent être trouvés dans [Jacobs, 1992] [Gros et Mohr, 1992] [Startchik *et al.*, 1994] [Pope et Lowe, 1996] [Beis et Lowe, 1997] .

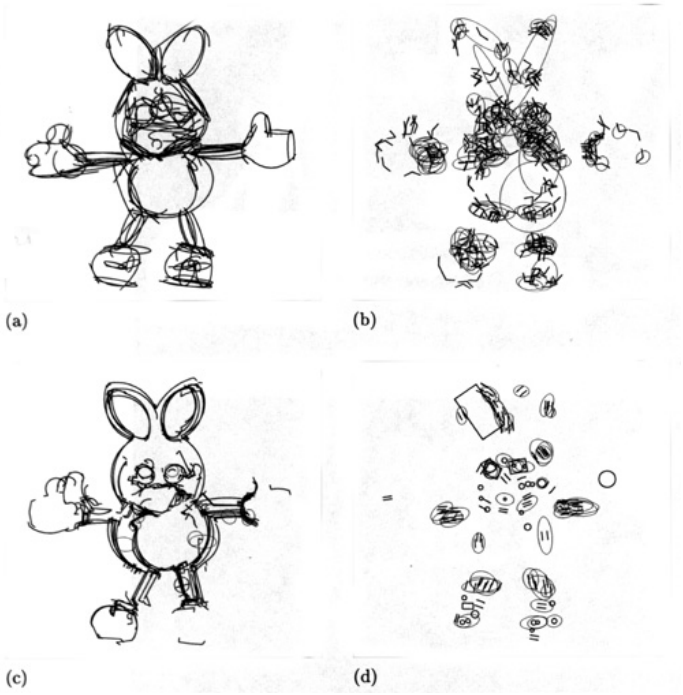

Figure 2.18 - *Graphes de caractéristiques visuelles construits lors de l'apprentissage du modèle - (a) segments et arcs, (b) jonctions, (c) groupes de segments, (d) segments parallèles.*

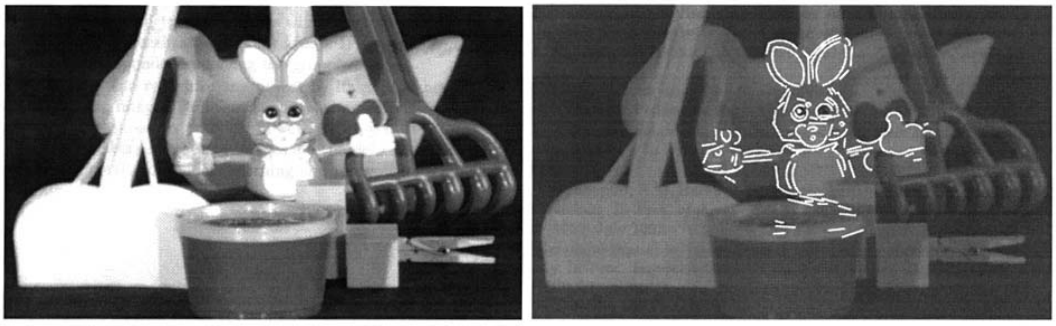

Figure 2.19 - *Reconnaissance de l'objet malgré d'importantes occlusions - Méthode de Pope et Lowe, 1993*

Parmi les représentations intermédiaires les plus utilisées, on peut citer les *Hash tables*, très efficaces pour des espaces de paramètres à faibles dimensions. En particulier, les techniques de *Geometric Hashing* représentent une approche unifiée des problèmes de représentation et mise en correspondance d'objets. L'indexation est réalisée dans ce cas, à partir de caractéristiques invariantes pour un certain nombre

de transformations géométriques [Wolfson, 1990] . Pour des espaces de paramètres de dimensions plus élevées, [Kropatsch, 1995] et [Pope et Lowe, 1993] rappellent l'efficacité particulière de structures de graphes. Ces derniers proposent par exemple un "Modèle d'Apparence" ; un graphe mesurant les relations topologiques entre différentes caractéristiques des images de l'objet, ainsi que des probabilités de positions, orientations et importance selon différentes vues. Cette représentation est très efficace pour retrouver un objet dans un environnement complexe, présentant des occlusions multiples (figures 2.18 et 2.19).

2.6 Conclusion

Nous venons de présenter dans ce chapitre les différentes étapes de l'analyse de scènes à partir des contours, depuis l'acquisition des images jusqu'aux représentations utiles pour les tâches de haut niveau d'interprétation.

La diversité des travaux abordés au cours de cette étude reflète bien l'étendue des difficultés que pose l'exploitation des contours d'une scène par un système de vision artificielle. L'accumulation des problèmes propres à chaque niveau de traitement rend l'extraction d'informations précises particulièrement hasardeuse.

Le choix d'une définition pour les contours recherchés impose au système des contraintes sur le type de scène ou d'images utilisables. La détection des contours apporte sa part de fausses détections liées à des ambiguïtés sur la définition des contours et différentes sources de bruit. L'interprétation des contours pose enfin des problèmes aussi délicats que la distinction entre segments et courbes sur une image nécessairement discrétisée, la localisation de points particuliers, de calculs de courbure où du choix d'une échelle d'observation.

Le but de notre travail est d'extraire les structures les plus régulières à partir des contours d'une image d'intensité. Pour faire face aux ambiguïtés introduites par la détection de contours, nous privilégions autant que possible l'utilisation de critères génériques d'organisation. Ces critères nous permettent d'établir un ensemble d'hypothèses sur les éléments dominants (segments, arcs et points d'intérêt) de la structure des contours. Le rôle de ces éléments est d'attirer l'attention sur les structures globales de la scène afin de réserver les méthodes de structuration plus précises aux éléments les plus probables.

Nous montrons dans le chapitre suivant comment ces principes génériques sont directement dérivés des principes de groupement perceptuel mis en évidence par l'école Gestaltiste de psychologie visuelle, avant d'aborder en détail les différentes étapes de notre approche.

Chapitre 3

Groupement perceptuel en vision par ordinateur

Comme nous l'avons vu dans les chapitres précédents, la structuration d'information à partir d'images représente une tâche extrêmement combinatoire. Parmi les nombreuses tentatives pour réduire cette complexité, l'introduction d'idées issues d'observations de la vision humaine semble une approche naturelle malgré les difficultés de compréhension de la vision biologique.

Ce chapitre est consacré aux méthodes dites de "groupement perceptuel". Le groupement perceptuel désigne la capacité de la vision humaine à former des groupements pertinents à partir d'une image sans aucun soucis d'interprétation.

Avant de passer en revue de quelle manière certains concepts de l'école Gestaltiste de perception ont pu être appliqués à la vision par ordinateur, et d'exposer les grandes lignes de notre approche, nous commençons par détailler les principes de cette théorie.

3.1 Principes d'organisation perceptuelle

L'organisation perceptuelle, telle que définie dans le chapitre 1, page 12, présente de nombreux attraits pour la vision par ordinateur. Les principes de groupement perceptuel permettent de réduire la complexité de la tâche de perception visuelle. L'organisation de données brutes en structures de plus haut niveau, invariantes selon certains points de vue, permet de faciliter la reconnaissance de formes.

Les techniques issues de ces principes sont génériques, appliquables à des données bruitées ou peu fiables, ce qui assure une certaine stabilité aux systèmes de vision ainsi qu'un champ d'application étendu.

Ils peuvent intervenir à divers niveaux d'abstraction, depuis les données sensorielles brutes jusqu'à des représentations de haut niveau. Ils apportent, par leur aspect générique, une alternative à l'application directe de détecteurs de formes spécialisées. Ils permettent enfin d'indexer plus facilement des modèles de scènes par

détection de formes caractéristiques au lieu d'une recherche exhaustive coûteuse.

Toutefois, avant d'aborder les problèmes que pose l'organisation de contours en structures linéaires, il est nécessaire de passer en revue ces principes d'une manière plus précise.

– *Champs perceptuels*

Historiquement, la théorie Gestaltiste de la perception visuelle est liée aux travaux de trois hommes : Wertheimer (1880-1943), Köhler (1887-1967) et Koffka (1886-1941). Leurs travaux ont mis en évidence le caractère dynamique de la perception visuelle, et sa tendance à chercher des solutions simples et cohérentes.

Cette théorie, en particulier dans sa forme liée à la perception visuelle, fut largement influencée par les théories en vogue au début du siècle à propos de phénomènes physiques. Faraday en électricité, Helmholtz pour le magnétisme et Hertz à propos de la gravitation, tous ont eu en commun l'idée d'un "champ" de forces reliant des éléments entre eux par des phénomènes d'attraction ou répulsion.

Köhler tenta de rendre compte des phénomènes observés par les Gestaltistes en des termes similaires. Il définit la notion de "champs perceptuel" présents dans le cerveau. Ces champs de force seraient responsables de l'impression de groupement spontané que donnent des éléments visuels présentant des propriétés similaires. Soumis à ces champs, des éléments visuels semblables entreraient en résonance et donneraient alors l'impression de s'attirer mutuellement selon des groupements.

L'existence de ces champs perceptuels n'a pourtant jamais pu être démontrée clairement. Si cette explication a, depuis, été occultée par de nombreuses découvertes en neuro-physiologie, cela n'enlève rien à la validité des observations des Gestaltistes, encore débattues de nos jours tant au niveau psychologique que neuro-physiologique [Pomerantz, 1981] [Kovács, 1996] [Robert, 1997] .

A défaut d'explications irréfutables, démontrables et prévisibles, la théorie de la perception Gestaltiste reste une théorie descriptive, plus attachée à décrire ce qu'on voit qu'à expliquer comment on peut voir.

– *Séparation entre "figure" et "fond".*

Quelques expériences simples montrent que figure et fond jouent des rôles extrêmement liés. L'un ne peut exister sans l'autre. Dans la figure 3.1, un disque blanc tracé sur un triangle noir est, en général, rarement perçu comme un "trou" découpé dans la surface du triangle. Il apparaît spontanément comme un disque (figure) posé sur un triangle (fond). Pourtant, un faible effort de concentration suffit à inverser les rôles. Un exemple parfait de cette relation est donné par le symbole réversible du Yin et du Yang ou encore par l'illusion classique du vase et des deux visages.

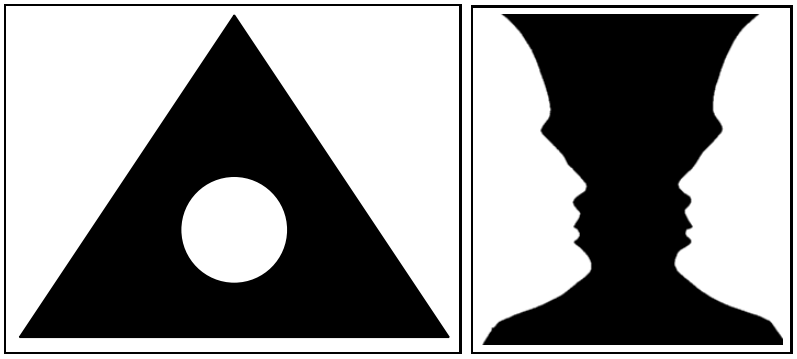

Figure 3.1 - *Séparation entre "figure" et "fond". La figure de gauche représente-t-elle un disque blanc sur un triangle noir? ou bien un triangle percé d'un cercle? La figure de droite représente-t-elle un vase noir ou bien deux visages blancs?*

Au delà de la distinction entre objet et fond se pose le problème de la définition même de ce qu'on voit. Qu'est-ce qui permet, parmi la myriade d'informations visuelles qui nous assaille en permanence, d'avoir une perception stable du monde ? ou en d'autres termes, quelle qualité particulière manifeste un objet pour qu'on puisse le désigner comme un objet cohérent ?

Ces questions conduisent au concept principal de la théorie du Gestalt. On pourrait le résumer par "Les propriétés du tout ne sont pas le résultat de la somme des propriétés des parties" et non, "Le tout est plus grand que la somme de ses parties" comme l'usage en a fait la déformation. L'idée n'est pas d'attribuer une qualité supérieure au tout, mais de remarquer des propriétés différentes entre un tout, et chacune de ses parties prise séparément. Ainsi, un rectangle présente une certaine qualité de "rectangularité", qu'il est impossible de percevoir en observant chacun des segments qui le composent.

– *Prägnanz et Gestaltqualität*

Deux termes ont été définis pour désigner différents aspects de ce principe. D'une part, la "prégnance" ou idée de "bonne forme", de l'allemand *Prägnanz*. Ce terme rend compte d'une idée de saillance, de mise en valeur de certaines propriétés de façon immédiate, avant tout processus d'interprétation. Cette saillance ne relève pas uniquement de la régularité géométrique. Elle met aussi en jeu des notions de stabilité, de simplicité, de cohésion et même, d'habitude. En effet, en cas d'ambiguité entre deux formes possibles issues d'une même image, le choix se portera plus facilement vers la forme dont on a l'habitude.

Le second terme représente la qualité de "forme" qu'on peut attribuer à un ensemble d'éléments, ou en allemand *Gestaltqualität*. Ce terme implique qu'il existe entre ces éléments, une organisation "ordonnée, commandée par certaines règles et non accidentelle", pour reprendre les termes de Wertheimer.

Cette organisation intervient grâce à un ensemble de groupements entre éléments visuels selon des règles simples, effectués de manière immédiate, avant interprétation. La proximité représente une loi majeure pour ces groupements, mais nous verrons plus loin que d'autres lois existent, coopèrent ou s'opposent.

3.2 Règles de groupement

La notion de groupement perceptuel peut donc être définie comme l'organisation de données sensorielles en groupes correspondant à des causes communes. Elle peut être aussi interprétée dans une certaine mesure comme un besoin fondamental d'organiser notre perception du monde de manière à pouvoir l'appréhender.

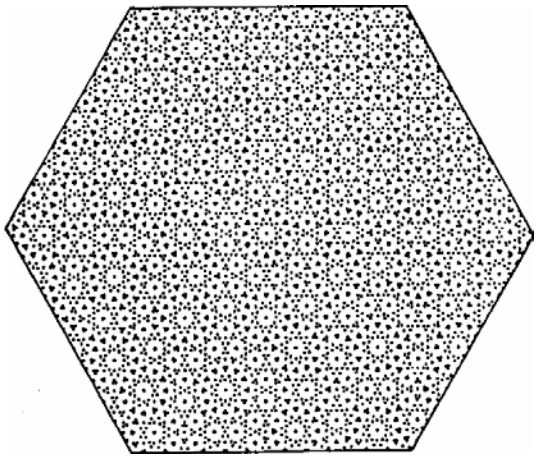

Figure 3.2 - *Les motifs de Marroquin révèlent le côté dynamique et continu des processus de groupements perceptuels.*

Les motifs de Marroquin illustrent bien comment ce phénomène de groupement intervient immédiatement lors d'un surplus d'information visuelle. Nous faisons face à ce surplus soit en filtrant délibérément une certaine part de cette information, soit en la simplifiant à partir de groupements perceptuels. Nous serions donc à la recherche de stabilité, de simplicité et de cohésion pour percevoir le monde de manière non chaotique. Cette simplification va même jusqu'à ajouter des éléments en cas de besoin pour assurer une stabilité, comme le montrent des exemples de contours subjectifs.

Les Gestaltistes ont catégorisé les différentes propriétés visuelles des groupements selon les règles suivantes [Wertheimer, 1923] .

– **Proximité**

C'est la relation de groupement la plus naturelle. Plus deux éléments visuels sont proches, plus ils ont de chances d'appartenir à un même groupe ou motif. La notion de proximité est cependant soumise à la définition d'une échelle d'observation.

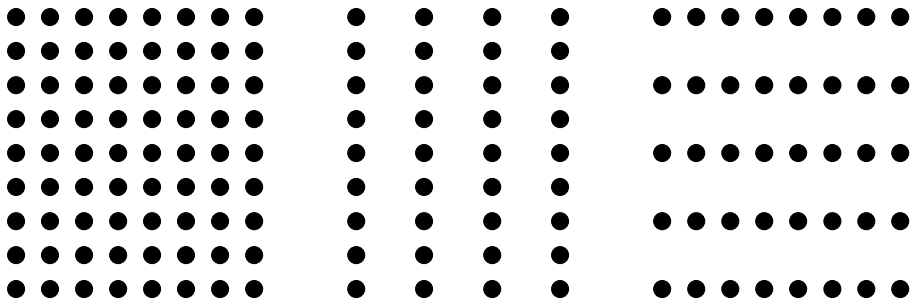

Figure 3.3 - *Groupement par* **proximité** - *Toute chose égale par ailleurs, les éléments visuels de cette figure sont groupés par lignes ou par colonnes selon leurs distances respectives.*

– **Continuité**

Les groupements sont favorisés par le minimum de changements. Cette propriété favorise l'émergence de formes lisses et régulières. Elle assure une plus grande stabilité aux structures perçues.

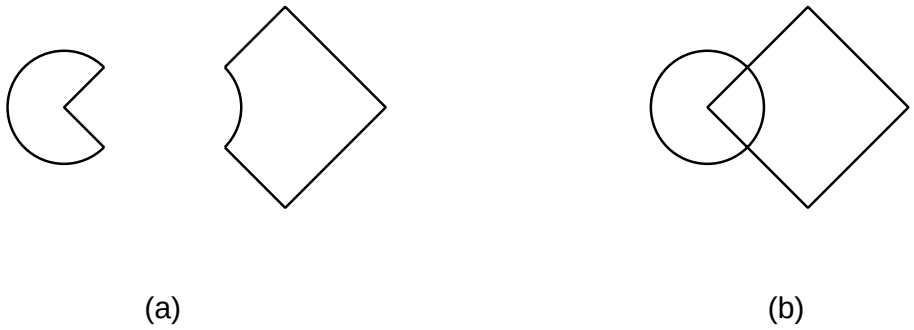

Figure 3.4 - *Groupement par* **continuité** - *Les figures distinctes de l'exemple (a) s'effacent au profit de figures plus continues dans l'exemple (b). Il est ainsi difficile de voir dans (b) autre chose qu'un cercle complet et un carré complet qui se superposent.*

– **Similarité et symétrie**

Les groupements tendent à respecter des propriétés visuelles communes entre objets et associent entre eux les éléments visuels appartenant à un même objet, une même source, un même mouvement. Dans la pratique, la similarité s'avère difficile à mesurer. Elle peut affecter la forme, la couleur, la texture, l'intensité, l'orientation. Cependant, les Gestaltistes ne sont pas très clair sur l'importance relative de ces propriétés. Selon quels critères privilégier des groupements selon la forme plutôt que selon la couleur ou l'orientation ?

La similarité peut être considérée comme un certain type de proximité entre propriétés d'éléments visuels. Une proximité de forme ou de couleur au lieu de proximité spatiale en quelques sortes.

La symétrie est aussi un cas particulier de similarité. Qu'elle soit radiale, axiale, ou éventuellement "penchée", la symétrie renforce l'aspect visuel d'une forme, et sa prédominance par rapport à d'autres. Dans un environnement visuel quelconque, la symétrie est un bon indicateur de portions planes. Ce qui explique utilisation fréquente des symétries pour retrouver une forme à partir d'une représentation de contours.

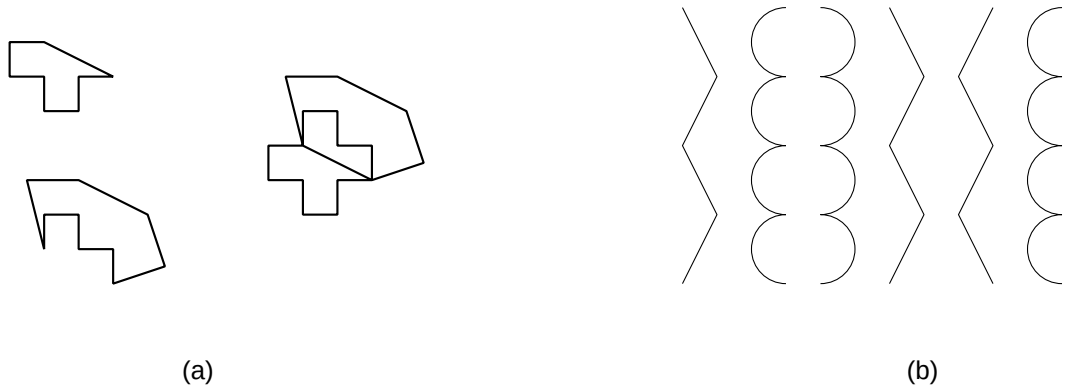

Figure 3.5 - *Groupement par* **symétrie** - *L'exemple (a) montre l'importance de la symétrie dans l'apparition de formes saillantes. L'exemple (b) montre comment l'importance de la proximité est atténuée par l'existence de symétries.*

Enfin, lorsqu'il s'agit de propriétés liées au mouvement, ou à la vitesse d'éléments visuels, on parle de règle de "cause commune" (*common fate*). Les éléments présentant les mêmes caractéristiques de déplacement sont groupés entre eux.

– **Fermeture**

Une grande importance est attribuée à des formes aussi complètes que possible. Les groupements reflètent, de préférence, des formes fermées, convexes, symétriques. La fermeture est l'une des caractéristiques proposées par les Gestaltistes pour distinguer une figure, stable et structurée, par rapport au fond, souvent "ouvert" et peu organisé. Lorsque une figure et le fond partagent un même contour, celui-ci est attribué à la forme et non au fond. Par extension de ces observations, des formes partiellement ressortent d'une façon plus vive. On peut de la même manière considérer la complétion de parties manquantes, d'occlusions et les contours fictifs comme des cas particuliers de fermeture.

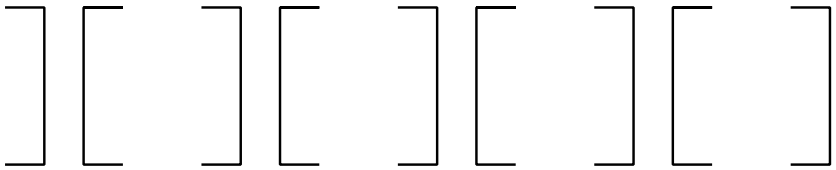

Figure 3.6 - *Groupement par* **fermeture**

– **Familiarité et contexte**

Comme nous l'avons évoqué précédemment, lorsque la situation se présente, des groupements semblables à des objets familiers sont réalisés de préférence. Les Gestaltistes ont souligné de même l'importance de l'attention et de l'intention de l'observateur dans les groupements. De même, le contexte de l'environnement observé joue un certain rôle dans ces phénomènes, comme le montre l'ambiguité entre la lettre "B" et le nombre "13" dans la figure 3.7.

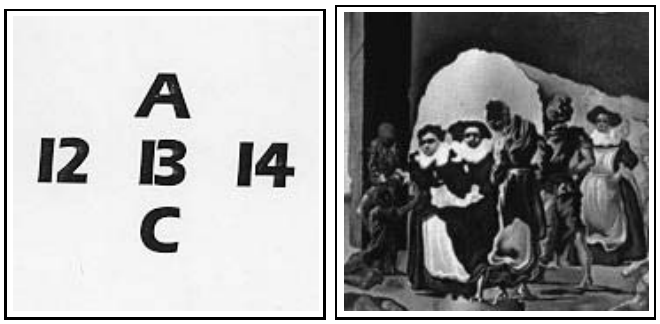

Figure 3.7 - *Groupement par* **contexte** *et par* **familiarité** *- Selon le sens de lecture, les éléments visuels "1" et "3" sont groupés pour former une lettre ou bien séparés pour former un nombre. L'autre figure représente une ambiguïté entre un groupe de personnes et un visage. Le visage est d'autant mieux perçu que son modèle, un buste célèbre de Voltaire, est connu des observateurs.*

Les interactions entre différentes règles de groupements sont multiples, et souvent concurrentes ou contradictoires. D'un côté, la tâche de groupements peut être facilitée par la redondance d'informations similaires. En effet, lorsque des groupements selon des critères différents se confirment mutuellement, le nombre de possibilités de groupements se réduit d'autant, ainsi que les incertitudes sur l'image.

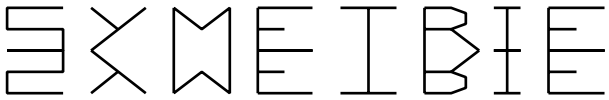

Figure 3.8 - *Le sens du mot "SYMETRIE" disparaît devant l'influence du groupement par continuité et par symétrie.*

D'un autre côté, il n'existe aucune règle claire qui permette de prévoir la prédominance de certains principes de groupement sur d'autres. En pratique, la continuité, la proximité et la fermeture jouent un rôle plus important que la symétrie. Les groupements par continuité, symétrie et fermeture sont même plus importants que l'interprétation, comme on peut le constater en plaçant une phrase face à son symétrique. Les formes qui émergent occultent complètement le sens de la phrase (Figure 3.8).

Kanizsa (1979) a montré que lorsqu'ils doivent faire des distinctions entre figure et fond, les sujets privilégient la continuité de direction et la convexité par rapport à la symétrie (Figure 3.4).

On peut enfin exprimer plusieurs de ces lois en fonction des autres. Par exemple, la fermeture peut être vue comme une conjugaison de groupements par proximité, continuité et similarité. De même, les contours fictifs peuvent être considérés comme une "continuité de discontinuités".

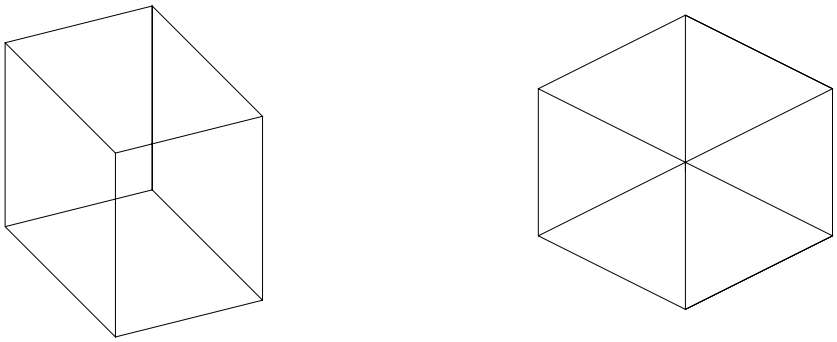

Figure 3.9 - *Principe de simplicité - En l'absence d'autres indices, la figure de gauche apparaît comme la projection 2D d'un cube 3D alors que celle de droite apparaît comme un motif uniquement 2D.*

Malgré l'absence de loi prédictive et explicative de ces phénomènes, il est possible de trouver un dénominateur commun à ces phénomènes de groupements. Les notions de *Prägnanz*, de stabilité, de complexité minimum ont toutes en commun un certain principe de simplicité.

" Toute chose égale par ailleurs, la réponse perceptuelle à un stimulus sera celle qui nécessitera le minimum d'information pour la décrire" Hochberg (1957)

Les principes de groupements peuvent être exprimés en termes de minimum de changement. La proximité reflète un minimum de distance spatiale, la similarité, un minimum de changement de propriété, la fermeture, un minimum de discontinuités et la continuité, un minimum de changement brusque. Même la familiarité peut être vue comme la recherche d'un minimum de surprise, de forme inconnue. En cas d'ambiguité entre plusieurs interprétations, la plus simple est souvent privilégiée avec une préférence pour les interprétations tri-dimensionnelles comme le montre la figure 3.9.

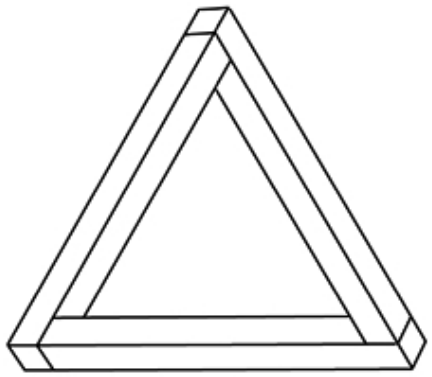

Figure 3.10 - *Triangle de Penrose. L'impression d'un "Tout" cohérent apparaît bien avant de remarquer que cette figure est physiquement impossible. La structure de chaque sommet, observée indépendamment des autres, est cohérente localement. L'agencement de chaque sommet est cohérent deux à deux, ce qui renforce l'illusion. L'instabilité de la figure est pourtant secondaire devant l'illusion d'un objet unique, et n'intervient que lorsqu'on interprète la figure plus en détail.*

Il est aussi intéressant de noter que notre tendance à grouper naturellement des éléments visuels selon des assemblages stables et continus, fait automatiquement ressortir ce qui n'obéit pas à ces règles de groupements. L'oeil est attiré plus particulièrement par les discontinuités car elles correspondent à des zones qui nécessitent une attention particulière. D'où la possibilité de ne pas "remarquer" quelque chose si sa présence est trop improbable et sa forme pas assez distincte d'un environnement homogène. L'efficacité du camouflage en est un parfait exemple.

Pour Kanizsa, ces mécanismes de groupements font partie d'un processus en deux étapes. Dans un premier temps, le champ visuel serait divisé en régions présentant des régularités temporelles et spatiales. Un "Tout" serait ensuite composé, par déduction des parties manquantes, complétion des discontinuités et enfin, vérification sommaire de la cohésion de l'ensemble. Différentes figures impossibles permettent d'avoir une bonne idée de ce fonctionnement (Figure 3.10).

Ce type de mécanisme est intéressant en vision par ordinateur, lorsqu'on sait combien il est difficile d'obtenir une segmentation claire à cause du bruit de l'image ou des occlusions. Modéliser ces phénomènes permet de rejeter le plus tard possible l'intervention de niveaux d'interprétation. Ceux-ci n'ont plus qu'à s'assurer de la cohérence globale entre structures construites à l'aide de groupements perceptuels.

3.3 Application à la vision par ordinateur

Il ressort des exemples précédents que le groupement perceptuel constitue un ensemble d'heuristiques robustes pour la vision par ordinateur. En effet, les parties d'un même objet sont susceptibles d'être proches (groupement par *proximité*) et de partager la même texture de surface (groupement par *similarité*). De plus, les occlusions sont susceptibles d'engendrer des parties semblables de part et d'autre (groupement par *continuité*) et les objets ont en général une forme fermée (groupement par *fermeture*).

Pourtant, malgré ses nombreux avantages, le groupement perceptuel joue un rôle plutôt tardif dans l'historique de la vision par ordinateur. En effet, des concepts tels que la "structure", la "saillance d'une forme", ou un "principe de simplicité" sont difficiles à définir clairement, et encore plus difficiles à modéliser à l'aide d'algorithmes. La théorie de la perception Gestaltiste reste souvent vague bien que traitant de phénomènes indéniables. Elle ne permet pas, en tout cas pour l'instant, une interprétation physique directe contrairement à des tâches telles que la détection de contours par exemple.

Pendant des années, ce côté important de la vision humaine a été laissé de côté au profit de méthodes plus quantitatives d'analyse d'images. Les principes de groupements perceptuels ont d'abord été utilisés de manière isolée, pour répondre à des problèmes bien spécifiques tels que la détection de formes géométriques, l'approximation polygonale, ou la perception d'orientations privilégiées pour détecter des structures linéaires dès les premiers niveaux de segmentation [Zucker, 1983] .

Marr fut l'un des premiers à suggérer l'importance de groupements perceptuels en vision artificielle. Son "ébauche préliminaire" (*primal sketch*) est un niveau de représentation qui implique l'utilisation d'informations provenant à la fois des contours et du regroupement de primitives entre elles (courbes ou lignes). Il propose de constituer des groupements selon des caractéristiques de l'image telles que des courbes continues, segments parallèles ou colinéaires, ou textures semblables en intensité, taille, orientation et densité. Cet aspect est resté malheureusement peu développé

dans cette théorie. Admettant la complexité du problème de séparation figure-arrière plan, Marr suggère de se concentrer plutôt sur des problèmes dérivant d'une théorie mieux définie, comme la détection de contours ou la reconstruction de surface à partir d'illumination.

Afin d'appliquer les principes d'organisation perceptuelle à la vision par ordinateur, il est donc nécessaire de formaliser ces phénomènes d'une manière plus claire. Comme le montrera le sous-chapitre 3.4, chaque application procède de son propre formalisme. Le principe de simplicité a été ainsi défini comme la recherche de modèles simples en des termes de théorie de l'information, optimisation d'une mesure de qualité ou d'énergie, ou encore par une approche fréquentielle.

3.3.1 Principe de "non-accidentalité"

L'une des tentatives d'explication les plus utilisées, en grande partie à cause de son élégance, est connue sous les différents termes de principe de "non-accidentalité", de "coïncidence", ou encore, principe de "cause commune". Elle fut introduite en vision par ordinateur par Witkin, Tenenbaum et Lowe afin de rendre compte de la préférence de perception de certaines formes par rapport à d'autres. Une forme dans une image a une plus grande chance de signifier quelque chose si elle est peu susceptible de se produire par accident.

[Witkin et Tenenbaum, 1983] furent parmi les premiers à souligner l'importance de concepts issus de l'école Gestaltiste afin de rendre compte de la structure présente dans une image. Ils définissent la structure d'une image par la présence d'organisation cohérente et régulière, présente à différentes échelles. Cette notion est indépendante de toute interprétation, et plus difficile à formaliser que d'autres problèmes, mieux délimités, tels que la recherche de cercles ou de toute autre forme bien définie. Leur tentative de comprendre la contribution de ce type de structure primitive dans la tâche de la vision artificielle aboutit à une définition du rôle fonctionnel de l'organisation perceptuelle en vision par ordinateur. Ils attribuent aux groupements perceptuels un rôle de "précurseurs sémantiques" [Witkin et Tenenbaum, 1986] , dont la fonction serait alors de détecter la présence possible d'éléments visuels dignes d'intérêt.

En effet, des relations régulières ont en général peu de chance d'apparaître par hasard. Par conséquent, leur apparition reflète certainement une cause commune, suffisamment importante en tout cas, pour en tenir compte. Les groupements sont donc utiles pour déclencher une interprétation plus approfondie, en fonction du contexte par exemple. De la même manière, [Lowe, 1985] propose une conception probabiliste du groupement perceptuel.

En partant de l'observation que la vision peut être aisément leurrée, par des illusions ou par l'efficacité du camouflage, Lowe suggère que la possibilité d'une mauvaise interprétation d'image en vision par ordinateur doit être non seulement acceptée mais doit participer au processus d'interprétation. Au lieu de chercher une interprétation exacte ou correcte, il s'agit de rechercher l'interprétation la plus

probable. La tâche de reconnaissance visuelle est définie en tant qu'optimisation des probabilités d'identifier des propriétés visuelles correctes dans l'image. Il rejoint ici le principe de simplicité ou de *Prägnanz*. Les groupements apparaissent donc importants lorsque leur probabilité d'apparaître par accident est faible.

D'une façon plus formelle, [Sarkar et Boyer, 1993b] proposent l'interprétation suivante.

Soit un ensemble d'objets. Si on définit la "Causalité" par l'appartenance de ces objets à une cause commune, et "l'Organisation" par le niveau de structure de chaque objet, on peut évaluer l'importance d'un groupement en appliquant la loi de Bayes de la façon suivante :

$$P(\text{Causalité}|\text{Organisation}) = \frac{P(\text{Organisation}|\text{Causalité}).P(\text{Causalité})}{P(\text{Organisation})}$$

Ce qui peut être interprété de la façon suivante : l'importance d'un groupement est d'autant plus grande que sa probabilité d'apparition accidentelle est faible et que la probabilité de provenir d'un objet unique est grande.

En effet, les différentes probabilités du terme de droite sont respectivement :

- $P(\text{Causalité})$

  Probabilité que les objets considérés proviennent tous d'un même ensemble (parties d'un seul objet). Dans une scène purement chaotique, cette probabilité serait nulle. Elle est en général relativement élevée.

- $P(\text{Organisation})$

  Probabilité de trouver de l'organisation au sein du groupe des objets considérés. C'est la probabilité d'un groupement accidentel. Elle décroît avec le nombre d'objets présents dans le groupe.

- $P(\text{Organisation}|\text{Causalité})$

  Probabilité d'observer une organisation dans l'ensemble d'objets sachant qu'ils ont une origine commune. En général, elle a aussi une valeur relativement forte.

Le principe de "non-accidentalité" est commun à de nombreuses méthodes de groupement perceptuel en vision artificielle. Pour mettre en application les principes de groupement perceptuel, le choix d'un formalisme ne suffit évidemment pas. Il reste encore à fixer les règles de groupements à appliquer et le type d'éléments visuels à grouper.

3.3.2 Choix des règles de groupements

L'une des fonctions principales du groupement perceptuel est de faciliter la reconnaissance de formes, ou des parties d'objets entre images, ou bien entre images et modèles. Les structures issues du groupement doivent donc exhiber des propriétés

visuellement stables selon un certain nombre de vues et attirer l'attention sur les éléments de l'image dignes d'analyse.

Lowe définit ainsi la notion d'invariants, propriétés visuelles plus ou moins stables pour des projections d'une même scène selon différents points de vues. Seuls les éléments visuels présents sur de nombreuses projections de la scène sont susceptibles de provenir d'un phénomène physique commun dans la scène tridimentionnelle. En d'autres termes, les structures 2D obtenues par groupement perceptuel doivent être utiles pour établir des hypothèses solides sur des structures 3D.

Par exemple, les segments de l'image ont de fortes chances de correspondre à des arêtes dans la scène 3D. De même, les jonctions entre courbes ou segments dans l'image ont de fortes chances de correspondre à des sommets ou des occlusions en 3D. Il existe toujours une possibilité d'accidents visuels, où ce genre de configuration pourrait ne correspondre à rien d'autre qu'un alignement fortuit. Ce type d'accident peut entraîner une hypothèse incohérente avec la représentation en cours de construction ou avec un modèle, auquel cas il est toujours possible d'ignorer cette hypothèse, ou au moins de réduire sa crédibilité.

La complexité d'un groupement 2D ou la redondance d'indices ne font qu'augmenter la probabilité de structures 3D particulières. Il en est ainsi pour des structures régulières telles que des courbes continues, des segments parallèles, des arrangements convexes ou symétriques, encore, la présence de motifs similaires répétitifs de couleurs, formes ou textures si on admet des hypothèses plus complexes.

Il est intéressant de noter que ces hypothèses correspondent aux lois de groupements mises en évidence en vision humaine, ce qui assure une certaine cohérence avec les observations réelles. L'utilisation de groupements invariants permet de s'affranchir d'hypothèses sur le contenu de la scène par opposition aux techniques d'étiquetage de représentations par contour vues dans le chapitre précédent.

On peut noter enfin que les concepts avancés par les Gestaltistes n'ont été appliqués qu'en partie à la vision par ordinateur. En plus de la simplicité et de la régularité, des notions telles que l'attention, l'intention, ou le contexte jouent des rôles importants dans la vision "Gestaltiste". Ce manque d'intérêt est principalement dû à la difficulté de formalisation. A ce titre, on peut espérer que les formalismes établis pour rendre compte du concept de *Prägnanz* ne soient qu'un premier pas vers l'introduction d'autres concepts Gestaltistes en vision par ordinateur.

3.3.3 Choix des primitives à grouper

Dans la pratique, le problème de groupement est le plus souvent présenté comme l'aggrégation de points d'intérêts selon des structures de plus haut niveau (chaînes ou bien régions). D'autres approches existent, selon la nature et la complexité des primitives utilisées pour le groupement. Ces primitives peuvent être regroupées selon des formes indéterminées (telles que des chaînes, des courbes quelconques ou bien des régions) ou bien selon des formes paramétriques (cercles, polygones, ellipses). Peu de travaux ont exploité la possibilité de grouper des objets plus complexes entre

eux (tels que le groupement de polygones ou d'ellipses selon des rubans dans une séquence vidéo).

Un système de classification des différentes méthodes de groupement perceptuel en vision par ordinateur a été défini par [Sarkar et Boyer, 1993b] . Leur nomenclature classe ces méthodes selon la dimension des groupement effectués (signal, primitive, structure ou assemblage) et la dimension du domaine d'application (2D, 3D, 2D+temps, 3D+temps).

En se conformant à cette nomenclature, les groupements "$3D$" correspondent à des cartes de profondeurs issues d'acquisition d'images par des capteurs actifs (lasers par exemple) ou bien par triangulation en vision stéréoscopique. Etant donné que ce mode d'acquisition ne donne qu'une information sur les surfaces visibles, ces groupements sont aussi désignés par le terme de groupements "$2D\ \frac{1}{2}$" en référence au niveau de représentation intermédiaire de Marr. Sarkar et Boyer proposent ensuite de classer à part les méthodes associées à des séquences d'images, avec les catégories "$2D + temps$" et "$3D + temps$".

Dans le cadre de notre étude, nous nous intéresserons en particulier aux méthodes "$2D$" de groupements perceptuels, c'est à dire, les méthodes associées à des images d'intensité lumineuse. Notons aussi que parmi les méthodes décrites ci-après, un certain nombre pourrait figurer dans le chapitre précédent. C'est particulièrement vrai pour les groupements de bas niveau ("signal" et "primitive"). Les méthodes de segmentation et structuration vues précédemment peuvent en effet être utilisées pour fournir des éléments de groupements aux méthodes de niveau supérieur d'organisation. Les méthodes sont abordées dans ce chapitre en fonction des principes de groupement perceptuel qu'elles exploitent. Pour rester dans le cadre de notre étude, nous nous focalisons en particulier sur les groupements de structures linéaires.

Une bonne illustration de cette distinction est la transformée de Hough. Bien qu'elle soit particulièrement efficace pour détecter des droites, cette méthode ne permet pas de donner un caractère "perceptuel" aux droites détectées. Elle ne tient pas compte de notions de qualité visuelle, de proximité ou de continuité par exemple. En revanche, elle peut parfaitement être utilisée comme pré-traitement pour une méthode de groupement de droites qui se chargerait, elle, de leur attribuer une qualité visuelle.

3.4 Techniques de groupement de contours

La détection de structures perceptuellement importantes pose de nombreux problèmes de complexité et de stabilité. Les méthodes de groupement perceptuel font appel à de multiples techniques, souvent adaptées au type de données et de problème envisagé.

Encore une fois, le but n'est pas ici de construire une représentation complète de l'image mais d'élaborer des hypothèses fortes à partir de structures visuelles présentes dans l'image. Le principe de ce type d'approche consiste donc, en général, à

élaborer des hypothèses de plus en plus structurées afin d'aboutir à des représentations de haut niveau. Il est donc nécessaire de disposer de méthodes pour comparer ces hypothèses, les combiner, afin d'en rejeter certaines, d'en déduire de nouvelles ou de résoudre les conflits éventuels entre hypothèses.

Ces hypothèses devront ensuite servir de contraintes pour des niveaux supérieurs d'interprétation de même que les hypothèses de formes que nous pouvons faire à partir de dessins d'objets restent cohérentes avec les éléments visuels présents dans ces dessins.

Ici encore, on ne s'intéressera qu'à des exemples de groupements de structures curvilignes, représentatifs de chacune de ces méthodes. Pour plus de détails et de références, on pourra se reporter à l'étude de [Sarkar et Boyer, 1993b] sur ces différentes techniques.

3.4.1 Approches algorithmiques

Les approches algorithmiques du groupement perceptuel consistent à rechercher les groupements intéressants parmi tous les groupements possibles, en appliquant un certain nombre de règles ou d'heuristiques pour simplifier la complexité combinatoire du problème. Ces règles peuvent être appliquées de manière automatisée, à l'aide de systèmes experts par exemple. Ou bien de manière explicite, en ne recherchant que certaines configurations dans un voisinage limité autour des primitives visuelles à grouper.

La démarche de [Lowe, 1985] tombe dans cette catégorie. Son système SCERPO constitue des groupements de manière simple et efficace. Les structures obtenues sont des paires de segments parallèles, proches ou colinéaires. La recherche est effectuée sur un voisinage limité autour de chaque segment pour des raisons de complexité. Des mesures locales de probabilités de groupements tiennent compte de la densité de segments dans le voisinage et de la probabilité de non-accidentalité de chaque configuration. Un groupe est d'autant plus significatif qu'il a peu de chance de se produire. Les groupements sont finalement utilisés pour élaborer des hypothèses de structures 3D et faciliter la reconnaissance de formes dans l'image à partir de modèles.

[Horaud *et al.*, 1990] proposent une approche hiérarchique de groupements d'éléments de contours selon des entités de plus en plus complexes. Chaque relation de groupement entre segments est modélisée par un graphe de relations. Ainsi, un rectangle est représenté par un graphe entre deux groupements de segments parallèles, chacun étant un graphe entre deux segments. Ce type de graphe perceptuel présente l'intérêt d'être relativement proche d'une représentation "fil de fer" qu'on pourrait obtenir à partir d'un modèle, l'idée étant de faciliter encore une fois la reconnaissance de formes à partir de modèles.

On pourra trouver dans [Jacobs, 1996] une étude de fond sur la complexité et l'efficacité d'une recherche intensive de groupements. Pour détecter les meilleurs groupement de segments en ensembles convexes, Jacobs justifie l'utilité d'une re-

cherche quasi-exhaustive. La complexité de la recherche est contrôlée par l'usage des propriétés "non-accidentelles" des ensembles recherchés. Les groupes convexes de segments, avec peu de discontinuités, ont peu de chances d'apparaître par accident. Les segments sont ajoutés, de proches en proches, et chaque groupement est évalué pour arrêter la recherche s'il devient trop improbable. Cette démarche a été appliquée avec succès pour l'indexation automatique de modèles et la reconnaissance de formes 3D à partir de caractéristiques 2D.

3.4.2 Méthodes d'optimisation

Comme nous l'avons vu en début de ce chapitre, les tentatives des Gestaltistes pour expliquer le groupement perceptuel par champs perceptuels semblables à des champs électromagnétiques furent mises en défaut par des découvertes en neurophysiologie. Pourtant, cette intuition a inspiré de nombreuses approches de groupement par optimisation d'énergie. Selon ce formalisme, l'organisation perceptuelle est représentée sous la forme d'une mesure de qualité visuelle qu'on cherche à optimiser. Cette fonction de qualité reflète en quelque sorte la saillance d'un groupement, ou sa probabilité de se produire ou non par accident.

Nous n'avons trouvé que peu de travaux exprimant directement l'organisation perceptuelle à partir de modèles physiques de champs électromagnétiques. Un exemple est cependant proposé par [Pun, 1992] pour des groupements par proximité et orientation.

La plupart des autres approches diffèrent par le choix de la mesure de qualité. Selon les modèles mathématiques utilisés pour l'optimisation, la mesure de qualité peut être décrite en des termes d'énergie ou de probabilités. Elle offre l'avantage certain de pouvoir rendre compte de groupements globaux à partir de l'optimisation de mesures locales et des contributions des éléments visuels voisins sur ces mesures. Toutes ont pour point commun l'utilisation de réseaux d'éléments localement connectés.

[Grossberg et Mingolla, 1985] furent parmi les premier à souligner le besoin de faire appel à des techniques adaptées aux coopérations d'éléments en réseaux pour des tâches de groupement. Ils proposent en particulier une théorie et un ensemble de règles génériques capables d'expliquer comment des mesures locales et des structures globales peuvent travailler ensemble pour segmenter une scène. Cette approche leur permet enfin d'apporter un éclairage nouveau sur un grand nombre de phénomènes psycho-visuels tels que les contours fictifs ou les frontières entre différentes textures.

Nous présentons ici quelques exemples parmi les méthodes d'optimisation les plus utilisées.

- Etiquetage et relaxation

  Appliquées au groupement perceptuel, les techniques d'étiquetage consistent à définir des classes d'éléments visuels appartenant à des structures globales. Elles offrent l'avantage de trouver des solutions qui répondent globalement

aux contraintes imposées par les éléments visuels. Ce qui les rend robustes aux perturbations locales.

La relaxation est l'une des techniques d'étiquetage les plus utilisées. Elle est appliquée, en général, directement aux éléments visuels à grouper, détectés préalablement par des méthodes classiques de segmentation. La probabilité de groupement entre chaque élément est estimée à l'aide d'une mesure de compatibilité entre la courbure ou l'orientation des éléments. Un exemple type de ce genre d'approche est la mise en valeur de structures linéaires par relaxation d'un champ de vecteurs [Duncan et Birkhölzer, 1992] .

– Programmation dynamique

L'un des attraits de la programmation dynamique pour le groupement perceptuel est de pouvoir construire des structures optimisées globalement, à partir de mesures locales.

La programmation dynamique a été appliquée, entre autres choses, à l'extraction de structures linéaires adaptées aux images satellitaires. Par exemple, [Merlet et Zerubia, 1996] proposent une méthode de groupement récursif sur des critères de courbures et contrastes. Cette méthode consiste à chercher, dans une image aérienne, des chemins minimisant une fonction d'énergie entre deux ensembles de points sur l'image. Une série d'images de potentiels optimise la répartition de la fonction d'énergie sur les chemins possibles entre ces points. Les structures linéaires sont finalement obtenues par un suivi de chemins d'énergie minimale dans l'image de potentiel optimisée.

Un autre exemple d'application des principes de programmation dynamique, proposé par [Shashua, 1988] sera abordé d'une manière plus détaillée dans le chapitre suivant.

– Optimisation de réseaux de neurones

Les réseaux de neurones sont une autre structure utile pour optimiser et propager des contraintes. Les objets à grouper correspondent aux noeuds du réseau, et la compatibilité des relations entre eux correspondent aux poids des connexions.

Ils peuvent contribuer de manière indirecte au groupement perceptuel, en permettant la sélection d'hypothèses parmi un grand nombre de possibilités. A titre d'exemple, [Mohan et Nevatia, 1992] optimisent un réseau de Hopfield pour extraire les axes de symétrie répondant le mieux à certaines contraintes parmi toutes les symétries possibles.

L'approche neuronale peut aussi jouer un rôle central dans le groupement. Ainsi, [Mangin *et al.*, 1992] [Mangin, 1994] proposent une approche selon deux résolutions différentes. En basse résolution, les règles de proximité, de bonne

continuation et de parallélisme sont appliquées selon des processus de coopération locale entre éléments de contours. L'organisation globale ainsi obtenue est utilisée pour améliorer l'image initiale en haute résolution. Cette approche permet de mettre en oeuvre des règles de groupement complexes telles que le parallélisme, d'une manière massivement parallèle.

[Huddleston et Ben-Arie, 1993] suggèrent la mise en pratique de leur Transformée de Hough Distribuée (Distributed Hough Transform) à l'aide d'un réseau de neurones. Cette variante "perceptuelle" de la Transformée de Hough met en jeu des mesures de probabilité de non-accident, d'invariance de point de vue, et surtout, de symétrie circulaire. Elle consiste à estimer un cercle tangent à chaque élément de contour et accumuler le nombre d'éléments qui confirment chaque hypothèse. Ces éléments de contours ayant déjà leur propre information d'orientation, les hypothèses ne sont représentées que par la courbure du cercle tangent à cet élément. Ceci permet de réduire à 1 (au lieu de 3) la dimension de l'espace de paramètres et de simplifier la transformée. Cette approche revient à contraindre l'espace des paramètres d'une transformée de Hough avec des relations perceptuelles mesurées sur l'espace des données.

Il existe bien entendu d'autres approches d'optimisation moins répendues, comme par exemple, l'extraction de primitives géométriques à l'aide d'algorithmes génétiques [Roth et Levine, 1992] .

3.4.3 Théorie des graphes

Les graphes offrent un formalisme mathématique fort pour modéliser la structure sous forme de relations entre éléments visuels. Ils permettent une représentation hiérarchique des groupements, à différents niveaux d'organisation ou d'échelle. Il n'est donc pas surprenant que certaines des méthodes vues précédemment constituent des groupements sous forme de graphes. Les propriétés des graphes sont souvent mises à profit pour des tâches de groupement. Ainsi, [Shiu, 1990] propose d'organiser les groupements par proximité, colinéarité et parallélisme selon un graphe de connectivité. Il définit ensuite des règles de parcours du graphe pour en rechercher certaines propriété, en particulier les cycles. On retrouve des idées semblables pour des méthodes de recherche de groupements convexes [Wong *et al.*, 1991] et de polygones.

L'utilisation des graphes peut aussi faire partie intégrante de la méthode même de groupement. Par exemple, [Cox *et al.*, 1993] utilisent un graphe d'hypothèses Bayesiennes pour rechercher des courbes lisses dans un ensemble d'éléments de contours. Après détection, les contours sont prolongés progressivement à l'aide d'un filtre de Kalmann appliqué à un modèle local de courbure. La définition du filtre tient compte d'un modèle de bruit lié à la détection de contours. Un graphe d'hypothèses est ainsi construit, permettant d'évaluer l'appartenance ou non de nouveaux segments à une courbe en cours de construction. Cette démarche présente l'originalité de lier segmentation et groupement de contours.

3.4.4 Autres approches

En dehors des ces trois principales approches, il existe d'autres formalismes possibles, mais moins répendus, pour obtenir des structures linéaires par groupement perceptuel.

- *Classification et indexation*

  L'utilisation de techniques de classification et d'indexation pour grouper efficacement des éléments visuels présentant des propriétés semblables est assez récente. [Havaldar *et al.*, 1996] en donnent un exemple avec une approche hiérarchique pour la reconnaissance de formes génériques par organisation perceptuelle. Après détection de contours, des segments sont groupés selon différentes relations : continuité, symétrie, parallélisme, co-circularité et fermeture. La définition d'une indexation par proximité leur permet de fixer efficacement l'intervalle d'indexation. Les hypothèses émises par l'étape de groupement sont validées par la configuration des jonctions qui les délimitent. Les groupements ainsi obtenus permettent de définir des relations topologiques représentatives de la structure des objets présents dans la scène et de les identifier par comparaison avec des relations extraites de la même manière à partir de modèles.

- *Théorie de transformations*

  L'organisation perceptuelle peut être perçue comme l'application locale de transformations entre éléments visuels. Par exemple, la loi de proximité peut être définie comme le meilleur groupement possible par application de translations sur des éléments visuels. De la même manière, une loi de similarité peut être définie à partir de transformations exploitant des rotations, dilatations, translations et réflections.

  Ainsi la structure d'une scène peut être, plus généralement, détectée à l'aide de la projection des éléments visuels de l'image dans un espace dont les paramètres seraient, en plus de leur position dans l'image, définis par l'orientation, la couleur ou bien encore la taille. Extraire des groupements perceptuels reviendrait à grouper les éléments visuels dans cet espace. Ce type d'approche, semblable à une transformée de Hough généralisée à des termes plus complexes que les seuls paramètres de formes, est abordé plus en détail par [Sarkar et Boyer, 1993b] . Dans un registre semblable, [Palmer, 1983] suggère que la qualité d'une forme dépend de son degré d'invariance par un certain nombre de transformations. Les bonnes formes sont celles qui conserveraient alors le plus grand nombre d'invariants par le groupe de transformation des similarités Euclidiennes. On retrouve ici l'idée d'invariants selon des points de vues différentes évoquée par Lowe.

– *Simplicité de codage*

Une autre manière de considérer le groupement perceptuel est d'exprimer le principe de simplicité en des termes propres à la Théorie de l'Information. Parmi différents modèles possibles pour construire des groupements, notre choix se porterait sur le modèle le plus simple. Un bon groupement est ici un groupement régulier, prévisible, donc simple à modéliser.

D'une manière formelle, la théorie de l'information propose des méthodes d'évaluation de la simplicité d'un modèle, en définissant la notion de Longueur Minimale de Description [Lindeberg et Li, 1997] . Le choix du modèle le plus simple peut être aussi réalisé par application de modèles successifs, accompagnée d'une mesure d'erreur entre les modèles et la structure réelle des données [Leonardis et Bajcsy, 1992] . L'inconvénient de la méthode est qu'elle reste limitée à des groupements selon des formes paramétriques.

L'étendue des différentes approches présentées jusqu'ici confirme bien le rôle croissant du groupement perceptuel en vision par ordinateur, en particulier à l'aide de représentations hiérarchiques et de méthodes d'optimisation.

Malgré des difficultés de formalisation pour des règles de groupement telles que la similarité ou la *Pragnänz*, les avantages apportés par ces techniques ne peuvent pas être négligés. La formalisation des propriétés visuelles caractéristiques de la structure, ainsi que des mécanismes de groupement devient de plus en plus aboutie.

On peut espérer, à l'avenir, voir l'exploitation de nouvelles propriétés visuelles ainsi que l'introduction d'autres phénomènes visuels importants pour le groupement perceptuel mais toujours aussi difficiles à modéliser : l'attention, l'intention et l'apprentissage. L'apparition de nouvelle méthodes génériques de groupement devrait assurer, en particulier, une plus grande coopération et intégration entre ces diverses approches.

3.5 Principes de notre méthode

A partir de l'étude de la perception visuelle, nous avons souligné les différentes sources d'ambiguités qui font de la vision par ordinateur un problème d'une extraordinaire complexité. Dans le même temps, nous avons vu dans quelle mesure les différentes théories de la perception peuvent contribuer aux méthodologies de la vision par ordinateur.

En particulier, nous nous sommes intéressés aux problèmes posés par l'interprétation de scènes de contours ainsi qu'à différentes approches proposées en vision artificielle pour traiter ce type de scènes. Enfin, nous avons étudié le rôle du groupement perceptuel pour la réduction de cette complexité.

L'objectif de notre travail est d'extraire à partir des contours d'une image d'intensité, des primitives géométriques utiles à la construction de représentations de haut niveau. En raison des nombreuses sources d'erreurs possibles, nous choisissons

d'extraire ces primitives géométriques à l'aide de méthodes de groupement perceptuel. Ces méthodes permettent en effet une perception qualitative des structures linéaires de la scène, et par conséquent, plus robuste.

Nous proposons une approche de groupement hiérarchique pour extraire les éléments visuels les plus importants et fournir une base de représentation pour des processus de haut niveau d'interprétation. Cette méthode procède en trois niveaux d'organisation.

- *Groupements élémentaires*

  Le premier niveau organise d'abord les éléments de contours en *groupements élémentaires*, représentant les structures curvilignes perceptuellement importantes dans l'image. Le rôle de cette étape est de réduire la complexité des niveaux suivants en estimant de manière robuste les contours les plus réguliers. Cette étape peut être perçue comme une sorte de filtrage des contours à analyser selon des critères "perceptuels". Les contours les plus réguliers sont préservés, tout en comblant les discontinuités et en ignorant les structures bruitées.

  A cette fin, nous choisissons une méthode de groupement par optimisation. Une mesure de saillance structurelle, inspirée des réseaux de saillance de Shashua et Ullman, nous permet en effet de définir des structures d'intérêt pour les niveaux suivants, un peu à la manière des précurseurs sémantiques de Witkin et Tenenbaum. Le chapitre 4 aborde en détail un nouveau formalisme générique pour l'optimisation de ce type de mesure, ainsi que le groupement et la sélection des contours les plus saillants.

  Le résultat est l'ensemble des structures les plus importantes en regard de cette mesure. Ces structures forment des chaînes élémentaires à partir desquelles sont extraites les primitives géométriques recherchées. Ces chaînes ont ainsi une fonction de centre d'attention pour les reste des groupements.

- *Groupements intermédiaires*

  Le niveau suivant extrait, à partir des groupements élémentaires, des éléments de représentation sous forme de primitives géométriques. Ces *groupements intermédiaires* sont représentatifs des portions rectilignes des groupements élémentaires (segments), des portions curvilignes (arcs) et de points caractéristiques comme les coins ou les points d'inflexion.

  Dans un premier temps, des primitives élémentaires sont extraites à partir de chaque chaîne, à l'aide de méthodes spécifiques, adaptées à la nature des objets groupés. Cette approche permet de traiter en parallèle la détection de segments, d'arcs et de points d'intérêt. Elle permet de plus une perception de structures à différentes échelles.

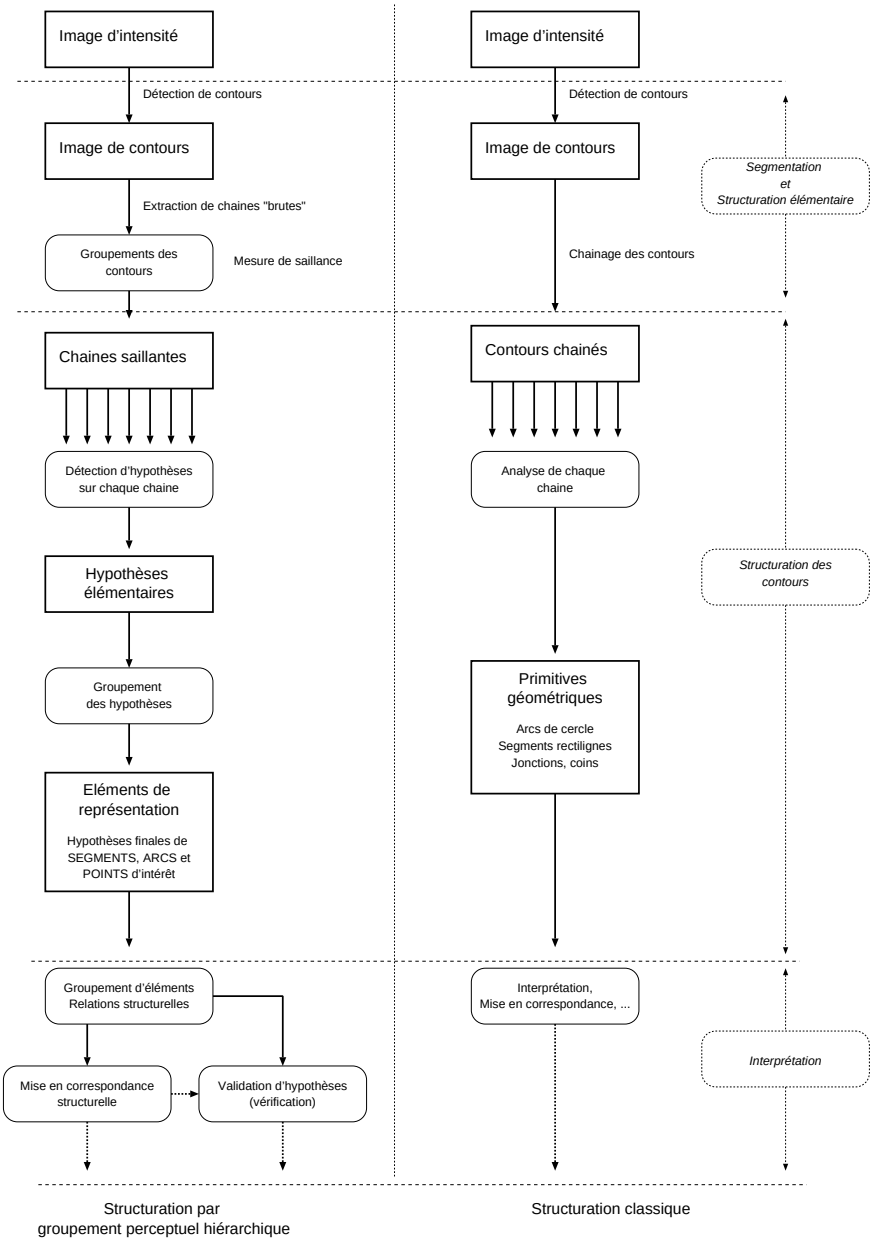

Figure 3.11 - *Extraction d'éléments de représentation par groupement perceptuel hiérarchique. Comparaison avec l'approche classique de structuration de contours.*

Les primitives élémentaires sont ensuite mises en commun et organisées à l'aide de méthodes algorithmiques de groupements, plus adaptées à ce type de structures.

Le chapitre 5 expose de quelle manière ces primitives sont détectées puis organisées selon un ensemble d'hypothèses de segments, d'arcs et de points d'intérêt. Ces hypothèses sont représentatives des structures curvilignes de la scène.

– *Groupements de haut niveau*

Finalement, ces hypothèses sont utilisées afin d'établir des *groupements de haut niveau*. Le chapitre 6 donne un exemple d'utilisation de ce type d'hypothèses au sein d'une application de mise en correspondance structurelle entre deux images.

Une relation de proximité entre les points d'intérêt et les extrémités des hypothèses de segments nous permet de construire un ensemble de jonctions sous la forme de groupements structurels entre primitives. Les jonctions ainsi détectées à partir de deux images sont ensuite mises en correspondance à l'aide d'un algorithme robuste de relaxation.

La figure 3.11 permet de comparer cette approche avec les étapes d'une structuration classique de contours. La principale différence est l'étape de groupement de contours. Contrairement à un chaînage classique, cette étape élimine une majorité de structures irrégulières, susceptibles d'être bruitées. De plus, elle permet la fermeture des discontinuités et produit, par conséquent, des chaînes plus longues et moins nombreuses que l'approche classique.

Cette sélection de groupements saillants a pour conséquence de produire une certaine quantité de redondance. Comme on pourra le voir au chapitre suivant, les groupements saillants partagent fréquemment des contours communs. Ainsi, au lieu de travailler sur des chaînes de contours uniques, l'étape suivante de structuration doit tenir compte de ces redondances.

C'est pourquoi, au lieu de chercher directement des primitives géométriques précises, notre approche travaille sur des hypothèses de primitives, en tolérant une certaine part d'erreurs et de redondances. Le rôle des groupements intermédiaires est de simplifier ces hypothèses au maximum et de ne garder que les plus représentatives de la scène observée. La recherche de primitives précises est rejetée le plus tard possible dans la chaîne de traitements, afin de ne manipuler qu'un faible nombre de primitives à vérifier.

Par comparaison, le système SCERPO de [Lowe, 1985] isolait le groupement perceptuel comme une étape intermédiaire de traitement entre segmentation en lignes et mise en correspondance. Notre approche permet d'étendre l'influence du groupement perceptuel en l'appliquant dès l'image de contours jusqu'à un niveau proche de l'interprétation. Elle est en cela plus proche des graphes perceptuels de [Sarkar et Boyer, 1994] , et des modèles d'apparence de [Pope et Lowe, 1993] .

Nous présentons plus en détails chaque niveau de groupement de notre méthode dans la seconde partie de cette thèse.

Deuxième partie

Analyse de scènes de contours par groupement perceptuel

Chapitre 4

Saillance structurelle et groupements élémentaires

Dans ce chapitre, nous abordons le premier niveau de groupement de notre approche. Il s'applique dès la détection de contours. Sa fonction est de faire ressortir les structures linéaires importantes présentes dans l'image de contours et de produire un ensemble de chaînes, ou groupements, correspondant à ces structures. Ces groupements servent de point de départ à l'extraction d'éléments de représentation réalisée aux niveaux suivants, en réduisant ainsi la complexité de la recherche visuelle aux seules structures d'intérêt de l'image.

4.1 Saillance structurelle

On peut définir deux sortes de saillances relatives à la perception. Un élément visuel peut présenter une saillance qui lui est propre, comme par exemple, un point blanc parmi un ensemble de points gris. Cette saillance *locale* est à distinguer de la saillance *globale* d'un groupe d'éléments visuels, qui traduit la structure d'ensemble de ce groupe. Cette saillance structurelle est voisine de l'idée de "bonne forme" de la Gestalt.

Le principal problème de ce niveau préliminaire est la définition d'une mesure de saillance sur les groupements possibles entre éléments de contours. Pour reprendre la définition rencontrée précédemment, il s'agit d'évaluer, pour chaque élément de contour, sa possible appartenance à une structure plus globale. On retrouve, en d'autres termes, le problème de séparation entre figure et fond, ou encore, entre forme et bruit.

Dans notre cas, mesurer la saillance d'éléments de contours revient donc à favoriser les éléments appartenant à des formes linéaires tout en pénalisant les éléments perturbateurs. En pratique, ce problème se traduit par la définition d'une fonction de coût, ou de qualité, pour un arrangement donné d'éléments de contours. On peut aussi rapprocher cette fonction de qualité de la probabilité de groupement accidentel

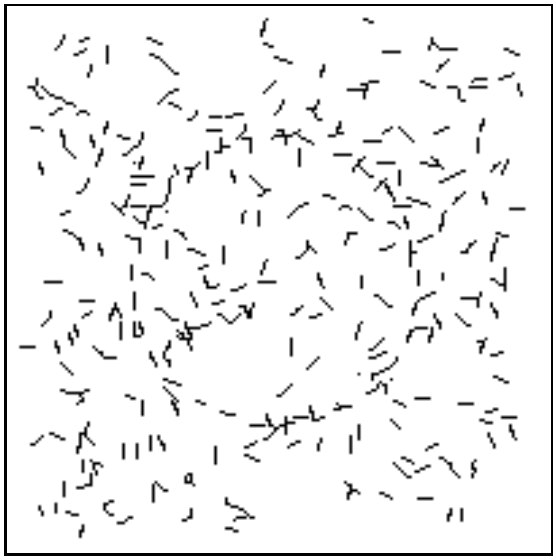

Figure 4.1 - *Exemple de groupement saillant de segments dans une scène bruitée. Une mesure de saillance structurelle doit attribuer un score important aux segments placés sur le tracé du cercle.*

suggérée par Lowe. Rechercher la valeur du meilleur groupement en chaque élément de contour devient alors un problème d'optimisation combinatoire de la fonction de qualité sur tous les groupements possibles. En outre, ce problème est assimilable à la recherche de sous graphes particuliers parmi l'ensemble de combinaisons possibles entre éléments de contours. Il s'agit donc d'un problème de classe NP-complet.

Le groupement perceptuel à l'aide d'une mesure de saillance soulève ainsi trois questions. Selon quels critères évaluer la qualité d'un groupement ? Comment optimiser, en chaque élément de contour, la mesure de saillance ? Et enfin, une fois définie la saillance de chaque élément, comment sélectionner les meilleurs groupements ?

La méthode que nous appliquons pour extraire les structures importantes d'une image est inspirée des réseaux de saillance de [Shashua et Ullman, 1988] . Cette méthode tient à la fois de la programmation dynamique, de la relaxation et des réseaux de neurones. La mesure de saillance de Shashua est une fonction de qualité calculée le long d'un chemin reliant les éléments visuels à grouper. Ces chemins étant inconnus avant le calcul de cette fonction, celle-ci est construite à l'aide d'une approche itérative semblable à la programmation dynamique. Les éléments visuels à grouper forment un réseau de pixels localement connectés. La mesure de saillance est définie à partir de mesures locales entre chaque élément et ses voisins. En fin d'optimisation, les structures globales sont obtenues en parcourant les éléments du réseau selon les connexions de qualité maximale. L'annexe A présente en détail les principes d'utilisation de ce type de réseaux tels que définis par Shashua et Ullman, ainsi que les principales mesures de saillance définies depuis.

Nous définissons à présent un formalisme générique pour construire ce type de réseau et en utiliser les résultats à des fins de groupement perceptuel. Nous appliquerons ensuite ce nouveau formalisme au groupement de pixels et de chaînes de pixels afin de le comparer à la méthode d'origine.

4.2 Méthodologie de groupement par réseaux de saillance

Ce chapitre est composé de trois parties. Nous définissons dans un premier temps des principes génériques pour la construction de réseaux de saillance et pour l'extraction des meilleurs groupements après optimisation d'une mesure de saillance. Ces principes sont ensuite appliqués au groupement de pixels, puis de chaînes de pixels.

Comme nous le verrons par la suite, l'adaptation de cette méthode à différentes primitives conduit nécessairement à exploiter des propriétés caractéristiques de celles-ci. Cette approche générique permettra donc de séparer la méthodologie de groupement à partir de réseaux de saillance et son application au groupement de pixels puis de chaînes de pixels.

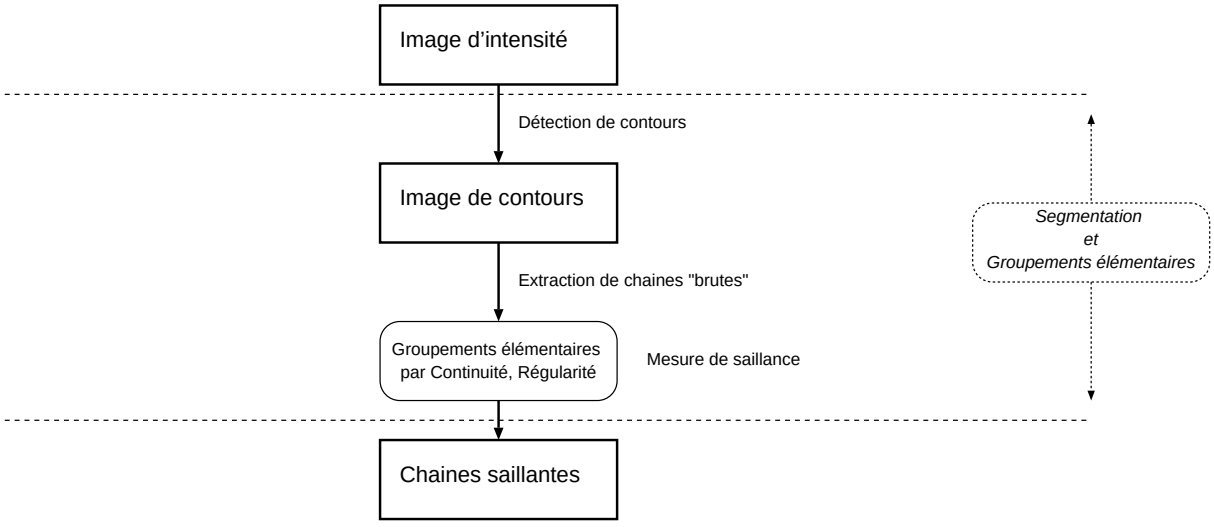

Figure 4.2 - *Principes du premier niveau de groupements. Le but est d'obtenir un nombre réduit de groupements de contours saillants, par rapport au nombre initial d'éléments de contours.*

4.2.1 Définitions

Soit $\mathcal{P}$ un ensemble de primitives préalablement détectées dans une image. La nature de ces primitives n'est pas limitée aux seuls points de contours, comme dans la méthode originalement proposée par Shashua et Ullman. Il peut s'agir éventuellement de segments élémentaires ou bien encore d'objets plus complexes.

Une définition générique d'un réseau de saillance est un graphe d'éléments localement connectés, représenté par un quadruplet $(\mathcal{P}, \mathcal{V}, \mathcal{F}, \mathcal{S})$. Les noeuds de ce graphe sont les primitives de $\mathcal{P}$. Ses arcs sont des "éléments de connexion" entre ces primitives, définis à l'aide d'un voisinage $\mathcal{V}$. On appelle "groupement" un chemin, éventuellement cyclique, parmi les noeuds du graphe.

Deux mesures sont établies à partir de ce graphe. La fonction de qualité $\mathcal{F}$ permet d'évaluer la compatibilité d'un groupement avec un ensemble de relations structurelles sur les primitives qui le composent. La mesure de saillance $\mathcal{S}$ permet, quand à elle, d'évaluer à l'aide du graphe la qualité du meilleur groupement passant par chaque primitive.

On notera enfin N_p, le nombre de primitives de $\mathcal{P}$ et N_v, le nombre moyen de voisins autour de chacune de ces primitives. Pour plus de détails sur l'origine de ces notations, on pourra se reporter à l'annexe A sur les réseaux de saillance tels que définis par Shashua.

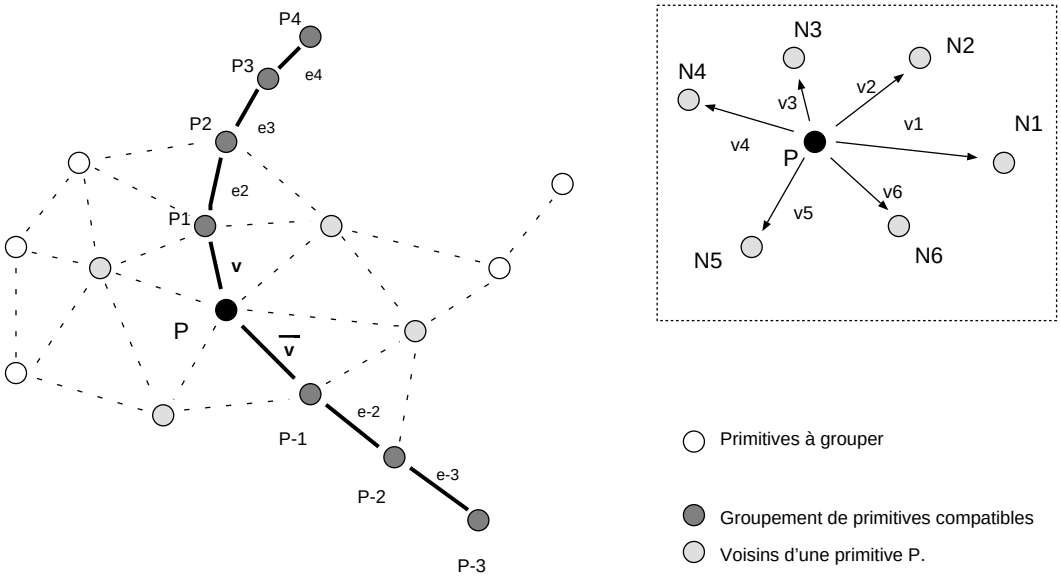

Figure 4.3 - *Exemple de groupement de primitives compatibles dans un réseau localement connecté. Le groupement traverse P selon les "directions" des éléments $v = e_1$ et $\overline{v} = e_{-1}$. La mesure de saillance pour P est la qualité maximale des groupements possibles traversant P selon deux de ses voisins.*

4.2.2 Voisinage

Le choix du voisinage $\mathcal{V}$ conditionne la finesse des groupements tout en pesant lourdement sur la complexité de l'algorithme d'optimisation.

- Influence du nombre de voisins

  Des voisins trop peu nombreux ou trop proches d'une primitive donnée rendront la forme des groupements trop dépendante de configurations locales. A l'inverse, un maillage trop dense peut conduire à un besoin de place mémoire prohibitif lorsque les primitives sont trop nombreuses.

  Un ensemble d'orientations autour de pixels est un bon exemple de ce dilemme. Un voisinage trop réduit, en 8 connexité par exemple, ne couvre pas assez d'orientations pour détecter une grande variété de courbes, alors qu'un voisinage de 180 orientations imposerait de mémoriser, pour chaque pixel, 180 qualités possibles et rendrait la méthode impossible à appliquer à des images de taille raisonnable (512×512 pixels par exemple).

- Influence de la forme du voisinage

  Une part importante d'opérations inutiles peut être évitée en choisissant un voisinage spécifique au type de primitive utilisé, en ne conservant que les voisins les plus susceptibles d'être groupés efficacement. Ceux-ci peuvent alors être déterminés à l'aide de procédures telles qu'un rayon ou un cône de recherche ou encore, une triangulation de Delaunay. Le nombre de voisins peut donc être éventuellement variable.

D'un point de vue de groupement perceptuel, le choix d'un ensemble de voisins revient à établir un groupement par proximité entre les primitives. Un voisinage statique correspond à un critère de proximité seul, alors qu'un voisinage dynamique correspond à des critères de proximité et de similarité conjugués.

Les deux applications de groupement données en fin de ce chapitre ont chacune un voisinage spécifique. Le groupement de pixels de contours porte sur un voisinage statique de 16 éléments de contours, alors que le groupement de chaînes utilise un voisinage dynamique, construit à l'aide de cônes de recherche dans la direction des extrémités de chaque chaîne.

Une dernière remarque peut être formulée sur la nature des primitives groupées. Les primitives présentes dans un voisinage peuvent être de nature éventuellement différente. C'était déjà le cas, par exemple, avec les éléments d'orientation "réels" et "virtuels" définis par Shashua. La seule contrainte sur le choix d'un voisinage est imposée par les relations structurelles de la fonction de qualité ; les primitives doivent être au moins comparables.

4.2.2.1 Relations structurelles et fonction de qualité

La fonction de qualité $\mathcal{F}$ est une mesure de compatibilité entre les primitives d'un groupement et les éléments de connexion qui les séparent.

Shashua et Ullman ont défini leur fonction de qualité comme une somme de contributions locales de chaque élément d'orientation, pondérée par des contraintes antagonistes. D'une part, une évaluation de la courbure globale d'un groupe de pixels assure à la fonction une décroissance monotone sur le nombre d'éléments du groupe. D'autre part, la proportion de discontinuité le long du groupe assure à la fonction une croissance monotone sur sa longueur. Les groupes assurant un bon compromis entre ces contraintes sont jugées de "bonne" qualité.

Cependant, le choix de combiner les contraintes de manière multiplicative présente deux inconvénients majeurs. Il rend tout d'abord la fonction de qualité trop sensible aux ordres de grandeur des variations de chacun des deux poids. Cet effet se traduit en pratique par une convergence trop rapide vers des groupements de taille réduite et de qualité localement élevée, par rapport à une convergence trop lente des groupements plus étendus et pourtant de meilleure qualité globale. La seconde restriction concerne l'extension éventuelle de la fonction de qualité à d'autres types de relations entre primitives, en rendant de plus en plus difficile à contrôler les influences respectives de chacune des contraintes.

Ce type d'interaction entre contraintes n'est pas sans rappeler les énergies internes et externes impliquées dans la définition des contours actifs ou 'snakes'[1] . En effet, d'une manière semblable à la fonction de qualité de Shashua, l'énergie interne d'un contour actif est constituée d'un terme de tension (terme du premier ordre représentant la longueur du snake) et d'un terme d'élasticité (terme du second ordre représentant la courbure du snake). A l'inverse de la fonction précédente, l'énergie d'un snake est composée de contraintes additives du type :

$$E_{snake} = \sum_{contour} E_{interne} + \sum_{contour} E_{externe} + \sum_{contour} E_{contraintes}$$

Afin de permettre un meilleur contrôle de l'influence de chaque contrainte et un meilleur équilibre de ces influences, nous proposons d'exprimer la fonction de qualité à l'aide d'un formalisme inspiré de celui des contours actifs. Les relations structurelles qui composent cette fonction expriment ainsi deux types d'influences :

- Les influences externes aux groupements sont des contraintes imposées par l'image sur le parcours d'un groupe. Elles peuvent correspondre à la proportion de discontinuités d'une courbe, une somme d'intensités ou de gradients, ou bien encore, à une différence d'orientations entre éléments de contours consécutifs et le tracé de la courbe.

1. Au sujet des 'snakes', voir le sous-chapitre 2.3.2, page 57

– Les influences internes sont des fonctions de régularisation, définies par la forme du groupement. Pour des structures curvilignes, ces relations sont liées à une mesure de courbure ou de co-circularité le long du groupement.

A ces influences s'ajoutent un certain nombre de contraintes éventuelles, imposées par exemple par les contours de l'image ou encore des zones interdites pour le contour.

Nous proposons comme fonction de qualité une combinaison linéaire d'influences normalisées. Soit N_r le nombre de relations structurelles $R_k(\cdot)$ qui composent $\mathcal{F}$. L'expression de la qualité d'un groupement γ est de la forme :

$$\mathcal{F}(\gamma) \;=\; \sum_{k=1}^{N_r} \alpha_k R_k(\gamma) \quad avec \quad \forall k,\; 0 < \alpha_k < 1 \quad et, \quad 0 < R_k(\gamma) < 1 \tag{4.1}$$

Les paramètres α_i permettent un contrôle précis de l'influence de chaque relation sur la fonction de qualité finale. La normalisation de chaque relation assure une contribution finie de la part de chaque terme.

A la différence des contours actifs, nous choisissons de maximiser notre fonction de qualité au lieu de réduire une énergie. Il est évidemment possible d'exprimer ces fonctions de manières différentes tant que le principe d'opposer des influences internes et externes au groupement est conservé.

4.2.2.2 Mesure de saillance

A partir de la fonction de qualité, la saillance d'une primitive P est définie comme la qualité du meilleur groupement partant de P dans la direction de chaque élément $v_i \in \mathcal{V}(P)$:

$$\mathcal{S}(\,P\,) = \mathop{\mathbf{Max}}_{\Gamma_P \in \delta^n(P)} S^n(\Gamma_P) \tag{4.2}$$

où $\delta^n(P)$ est l'ensemble de tous les groupes possibles[2] de longueur n passant par la primitive P. Plus précisément, un groupe Γ_P traverse P selon deux "directions", définies par les éléments v et $\overline{v}$. La saillance du groupement est donc une fonction bilatérale de chaque relation structurelle :

$$S^n(\,v, \overline{v}\,) \;=\; \mathcal{F}(\Gamma_P(v)) + \mathcal{F}(\Gamma_P(\overline{v})) \tag{4.3}$$

$$S^n(\,v, \overline{v}\,) \;=\; \sum_{k=0}^{N_r} \alpha_k (\, R_k^n(\,v\,) + R_k^n(\,\overline{v}\,) + H_k(\,P, v, \overline{v}\,)\,) \tag{4.4}$$

Les termes $R_k^n(\,v\,)$ et $R_k^n(\,\overline{v}\,)$ sont les expressions récursives de la fonction de qualité de chaque branche de Γ_P. Elles ne dépendent que du choix des éléments de départ v et $\overline{v}$ de chaque branche.

2. Par extension, $\delta^n(v_i)$ est l'ensemble des groupes de longueur n partant de P dans la direction de v_i. Si P_i est la primitive reliée à P par l'arc v_i, alors $\delta^1(v_i)$ s'écrit $\delta(v_i)$ et correspond au voisinage de P_i.

Le rôle des fonctions $H_k(\cdot)$ est de corriger les éventuels artefacts dus à la présence de termes communs dans la somme des contributions latérales. La définition de ces fonctions sera plus explicite en fin de chapitre avec les applications au groupement de pixels et de chaînes.

Une expression récursive de chaque relation structurelle est nécessaire pour optimiser la fonction de qualité à l'aide du graphe. Nous supposons que ces relations sont des fonctions extensibles répondant à la définition A.3 de la page 259, et que leur expression récursive s'écrit sous la forme :

$$R_k{}^{(n+1)}(v) = Q_k(v) + \rho_k \cdot \{P_k(v, e_j) \cdot R_k{}^{(n)}(e_j)\} \tag{4.5}$$

où $Q_k(v)$ est la contribution locale de l'élément v en P et ($0 < \rho_k < 1$) l'atténuation de la contribution de chaque élément avec la distance.

Le terme :

$$\{P_k(v, e_j) \cdot R_k{}^{(n)}(e_j)\}$$

est la contribution en v de la part de son voisin e_j. Ce voisin doit être choisi de manière à maximiser $S^n(v, \overline{v})$ afin de conserver la meilleure valeur de la fonction de qualité en permanence.

La fonction $P_k(v, e_j)$ permet d'accorder plus ou moins de crédibilité à une contribution en fonction de la configuration locale entre v et e_j. Les fonctions P_k et Q_k sont extraites de l'expression récursive de la fonction de qualité.

Il est important de noter que cette forme récursive n'est pas la seule possible. Le mécanisme d'optimisation reste le même pour toute fonction vérifiant la propriété d'extensibilité.

Cette définition permet de réduire l'espace de recherche des groupes de longueur n partant de l'élément v à $(N_v - 1) \cdot N_p$ possibilités[3] parmi $(N_v - 1)^{N_p}$.

4.2.3 Optimisation par programmation dynamique

L'optimisation de la mesure de saillance revient à effectuer une opération de relaxation sur l'expression récursive de chacune des relations $R_k^n(\cdot)$. Elle procède en deux temps. Une première étape définit, dans le voisinage de chaque primitive, les paires d'éléments de connexion qui génèrent les meilleurs groupements. Ces paires sont ensuite utilisées pour mettre à jour les valeurs de l'expression récursive de saillance des primitives.

L'algorithme 4.1 donne une vue d'ensemble de l'optimisation du réseau pour une itération donnée n. Une version plus détaillée de cet algorithme est donnée page 112.

3. Si N_v est le nombre moyen de voisins dans $\mathcal{V}(P)$, $(N_v - 1)$ est le nombre d'éléments susceptibles de prolonger un groupe arrivant sur un noeud depuis l'élément v.

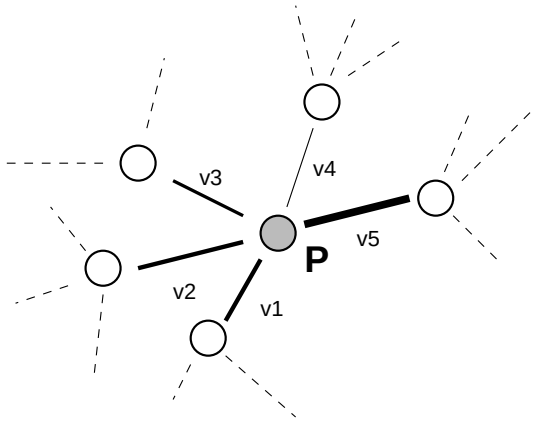

Figure 4.4 - *Exemple d'une primitive P et de 5 éléments de connexion. La largeur des connexions correspond à la saillance de la courbe partant de P dans la direction de l'élément. Ici, v_5 est l'élément le plus saillant, et v_4 le moins saillant.*

Algorithme 4.1 : Optimisation de réseau de saillance

début
- % Appariement des voisins
- %
- **pour** *Chaque primitive* $P \in \mathcal{P}$ **faire**
  - **pour** *Chaque voisin* $v_i \in \mathcal{V}(P)$ **faire**
    - Définir la paire de voisins (v_i, v_j) pour laquelle la saillance $S^n(\, v_i, v_j \,)$ est maximale
    - $v_j = \phi(v_i), \quad v_j \notin \delta^*(v_i)$
- %
- % Mise à jour des valeurs de saillance
- %
- **pour** *Chaque primitive* $P \in \mathcal{P}$ **faire**
  - **pour** *Chaque voisin* $v_i \in \mathcal{V}(P)$ **faire**
    - **pour** *Chaque relation structurelle* $R_k, k \in [1, N_r]$ **faire**
      - Mettre à jour ${R_k}^{(\, n+1)}(\, v_i \,)$ en fonction de ${R_k}^{(\, n)}(\, \phi(v_i) \,)$

fin

Chaque élément v_i dans le voisinage d'une primitive donnée, est associé à la variable d'état $R_k^n(\, v_i\,)$. Cette variable représente la saillance du meilleur groupe de longueur n partant de P dans la direction de v_i. Elle est initialisée par la contribution locale de v_i :

$$R_k^{(0)}(v_i) \;=\; Q_k(\, v_i\,), \quad \forall k \in [1, N_r]$$

D'après l'équation 4.5, chaque variable d'état est mise à jour en trouvant la paire d'éléments $(\; v_i, e_j\;)$, $e_j \in \delta(v_i)$ contribuant le plus à l'état de l'élément v_i.

Appariement d'éléments

D'une manière générale, la recherche des meilleures paires d'éléments se traduit par la définition d'une application entre voisins d'une même primitive.

$$\begin{array}{rccccc} \phi \;:\; & \mathcal{V}(P) & \longmapsto & & \mathcal{V}(P) \\ & v_i & \longrightarrow & v_j, & v_j \notin \delta^*(v_i) \end{array} \tag{4.6}$$

où $\delta^*(v_i)$ est un ensemble d'éléments voisins de P pour lesquels l'appariement n'est pas souhaité. Ces limitations d'appariement dépendent surtout des propriétés des primitives utilisées. A défaut d'une définition plus précise, $\delta^*(v_i)$ est, au minimum, composé de l'élément v_i afin d'éviter tout retour en arrière le long d'un même groupement. La suite de ce chapitre donne deux exemples de limitations de voisinage.

Dans la méthode d'origine, cette application associe simplement à chaque élément le voisin qui apporte la meilleure contribution à sa variable d'état. Or, il est souvent nécessaire d'interdire certaines paires pour éviter des groupements indésirables. La figure 4.4 est un exemple de primitive avec son voisinage immédiat. Dans cet exemple, l'élément v_5 apporte la meilleure contribution, v_1, v_2 et v_3 apportent une contribution semblable et v_4 fournit la plus faible contribution. Le rôle du voisinage inhibiteur δ^* est de s'assurer, par exemple, que les angles entre primitives ne sont pas trop aigus. L'élément v_4 sera donc apparié avec v_1, v_2 ou v_3.

Le choix de la meilleure paire d'éléments peut exercer une forte influence sur le résultat de l'optimisation. Ainsi dans [Shashua et Ullman, 1991] , le choix de paires disjointes autour de chaque primitive permet de forcer la convergence vers une partition de l'image sous forme de groupes optimisés. Cependant, forcer l'application ϕ à être bijective coûte que coûte représente une contrainte trop restrictive car elle interdit toute intersection ou jonction impaire.

Nous proposons une approche intermédiaire en s'assurant simplement que les paires réversibles[4] qui existent ont en permanence la meilleure qualité possible. L'appariement consiste alors à rechercher des paires de contributions maximales, puis à rectifier les paires réversibles si elles existent.

4. Une paire (v_i, v_j) est dite réversible si elle vérifie la propriété: $\phi(v_i) \;=\; v_j \;\; et \;\; \phi(v_j) \;=\; v_i$.

Plus formellement, notons $\mathcal{C}^n(\, v_i, v_j\,)$, la saillance du meilleur groupement passant par P selon les directions v_i et v_j. La valeur de $\mathcal{C}^n(\cdot)$ est donc fonction des contributions de l'élément v_j à l'état de v_i selon les différentes relations de la fonction de qualité (cf. équation 4.4).

Les paires autour d'une primitive sont initialisées par la relation suivante :

$$\phi(v_i) \;=\; v_j \quad \Longrightarrow \quad \mathcal{C}^n(\, v_i, v_j\,) \;=\; \underset{v_j \notin \delta^*(P)}{\mathbf{Max}} \{S^n(\, v_i, v_j\,)\} \tag{4.7}$$

v_j est l'élément qui, parmi les appariements possibles de v_i, fournit une contribution maximale à v_i.

Afin de privilégier les paires réversibles, il suffit de comparer la contribution de chaque paire avec la contribution des paires inverses éventuelles et de sélectionner, si elle existe, la paire réversible qui apporte la meilleure contribution.

Soit $\{\overline{v_1}, \cdots, \overline{v_m}\}$, l'ensemble des éléments dont l'image par ϕ est l'élément v_j :

$$\forall e \in \{\overline{v_1}, \cdots, \overline{v_m}\}, \qquad \phi(e) \;=\; v_j$$

Si cet ensemble n'est pas vide, la contribution de la paire initiale $(v_j, \phi(v_j))$ est comparée à la contribution qu'apporterait chaque paire inverse (v_j, e). Si l'une de ces contributions est meilleure, la paire $(v_j, \phi(v_j))$ est remplacée par la paire (v_j, e).

Si :

$$\exists e \in \{\overline{v_1}, \cdots, \overline{v_m}\} \quad / \quad \mathcal{C}^n(\, v_j, e\,) > \mathcal{C}^n(\, v_j, \phi(v_j)\,)$$

Alors, on pose $\phi(v_j) \;=\; e$ et la paire $(v_j, \phi(v_j))$ est inversible.

Mise à jour de la saillance

Les contributions de chaque élément sont diffusées au travers du réseau à l'aide de l'expression récursive de chaque relation structurelle.

Soit, pour chaque paire $(v_i, \phi(v_i))$:

$$R_k{}^{(\,n+1)}(\, v_i\,) \;=\; Q_k(\, v_i\,) + \rho_k \cdot [P_k(\, v_i, \phi(v_i)\,) \cdot R_k{}^{(\,n)}(\, \phi(v_i)\,)], \quad \forall k \in [1, N_r]$$

A chaque itération, les contributions prennent en compte l'influence de primitives de plus en plus distantes. Au cours de l'optimisation, les primitives situées le long de groupements visuellement importants sont alimentées en permanence par les contributions importantes des autres primitives de ces groupements. A l'inverse, les primitives plus isolées reçoivent peu de contributions, comme le montre l'exemple des figures 4.5 à 4.7. L'image de départ est une ellipse pour laquelle 5% de pixels ont été supprimés puis 10% de pixels de bruit blanc ont été ajoutés. Les régions noires sur l'image 4.5 (à droite) représentent les parties du réseau inutiles lors de la première itération (et par conséquent, non initialisées).

Le résultat de l'optimisation est une évaluation de la saillance de chaque primitive du réseau. En plus de cette *carte de saillance*, la mise a jour constante des meilleures paires d'éléments (v_i, v_j) autour de chaque primitive P permet de conserver les directions à emprunter pour suivre le meilleur groupe partant de P dans la direction de v_i. Cette information est fondamentale pour l'extraction des groupes les plus importants.

Algorithme 4.2 : Appariement et mise à jour des valeurs de saillance

début
- % Appariement des voisins
- %
- **pour** *Chaque primitive* $P \in \mathcal{P}$ **faire**
  - **pour** *Chaque voisin* $v_i \in \mathcal{V}(P)$ **faire**
    - Initialisation de la paire $(v_i, \phi(v_i))$ pour laquelle
    - la saillance $S^n(\, v_i, v_j\,)$ est maximale
    - $\phi(v_i) \;=\; v_j \quad \Longrightarrow \quad \mathcal{C}^n(\, v_i, v_j\,) \;=\; \mathbf{Max}_{v_j \notin \delta^*(P)}\{S^n(\, v_i, v_j\,)\}$
  - %
  - **pour** *Chaque voisin* $v_i \in \mathcal{V}(P)$ **faire**
    - Recherche des paires inversibles et mise à jour si elles existent
    - **pour** *Tout voisin* $e \notin \delta^*(v_i)$ *tel que* $\phi(e) = v_i$ **faire**
      - % Si la paire réversible génère une meilleure saillance que la paire initiale
      - % alors remplacer la paire initiale par la paire réversible
      - **si** $(\mathcal{C}^n(\, e, v_i\,) > \mathcal{C}^n(\, v_i, \phi(v_i)\,))$ **alors**
        - $\phi(v_i) = e, \quad e \notin \delta^*(v_i)$
- %
- % Mise à jour des valeurs de saillance
- %
- **pour** *Chaque primitive* $P \in \mathcal{P}$ **faire**
  - **pour** *Chaque voisin* $v_i \in \mathcal{V}(P)$ **faire**
    - **pour** *Chaque relation structurelle* $R_k, k \in [1, N_r]$ **faire**
      - $R_k{}^{(\,n+1)}(\, v_i\,) = Q_k(\, v_i\,) + \rho_k \cdot [P_k(\, v_i, \phi(v_i)\,) \cdot R_k{}^{(\,n\,)}(\, \phi(v_i)\,)]$

fin

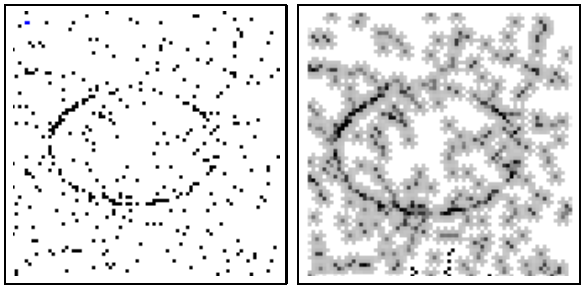

Figure 4.5 - *Exemple d'évolution de la carte de saillance - L'image de départ, à gauche, est une image de* 80×80 *pixels. Le réseau de saillance est défini à l'aide d'un voisinage à 16 éléments tel que défini dans l'application au groupement de pixels. La figure de droite montre l'état initial du réseau* ($n = 0$).

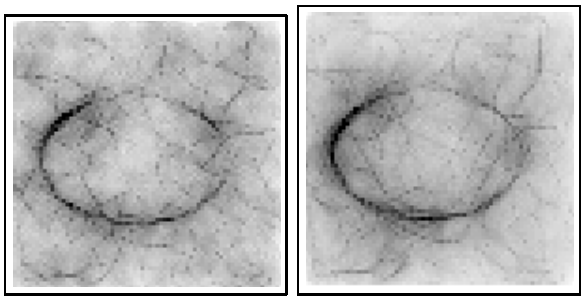

Figure 4.6 - *Exemple d'évolution de la carte de saillance pour 5 et 10 itérations du réseau. L'intensité minimale correspond à un maximum de saillance.*

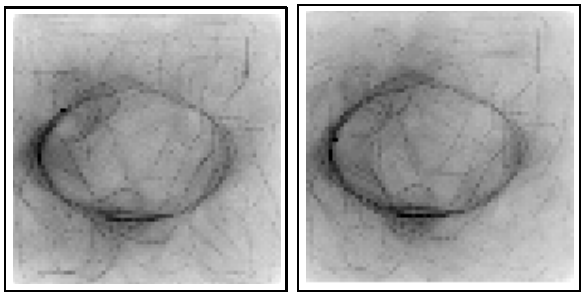

Figure 4.7 - *Exemple d'évolution de la carte de saillance pour 15 et 20 itérations. Seuls les points situés dans le voisinage direct de structures linéaires conservent une saillance élevée. Les autres pixels, plus isolés, sont atténués.*

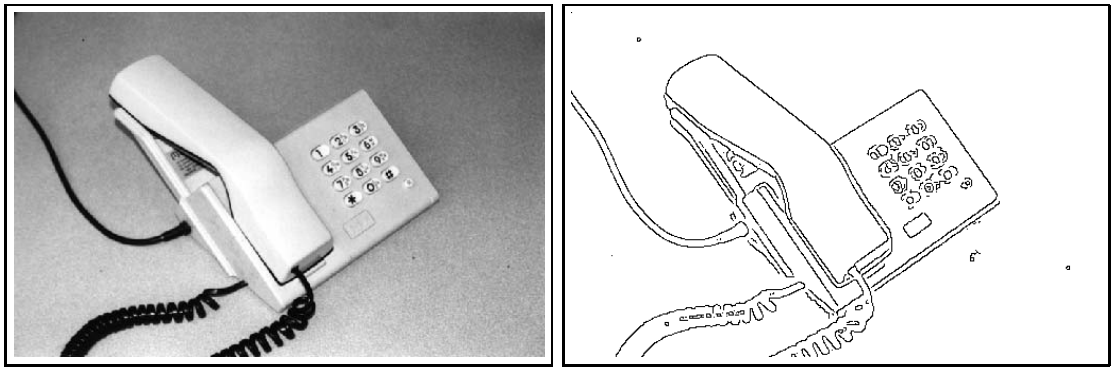

Figure 4.8 - *Exemples de groupements individuels - Image d'intensité et détection de contours.*

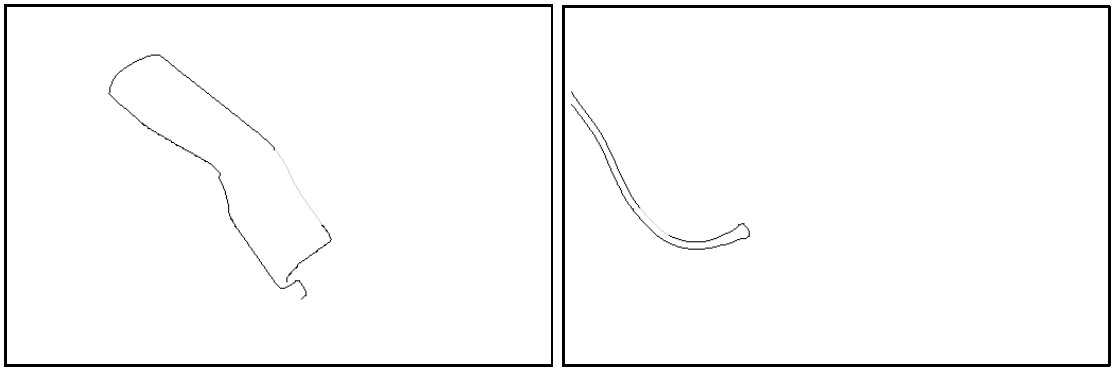

Figure 4.9 - *Exemples de groupements individuels à partir de chaînes. Le groupement obtenu après suivi des éléments de connexion est en noir. La chaîne blanche représente le point de départ du groupement. Ces deux groupements délimitent les contours d'objets bien distincts. Ils illustrent bien l'intérêt d'organiser les contours selon des critères de régularité.*

4.2.4 Groupement à partir d'une carte de saillance

L'une des principales conséquences de l'optimisation est de réduire le nombre de groupements possibles à un seul parcours pour chaque primitive P du réseau et chaque direction de départ v_i à partir de cette primitive, soit $N_v \cdot N_p$ parcours possibles dans le graphe.

La définition des meilleures paires d'éléments permet de réduire ce nombre à une moyenne de $p \cdot N_p$ parcours possibles, si p désigne le nombre moyen de paires définies autour de chaque primitive. Chaque paire définit les directions selon lesquelles les meilleurs groupes traversent une primitive. En privilégiant le couple apportant la

contribution la plus forte, on peut réduire le nombre de parcours possibles à un seul groupement par primitive, soit N_p groupements. Cette réduction de complexité rend possible une recherche exhaustive des groupements optimisés afin de sélectionner les meilleurs d'entre eux.

4.2.4.1 Extraction de groupes individuels

L'algorithme de reconstitution des meilleurs groupes revient simplement à suivre, de proche en proche, les paires d'éléments tant que celles ci sont définies et que certaines conditions d'arrêt n'ont pas été rencontrées. Ce suivi peut être effectué à partir de chaque primitive, selon les directions de sa meilleure paire d'éléments.

Conditions d'arrêt du suivi

Les conditions d'arrêt d'un suivi peuvent être imposées par la structure du réseau, par la forme des groupes extraits ou bien encore, par la méthode même d'optimisation de ces groupements.

La contrainte la plus évidente imposée par le réseau correspond à l'absence de paire permettant de prolonger un groupement arrivant en une primitive P par un élément v_i. Il s'agit du cas où $\phi(v_i) = \emptyset$ (figure 4.11, à droite).

Ce type de suivi des primitives de proche en proche privilégie l'extraction de structures linéaires. Il est donc inutile de poursuivre plus loin un groupement en cas de boucle. En effet, si la boucle se referme sur le point de départ du groupement, il s'agit bien d'une structure circulaire. Si par contre il se referme plus loin dans le groupe, il s'agit d'une intersection ou d'une jonction, et le groupement ne possède pas assez d'information globale à ce niveau pour décider de poursuivre ou non le suivi.

La contrainte imposée par la méthode d'optimisation concerne la longueur des groupes reconstitués. Après n itérations, la saillance de chaque primitive est exprimée pour des groupes de longueur n partant dans une direction donnée. Poursuivre un parcours au delà de n noeuds du graphe reviendrait à ajouter des primitives peu significatives pour la primitive de départ.

En effet, pour un parcours $\{e_1, e_2, \cdots, e_n\}$, l'élément e_2 a contribué n fois à l'état de e_1 au cours de l'optimisation. De même, l'élément e_3 a contribué $(n-1)$ fois en e_1, jusqu'à e_n qui n'apporte qu'une seule contribution en e_1.

Afin de tenir compte de cette observation, l'ajout d'une nouvelle primitive dans un parcours est soumis à un test de "crédibilité". Ce test représente en quelque sorte la probabilité pour que la contribution apportée par e_t soit significative pour l'élément de départ e_1. Le suivi est interrompu lorsque la crédibilité devient inférieure à un seuil (proche de 0).

Soit $p(\cdot)$ cette mesure, définie de la manière suivante :

$$p(t) \;=\; \exp^{-\lambda\frac{t^2}{\sigma^2}} \quad avec \quad \sigma \;=\; \frac{n}{2} \;\; et \;\; t \in [1,n]$$

λ est un coefficient de dégradation qui permet éventuellement de moduler la décroissance de $p(\cdot)$ (figure 4.10). Ce coefficient est utile, par exemple, pour limiter la taille des groupements dont la qualité est trop faible.

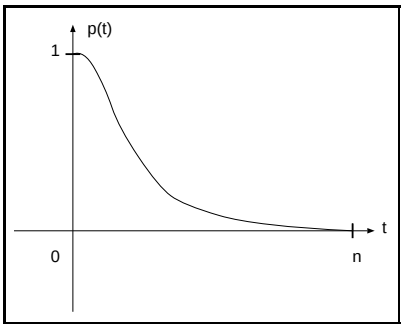

Figure 4.10 - *Allure de la fonction de "crébilité" pour les éléments d'un groupement de longueur n.*

Analyse des groupes individuels

Comme le montrent les exemples des figures 4.8 à 4.11, le suivi des meilleures paires est insuffisant pour extraire des structures cohérentes depuis une image. Les principaux problèmes apparaissent autour des intersections entre structures linéaires. En effet, chaque branche d'une intersection se traduit par une forte contribution provenant de sa direction.

Un suivi au niveau des paires d'éléments permet, au mieux, de déterminer dans quelles directions se trouvent des structures d'intérêt. Le long d'une même structure, la présence de deux contributions de forte qualité permet de poursuivre le parcours sans difficulté. Dans le cas de jonctions entre structures, le suivi ne dispose pas d'assez d'informations globales pour décider de la meilleure direction à suivre - ce qui explique les changements de direction observables dans les deux exemples. Le cas le plus extrême se rencontre sur les "plateaux" de la carte de saillance, où les valeurs des contributions sont du même ordre dans chaque direction. C'est le cas de zones du réseau où la distribution de primitives est relativement homogène.

Les groupements individuels sont donc insuffisants pour extraire à coup sûr des structures cohérentes, mais ils n'en perdent pas pour autant leur caractère saillant. On peut raisonnablement les considérer comme des fragments dont la somme couvre les structures saillantes de la scène.

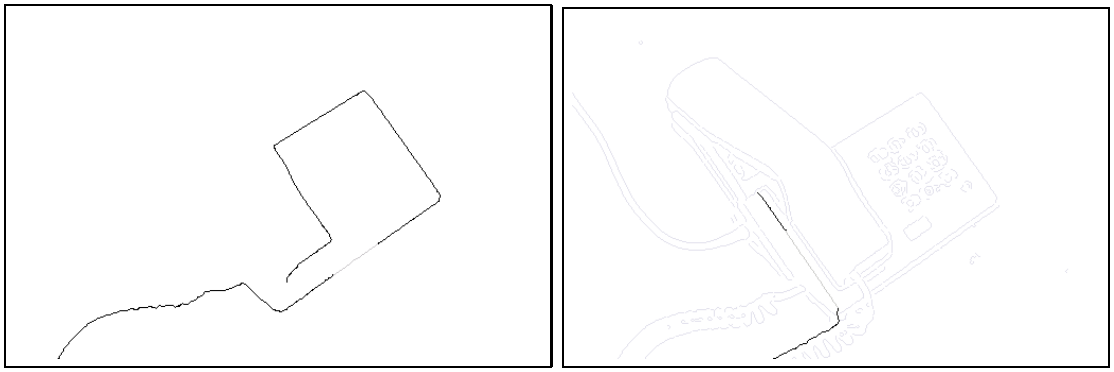

Figure 4.11 - *Exemples de groupements individuels. Ces deux exemples illustrent des situations de groupements incomplets. Dans la figure de droite, le suivi des éléments du réseau de saillance est interrompu par le critère de distance. Passé une certaine distance de la chaîne de départ, ajouter de nouveaux éléments à un parcours n'est plus utile. La figure de droite illustre un groupement interrompu par manque de connexions valides entre chaînes.*

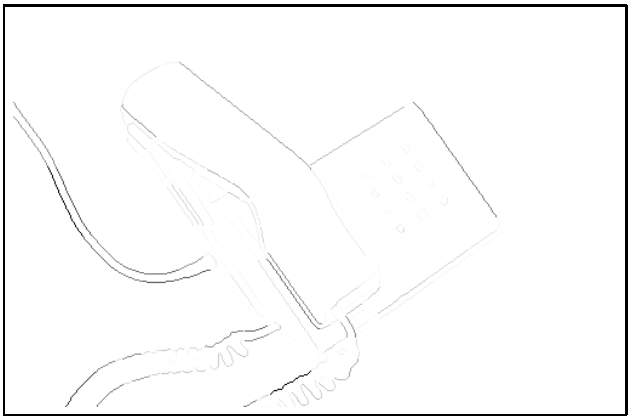

Figure 4.12 - *Exemple de carte saillance locale pour des chaînes de pixels. Les chaînes les plus saillantes sont en noir. La valeur de saillance locale d'une chaîne dépend de la répartition du gradient de l'intensité lumineuse le long de cette chaîne. Dans le cas de chaînes, ce critère permet de privilégier les départs de suivi depuis les chaînes les plus longues.*

4.2.4.2 Sélection des groupements

Afin d'obtenir la meilleure couverture possible des structures saillantes, tout en laissant de côté les mauvais groupements, il est nécessaire de définir des critères de sélection sur les groupements.

– *Saillance locale*

Le premier critère dans le choix du point de départ concerne les caractéristiques propres de chaque primitive. Un groupement a en effet plus de chance d'être saillant s'il part d'une primitive dont la saillance locale est importante. Le choix d'une mesure de saillance locale est intimement lié à la nature des primitives et sera, par conséquent, abordé dans la dernière partie de ce chapitre.

– *Saillance globale*

Les groupements peuvent être également caractérisés par leur saillance globale. Afin de tenir compte des éventuelles variations de longueur de parcours, la saillance d'un groupe est définie par la somme des saillances de chaque primitive rencontrée dans son parcours. Soit, pour un groupement Γ_P constitué des primitives $\{P_{-n}, \cdots, P-1, P, P_1, \cdots, P_n\}$:

$$\mathcal{Q}(\Gamma_P(v, \overline{v})) \;=\; \sum_{k=0}^{n-1} S^n(e_{-k}, \phi(e_{-k})) \;+\; \sum_{k=0}^{n-1} S^n(e_k, \phi(e_k))$$

en reprenant les notations de la relation 4.4.

Ce critère permet de séparer les groupes visuellement importants des groupes isolés et peu saillants. Il n'est pourtant pas suffisant pour départager les meilleurs groupes entre eux. Toute primitive se trouvant à proximité d'une structure importante sera forcément le point de départ d'un groupement à forte saillance, qu'elle appartienne ou non à cette structure.

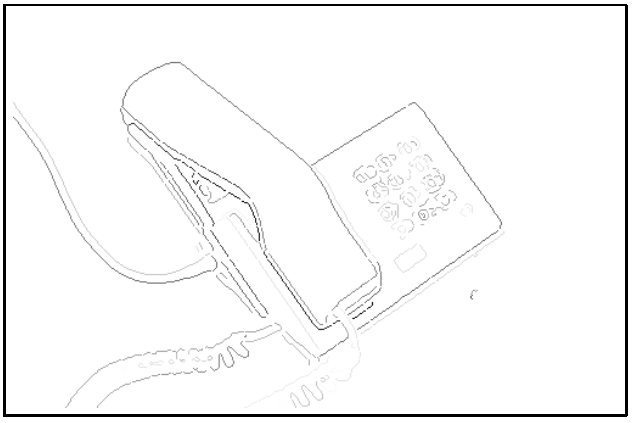

Figure 4.13 - *Exemple de carte de saillance globale pour des chaînes de pixels. Ce critère met en valeur les chaînes appartenant à des structures régulières, mais attribue également une forte saillance aux chaînes voisines de chaînes saillantes.*

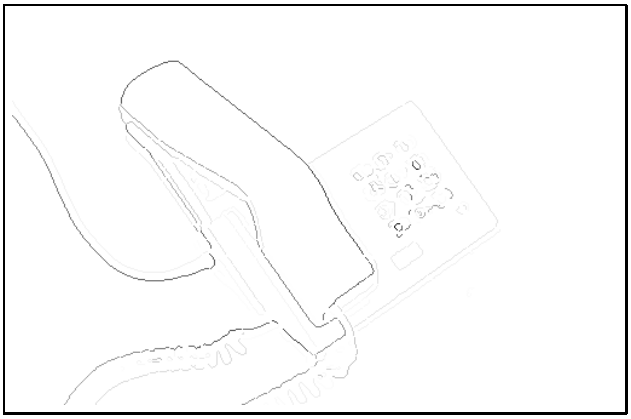

Figure 4.14 - *Exemple de carte d'accumulation pour des chaînes de pixels. Ce dernier critère élimine bien l'effet de voisinage de la mesure de saillance mais ignore également certaines chaînes longues pour lesquelles peu de groupements ont apporté leurs votes.*

– *Accumulation*

Afin de mieux séparer les véritables groupements saillants de ces groupements "parasites", nous introduisons un troisième critère lié au nombre de groupes traversant chaque primitive. En effet, une primitive appartenant à une réelle structure saillante sera traversée par un nombre important de groupes, alors qu'une primitive proche de cette structure ne sera traversée que par des groupements "parasites", en général moins nombreux.

Lors du parcours de tous les groupements possibles en préliminaire à la sélection, chaque groupe vote pour les primitives qu'il traverse. Il suffit ensuite de sélectionner les primitives ayant reçu le plus grand nombre de votes.

Les groupes correspondant à l'un de ces trois critères au moins sont sélectionnés. Les seuils de sélection de ces critères sont laissés à la discrétion de l'utilisateur afin d'adapter la méthode aux particularités de la scène observée comme par exemple, le rapport signal sur bruit entre les primitives provenant d'une fausse détection et celles appartenant réellement à des structures globales de la scène.

Dans la pratique, le critère prédominant est le seuil d'accumulation. Ce seuil permet de couvrir la majeure partie des structures saillantes de la scène. Les deux autres critères viennent en complément à cette première sélection. Le cas échéant, il est parfois nécessaire de terminer la sélection par des retouches manuelles. L'intérêt de ces critères est de simplifier la tâche de la sélection en la rammenant au choix de trois seuils au lieu d'une recherche manuelle exhaustive.

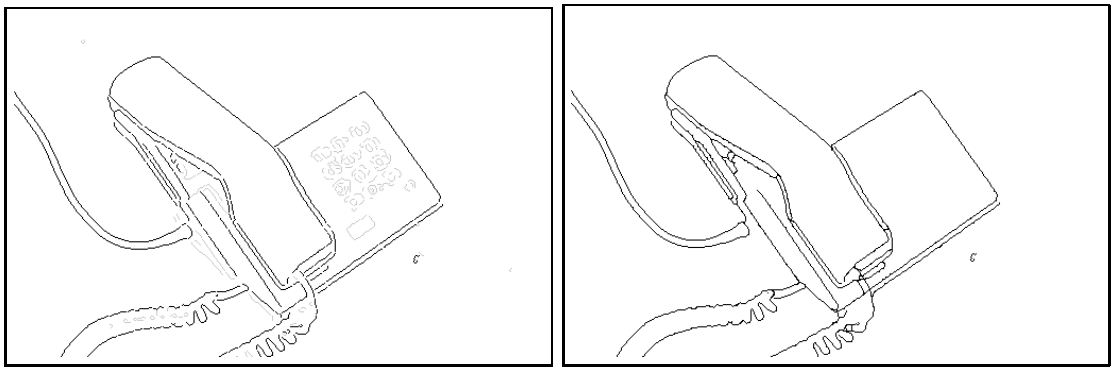

Figure 4.15 - *Sélection finale des chaînes servant de point de départ au suivi des éléments de connexion (à gauche) et superposition des groupements (à droite). A partir des 560 chaînes de contours, 90 groupements ont été sélectionnés.*

4.2.5 Conclusion et applications

En résumé, notre méthode de groupement perceptuel par réseaux de saillance se distingue de l'approche initiale de Shashua et Ullman par plusieurs aspects.

Le formalisme générique nous a permi de séparer la stratégie de groupement par réseau de saillance des contraintes liées au choix d'une primitive spécifique. Ce formalisme permet d'envisager l'extension de cette approche à tout type de groupement par alignement d'éléments consécutifs.

Nous avons introduit une famille de fonction de qualité proche des fonctions d'énergie des contours actifs, qui permet ainsi un meilleur contrôle de l'influence de chaque relation structurelle sur l'aspect des courbes obtenues par groupement.

C'est également par le biais de la fonction de qualité que ce niveau de groupement pourrait être utilisé en coopération avec d'autres sources d'information visuelle que les contours. Ainsi, une mesure de saillance entre éléments de contours pourrait être renforcée par la connaissance d'une carte de régions ou de textures afin de privilégier la saillance de contours appartenant à une même région.

Notre algorithme d'optimisation, en recherchant à chaque itération les meilleurs appariements autour de chaque primitive, permet de guider la convergence vers des solutions plus stables tout en autorisant l'existence de jonctions multiples.

Enfin, les critères de reconstruction et de sélection des groupements nous permettent de mettre en valeur les courbes les plus importantes de la scène et non uniquement la meilleure courbe possible. De même que pour la définition de la fonction de qualité, les critères de suivi et de sélection sont ouverts à la coopération avec d'autres indices visuels. Pour reprendre l'exemple d'une carte de régions, les groupements respectant le mieux les frontières entre régions pourraient être ainsi privilégiés.

Paramètres

Indépendamment du type de primitive choisie, les paramètres utilisés par cette méthode de groupement sont les suivants :

- $0 < \alpha_i < 1$, poids de chaque terme dans la fonction de qualité.
- $0 < \rho_i < 1$, atténuation des contributions avec la distance.
- $n_{max} < N_p$, longueur maximale admissible pour un groupement.
- Seuils d'arrêt du suivi et de sélection des groupements.

Le rôle de chaque paramètre est abordé plus en détail dans la présentation de chaque application. Une fois ces paramètres fixés pour une catégorie de courbes, la méthode peut être appliquée à différentes situations sans avoir pour autant à redéfinir ces paramètres. Pour illustrer cette stabilité des paramètres, les résultats présentés pour le groupement de pixels et le groupement de chaînes, ont été obtenus dans chaque cas avec le même jeu de paramètres.

Complexité.

Nous nous plaçons dans un cas général, où le nombre de voisins autour de chaque primitive peut être éventuellement variable. La complexité algorithmique est directement liée au nombre de primitives N_p et au nombre moyen de voisins N_v autour de chaque primitive. Elle est de l'ordre de $\mathcal{O}(n \cdot N_p \cdot N_v^2)$ pour une optimisation de n itérations. La sélection des meilleurs groupements est, quand à elle, de l'ordre de $\mathcal{O}(N_p)$, les possibilités ayant été réduites à un seul parcours privilégié par primitive.

Nombre d'itérations

L'optimisation permet de diffuser l'influence de chaque primitive dans le réseau un peu plus loin à chaque itération. Le nombre d'itérations nécessaires à un groupement dépend essentiellement de la longueur des groupements recherchés. Cette longueur étant inconnue au préalable, la limite est fixée au nombre maximal d'éléments admissible pour un groupement. Ce nombre est limité, au pire, par le nombre de primitives du réseau. En effet, au delà de N_p éléments, un groupement est forcément une boucle et ne peut donc plus être prolongé.

Cette méthodologie de groupement perceptuel à partir de réseaux de saillance a été testée sur des groupements de pixels de contours [Montesinos et Alquier, 1996] , puis étendue au groupement de chaînes de pixels [Alquier et Montesinos, 1997] .

Le groupement de pixels utilise un voisinage statique de 16 voisins. Sa fonction de qualité implique des termes de courbure et de co-circularité pour les influences internes et des termes de continuité d'intensité et d'orientation pour les influences externes.

Par comparaison, le groupement de chaînes de pixels dispose d'un voisinage dynamique autour des extrémités des chaînes. Les influences externes de sa fonction de qualité est composée d'un terme de continuité et d'orientations respectives des extrémités. Ses influences internes impliquent une mesure de courbure et de co-circularité.

Chacune de ces applications fait l'objet d'une présentation plus détaillée dans la dernière partie de ce chapitre.

4.3 Application au groupement de pixels

Afin de mieux comparer notre approche des réseaux de saillance avec la méthode originalement proposée par Shashua et Ullman, la première application porte sur le groupement de pixels. Le but du groupement est ici d'extraire les structures curvilinéaires visuellement importantes dans l'image, tout en fermant les contours incomplets et en ignorant les fausses détections.

4.3.1 Primitive "pixel"

L'ensemble des primitives choisi pour cette application est constitué des pixels d'une image de détection de contours. La caractéristique principale de ces pixels est leur intensité, fixée à 1 pour la détection d'un point de contours et à 0 pour une primitive "virtuelle". Tous les pixels de l'image sont pris en compte dans ce réseau.

4.3.2 Voisinage statique

Les éléments de connexion représentant le voisinage d'une primitive constituent l'ensemble des orientations possibles pour des courbes la traversant. Ce voisinage définit un nombre constant d'orientations autour de chaque pixel.

Comme nous l'avons déjà évoqué dans la présentation de la méthode, cette discrétisation des orientations doit être suffisamment fine pour ne pas limiter la forme des courbes détectées. De plus, la notion d'échelle étant absente à ce niveau de détection, les changements d'orientations locales le long d'une courbe entraînent l'apparition "d'escaliers" qui pénalisent un peu plus sa mesure de qualité.

Les groupements détectés seront donc plus ou moins grossiers selon le nombre de voisins choisi. L'aspect des groupements peut être affiné en cours d'optimisation, comme suggéré dans [Shashua, 1988] , ou bien en fin d'optimisation, en modélisant les groupements à l'aide de courbes polynômiales.

Un nombre d'orientations trop élevé peut, à l'inverse, imposer de sérieuses limitations sur la taille des images envisageables pour la méthode. En effet, l'optimisation nécessite de mémoriser, pour chaque pixel, N_v paires d'éléments, soit $2 \cdot N_v$ variables d'état pour la mise à jour de la mesure de saillance , sans compter les ressources nécessaires à la manipulation du réseau de saillance.

Cette restriction est moins importante que la précédente sachant que la méthode d'optimisation est parallélisable. Pour une itération donnée, les primitives ne dépendent que des valeurs des contributions de l'itération précédente. Chaque primitive pourrait donc être associée à un processeur particulier et la mise à jour de la mesure de saillance pourrait ainsi être réalisée simultanément sur toutes les primitives.

Cependant, en l'absence d'implémentation sur machine parallèle, la manipulation des voisinages peut être simplifiée en ne traitant que les pixels utiles pour une itération donnée. Il est en effet inutile de mettre à jour la saillance de primitives "virtuelles" lorsque les éléments de leur voisinage ont une contribution nulle. C'est le cas de tout pixel qui ne se trouve pas dans le voisinage immédiat d'un pixel de contour.

Il suffit alors d'inhiber ces pixels "virtuels" par défaut, et de ne les activer qu'au voisinage d'un pixel recevant au moins un contribution non nulle. A chaque itération, de plus en plus de pixels virtuels sont ainsi activés, ce qui assure la méthode de ne prendre en compte que les pixels utiles. Cette heuristique permet en pratique de gagner jusqu'à 40% de temps de calcul, en particulier sur des images de grande taille avec des zones de faibles densité de pixels[5].

En tenant compte de ces considérations, le voisinage choisi pour le groupement de pixels est un voisinage à 16 éléments d'orientation tel que le montre la figure 4.16.

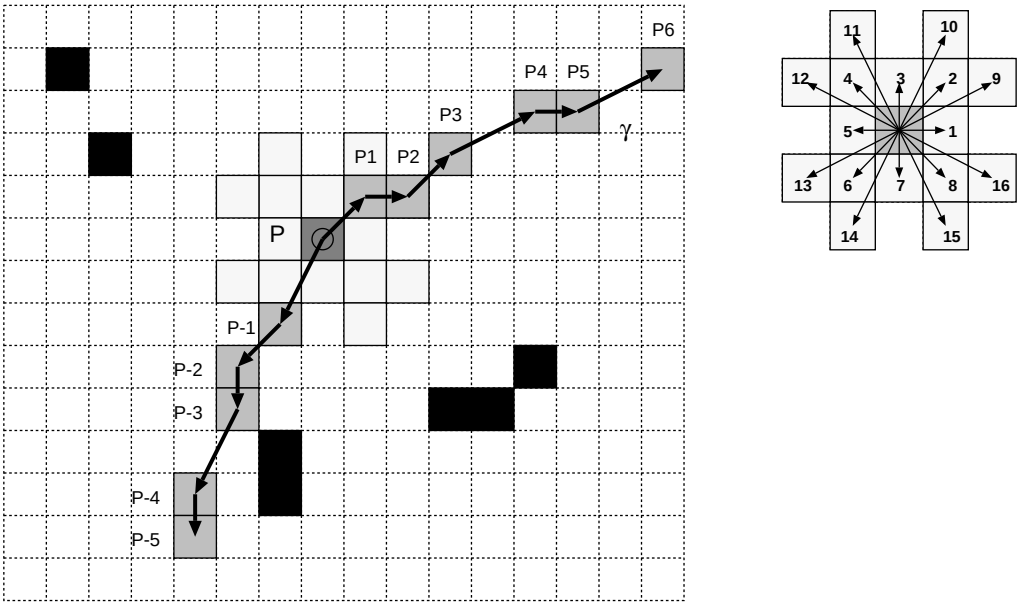

Figure 4.16 - *Voisinage de pixel à 16 éléments d'orientation. Ce voisinage permet à la fois des connexions rapprochées (voisins 1 à 8) et plus distantes (voisins 9 à 16).*

5. Voir à ce titre les résultats de groupement sur des images satellites.

4.3.3 Fonction de qualité

La fonction de qualité reprend les termes de continuité et de courbure définis par Shashua, auxquels sont associés des termes de co-circularité et d'orientations, introduits par [Montesinos et Blanc, 1994] Pour reprendre un formalisme de contours actifs, les relations de continuité et d'orientations représentent des contraintes imposées par l'image sur le parcours des courbes à détecter. Les termes de courbure et de co-circularité représentent, à l'inverse, des relations structurelles internes à ces courbes.

Soit, pour une courbe γ composée des pixels $\{P, P_1, \cdots, P_n\}$ et des éléments d'orientation $\{e_1, e_2 \cdots, e_n\}$:

$$\mathcal{F}(P, e_1) = \begin{cases} \alpha_c \cdot \mathcal{C}(\, P, e_1\,) + \alpha_k \cdot \mathcal{K}(\, P, e_1\,) \quad (\, influences \;\; internes\,) \\ \\ + \quad \alpha_g \cdot \mathcal{G}(\, P, e_1\,) + \alpha_o \cdot \mathcal{O}(\, P, e_1\,) \quad (\, influences \;\; externes\,) \end{cases} \tag{4.8}$$

avec ($\forall i \in [\, c, k, g, o\,] \quad 0 \; < \; \alpha_i \; < \; 1$) , paramètres déterminant l'importance de chaque terme au sein de la fonction de qualité.

Ces paramètres permettent de moduler l'importance de chaque relation structurelle au sein de la fonction de qualité. Ils sont respectivement liés à la courbure (α_c), la co-circularité (α_k), la continuité d'intensité (α_g) et celle d'orientation (α_o). Ainsi, une forte valeur pour (α_g) favorisera les courbes passant par le plus grand nombre de pixels de contours alors qu'une forte valeur de (α_k) donnera un avantage aux cycles par rapport à des courbes ouvertes. Le réglage de ces paramètres pour la détection d'un certain type de courbe peut prendre un certain nombre d'essais. Par contre, une fois réglés, les paramètres montrent une bonne stabilité pour la détection d'un même type de courbe sur des images différentes.

A titre d'exemple de cette stabilité, les résultats présentés pour le groupement de pixels ont été réalisés avec les mêmes paramètres :

$$\alpha_c \; = \; 0.5 \; , \; \alpha_k \; = \; 0.2 \; , \; \alpha_g \; = \; 0.9 \; , \; \alpha_o \; = \; 0.2$$

Comme tous les autres paramètres définis par la suite dans ce mémoire, ces valeurs ont été définies de manière empirique, à la suite d'essais sur différentes scènes. La robustesse de la méthode permet néanmoins de conserver le même jeu de paramètres une fois celui-ci fixé pour une classe de groupements donnée. Cette robustesse permet d'envisager, pour un certain nombre de ces paramètres, un apprentissage supervisé éventuel, en comparant les groupements obtenus avec les expressions paramètriques d'un jeu de courbes de référence. Cependant, au lieu de rechercher, dans le détail, un apprentissage particulier, nous avons privilégié une approche globale du problème de groupement, depuis la détection de contours jusqu'à l'interpétation.

Détaillons à présent chacun de ces termes :

– **Courbure globale et co-circularité:** *Relations structurelles internes*

L'une des principales différences de cette fonction de qualité avec l'expression d'origine est la dissociation des termes de courbure et de continuité. Le terme de courbure est donc uniquement composé de l'évaluation de courbure globale défini par la relation A.1, page 258.

Soit $C_{1,n}$ l'évaluation de la courbure globale entre les pixels P_1 et P_n :

$$C_{1,n} \;=\; \prod_{k=1}^{n-1} f(e_k, e_{k+1}) \;=\; \exp^{-\sum_{k=1}^{n-1} \frac{2\theta_k}{\Delta S_{k,k+1}} tan\frac{\theta_k}{2}}$$

Le terme de courbure globale pour γ est donc de la forme :

$$\mathcal{C}(P, e_1) \;=\; \mathcal{C}_0(P, e_1) \;+\; \mathcal{C}_0(P, e_1) \cdot \sum_{i=1}^{n-1} \rho_c^i \cdot f(e_i, e_{i+1}) \cdot C_{1,i} \qquad (4.9)$$

où $\mathcal{C}_0(P, e_1)$ représente la courbure locale au point P.

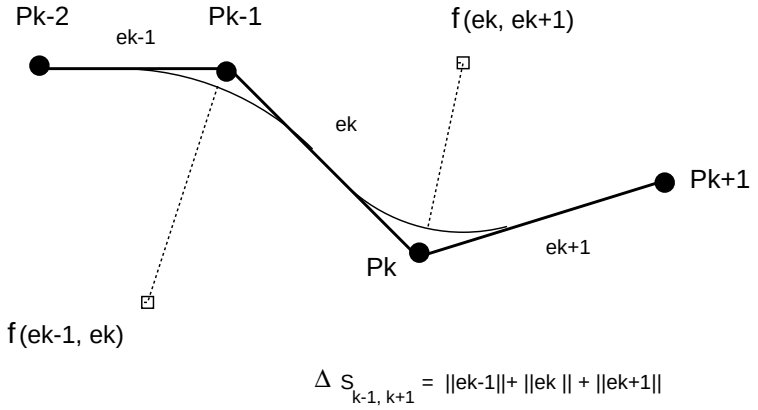

Figure 4.17 - *Notations pour l'estimation de la co-circularité entre trois éléments de connexions e_{k-1}, e_k et e_{k+1}.*

L'expression théorique de la co-circularité locale est une dérivée seconde de l'orientation en un point de la courbe, par rapport à l'abscisse curviligne de ce point, ($d^2\theta/ds^2$). La discrétisation des parcours du graphe conduit à estimer la co-circularité à partir de la différence du terme de courbure entre trois éléments d'orientation consécutifs, soit :

$$\kappa(e_{j-1}, e_j, e_{j+1}) \;=\; \frac{\mid \mathrm{sgn}(\theta_{j-1})\; f(e_{j-1}, e_j) \;+\; \mathrm{sgn}(\theta_j)\; f(e_j, e_{j+1}) \mid}{2\; \Delta S_{j-1,j+1}}$$

Le terme de co-circularité globale est alors de la forme :

$$\mathcal{K}(P, e_1) \;=\; \mathcal{K}_0(P, e_1) \;+\; \sum_{i=2}^{n-1} \rho_k^i \cdot \kappa(e_{i-1}, e_i, e_{i+1}) \qquad (4.10)$$

Les valeurs initiales de chaque terme, $\mathcal{C}_0(e_1)$ et $\mathcal{K}_0(e_1)$ représentent la contribution locale en termes de courbures et de co-circularité, de l'élément e_1 pour le pixel P. Comme elles dépendent de la paire d'éléments $(e_1, \overline{e_1})$ définie autour de P, leur expression sera fixée lors de la définition de la mesure de saillance.

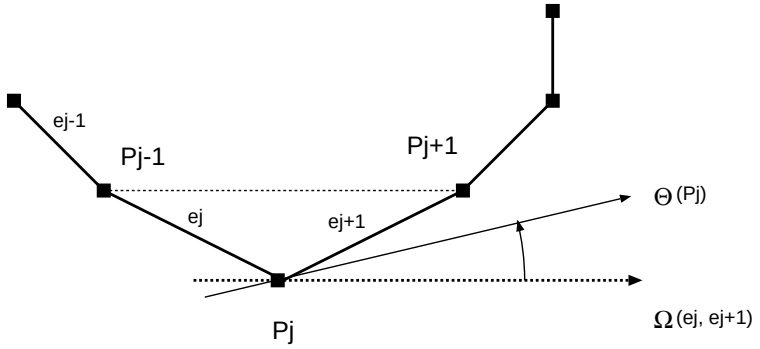

Figure 4.18 - *Notations pour le terme d'orientation entre deux éléments e_j et e_{j+1}.*

– **Continuité d'intensité et d'orientation:** *Influences externes*

Afin de pénaliser les courbes discontinues, nous définissons un terme de continuité d'intensité comme la somme des intensités des pixels d'un groupe pondérées par un facteur d'atténuation. L'intensité en chaque pixel est définie par :

$$\sigma_j \;=\; \begin{cases} 1, & \text{si } P_j \text{ est un pixel de contour} \\ 0, & \text{sinon} \end{cases}$$

Le terme recherché est donc :

$$\mathcal{G}(P, e_1) \;=\; \mathcal{G}_0(P, e_1) \;+\; \sum_{i=1}^{n} \rho_g^i \cdot \sigma_i \qquad (4.11)$$

Provenant directement de la détection de contours, ce terme d'intensités est trop dépendant de mesures locales. Une mesure plus globale, comme l'estimation de la tangente en chaque point de contour[6], permet d'assurer à l'optimi-

6. [Parent et Zucker, 1989] montrent comment la direction de la tangente le long d'un contour peut être estimée de manière relativement simple à l'aide d'un jeu de filtres directionnels constitués de différences de gaussiennes.

sation de courbes une plus grande robustesse face au bruit. Le terme d'orientation qui en découle mesure la différence entre cette tangente $\Theta(P_j)$ et une estimation locale entre éléments de connexion $\Omega(e_j, e_{j+1})$:

$$\Phi_j = \exp^{(\,-Tan|\Theta(P_j)-\Omega(e_j,e_{j+1})|\,)}$$

Les angles $\Omega(e_j, e_{j+1})$ sont pré-calculés de la manière suivante :

$$\Omega(e_j, e_{j+1}) \;\simeq\; \text{orient}(\overrightarrow{P_{j-1}, P_{j+1}})$$

Ce critère est maximal lorsque les deux orientations coïncident.

Le terme de continuité d'orientations est finalement :

$$\mathcal{O}(P, e_1) \;=\; \mathcal{O}_0(P, e_1) \;+\; \sum_{i=1}^{n-1} \rho_o^i \cdot \Phi_i \tag{4.12}$$

4.3.4 Mesure de saillance

En reprenant la relation 4.4, page 107, la mesure de saillance à optimiser dépend de la définition d'une paire d'éléments $(v, \overline{v})$ autour d'un pixel P. Elle s'exprime sous la forme d'une somme de qualités des branches partant de P dans chaque direction de la paire d'éléments.

Si on note Γ_P la courbe correspondant à ces deux branches, sa saillance est définie par :

$$S^n(\Gamma_P) \;=\; \left\{ \begin{array}{lll} & \alpha_c \cdot (\mathcal{C}^n(\,P, v\,) + \mathcal{C}^n(\,P, \overline{v}\,) + H_c(P)) \\ + & \alpha_k \cdot (\mathcal{K}^n(\,P, v\,) + \mathcal{K}^n(\,P, \overline{v}\,) + H_k(P)) \\ + & \alpha_g \cdot (\mathcal{G}^n(\,P, v\,) + \mathcal{G}^n(\,P, \overline{v}\,) + H_g(P)) \\ + & \alpha_o \cdot (\mathcal{O}^n(\,P, v\,) + \mathcal{O}^n(\,P, \overline{v}\,) + H_o(P)) \end{array} \right. \tag{4.13}$$

En tenant compte de la paire $(v, \overline{v})$, les valeurs initiales de chaque terme de la fonction de qualité sont respectivement :

$$\left\{ \begin{array}{ll} \mathcal{C}_0(\,P, v\,) \;=\; f(\overline{v}, v) & \mathcal{C}_0(\,P, \overline{v}\,) \;=\; f(v, \overline{v}) \\ \mathcal{K}_0(\,P, v\,) \;=\; \kappa(\overline{v}, v, e_2) & \mathcal{K}_0(\,P, \overline{v}\,) \;=\; \kappa(\overline{e_{-2}}, \overline{v}, v) \\ \mathcal{G}_0(\,P, v\,) \;=\; \sigma_P & \mathcal{G}_0(\,P, \overline{v}\,) \;=\; \sigma_P \\ \mathcal{O}_0(\,P, v\,) \;=\; \Phi_P & \mathcal{O}_0(\,P, \overline{v}\,) \;=\; \Phi_P \end{array} \right. \tag{4.14}$$

Les termes de courbure et de co-circularité sont établis sur les éléments du voisinage direct du pixel P. On note σ_P l'intensité du pixel P et Φ_P la tangente locale de la courbe en P :

$$\Phi_P \;=\; \exp^{(\,-Tan|\Theta(P_j)-\Omega(\overline{v},v)|\,)} \;=\; \exp^{(\,-Tan|\Theta(P_j)-\Omega(v,\overline{v})|\,)}$$

Ce qui conduit aux fonctions de correction suivantes :

$$\forall P \in \mathcal{P} \quad \begin{cases} H_c(P) \;=\; -f(\overline{v},v) \\ H_k(P) \;=\; 0 \\ H_g(P) \;=\; -\sigma_P \\ H_o(P) \;=\; -\Phi_P \end{cases} \tag{4.15}$$

Leur rôle est d'éliminer les termes communs à chaque contribution latérale en P.

4.3.5 Optimisation

Les particularités de l'algorithme d'optimisation dues aux choix des pixels comme primitives portent essentiellement sur la définition des paires interdites.

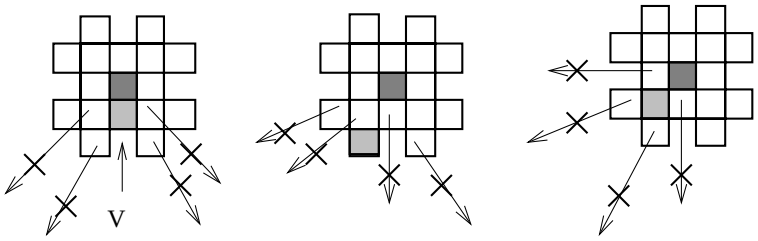

Figure 4.19 - *Paires interdites pour les trois types d'éléments v du voisinage d'un pixel. Les éléments e_j interdits correspondent dans chaque cas à une valeur $f(v,e_j) < 0.05$.*

Lors de la recherche d'un candidat pour l'appariement avec un élément v, les paires formant des angles trop aigus sont rejetées pour éviter la formation de boucles ou de rebroussements lors du suivi. Les éléments immédiatement voisins de v forment une sorte d'angle mort interdisant toute connexion. En pratique, le choix des candidats est établi par simple seuillage sur la valeur de $f(v,e_j)$.

$$f(v,e_j) < 0.05 \quad \Longrightarrow \quad paire\ (v,e_j)\ interdite$$

L'expression récursive des termes de la mesure de saillance se dérivent naturellement de leur définition. Soit, pour l'une des deux branches d'un groupement traversant un pixel P selon la paire $(\overline{v}, v)$:

– Initialisation : $n = 0$

$$\begin{cases} \mathcal{C}^0(P,v) &= f(\overline{v},v) \\ \mathcal{K}^0(P,v) &= \kappa(\overline{v},v,\phi(v)) \\ \mathcal{G}^0(P,v) &= \sigma_P \\ \mathcal{O}^0(P,v) &= \Phi_P \end{cases} \tag{4.16}$$

– Mise à jour : $n \neq 0$

$$\begin{cases} \mathcal{C}^{n+1}(P,v) &= 1 + \rho_c \cdot f(v,\phi(v)) \cdot \mathcal{C}^n(P_v,\phi(v)) \\ \mathcal{K}^{n+1}(P,v) &= \kappa(\overline{v},v,\phi(v)) + \rho_k \cdot \mathcal{K}^n(P_v,\phi(v)) \\ \mathcal{G}^{n+1}(P,v) &= \sigma_P + \rho_g \cdot \mathcal{G}^n(P_v,\phi(v)) \\ \mathcal{O}^{n+1}(P,v) &= \Phi_P + \rho_o \cdot \mathcal{O}^n(P_v,\phi(v)) \end{cases} \tag{4.17}$$

avec P_v, pixel relié à P par l'élément v. Ces variables d'état sont définies de manière symétrique pour la branche du groupement arrivant en P par l'élément $\overline{v}$.

Les poids $\rho_i \in [0\,,\,1]$ constituent le second jeu de paramètres liés à l'optimisation. Ils représentent l'atténuation avec la distance des contributions d'éléments éloignés. Bien qu'ils puissent être éventuellement différents, ils ont été fixés en pratique à une valeur unique, $\rho = 0.95$ pour tous les termes.

4.3.6 Extraction et sélection des meilleurs groupes

La prise en compte de tous les pixels de l'image assure au réseau une mesure de saillance dense sur l'image. Si elle permet de fermer les discontinuités le long de structures importantes, cette densité pose quelques problèmes pour le suivi et la sélection des meilleurs groupements.

– Suivi

Au voisinage d'une structure saillante, les pixels du réseau reçoivent des contributions importantes dans la direction de cette structure.

Supposons qu'une courbe visuellement importante se termine brutalement en un point P de l'image. Les pixels environnant reçoivent une forte contribution

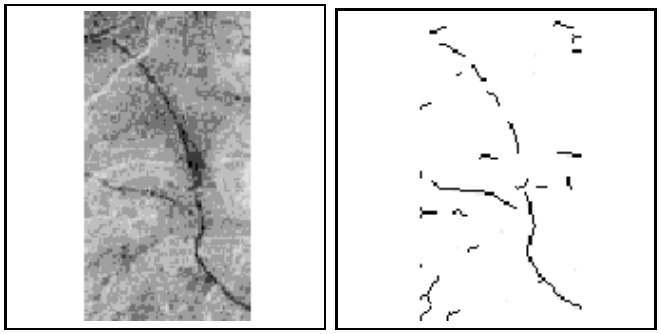

Figure 4.20 - *Détection des lignes de crête d'un fragment d'image satellite infrarouge.*

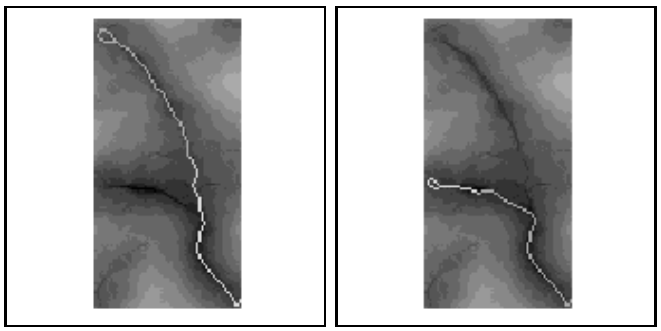

Figure 4.21 - *Exemples de "boucles" indésirables en fin de suivi, après groupement des pixels de lignes de crête.*

dans la direction de P et ce, jusqu'à une distance égale au nombre d'itérations. Un suivi de la courbe arrivant en P se trouverait devant des éléments de connexion de qualité équivalente pour peu qu'il n'y ait aucune autre structure importante dans le voisinage. Il en résulte un suivi d'éléments "virtuels" qui, s'éloignant de P finira fatalement par être attiré par la seule structure importante ; celle de P.

En conséquence de cette observation, le suivi de groupements individuels montre une certaine tendance à former localement des cycles en arrivant sur des zones clairsemées de l'image de contours. Ce problème est facilement résolu en éliminant du suivi les derniers éléments virtuels en cas de cycles.

– Sélection

Chaque pixel pouvant servir de point de départ au suivi d'un groupement, les structures globales de l'image génèrent des classes de groupements de qualité équivalentes. Ces classes sont constituées des groupements dont les points de départ se trouvent dans un proche voisinage d'une structure globale. Afin

de mieux départager les groupements autour d'une même structure, la qualité globale d'un groupement est pondérée par le rapport entre pixels réels et pixels "virtuels" le long du tracé. Les pixels "virtuels" jouant le rôle de discontinuités, on peut considérer ce poids comme un rapport signal sur bruit pour un groupement.

$$\mathcal{Q}(\Gamma_P(v,\overline{v})) \;=\; \frac{N_{\text{réels}}}{1+N_{\text{virtuels}}} \cdot [\sum_{k=1}^{n} S^n(e_{-k},\phi(e_{-k})) \;+\; \sum_{k=1}^{n} S^n(e_k,\phi(e_k))]$$

Enfin, pour compléter les critères de sélection de courbes, la saillance locale d'un pixel est définie comme la norme du gradient de l'intensité d'origine en ce point. Ce qui permet de privilégier les groupements dont les points de départ se trouvent sur un gradient élevé, donc sur un contour probable.

Dans la pratique, les seuils de sélection sont définis par rapport à un pourcentage de la valeur maximale de chaque critère.

4.3.7 Résultats et développements éventuels

Cette méthode a été appliquée à différentes situations, d'abord sur des images synthétiques pour en tester les performances dans des situations extrêmes, puis sur des scènes réelles pour en démontrer l'utilité. Dans tous les cas, les résultats ont été obtenus dans des conditions identiques, avec le même jeu de paramètres[7].

4.3.7.1 Images synthétiques

Les images synthétiques suivantes ont été choisies afin de tester les performances du groupement de pixels dans différentes conditions de bruit : bruit blanc, bruit gaussien et bruit structuré.

Dans chaque cas, un rapport Signal sur Bruit (RSB) peut être défini par :

$$RSB_{dB} \;=\; -10\log\left(\frac{\sum_i\sum_j I(i,j)^2}{\sum_i\sum_j N(i,j)^2}\right)$$

où $I(i,j)$ est la fonction intensité de l'image non bruitée, et $N(i,j)$ celle de l'image altérée.

Les groupements obtenus dans chaque situation ont été sélectionnés manuellement afin de montrer l'existence de groupements correspondant à la forme recherchée.

7. Implémentation en C sur une station de travail SPARC-5.

– *Paramètres de la fonction de qualité*

Le premier jeu de tests illustre le rôle de chaque terme de la fonction de qualité. Dans un premier temps, l'optimisation du réseau de saillance ne tient compte que d'un seul terme de qualité à la fois : termes d'intensité et d'orientation pour la figure 4.23 et termes de courbure et de co-circularité pour la figure 4.24). Dans chaque cas, un groupement représentatif de la fonction de qualité appliquée a été sélectionné manuellement et tracé par dessus la carte de saillance. Ces résultats montrent le rôle dominant des termes d'intensité et de courbure. Leur influence sur les connexions du réseau est suffisante pour guider correctement la reconstitution de groupements réellement saillants. Les termes d'orientation et de co-circularité jouent, quand à eux, un rôle d'appoint. Ces deux termes sont utiles pour corriger certaines situations ambigües mais leur influence doit rester limitée pour qu'ils ne perturbent pas trop le suivi des groupements. Une fonction de qualité composée uniquement des termes de courbure et d'intensité est ainsi suffisante à reconstituer une majeure partie de l'ellipse (figure 4.25), mais n'est pas assez flexible pour permettre des courbes fermées. Les termes de co-circularité (figure 4.26) et d'orientation (figure 4.27) apportent les contraintes manquantes pour reconstituer une plus large gamme de courbes.

– *Stabilité par rapport au bruit*

Les figures 4.28 à 4.30 représentent une ellipse pour laquelle 40% de pixels ont été supprimés aléatoirement. Malgré l'ajout de 5% à 20% de pixels de bruit blanc, une forme circulaire peut être détectée par le réseau. Les situations suivantes représentent respectivement un cercle et une ellipse (figure 4.31) perturbés par un bruit gaussien. Dans ce cas, la détection de contours est perturbée par le faible contraste entre les deux régions. Bien entendu, une meilleure détection de contours aurait pu être obtenue en changeant de paramètre d'échelle (σ pour un filtre gaussien), mais ce n'est pas le but recherché ici. L'intérêt de ces résultats est de montrer qu'un parcours cyclique peut être reconstitué malgré la présence de structures bruitées le long du contours. Les derniers tests, figures 4.34 et 4.35, illustrent l'influence d'un bruit structuré sur le groupement. Malgré des perturbations locales de la forme de la courbe extraite, l'important est de noter l'existence d'un groupement proche de la courbe initiale. La figure 4.35 montre en particulier un groupement selon une courbe de forme quelconque, indépendamment de tout modèle géométrique (cercle ou bien ellipse).

On peut noter enfin les temps de calculs élevés pour des images de taille relativement modeste, ce qui était à prévoir étant donnée la densité du réseau de pixels. Ces temps n'étant dépendant que de la taille de l'image (*i.e*, du nombre de primitives du réseau) et non de la complexité de la scène, il est toutefois possible de prévoir précisément la durée du traitement.

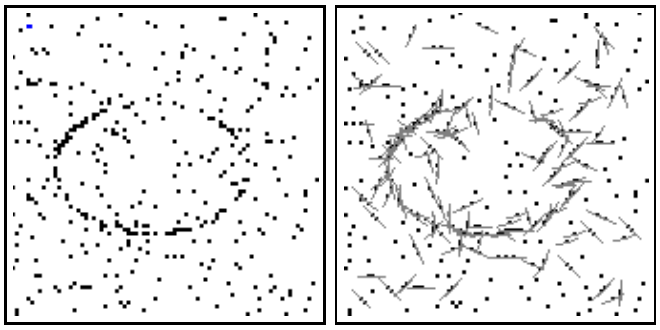

Figure 4.22 - *Ellipse* 80×80 *pixels avec* 5% *de bruit. Image d'intensité et évaluation des orientations locales.*

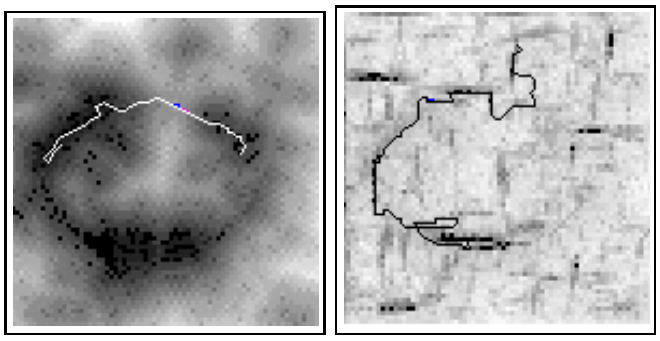

Figure 4.23 - *Optimisation du réseau de saillance et sélection manuelle du meilleur groupement. A gauche: Terme d'intensité seul* $\alpha_g = 0.9$. *A droite: Terme d'orientation seul* $\alpha_o = 0.9$. *La saillance maximale est en noir.*

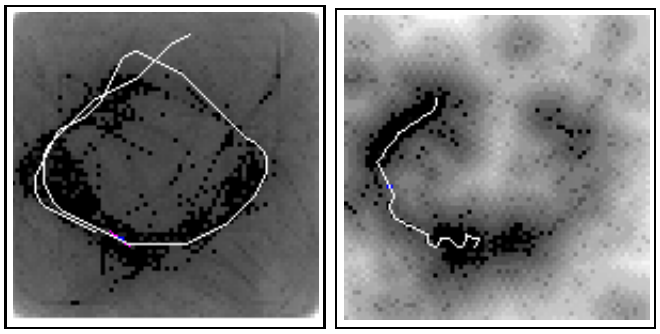

Figure 4.24 - *Optimisation du réseau de saillance et sélection manuelle du meilleur groupement (gauche). A gauche: Terme de courbure seul* $\alpha_c = 0.9$. *A droite: Terme de cocircularité seul* $\alpha_k = 0.9$.

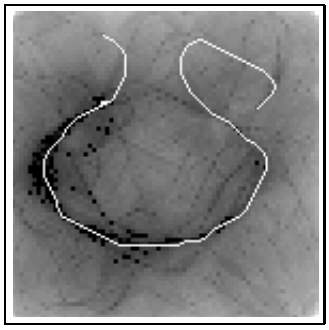

Figure 4.25 - *Optimisation du réseau avec les termes de courbure et d'intensité :* $\alpha_c = 0.6$, $\alpha_k = 0$, $\alpha_g = 0.9$, $\alpha_o = 0$.

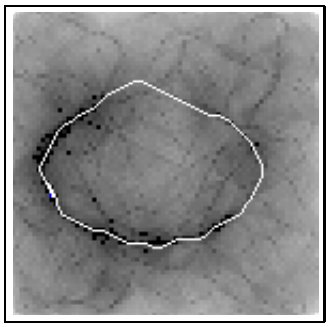

Figure 4.26 - *Optimisation du réseau avec les termes de courbure, d'intensité et de co-circularité :* $\alpha_c = 0.6$, $\alpha_k = 0.2$, $\alpha_g = 0.9$, $\alpha_o = 0$.

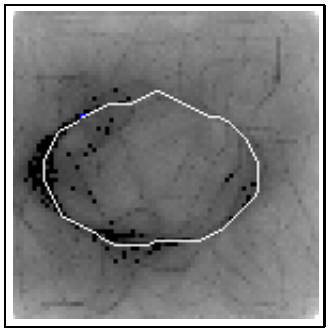

Figure 4.27 - *Optimisation du réseau avec tous les termes :* $\alpha_c = 0.6$, $\alpha_k = 0.2$, $\alpha_g = 0.9$, $\alpha_o = 0.2$.

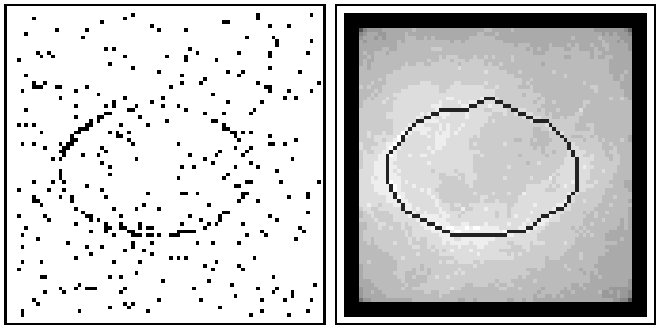

Figure 4.28 - *Ellipse* 80×80 *pixels avec* 5% *de bruit - 30 itérations (30 sec / itération)*

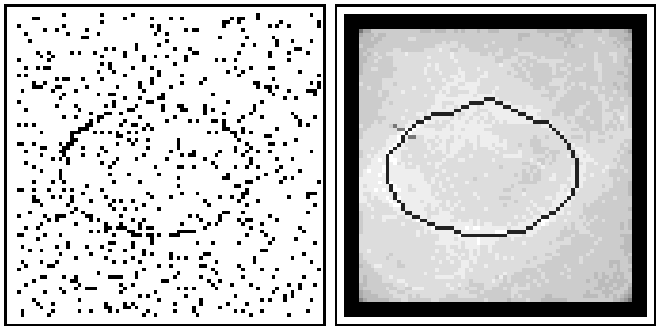

Figure 4.29 - *Ellipse* 80×80 *pixels avec* 10% *de bruit - 25 itérations (30 sec / itération)*

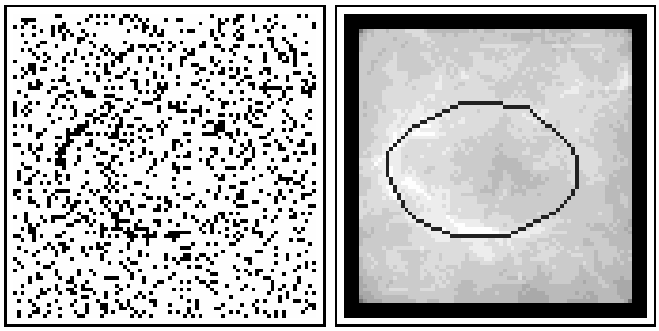

Figure 4.30 - *Ellipse* 80×80 *pixels avec* 20% *de bruit - 20 itérations (30 sec / itération). Le nombre d'itérations inférieur au cas précédent s'explique par la présence de pixels de bruit plus nombreux, qui renforcent accidentellement le parcours de la forme saillante.*

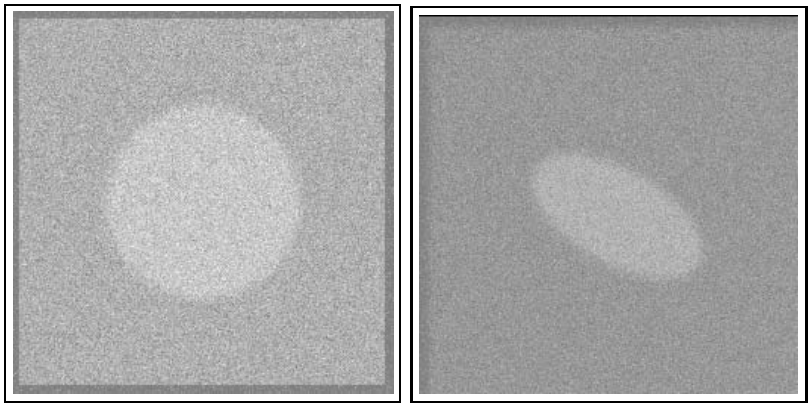

Figure 4.31 - *Cercle et ellipse* 256×256 *pixels, avec bruit gaussien*

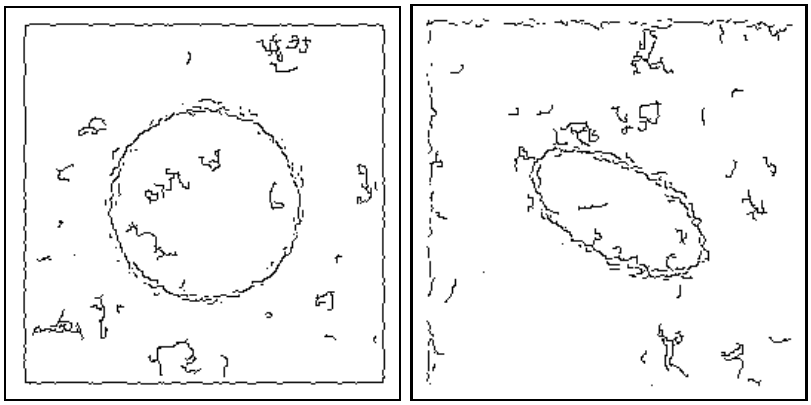

Figure 4.32 - *Détection de contours - Cercle* ($RSB = 5.8db$) *et Ellipse* ($RSB = 7.6db$)

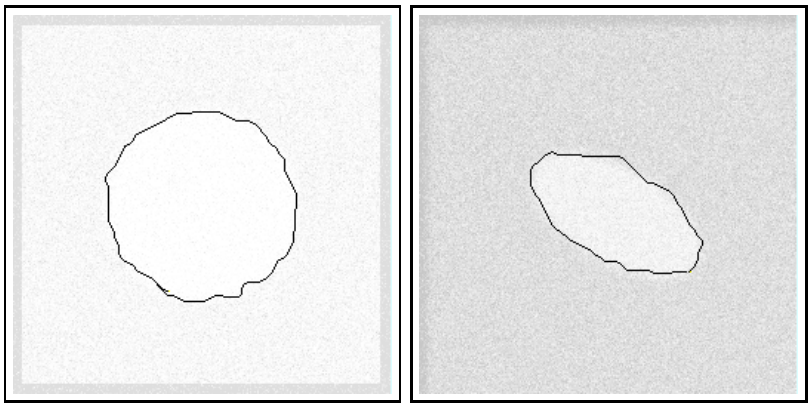

Figure 4.33 - *Groupement - 10 itérations (1 min / itération)*

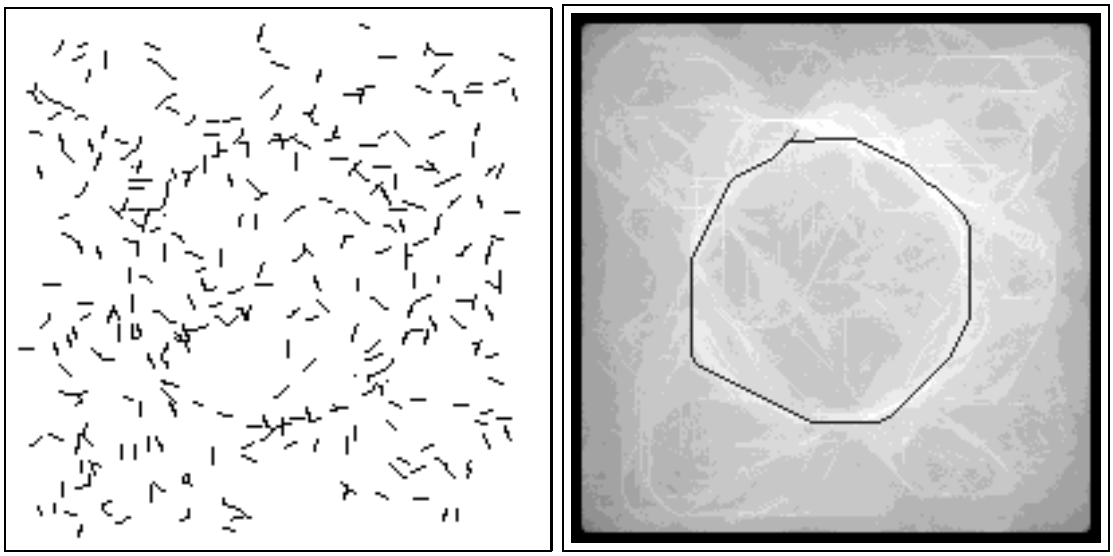

Figure 4.34 - *Cercle avec bruit directionnel. Les segments orientés aléatoirement perturbent la forme finale du groupement. Malgré ce défaut, celui-ci peut néanmoins servir de centre d'attention pour la recherche d'une forme plus précise.*

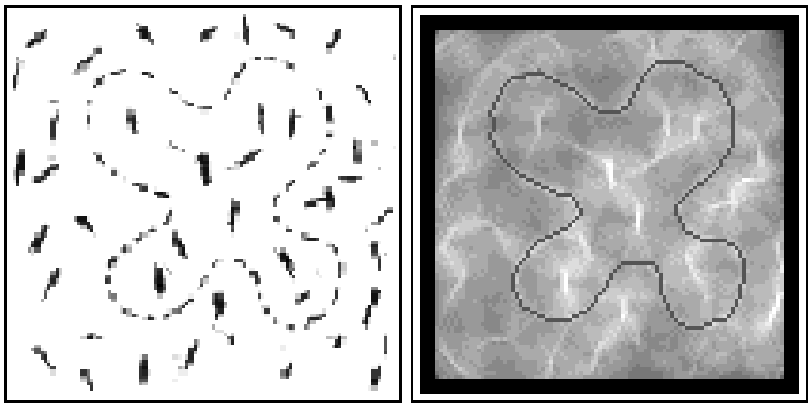

Figure 4.35 - *Courbe et bruit structuré. Ici encore, le groupement est choisi manuellement pour montrer l'existence de parcours corrects dans l'ensemble de groupements possibles sur l'image.*

4.3.7.2 Scènes réelles

Nous illustrons l'utilisation de notre méthode sur des images réelles au travers d'une application particulière d'extraction de réseaux fins. Un réseau fin est un ensemble de structures curvilinéaires tel qu'un réseau routier sur des images satellites ou bien des réseaux sanguins sur des images médicales. Les exemples utilisés ici sont respectivement une image SPOT infrarouge, figure 4.36, et une image d'angiographie du réseau vasculaire du cerveau, figure 4.38.

Dans chaque cas nous présentons l'image d'origine (image monochrome de l'intensité lumineuse) , l'image issue de la détection de réseau fin et finalement, les meilleurs groupements extraits après optimisation.

La méthode utilisée pour la détection de réseaux fins est une détection de contours de type "crêtes" telle que décrite en page 59. Le facteur d'échelle relatif à la largeur des crêtes désirées est le même pour chaque exemple ($\sigma = 1$).

Les paramètres d'optimisation et de calcul de la fonction de qualité sont les mêmes pour les trois images. Ces conditions d'utilisation illustrent bien la stabilité de la méthode face au choix des paramètres de la fonction de qualité. Une fois définis pour une certaine classe de courbes, ici, des réseaux fins, les même paramètres donnent de bons résultats pour de nouvelles images présentant ce type de courbes. Ils ne demandent qu'un ajustement éventuel pour affiner les résultats.

Dans ces trois exemples, les groupements ont été extraits à l'aide des seuils de sélection automatique. Ces seuils sont cependant moins stables que les paramètres de la fonction de qualité. Ils doivent être ajustés pour chaque image, afin d'obtenir les groupements les plus représentatifs de la scène.

Enfin, il est important de noter que l'extraction des meilleurs groupements n'obéit qu'à des critères de continuité géométrique, en dehors de toute connaissance particulière sur le type de scène observée. Dans le cas d'une application réelle, cette connaissance devrait intervenir à partir de la sélection des groupements.

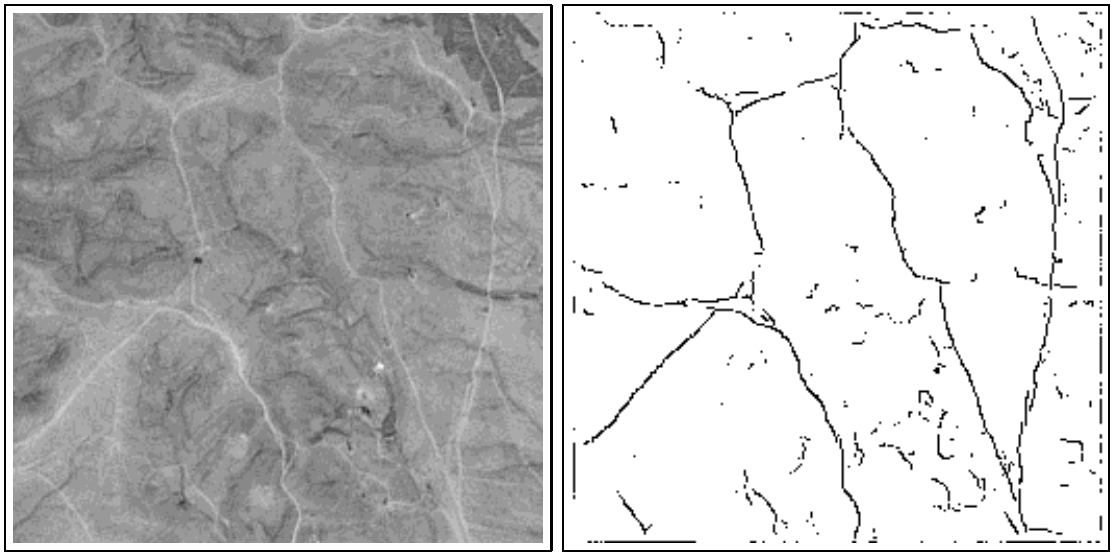

Figure 4.36 - *Image SPOT* 256×256 *pixels - Détection de réseau fin*

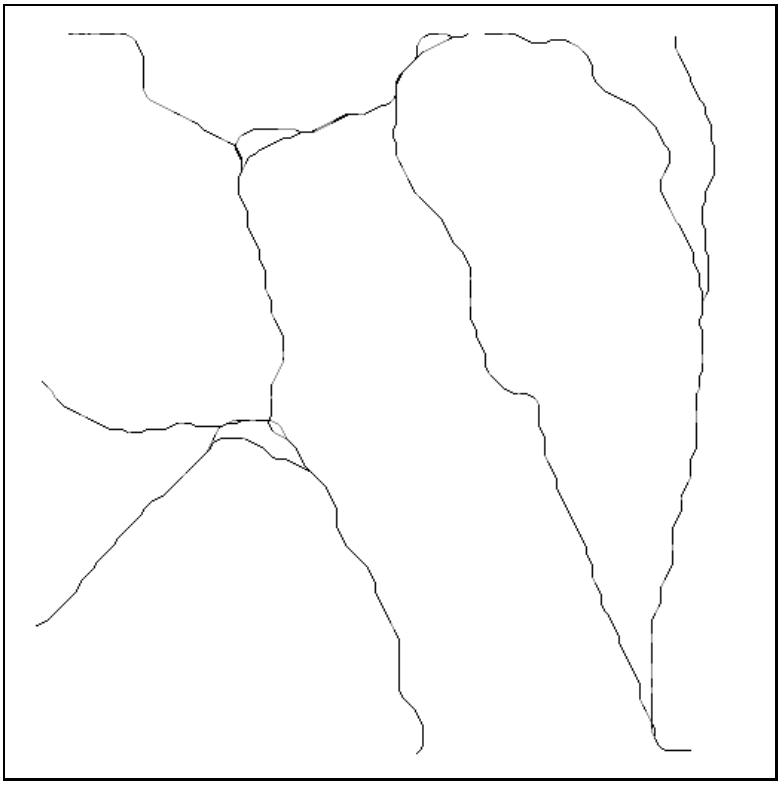

Figure 4.37 - *Détection de routes - Extraction de 16 groupements saillants - 18 itérations (40 sec / itération)*

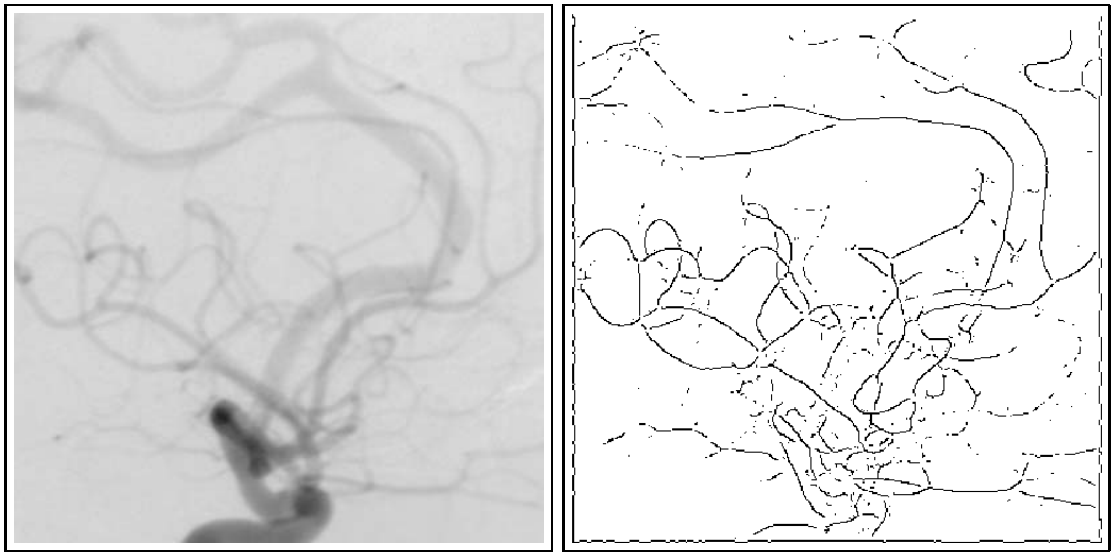

Figure 4.38 - *Angiographie du cerveau* 400×400 *pixels - Détection de réseau fin*

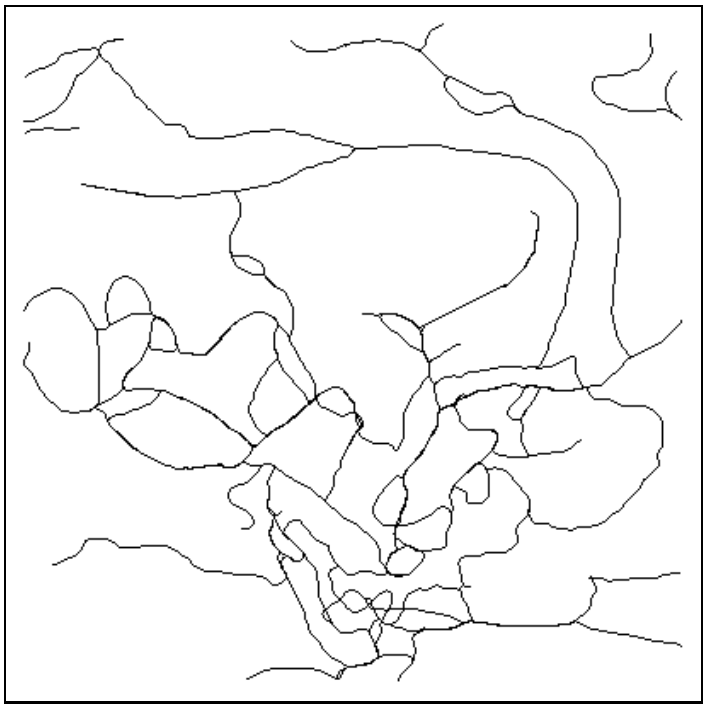

Figure 4.39 - *Extraction d'un réseau de 50 groupements - 30 itérations (2 min / itération)*

4.3.7.3 Conclusion

Malgré la qualité des résultats obtenus, un certain nombre de points mériteraient d'être développés.

Le choix d'un voisinage à 16 orientations s'est révélé être un bon compromis entre qualité des courbes obtenues et complexité de l'algorithme. L'estimation d'une orientation en certains points de contours permettrait d'améliorer ce voisinage en adaptant sa forme à l'environnement de chaque pixel. A ce titre, une version finie des champs d'extension de Guy et Medioni donnerait plus de choix d'orientations autour d'une direction privilégiée tout en interdisant des changements d'orientations trop abrupts.

D'autres extensions possibles à la méthode concernent le suivi et la sélection des meilleurs groupements. En particulier, la présence de paires ambigües pourrait être mise à profit pour autoriser des retours en arrière lors du parcours des paires optimisées et permettre ainsi une exploration plus poussée des meilleurs chemins du graphe.

En dépit des critères de sélection définis pour les points de départ des courbes, un certain nombre de redondances subsistent. La discrimination entre différentes classes de groupements pourrait être améliorée dans une étape supplémentaire. En guise d'exemple, les groupements sélectionnés à l'issue de l'optimisation peuvent servir d'initialisation à autant de contours actifs qu'il suffit ensuite de comparer. Cette méthode a été appliquée à l'extraction de routes en imagerie aérienne dans [Alquier, 1994] [Montesinos et Alquier, 1996] .

La principale limitation de cette méthode de groupement de pixels est bien entendu son coût élevé en temps et ressources de calcul. Chaque itération peut ainsi aller jusqu'à plusieurs minutes pour une image de 512×512 pixels. S'il serait intéressant d'implémenter la méthode sous une forme parallèle, d'autres solutions existent pour améliorer ses performances ; en particulier, comme le montre l'application suivante, par un choix de primitive plus approprié.

4.4 Application au groupement de chaînes

Nous l'avons vu au chapitre 2, l'utilisation d'éléments de contours sous forme de chaînes ou de segments rectilignes constitue une étape préliminaire à de nombreuses méthodes de structuration des images de contours. Ces segments étant considérablement moins nombreux que les pixels de l'image, ils forment de bons candidats pour réduire la complexité de notre méthode de groupement.

4.4.1 Primitive "chaîne"

Par soucis de clarté avec les chapitres suivants, nous appellerons "chaînes" les primitives décrites dans cette partie. Une chaîne est un ensemble de pixels, éventuel-

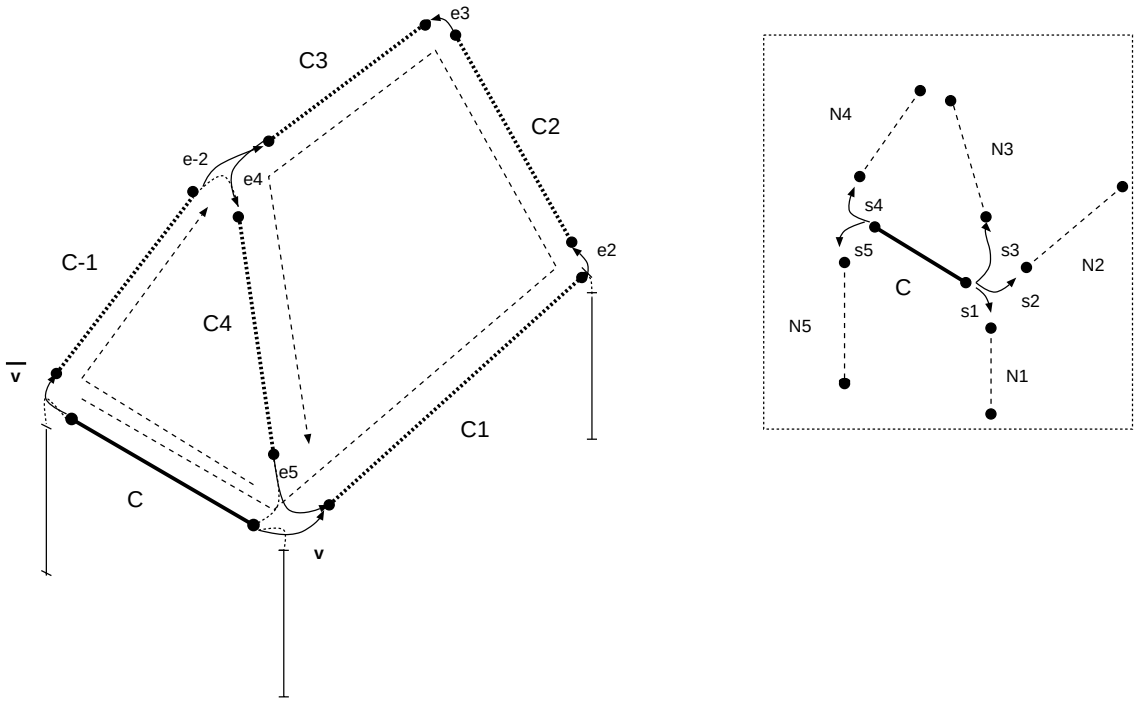

Figure 4.40 - *Exemple de voisinage de chaînes avec deux groupements, représentés par les séquences* $(C, v, C_1, e_2, C_2, \cdots, C_4, e_5)$ *et* $(C, \overline{v}, C_{-1}, e_{-2})$.

lement rectiligne, obtenu à partir de l'image de contours à l'aide d'un pré-traitement classique de chaînage en 8 connexité.

En termes de groupements perceptuels, cette étape préliminaire revient à grouper les pixels de contours selon un critère de proximité afin de former des chaînes de pixels contigus. Un critère de similarité (orientation commune et courbure faible) permet ensuite d'obtenir un ensemble de segments élémentaires. Ce dernier critère est délibérément choisi avec une tolérance très faible de manière à pouvoir reconstituer fidèlement les courbes saillantes de l'image.

A la différence des pixels, la primitive "chaîne" présente évidemment une structure de dimension plus élevée. Le réseau de saillance doit en conséquence tenir compte des extrémités de chaque chaîne, de leur longueur et éventuellement, leur orientation.

4.4.2 Voisinage dynamique

Nous définissons le voisinage d'une chaîne C par la réunion de chaînes détectées à proximité de chaque extrémité. Un exemple de construction de voisinage de chaînes est donné ci-après mais la définition reste ouverte à d'autres possibilités. Contrairement aux pixels, il est impossible de prévoir le nombre de chaînes pré-

sentes dans le voisinage d'une extrémité. La solution qui consisterait à admettre des primitives "virtuelles" se heurte à la difficulté de définition de telles primitives. De quelle longueur et selon quelle orientation doit-on les établir ? A quelle distance des extrémités ? C'est pourquoi nous définissons le voisinage de C comme un ensemble de chaînes "réelles" auxquelles chaque extrémité est susceptible d'être reliée.

– Construction d'un voisinage de chaînes

Une première conséquence de cette définition est un voisinage dynamique, défini à l'aide d'une procédure de recherche de voisins compatibles. A partir des extrémités de chaque chaîne, un cône de recherche est établi dans la direction de chaque extrémité. Sa longueur et l'évolution du diamètre en fonction de la distance permettent de modifier la forme de l'espace de recherche et d'adapter éventuellement le voisinage selon l'application désirée. Toute chaîne dont une extrémité se trouve dans la zone de recherche est testée pour être éventuellement incluse dans le voisinage.

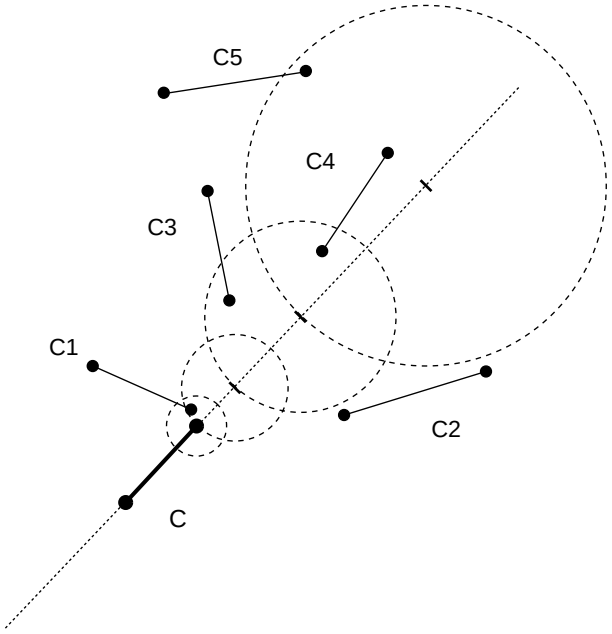

Figure 4.41 - *Cône de recherche pour la construction du voisinage de la chaîne C. La chaîne C_2 se trouve en dehors des zones de recherche successives, elle n'est pas incluse dans le voisinage de C. Les autres chaînes sont des candidats possibles et doivent passer le test de compatibilité afin d'être admises dans $\mathcal{V}(\mathcal{C})$.*

En pratique, le voisinage d'une chaîne est construit à l'aide de cercles centrés sur la direction de la chaîne, à partir de chaque extrémité. La distance d du centre de chaque cercle de recherche avec l'extrémité et leur rayon r sont

exprimés en fonction d'un paramètre t. Le choix de chacune de ces fonctions permet de changer la forme du voisinage de recherche. Par exemple :

$$\textit{Cercles concentriques :} \quad \begin{cases} d(t) \;=\; 0, \quad \forall t \\ \\ r(t) \;=\; a \cdot t, \quad t \;\in\; [0, t_{max}] \end{cases} \tag{4.18}$$

$$\textit{Cône de recherche simple :} \quad \begin{cases} d(t) \;=\; b \cdot t, \quad t \;\in\; [0, t_{max}] \\ \\ r(t) \;=\; a \cdot t, \quad t \;\in\; [0, t_{max}] \end{cases} \tag{4.19}$$

Dans le cas de notre application, les paramètres du cône de recherche sont définis de manière récursive :

$$\begin{cases} d(t+1) \;=\; b \cdot d(t), \quad t \;\in\; [1, 2, \cdots, t_{max}] \qquad\qquad d(0) \;=\; 1 \\ \\ r(t+1) \;=\; d(t), \quad t \;\in\; [1, 2, \cdots, t_{max}] \end{cases} \tag{4.20}$$

La recherche de nouveaux voisins est interrompue en cas de taille maximale de l'espace de recherche ($t \;=\; t_{max}$) ou bien lorsqu'un nombre limite de voisins est atteint. Ces conditions d'arrêt permettent d'éviter les voisinages trop gros par rapport à la taille de la chaîne de départ ainsi que des voisinages trop denses. Notons que les paramètres qui définissent la forme et la taille du voisinage pourraient être ajustés automatiquement en fonction des images observées, et en particulier, par rapport à la densité de chaînes autour de chaque extrémité. En effet, une mesure de densité de chaînes permettrait d'élargir le voisinage de recherche dans les zones de faible densité ou, à l'inverse, limiter les voisinages des chaînes de zones plus denses.

Afin d'être éventuellement incluse dans le voisinage d'une extrémité, une chaîne doit passer un test de compatibilité. Ce test permet d'éviter des configurations locales entre chaînes en contradiction avec le type de groupement recherché.

Soient deux chaînes C_0, C_1 et $S_{0,1}$ le segment rectiligne reliant une extrémité de chaque chaîne, X_0 et X_1. On note enfin $\lambda_{0,1}$ l'angle formé par $S_{0,1}$ et la direction de C_0, et $\lambda_{1,0}$ l'angle formé par $S_{0,1}$ et la direction de C_1.

$$\lambda_{0,1} \;=\; \widehat{\overrightarrow{X_0X_1}, \overrightarrow{T_0}} \qquad \lambda_{1,0} \;=\; \widehat{\overrightarrow{X_1X_0}, \overrightarrow{T_1}}$$

Dans ce cas, l'extrémité X_1 de la chaîne C_1 peut être connectée à l'extrémité X_0 de la chaîne C_0 si les deux angles vérifient la relation :

$$C_1 \;\in\; \mathcal{V}(C_0) \;\;\Longrightarrow\;\; \lambda_{0,1} < \lambda_{seuil} \;\; et \;\; \lambda_{1,0} < \lambda_{seuil} \tag{4.21}$$

L'angle limite λ_{seuil} est arbitrairement fixé à 70 degrés afin d'éviter les connexions trop abruptes. Notons que ce critère strict de compatibilité est donné

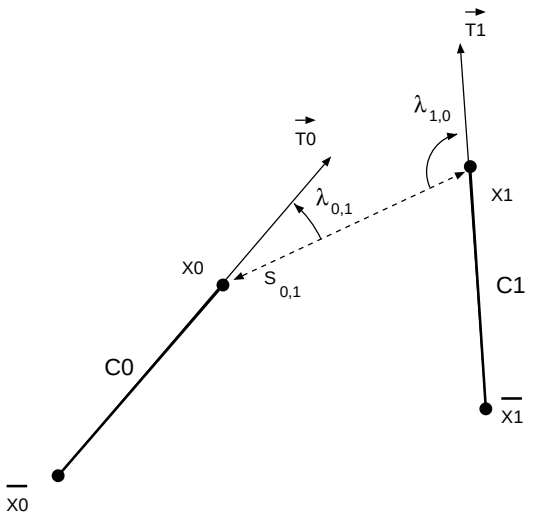

Figure 4.42 - *Test de compatibilité entre une chaîne C_0 et un voisin possible C_1. Ici, le test est négatif car $\lambda_{1,0} > \lambda_{seuil}$.*

à titre d'exemple. Il pourrait être utile, en effet, d'envisager des relations plus "floues" entre chaînes pour résoudre des situations ambigües (par exemple, dans la figure 4.41, les chaînes C_2 et C_3 jouent un rôle équivalent par rapport à C).

– Définition des éléments de connexion

Une autre conséquence de cette définition de voisinage concerne la nature des éléments de connexion entre chaînes, naturellement définis comme un chemin dans l'image reliant deux extrémités de chaque chaîne. Les connexions entre chaînes voisines d'une même extrémité sont interdites afin de forcer toute courbe arrivant par l'extrémité d'une chaîne à ressortir par l'autre extrémité. Cette contrainte correspond à la définition des paires interdites entre voisins d'une même chaîne.

Le modèle choisi pour relier deux chaînes entre elles est une simple courbe polynômiale d'interpolation ou *Cardinal-Spline*. Ce type de courbe est défini par un polynôme d'ordre 3. Soit $M(t)$ une telle courbe et $M'(t)$ sa dérivée, $t \in [0,1]$:

$$M(t) \;=\; \sum_{i=0}^{3} a_i \cdot t^i \quad \Longrightarrow \quad M'(t) \;=\; 3 \cdot a_3 \cdot t^2 \;+\; 2 \cdot a_2 \cdot t + a_1$$

Relier les extrémités de deux chaînes revient à trouver les coefficients de la courbe de manière à ce qu'elle passe par chaque extrémité X_0 et X_1, avec une tangente correspondant à l'orientation de chacune des chaînes, T_0 et T_1.

En écrivant le système pour $t = 0$ et $t = 1$, on obtient :

$$\begin{cases} M(0) = X_0 & M(1) = X_1 \\ M'(0) = T_0 & M'(1) = T_1 \end{cases} \tag{4.22}$$

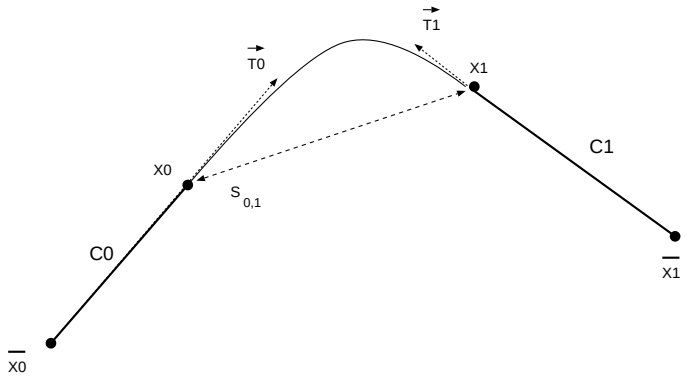

Figure 4.43 - *Elément de connexion entre deux chaînes - une courbe polynômiale définie par les extrémités X_0, X_1 et les tangentes T_0, T_1*

Soit, après substitution :

$$\begin{cases} a_3 = -2 \cdot [M(1) - M(0)] + M'(0) + M'(1) \\ a_2 = 3 \cdot [M(1) - M(0)] - 2 \cdot M'(0) - M'(1) \\ a_1 = M'(0) \\ a_0 = M(0) \end{cases} \tag{4.23}$$

Afin d'éviter des connexions entre chaînes trop petites, les chaînes réduites à un ou deux pixels sont éliminées immédiatement après l'étape d'approximation polygonale.

– Cas des jonctions en T

Relier les chaînes par leurs extrémités interdit la détection de jonctions multiples entre chaînes, à moins d'avoir découpé les chaînes autour de jonctions détectées au préalable. Un détecteur spécialisé de coins ou de jonctions tel que présenté dans la section 2.3.3 permettrait d'avoir un accès direct à ces points de coupure éventuels.

En l'absence de tels détecteurs, une recherche exhaustive des points de coupures possibles permet de préparer les chaînes à la construction du réseau de saillance.Cette recherche consiste à détecter la présence de chaînes traversant un voisinage réduit de chaque extrémité et de les découper dans le prolongement de l'extrémité comme le montre la figure 4.44.

Il appartient à l'algorithme de découpage autours des jonctions de limiter la taille des chaînes à découper afin de ne pas accroitre inutilement le nombre de primitives du réseau de saillance. Un nombre trop important de chaînes réduirait la vitesse d'optimisation du réseau. En outre, les petites chaînes étant peu fiables pour des mesures d'orientations ou de courbure, il n'est pas utile de fragmenter les éléments de contours en chaînes trop petites.

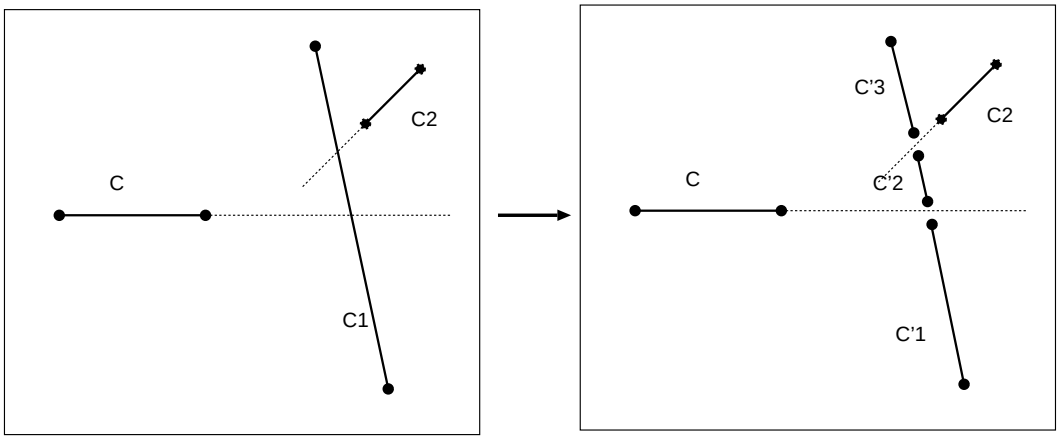

Figure 4.44 - *Le découpage de chaînes dans la direction des extrémités des chaînes voisines est indispensable pour permettre d'éventuelles jonctions en "T". La chaîne C_1 est ainsi remplacée par trois sous chaînes à cause de la proximité des chaînes C et C_2.*

Finalement, la construction du réseau de saillance pour les chaînes est constituée des étapes résumées par l'algorithme suivant :

Algorithme 4.3 : Préparation d'un réseau de saillance de chaînes

début
Chaînage de l'image de contours pour obtenir un ensemble de chaînes primaires.
Elimination des chaînes trop petites.
Estimation des orientations en bout de chaque chaîne.
Découpage autour de jonctions en T éventuelles.
Construction du voisinage aux extrémités de chaque chaîne
fin

4.4.3 Fonction de qualité

De la même manière que pour le groupement de pixels, la fonction de qualité utilisée pour le groupement de chaînes est une combinaison linéaire de quatre termes. Soit, pour une courbe γ composée des chaînes $\{C, C_1, \cdots, C_n\}$ et des éléments de connexion $\{s_1, s_2 \cdots, s_n\}$:

$$\mathcal{F}(C, s_1) = \begin{cases} \alpha_t \cdot \mathcal{T}(\, C, s_1\,) + \alpha_k \cdot \mathcal{K}(\, C, s_1\,) & (\, influences \;\; internes\,) \\ \\ + \quad \alpha_d \cdot \mathcal{D}(\, C, s_1\,) + \alpha_o \cdot \mathcal{O}(\, C, s_1\,) & (\, influences \;\; externes\,) \end{cases} \tag{4.24}$$

Cette fonction doit être forte pour des groupements de chaînes selon des courbes lisses, continues et peu sinueuses. Les termes de cette somme sont définis à l'aide des relations suivantes :

- **Relations structurelles internes**

  L'estimation de la courbure, et à plus forte raison de la co-circularité, se heurte à un difficile problème d'échelle lorsqu'il s'agit de chaînes de pixels [Worring et Smeulders, 1993] . Les approximations utilisées pour le groupement de pixels n'ont de sens que pour des évaluations locales, sur des portions élémentaires de courbes. Dans le cas de chaînes, nous privilégions la recherche de caractéristiques indirectement dérivées de la courbure.

  Afin de favoriser les courbes peu sinueuses par rapport à des groupements plus bruités, nous introduisons un premier terme de courbure. Ce terme est maximal lorsqu'un élément de connexion relie deux chaînes colinéaires et décroît en fonction des angles entre cet élément et chacune des chaînes :

$$\mathcal{T}(C, s_1) \;=\; \sum_{i=1}^{n} \rho_t^{i-1} \cdot \frac{2\pi - |\lambda_{i-1,i}| - |\lambda_{i,i-1}|}{2\pi} \tag{4.25}$$

  Le terme suivant est assimilable à un terme de co-circularité. Comme c'était le cas pour les pixels, ce terme apporte une information plus globale sur la forme d'un groupement. Il permet de favoriser les portions de courbes tournant d'une manière régulière dans un même sens.

  Ce terme est nul lorsque la connexion entre chaînes forme une inflexion, c'est à dire, lorsque les angles $\lambda_{0,1}$ et $\lambda_{1,0}$ sont de même signe.

  Dans le cas contraire, les cercles passant par trois sommets consécutifs de la connexion, respectivement $(\overline{X_0}, X_0, X_1)$ et $(X_0, X_1, \overline{X_1})$, permettent de définir :

$$K(C, C_1) \;=\; \begin{cases} \dfrac{\min(r_0, r_1)}{\max(r_0, r_1)} & si \quad sgn(\lambda_{0,1}) \cdot sgn(\lambda_{1,0}) > 0 \\ \\ 0 & sinon \end{cases} \tag{4.26}$$

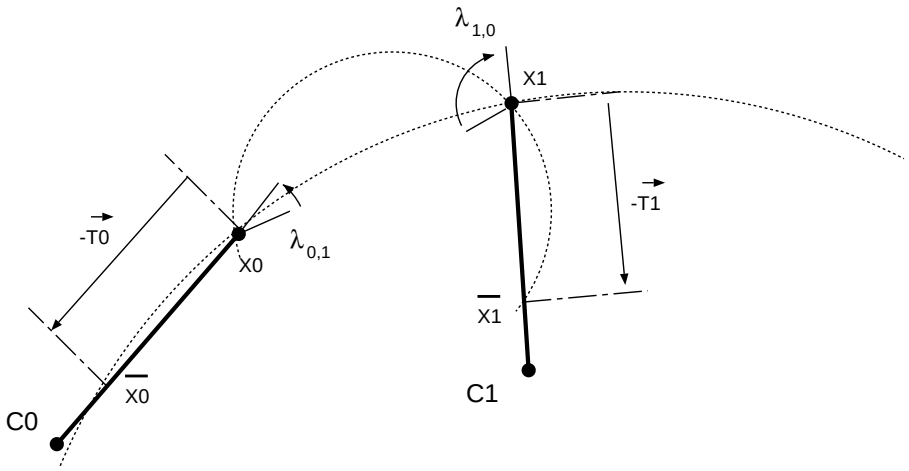

Figure 4.45 - *Estimation de la co-circularité de deux chaînes. Dans le cas où les deux angles $\lambda_{0,1}$ et $\lambda_{1,0}$ sont de signe opposé, la connexion entre les deux chaînes ne forme pas d'inflexion. La co-circularité est estimée par le rapport entre les rayons des cercles "porteurs".*

où r_0 et r_1 sont les rayons respectifs des deux cercles. Ce terme est maximal lorsque les deux cercles ont un rayon identique. Les points $\overline{X_0}$ et $\overline{X_1}$ sont définis à l'aide des vecteurs d'orientation en extrémité des chaînes, comme le montre la figure 4.45. Les deux chaînes sont d'autant plus co-circulaires que les deux cercles ont un rayon similaire.

Soit, pour le terme total :

$$\mathcal{K}(C, s_1) \;=\; \sum_{i=1}^{n} \rho_k^{i-1} \cdot K(C_i, C_{i-1}) \tag{4.27}$$

– **Influences externes**

Les contraintes imposées par l'image sont plus proches de celles utilisées pour les pixels. Leur rôle est de pénaliser les courbes discontinues et présentant des écarts angulaires trop importants.

Le premier terme externe récompense les connexions dont la longueur reste faible par rapport à la longueur des chaînes connectées, ce qui permet de récompenser les courbes contenant peu de discontinuités.

Soient L_0 et L_1 les longueurs respectives de chaque chaîne et $Ls_{0,1}$ la longueur de l'élément de connexion entre ces chaînes. Le terme recherché peut être défini par :

$$\mathcal{D}(C, s_1) \;=\; \sum_{i=1}^{n} \rho_d^{i-1} \cdot \frac{(L_{i-1} + L_i)}{(L_{i-1} + L_i + Ls_{i-1,i})} \tag{4.28}$$

La contribution de ce terme est proche de 1 lorsque la longueur de la connexion est faible devant les longueurs des chaînes. Sa valeur décroît ensuite en fonction de la longueur de connexion.

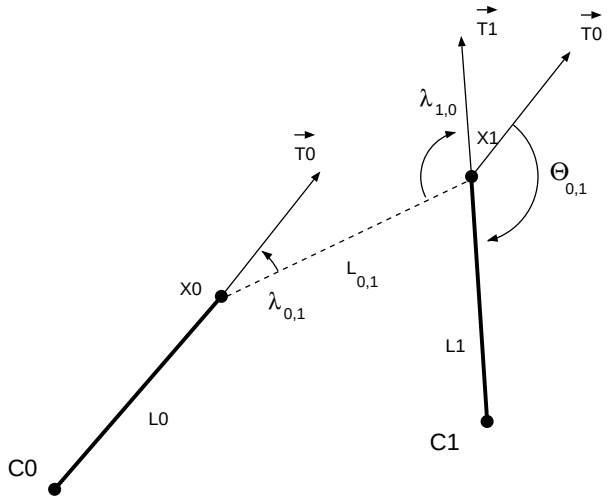

Figure 4.46 - *Mesures de longueurs et d'orientation entre deux chaînes et leur connexion.*

Pénaliser les écarts angulaires forts revient simplement à calculer cet écart. Soient $\overrightarrow{T_0}$ et $\overrightarrow{T_1}$ les orientations aux extrémités de chaque chaîne. L'écart angulaire entre ces tangentes est noté :

$$\Theta_{0,1} \;=\; (\overrightarrow{T_0}, \widehat{-\overrightarrow{T_1}})\ [\pi]$$

Ce qui conduit au terme d'orientations :

$$\mathcal{O}(C, s_1) \;=\; \sum_{i=1}^{n} \rho_o^{i-1} \cdot (1.0 - (\frac{\Theta_{i-1,i}}{\pi})) \tag{4.29}$$

Ce terme accorde de l'importance aux connexions lorsque les chaînes partagent une direction similaire. Sa valeur décroît lorsque l'écart angulaire entre les deux directions se rapproche de π.

Les termes ainsi définis permettent de contrôler le type de courbes désiré de la même manière que pour le groupement de pixels.

4.4.4 Mesure de saillance

En conservant toujours les mêmes notations, la mesure de saillance à optimiser est exprimée en fonction des courbes partant d'une chaîne C, selon les directions d'une paire d'éléments de connexion $(v, \overline{v})$. Soit, pour une courbe Γ_C :

$$S^n(\Gamma_C) \;=\; \left\{ \begin{array}{lll} & \alpha_t \cdot (\mathcal{T}^n(\,C,v\,) + \mathcal{T}^n(\,C,\overline{v}\,) + H_t(C)) \\ + & \alpha_k \cdot (\mathcal{K}^n(\,C,v\,) + \mathcal{K}^n(\,C,\overline{v}\,) + H_k(C)) \\ + & \alpha_d \cdot (\mathcal{D}^n(\,C,v\,) + \mathcal{D}^n(\,C,\overline{v}\,) + H_d(C)) \\ + & \alpha_o \cdot (\mathcal{O}^n(\,C,v\,) + \mathcal{O}^n(\,C,\overline{v}\,) + H_o(C)) \end{array} \right. \tag{4.30}$$

En tenant compte de la paire $(v, \overline{v})$, les valeurs initiales de chaque terme de la fonction de qualité sont, pour la branche de la courbe relative à v :

$$\left\{ \begin{array}{lll} \mathcal{T}_0(\,C,v\,) \;=\; \frac{2\pi - |\lambda_{0,1}| - |\lambda_{1,0}|}{2\pi} & & \mathcal{K}_0(\,C,v\,) \;=\; K(C,v) \\ \mathcal{D}_0(\,C,v\,) \;=\; \frac{(L_0+L_1)}{(L_0+L_1+Ls_{0,1})} & & \mathcal{O}_0(\,C,v\,) \;=\; (1.0 - (\frac{\Theta_{0,1}}{\frac{\pi}{2}})) \end{array} \right. \tag{4.31}$$

Les valeurs initiales des termes de l'autre branche sont symétriques. Contrairement à la fonction de qualité des pixels, aucun de ces termes ne dépend à la fois de v et de $\overline{v}$, ce qui conduit aux fonctions de correction suivantes :

$$H_i(C) \;=\; 0, \quad \forall i \;\in\; [t,k,d,o]$$

4.4.5 Optimisation et sélection des meilleures courbes

Comme nous l'avons signalé lors de la définition du voisinage d'une chaîne, la principale contrainte imposée sur le choix des paires d'éléments interdit toute connexion entre éléments partageant une même extrémité de la chaîne.

L'équation 4.31 donne directement la valeur initiale des variables d'état de chaque élément de connexion. La relation de récurrence pour chaque terme de la fonction de qualité est tout aussi immédiate : Soit, pour l'une des deux branches, avec $n \;\neq\; 0$:

$$\left\{ \begin{array}{lll} \mathcal{T}^{n+1}(C,v) \;=\; \mathcal{T}_0(C,v) \;+\; \rho_t \cdot \mathcal{T}^n(C_i, \phi(v)) \\ \mathcal{K}^{n+1}(C,v) \;=\; \mathcal{K}_0(C,v) \;+\; \rho_k \cdot \mathcal{K}^n(C_i, \phi(v)) \\ \mathcal{D}^{n+1}(C,v) \;=\; \mathcal{D}_0(C,v) \;+\; \rho_d \cdot \mathcal{D}^n(C_i, \phi(v)) \\ \mathcal{O}^{n+1}(C,v) \;=\; \mathcal{O}_0(C,v) \;+\; \rho_o \cdot \mathcal{O}^n(C_i, \phi(v)) \end{array} \right. \tag{4.32}$$

avec C_i, chaîne reliée à C par l'élément v. Les coefficients d'atténuation sont, ici aussi, identiques : $\forall i, \quad \rho_i \;=\; 0.9$.

Le nombre de chaînes dans une image étant considérablement plus faible que le nombre de pixels, l'optimisation d'un réseau de chaînes permet d'obtenir des temps de calculs considérablement réduits. En pratique, ces temps dont de l'ordre de la

seconde par itération, pour un résultat comparable au groupement de pixels. Les résultats présentés plus loin permettent de mieux comparer les deux méthodes de groupement.

L'algorithme d'extraction et de sélection des meilleures courbes présente peu d'adaptations spécifiques aux chaînes. L'absence de groupements avec des éléments virtuels permet d'éviter les cycles parasites en fin de courbe ouverte que présentait le groupement de pixels. La construction des chaînes et le découpage préalable autour des extrémités rend impossible toute intersection entre chaînes autrement que par le biais d'un élément de connexion. Cette propriété permet de limiter la recherche de cycles éventuels aux seules courbes de liaison entre chaînes.

Seules les définitions des critères de saillance locale et globale dépendent réellement du choix du type de primitive à grouper. Dans le cas des chaînes, la saillance locale est définie comme la somme des gradients le long d'une chaîne, normalisée par sa longueur. Soit, en notant $\{P_1, \cdots, P_m\}$ les pixels d'une chaîne et $I(P_i)$ l'intensité de l'image d'origine en P_i :

$$\mathcal{L}(\Gamma_C) \;=\; \frac{\sum_{k=1}^{n} \sqrt{(\frac{\partial I(P_k)}{\partial x})^2 \;+\; (\frac{\partial I(P_k)}{\partial x})^2}}{L_C \cdot G_{max}}$$

où L_C est la longueur de la courbe et G_{max} la valeur maximale de la norme du gradient sur toute l'image.

La mesure de saillance globale d'un groupement est calquée sur celle utilisée pour les pixels. La seule différence est la manière de mesurer la proportion de discontinuité le long de la chaîne. Soit L_{chaines} la somme des longueurs des chaînes du groupement et L_{liens} la somme des longueurs des éléments de connexion entre ces chaînes. La saillance globale est alors définie par :

$$\mathcal{Q}(\Gamma_C(v,\overline{v})) \;=\; \frac{L_{\text{chaines}}}{1+L_{\text{liens}}} \cdot [\sum_{k=0}^{n-1} S^n(\overline{s_k}, \phi(\overline{s_k})) \;+\; \sum_{k=0}^{n-1} S^n(s_k, \phi(s_k))]$$

4.4.6 Résultats et perspectives

Par soucis de comparaison, nous avons appliqué le groupement de chaînes au même type de scènes que pour le groupement de pixels. Les résultats ont été obtenus, ici encore, avec un même jeu de paramètres pour la fonction de qualité. Les seules variations de paramètres concernent la taille de l'espace de recherche pour la construction du réseau[8] D'autres types de scènes permettent ensuite d'illustrer le comportement de notre approche dans des situations variées.

4.4.6.1 Images synthétiques

Ces images confirment la résistance de la méthode au bruit structuré. Contrairement à la méthode précédente, le bruit ponctuel joue ici un rôle minime étant donné

8. Implémentation en C sur une station de travail Ultra-SPARC et sur PC.

l'élimination des chaînes de petite taille avant la construction du réseau. Dans chacune de ces scènes, les groupements ont été sélectionnés à l'aide des seuils de qualité. Les groupements solutions sont présentés en noir et les chaînes rejetées en gris.

Le première scène, figures 4.47 et 4.48, illustre les différentes étapes du groupement, depuis la construction du réseau jusqu'à l'extraction des meilleurs groupes. Le cercle bruité de la figure 4.50 montre l'intérêt de superposer les meilleurs groupes pour en extraire une forme saillante.

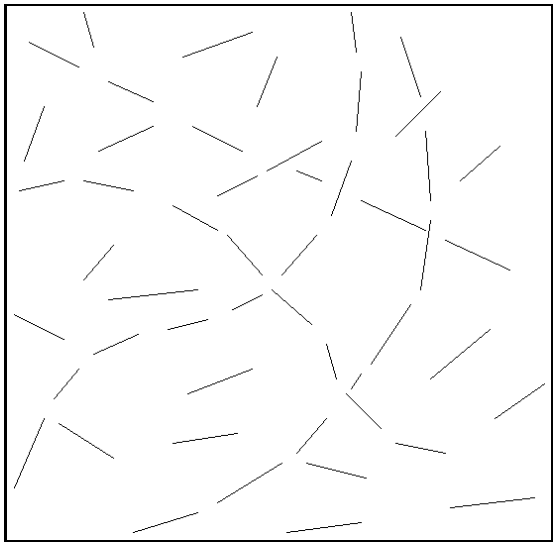

Figure 4.47 - *Chaines de départ*

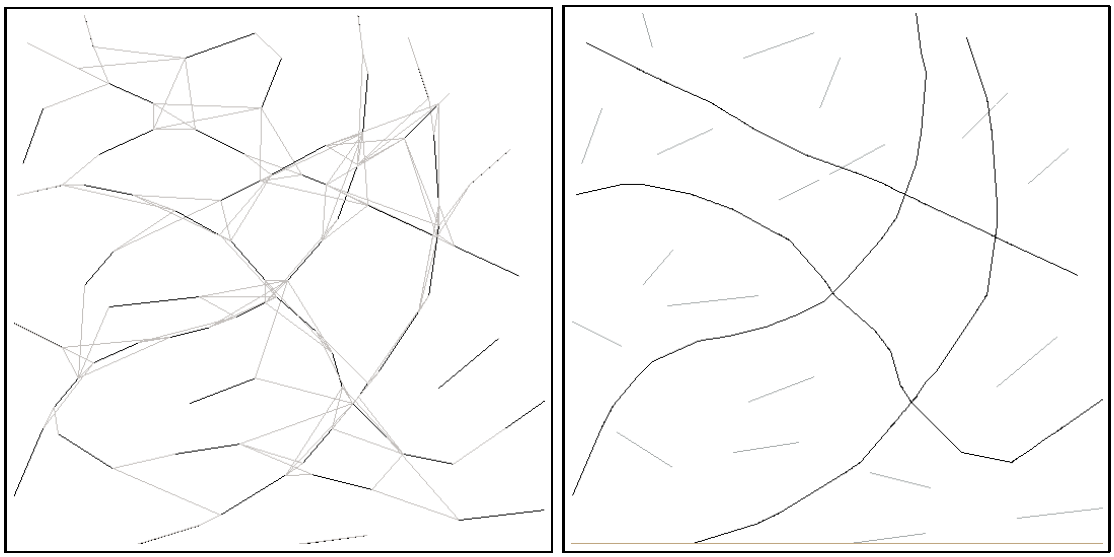

Figure 4.48 - *Graphe de connexions et Sélection automatique des 4 meilleurs groupes*

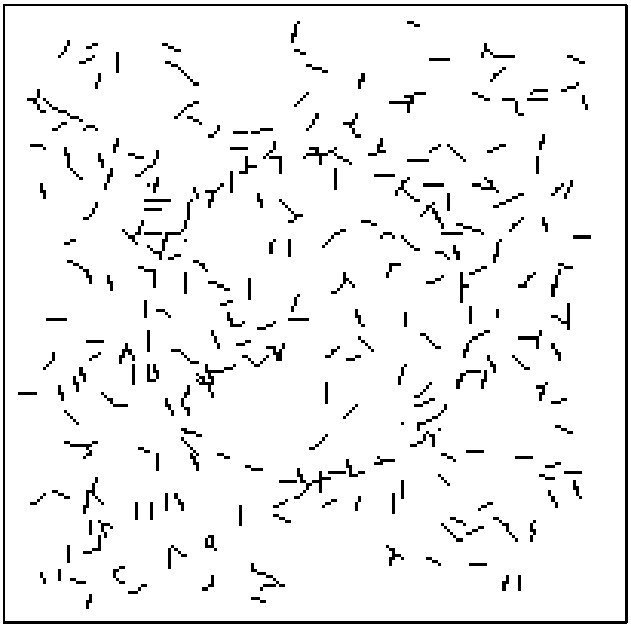

Figure 4.49 - *Cercle avec bruit directionnel - segments orientés aléatoirement*

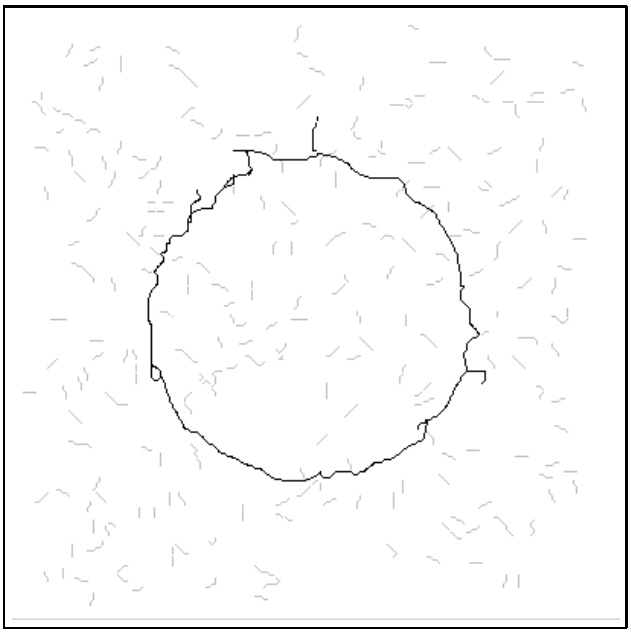

Figure 4.50 - *Superposition de 11 groupements, aprés seuillage selon les trois critères de sélection.*

4.4.6.2 Images réelles

Nous reprenons à la page 157 l'une des scènes de détection de réseaux fins utilisées pour le groupement de pixels. Le groupement de chaînes permet d'obtenir des résultats d'une qualité comparable dans un temps de calcul beaucoup plus court : 0.5 secondes par itération contre 40 secondes à 2 minutes par itération pour le groupement de pixels.

La différence d'aspect entre les courbes, plus lisses pour les groupements de pixels que de chaînes, s'explique par l'influence des éléments de connexion "virtuels". En effet, les voisins immédiats des pixels de contours permettent de lisser le tracé des groupements. Dans le cas des groupements par chaînes, l'extraction de courbes "lisses" est reléguée aux niveaux supérieurs d'organisation afin de permettre une interprétation des groupements selon plusieurs échelles de lissage.

On peut noter dans la figure 4.51 que les branches du croisement situé en bas, à gauche, ont été correctement reconstituées. L'utilisation des chaînes comme primitive de groupement permet de réduire considérablement les effets de "basculement" du suivi d'une structure à l'autre en cas de contours parallèles.

Les scènes des pages suivantes, pages 158 à 268, montrent comment les principes génériques de groupement peuvent être appliqués à des situations aussi différentes qu'une façade de bâtiment ou un objet manufacturé. Ces derniers exemples illustrent l'utilité du groupement perceptuel pour mettre en évidence des structures régulières, en général associées aux contours des objets, tout en écartant les courbes trop sinueuses et qui correspondent plus souvent à du bruit ou bien des surfaces texturées.

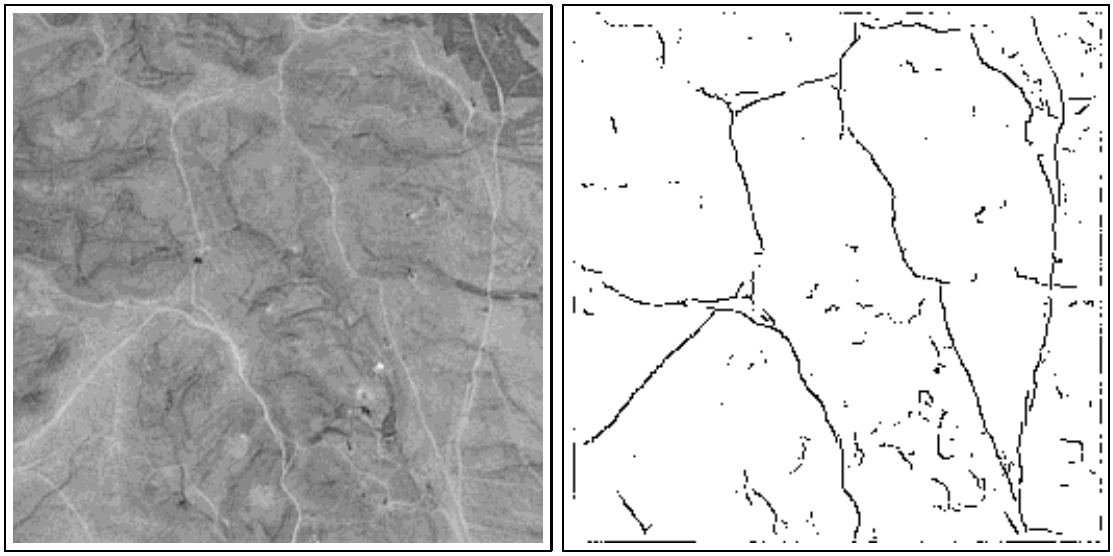

Figure 4.51 - *Image SPOT* 256×256 *pixels - Détection de réseau fin - 344 chaînes*

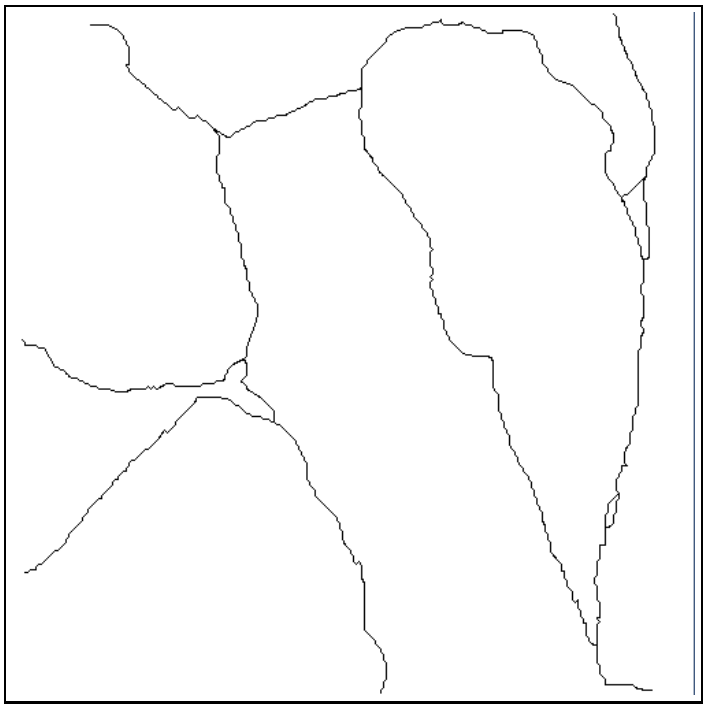

Figure 4.52 - *Détection de routes - Extraction de 9 groupements saillants - 50 itérations (0,3 sec / itération)*

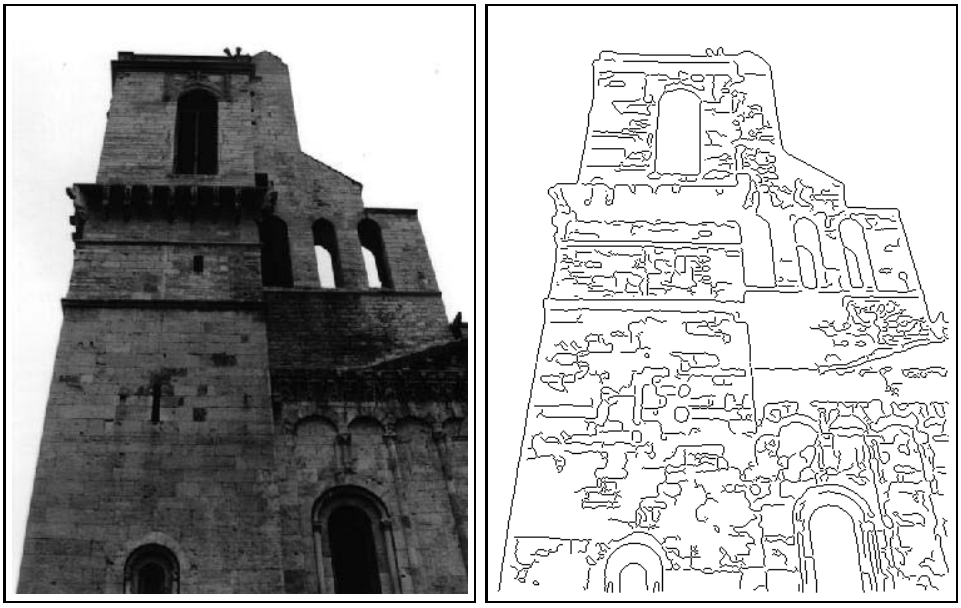

Figure 4.53 - *Cathédrale* 360×460 *pixels - Détection de contours*

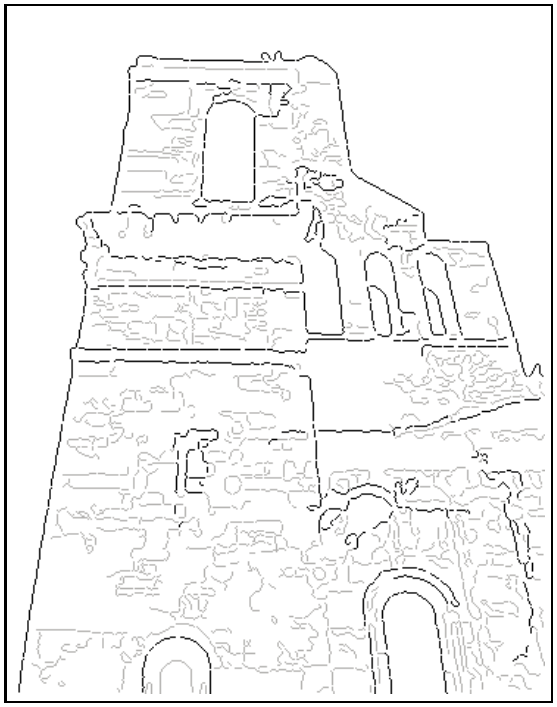

Figure 4.54 - *165 chaînes saillantes sur 2397 chaînes - 50 itérations (0,8 sec / itération)*

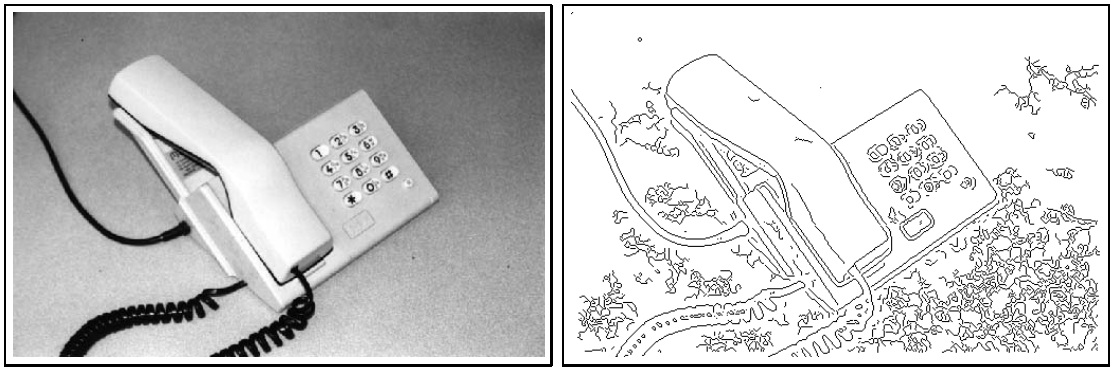

Figure 4.55 - *Téléphone* 500×328 *pixels - Détection de contours*

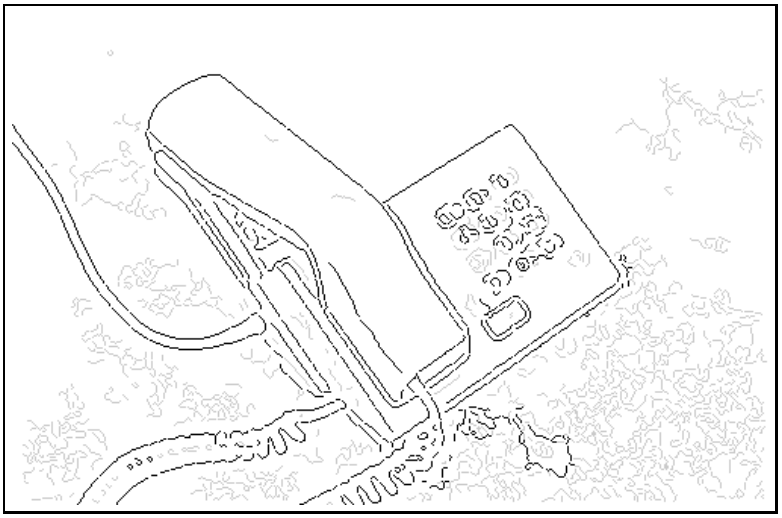

Figure 4.56 - *310 chaînes saillantes sur 2780 chaînes - 50 itérations (0,8 sec / itération)*

4.4.6.3 Conclusion

Nous venons de montrer les avantages qu'offrent les chaînes de contours pour le groupement à partir d'un réseau de saillance. Le nombre de chaînes étant considérablement inférieur au nombre de pixels d'une image, la complexité algorithmique de l'optimisation de la mesure de saillance est réduite de manière significative. Les temps de calculs sont ainsi considérablement réduits.

Les chaînes apportent des informations plus globales sur les contours, ce qui contribue à stabiliser d'autant plus les résultats du groupement. En particulier, le groupement par proximité des pixels d'une même chaîne permet d'éviter les phénomènes de "basculement" entre structures proches, constatés pour le groupement de pixels.

Enfin, le choix d'un voisinage variable permet d'adapter la densité du maillage établi entre les contours, ce qui permet de combler des discontinuités plus importantes que dans le cas de groupement de pixels.

Comme c'était le cas pour l'application au groupement de pixels, les extensions possibles au groupement de chaînes portent sur le type de voisinage et l'extraction finale des meilleures courbes. Dans chacun de ces exemples, la plus grande partie du temps de calcul est consacrée à la construction du voisinage. Bien que sa durée reste inférieure à 10 minutes (pour une image contenant environ chaînes 3000 chaînes), cette étape constitue un frein pour une application éventuelle sur des problèmes concrets.

Malgré ces limitations, le gain de temps obtenu pour le groupement de chaînes rend enfin possible l'utilisation des réseaux de saillance pour extraire rapidement les contours les plus réguliers dans des images de l'ordre de 500×500 pixels. Il en résulte un premier niveau de groupement perceptuel, répondant à des critères génériques de régularité, dont la fonction est d'attirer l'attention sur les contours les plus saillants.

Le produit de ce niveau préliminaire est un ensemble de chaînes saillantes. Ces chaînes sont considérablement moins nombreuses que le nombre d'éléments de contours initialement détectés. Elles représentent des fragments de contours, dont la somme couvre la majeure partie des structures linéaires de l'image. Ce sont les hypothèses à partir desquelles sont extraits les segments, arcs et points d'intérêt du niveau de groupement présenté au chapitre suivant.

Chapitre 5

Eléments de représentation et groupements intermédiaires

Nous venons de présenter une méthode permettant d'extraire les structures linéaires visuellement importantes dans l'image à l'aide d'un premier niveau de groupements élémentaires. Nous montrons dans ce chapitre comment extraire de ces groupements élémentaires, un ensemble d'hypothèses géométriques utiles aux niveaux supérieurs de traitement.

Dans un premier temps, nous présentons les principes de notre méthode de structuration hiérarchique, issus de l'analyse des groupements élémentaires. Cette méthode est ensuite détaillée pour chaque type d'hypothèse géométrique envisagée et illustrée par une application à différentes scènes réelles.

5.1 Structuration hiérarchique

Comme nous l'avons présenté au chapitre 2, l'extraction de structures géométriques élémentaires à partir de contours est réalisée en général soit directement sur l'image de détection de contours, soit après une étape de chaînage et d'approximation polygonale. Les structures recherchées sont des segments, des arcs et les jonctions qui les relient entre elles. Ces dernières peuvent être considérées soit comme des objets à part entière, définis en tant que points d'intérêt, soit comme des relations de connectivité entre segments et arcs.

En éliminant une majeure partie des structures irrégulières, le premier niveau de groupement produit un meilleur ensemble d'hypothèses de départ pour une recherche de caractéristiques géométriques plus poussée. Au contraire des chaînes utilisées dans les approches classiques de structuration de contours, les chaînes groupées sont plus régulières, continues et surtout, moins nombreuses.

La fonction du second niveau d'organisation est, par conséquent, d'extraire les meilleures structures géométriques possibles à partir des groupements issus du niveau précédent. Afin d'obtenir des éléments de représentation aussi stables que pos-

sibles, il est nécessaire d'étudier dans un premier temps la nature des groupements issus du réseau de saillance.

5.1.1 Analyse des groupements élémentaires

L'optimisation globale d'une mesure de saillance locale est intéressante d'un point de vue combinatoire, mais elle reste insuffisante à extraire des structures d'intérêt directement utilisables. Bien que répondant à un certain nombre de critères de régularité, les groupements obtenus n'en restent pas moins des chaînes de pixels difficilement comparables ou manipulables.

Ces chaînes présentent de plus un certain nombre de problèmes liés à leur nature même de groupements :

- *Groupements fragmentés*

  Les conditions d'arrêt variables le long du suivi d'un groupement ne garantissent pas la reconstruction de structures globales complètes. Le résultat obtenu correspond le plus souvent à une superposition de fragments de contours groupés. Cette superposition des groupements permet de couvrir les structures globales mais elle introduit aussi une certaine redondance. En particulier en cas d'occlusions ou d'intersections entre structures d'intérêt, les groupements sélectionnés se superposent et partagent de nombreuses portions de contours.

- *Comportement du suivi autour des jonctions*

  En cas de jonction, le suivi des noeuds du réseau se poursuit en direction de la branche de meilleure qualité. Le sens de parcours d'une structure le long d'un groupement peut être différent selon les valeurs des branches les plus saillantes. L'information globale est, à ce niveau, uniquement fondée sur des critères de régularité. Elle est insuffisante pour faire un choix fiable de la meilleure direction à prendre car elle ne tient pas compte d'informations structurelles que pourrait fournir, par exemple, un étiquetage des éléments de contours. Cette situation est résumée par la figure 4.11, page 117.

Ces problèmes rendent les groupements individuels instables et par conséquent, impropres à une utilisation directe par des tâches visuelles de plus haut niveau. Leur somme permet pourtant de couvrir l'ensemble des structures d'intérêt. Ces constatations, familières aux problèmes d'organisation perceptuelle, conduisent à l'idée de grouper selon des règles plus globales des hypothèses extraites à partir de chaque courbe saillante.

5.1.2 Principes de groupement intermédiaire

Ce niveau de groupement intermédiaire est donc composé de deux étapes. Dans un premier temps, des hypothèses sont extraites à partir de chaque groupement.

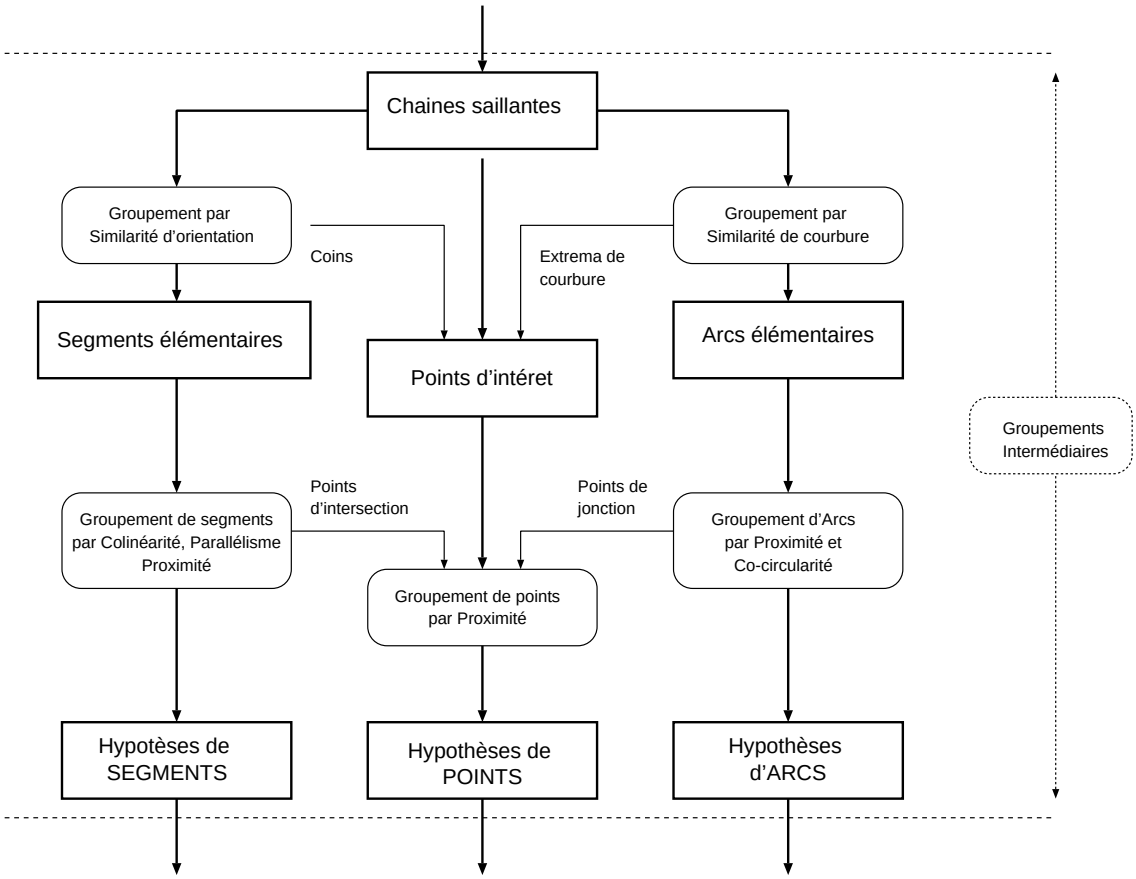

Figure 5.1 - *Principes du niveau intermédiaire de groupements. Le but est d'analyser les chaînes saillantes afin d'en extraire des hypothèses élémentaires de segments, d'arcs et de points d'intérêt. Ces hypothèses sont ensuite simplifiées par groupement.*

Ces hypothèses sont ensuite groupées entre elles afin d'obtenir un nouvel ensemble d'hypothèses simplifiées. Cette méthode, présentée d'abord de manière générale, est ensuite adaptée à chaque type d'hypothèse géométrique à l'aide d'algorithmes spécifiques.

Ce principe rejoint l'idée, suggérée entre autres par [Zucker *et al.*, 1989] , de séparer en deux parties distinctes la détection de primitives et leur groupement en structures de plus haut niveau. La principale différence de notre méthode est d'extraire ces primitives à partir de groupements préliminaires et de les grouper ensuite globalement. Par comparaison, la méthode proposée par Zucker consiste à grouper à l'aide de contours actifs, des segments élémentaires définis directement à partir de l'image de contours.

– *Segmentation des groupements élémentaires*

Cette première phase consiste à détecter, sur chaque groupement, les primitives qui le composent à l'aide de techniques classiques de structuration de contours telles que celles décrites au chapitre 2, section 2.4. Sans qu'elles soient pour autant exprimées en termes Gestaltistes, la plupart de ces méthode de segmentation répondent aux mêmes règles de groupements perceptuels lorsqu'elles sont appliquées à une chaîne de pixels. En effet, des techniques d'extraction de segments de droites par approximation polygonale [Gupta *et al.*, 1993] ou de détection de courbes à différentes échelles [Saund, 1991] impliquent nécessairement des groupements par proximité, continuité, ou similarité d'orientation.

Nous privilégions ici une approche modulaire, chaque groupement étant décomposé en segments rectilignes, en arcs et points d'intérêt indépendamment de la phase d'organisation. Ce type d'approche permet d'adapter sans difficulté la méthode à un plus large ensemble de situations en choisissant éventuellement la technique de structuration la plus adaptée au type de scène observée. En effet, la stratégie de groupement n'est pas remise en cause par un changement de la méthode de détection.

L'étude des différentes techniques de structuration de contours a mis en valeur l'importance de la notion d'échelle lorsqu'il s'agit de décomposer une chaîne de pixels en segments ou en arcs. Chaque hypothèse doit donc être associée à l'échelle à laquelle elle a été détectée.

Cette détection doit en outre permettre une certaine part de redondance afin de rendre possible plusieurs interprétations d'un même fragment de contour et de produire un ensemble d'hypothèses utiles aux niveaux supérieurs de représentation. Séparer ainsi l'extraction des hypothèses à partir des courbes saillantes et leur groupement permet enfin de garder cette approche ouverte à des coopérations éventuelles avec d'autres types de détections plus spécialisés. Par exemple, la détection de points d'intérêt peut être enrichie par un détecteur de coins spécialisé.

– *Fusion des hypothèses*

Parmi les différentes techniques de groupement présentées au chapitre 3, les plus adaptées à cette seconde étape intermédiaire sont les techniques algorithmiques de groupement ainsi que les approches fondées sur la théorie des graphes. Ces méthodes permettent une approche hiérarchique particulièrement utile pour représenter un ensemble d'hypothèses à différentes échelles et les relations qui les lient.

Le rôle de ce niveau de groupement est double. Il doit permettre d'une part, de réduire les redondances au sein d'un même jeu d'hypothèses pour une échelle donnée, et d'autre part, de définir un graphe de relations entre primitives.

5.2 Hypothèses "segments"

L'extraction de segments de droites à partir des chaînes saillantes correspond au problème classique de l'approximation polygonale d'une chaîne de pixels. Cette segmentation doit rester stable par rapport à des critères objectifs, liés à la discrétisation des chaînes, mais aussi critères subjectifs, relatifs à l'application.

Parmi les critères objectifs, on peut citer la résistance aux perturbations, l'approximation devant rester stable sinon invariante, en cas de modifications locales de la chaîne. L'invariance pour des transformations usuelles est tout aussi importante, en particulier face aux nombreux "escaliers" introduits le long d'une chaîne après une rotation ou une homothétie. L'approximation doit rester enfin invariante au fenêtrage, c'est à dire, l'approximation d'une partie de chaîne doit donner le même résultat si cette chaîne est prise individuellement. En conséquence de ce dernier critère, les extrémités d'une chaîne doivent être sans influence sur l'approximation.

Les critères subjectifs se retrouvent souvent sous la forme de seuils d'erreurs ou de longueurs, adaptés à l'application recherchée. Parmi les critères les plus généraux, on peut citer la conservation des mesures d'angles ainsi que le respect des droites en dépit de la discrétisation.

5.2.1 Détection des segments

La littérature concernant l'approximation polygonale de chaînes est abondante [Sklansky et Gonzalez, 1980] [Aoyama et Kawagoe, 1991] [Ray et Ray, 1992] . Elle reflète les deux approches citées précédemment pour la segmentation de chaînes, par fusion de points similaires ou par la détection de points de coupure.

On pourra se référer à [Garnesson et Giraudon, 1991] et [Rosin, 1997] pour des évaluations comparées des principales méthodes d'approximation polygonale.

Ces techniques donnent de bons résultats pour des chaînes constituées uniquement de segments. Des problèmes d'échelle et de localisation de points de coupure apparaissent lorsque les chaînes présentent des parties courbes. En particulier, la localisation de points particuliers, comme la séparation entre un segment tangent à une courbe doit faire l'objet de rectifications après détection.

Parmi les deux types d'approches décrites au chapitre 2, nous privilégions les approches par division afin d'obtenir à la fois un ensemble de segments et de points d'intérêts.

Algorithme de détection de segments

La technique de segmentation que nous retenons en particulier est l'application d'une méthode rapide et stable de division récursive. Cette méthode a été choisie pour sa simplicité et parce-qu'elle permet un contrôle direct de l'écart entre les segments et la chaîne, qui joue ici le rôle d'un facteur d'échelle.

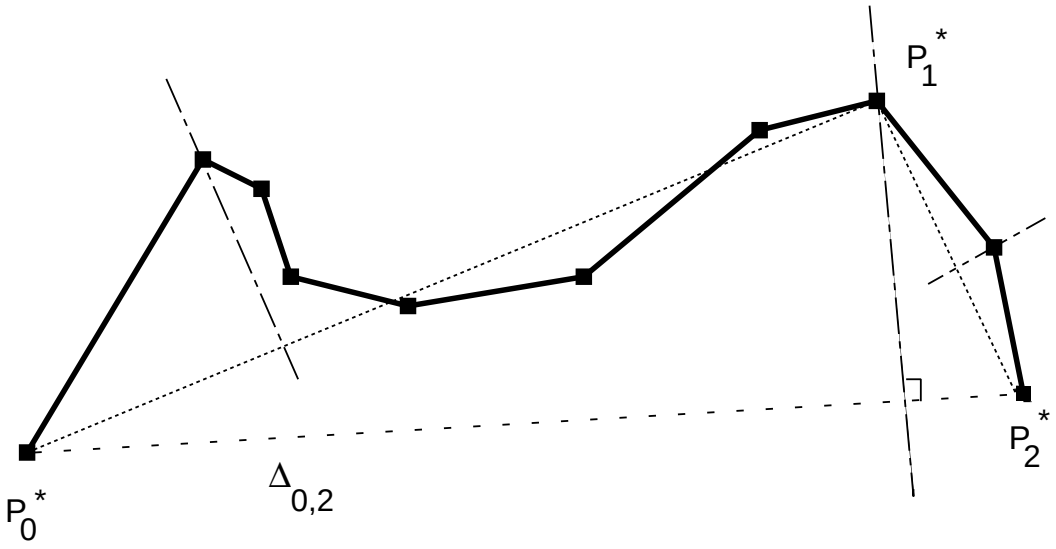

Figure 5.2 - *Initialisation du découpage récursif d'une chaîne de pixels. Après détection du point le plus éloigné de la droite Δ, l'opération est répétée à gauche et à droite du point de coupure jusqu'à ce l'écart maximal entre la chaîne et chaque segment soit inférieur à un seuil donné..*

Elle consiste simplement à rechercher le point de la courbe le plus éloigné de la droite formée par les extrémités et de reproduire récursivement cette opération à gauche et à droite de ce point de coupure.

Soit γ une chaîne constituée des points $\{P_0, \cdots, P_{n-1}\}$, on recherche le point P_i tel que :

$$dist(P_i, \Delta_0^{n-1}) \; = \; \mathbf{Max}_{k \in [1,n-2]} \; dist(P_k, \Delta_0^{n-1})$$

où : $\Delta_0^{n-1} \; = \; (P_0, \overrightarrow{P_0 P_{n-1}})$ est la droite reliant les deux extrémités de la chaîne.

Le pixel P_i est choisi comme point de coupure de la chaîne. Cette opération est répétée récursivement sur les portions de la chaîne composées des points $\{P_0, \cdots, P_i\}$ et $\{P_i, \cdots, P_{n-1}\}$.

Le processus s'arrête lorsque l'écart maximal est inférieur à un seuil $\epsilon^{\surd}$ représentant l'échelle d'approximation de la courbe. Les segments correspondent alors à la succession des points de coupure définis le long de la chaîne. Plus $\epsilon^{\surd}$ est petit, plus l'approximation est proche de la chaîne. Ce seuil varie en pratique entre $\frac{1}{2}$ et 10 pixels d'écart selon les échelles observées. Une autre condition d'arrêt concerne la longueur des segments, arbitrairement limitée à un minimum de 5 pixels.

En supposant que la chaîne soit découpée en k points de coupure P_i^*, ceux-ci

vérifient donc les propriétés suivantes :

$$\forall i \in [1,k] \quad \begin{cases} dist(P_i^*, \Delta_{i-1}^{i+1}) & = & \underset{P_j \in [P_{i-1}^*, P_{i+1}^*]}{\mathbf{Max}} dist(P_j, \Delta_{i-1}^{i+1}) \\ dist(P_i^*, \Delta_{i-1}^{i+1}) & > & \epsilon^{\surd} \\ \|\overrightarrow{P_i^* P_{i+1}^*}\| & > & 5 \end{cases} \tag{5.1}$$

On note P_0 et P_{k+1} les extrémités de la chaîne γ.

Cas particulier des boucles

Cette méthode suppose que la droite passant par les points P_0 et P_{n-1} existe. Dans le cas d'une boucle, les deux points sont confondus et cette droite ne peut pas être définie. Lorsque cette situation se produit, il est nécessaire de choisir un point de la boucle et de découper récursivement chacune des deux parties.

Le point de coupure recherché se doit de correspondre à un sommet de la chaîne et non à une perturbation accidentelle due à la discrétisation. Appliquer temporairement une autre approximation polygonale permet de sélectionner le meilleur sommet, en l'occurence, le sommet le plus éloigné du point de départ. Nous utilisons à cette fin la méthode de [Wall et Danielson, 1984] . Cette méthode rapide détecte les points de coupure le long de la chaîne en estimant la surface définie par le segment courant et la portion de contours que ce segment délimite.

Afin d'obtenir une approximation polygonale de la boucle présentant les mêmes caractéristiques que pour les chaînes ouvertes, l'approximation récursive est appliquée aux portions de la boucle définies par les points $\{P_0, \cdots, P^*\}$ et $\{P^*, \cdots, P_{n-1}\}$. Le point P^* est le point de coupure le plus éloigné de P_0.

Analyse des résultats

La complexité algorithmique de l'approximation récursive est comprise entre $\mathcal{O}(n^2)$ dans le cas le plus défavorable la chaîne est découpée en $(n-1)$ segments, et $\mathcal{O}(n \cdot \log m)$ dans le cas le plus favorable où sont créés m segments de longueur égale.

Cette méthode permet une meilleure détection des points de coupure que les méthodes de fusion. Les points de coupures sont autant de sommets probables et ajoutés en tant que tels à une liste de points d'intérêt. Cette liste est utilisée ensuite pour la détection des jonctions.

Utilisée seule, cette méthode est cependant sensible au bruit. Elle présente une certaine tendance à trop découper et, surtout, dépend du choix des extrémités de départ.

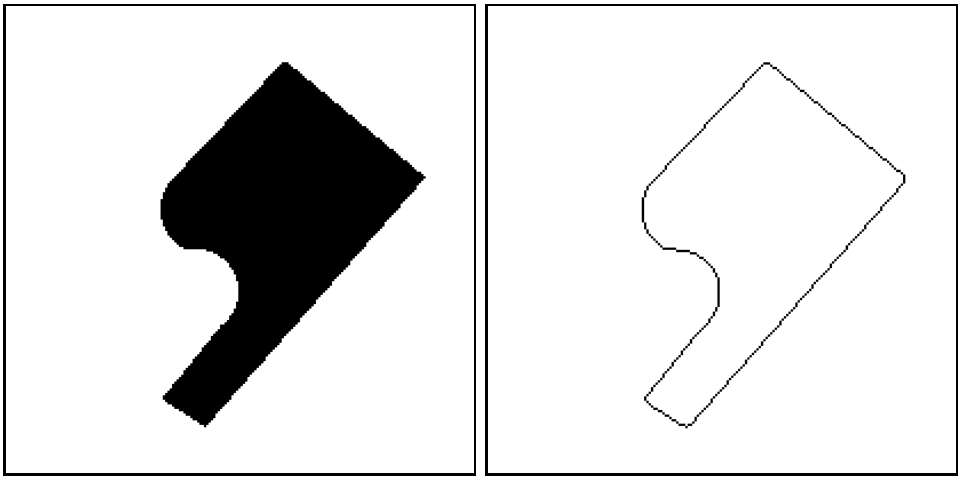

Figure 5.3 - *Scène de test et détection de contours.*

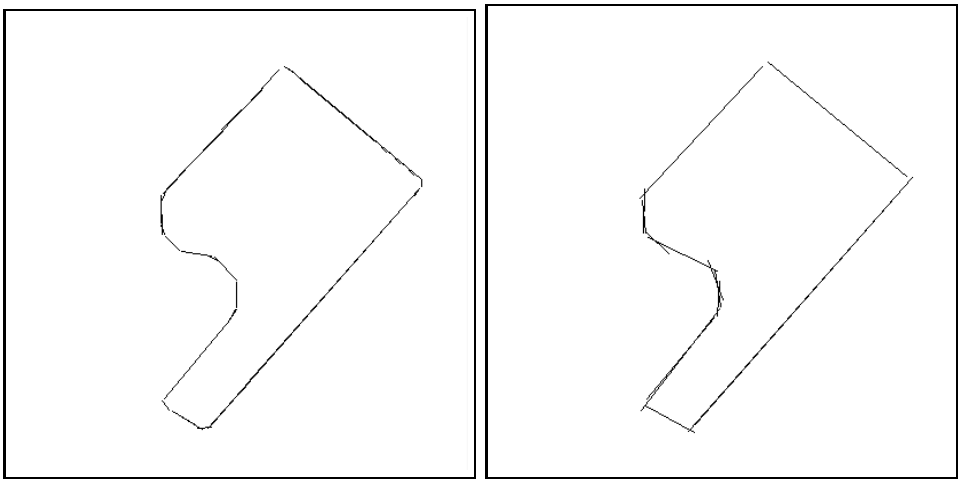

Figure 5.4 - *Détection de segments pour les écarts* $\epsilon^{\surd} = 1$ *et* $\epsilon^{\surd} = 11$. *Les segments ont été extraits de la seule chaîne saillante issue du groupement élémentaire.*

C'est pourquoi elle est souvent suivie d'une étape de fusion de segments et d'ajustement de points de coupure au sens des moindres carrés [Pavlidis, 1981] . Etant donné le grand nombre de groupements élémentaires à traiter, l'ajustement des points de coupure représente un coût trop important en temps de calcul. De plus, le groupement des segments similaires ferait perdre une partie de l'apport d'un tel ajustement. Dans le cadre de notre application, c'est précisément l'organisation perceptuelle des segments décrite dans le paragraphe suivant qui sert d'ajustement des hypothèses de segments. L'hypothèse d'insensibilité au fenêtrage décrite précédemment doit donc être envisagée pour les segments aprés groupement au lieu des hypothèses de segments extraites à

partir des groupements élémentaires.

Pour plus de stabilité, les segments $[P_k^* P_{k+1}^*]$ sont replacés par une approximation au sens des moindres carrés de la partie de la chaîne délimitée par P_k^* et P_{k+1}^*. Cette approximation produit en effet des segments moins sensibles à la localisation des extrémités, et par conséquent, plus stables pour le groupement de l'étape suivante. Les discontinuités entre segments que l'on peut observer dans la figure 5.3 (pour un écart de $\epsilon^{\surd} = 11$ pixels) sont dus à cette approximation supplémentaire. L'un des objectifs de la seconde étape est précisément de réduire ces discontinuités (cf. figure 5.7).

Le bruit local le long de la chaîne a été en outre considérablement réduit par le réseau de saillance. Enfin, la superposition des segments issus des différents groupements permet, lors de l'organisation perceptuelle des segments, de rectifier dans une certaine mesure les erreurs de détection des points de coupure.

La figure 5.3 donne un exemple de segments détectés à partir des groupements selon deux écarts $\epsilon^{\surd} = 1$ pixel et $\epsilon^{\surd} = 11$ pixels. Malgré les superpositions de segments le long des parties rectilignes, les orientations des parties droites sont bien conservées. Cet exemple illustre les problèmes rencontrés le long des parties courbes, avec en particulier l'accumulation de segments issus de groupements différents. Les nombreuses intersections entre segments sont dues à la mauvaise localisation des points de coupure lorsque la courbure des contours devient ambiguë.

5.2.2 Organisation perceptuelle des segments

Le groupement des segments doit remplir deux fonctions. D'une part, cette étape doit éliminer les segments redondants pour une échelle donnée en fusionnant les segments superposés ou juxtaposés. D'autre part, elle doit ajuster les extrémités des segments autour des intersections afin de préparer la détection des jonctions. Le but est d'obtenir au final des segments les plus longs possibles, avec le moins d'intersections possibles, tout en préservant la structure d'ensemble des contours.

D'un point de vue perceptuel, le groupement d'éléments de contours selon des segments de droites fait appel aux principes de proximité et de continuité. L'approche par groupements hiérarchiques proposée par [Lowe, 1985] ou encore, par [Dickson, 1991] est représentative d'un groupement perceptuel progressif, à différentes échelles de structures. L'organisation est réalisée d'abord à un niveau local, par similarité, continuité et proximité. Chaque groupe est progressivement remplacé par un nouvel élément, ce qui conduit à l'émergence de propriétés globales. Un graphe de relations entre éléments groupés permet une recherche éventuelle de structures à différentes échelles.

Principes du groupement de segments

La première partie du groupement compare les segments deux à deux. Afin d'agglomérer plus efficacement les segments entre eux, chaque segment est comparé de préférence avec les segments de longueur inférieure. De cette manière, les segments les plus petits sont éliminés au profit de segments plus grands, moins susceptibles d'être bruités.

Soient deux segments S_1 et S_2 tels que leurs longueurs respectives vérifient : $L_{S_1} > L_{S_2}$. Ces segments sont éventuellement fusionnés après évaluation de leurs positions, distances et orientations relatives par rapport aux critères suivants.

– *Colinéarité et alignement*

Les segments sont à la fois colinéaires et alignés s'ils partagent la même droite porteuse. Ce critère est rempli lorsque les distances entre chaque extrémité de S_2 et la droite porteuse de S_1 sont inférieures à un seuil $\epsilon^{\parallel}$ (fixé en pratique à 3 pixels).

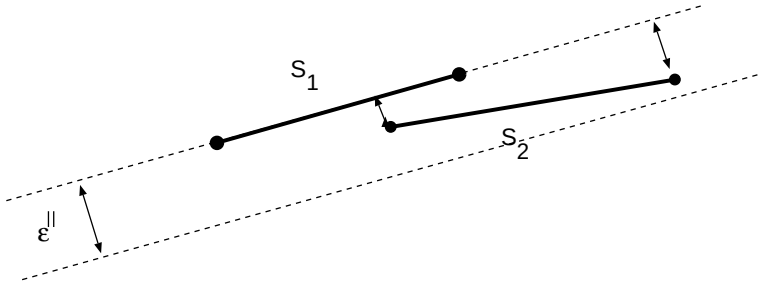

Figure 5.5 - *Critère de colinéarité et d'alignement entre deux segments.*

Soient $d^+_{S_1}(S_2)$ et $d^-_{S_1}(S_2)$ ces deux distances. Le critère d'alignement est alors :

$$\left.\begin{array}{rcl} d^+_{S_1}(S_2) & < & \epsilon^{\parallel} \\ et & & \\ d^-_{S_1}(S_2) & < & \epsilon^{\parallel} \end{array}\right\} \quad \Longrightarrow \quad S_1 \equiv S_2 \qquad (5.2)$$

Comme les autres distances de tolérance définies pour les groupements de primitives dans ce chapitre, $\epsilon^{\parallel}$ est paramétrable en fonction de la précision de détection des contours et des hypothèses élémentaires.

– *Proximité*

En supposant que les deux vecteurs soient colinéaires et alignés, le critère de proximité est fixé à l'aide des vecteurs de recouvrement entre segments.

On définit un vecteur de recouvrement entre deux segments S_1 et S_2 par la distance entre le milieu de S_1 et la projection de chaque extrémité de S_2 sur la droite porteuse de S_1.

Soient $\overrightarrow{v^+_{1,2}}$ et $\overrightarrow{v^-_{1,2}}$ les deux vecteurs de recouvrement associés à S_1 et S_2. Si l'un des deux vecteurs est de longueur proche de $\frac{\|S_1\|}{2}$, alors les segments partagent une extrémité. Pour tenir compte des erreurs de localisation des extrémités des segments, une zone de tolérance est admise autour des extrémités de chaque segment pour un groupement. Cette distance ϵ_d est fixée en pratique à 5 pixels.

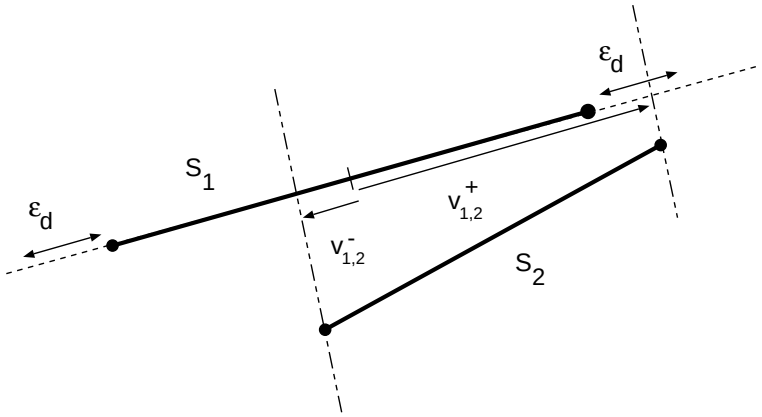

Figure 5.6 - *Critère de proximité entre deux segments.*

Le critère de proximité est donc :

$$\left.\begin{array}{rrcl} & \|\overrightarrow{v^+_{1,2}}\| & < & (\epsilon_d + \frac{\|S_1\|}{2}) \\ ou & & & \\ & \|\overrightarrow{v^-_{1,2}}\| & < & (\epsilon_d + \frac{\|S_1\|}{2}) \end{array}\right\} \quad \Longrightarrow \quad S_1 \equiv S_2 \qquad (5.3)$$

Les segments qui remplissent ces deux critères sont fusionnés. Cette opération est répétée jusqu'à ce qu'il n'y ait plus de groupements possibles entre segments superposés. De manière générale, et pour le reste de ce chapitre, on note le groupement de deux hypothèses par la relation $\equiv$. Ainsi, $[S_1 \equiv S_2]$ signifie que le segment S_2 est fusionné avec segment S_1.

Simplification des intersections

Comme le montre la figure 5.7, une rectification des extrémités des segments est nécessaire afin d'obtenir un ensemble d'hypothèses plus stables. Une certaine tolérance est admise sur la localisation des extrémités tant que les intersections ne forment que des jonctions en L ou en T. Le but de cette étape est de réduire au maximum le nombre d'intersections en X.

La méthode que nous proposons pour rectifier les extrémités consiste à réduire progressivement la longueur des segments pour lesquels il existe une intersection proche des extrémités. Les intersections sont évaluées par ordre

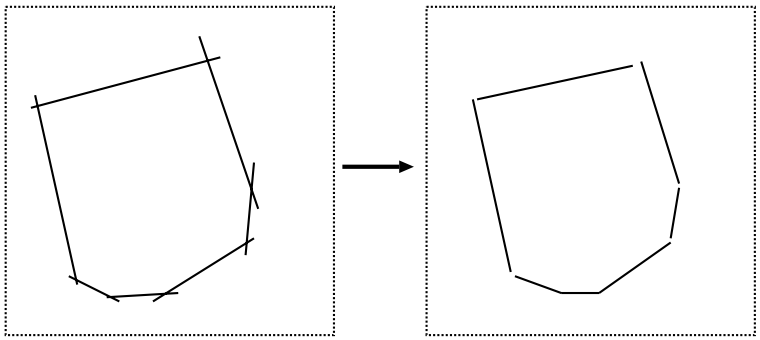

Figure 5.7 - *L'approximation des segments au sens des moindres carrés, et le groupement de segments par proximité introduisent des erreurs de localisation pour les extrémités des segments.*

de distance croissante avec les extrémités des segments. Pour des raisons d'efficacité algorithmique, la distance entre le point d'intersection et les extrémités d'un segment est exprimée par rapport au milieu de celui-ci. Ainsi, il n'est pas nécessaire de tester les deux extrémités à chaque fois.

Pour un segment S donné, de longueur L, on recherche le segment S_1 pour lequel l'intersection avec S est la plus proche de l'une des extrémités de S. On note I_{S,S_1} l'intersection entre ces deux segments. Le segment S_1, s'il existe, remplit les conditions suivantes :

$$\begin{cases} d_m & < & \frac{L}{2} \\ d_m & = & \underset{S_i \neq S \ , \ S \cap S_i \neq \emptyset}{\mathbf{Max}} \{ \| \overrightarrow{M_S, I_{S,S_i}} \| \} \end{cases} \tag{5.4}$$

avec : $d_m = \| \overrightarrow{M_S, I_{S,S_1}} \|$, distance entre le point d'intersection et le milieu du segment.

Une fonction d'énergie est définie à partir des intersections entre segments. Les contributions de chaque segment pour cette fonction d'énergie sont inversement liées à la distance entre l'intersection et les extrémités des segments. Elles permettent de définir dans quelle proportion rectifier les extrémités des segments.

Soit t la distance entre un point de S et son milieu M_S. La contribution de S à l'énergie d'une intersection est définie par :

$$E_S(t) = \exp\left(-\frac{t^2}{\sigma_L^2}\right)$$

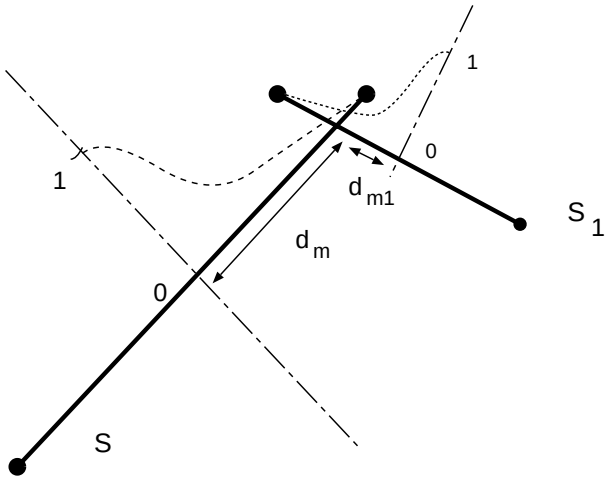

Figure 5.8 - *Energies d'intersection entre deux segments. Elles déterminent si la rectification de l'un ou l'autre segment doit avoir lieu, et si oui, dans quelle proportion.*

La constante σ_L est fixée de manière à ce que l'énergie soit presque nulle aux extrémités du segment, soit :

$$\sigma_L \;=\; \frac{1}{2} \cdot \frac{L}{2}$$

L'énergie associée à l'intersection I_{S,S_1} correspond à la somme des contributions des segments S et S_1, soit :

$$E_{S,S_1} \;=\; E_S(d_m) \;+\; E_{S_1}(d_{m_1})$$

En cas d'intersections proches des extrémités, l'énergie d'intersection est inférieure à un certain seuil. Dans ce cas, l'extrémité de S proche de l'intersection doit être rectifiée afin de réduire la distance avec le point d'intersection et ainsi, réduire la nouvelle énergie. La rectification est réalisée en réduisant la longueur du segment S à partir de l'extrémité proche de l'intersection.

$$0 \;<\; E_S(t) \;<\; E_{sup} \quad \Longrightarrow \quad \{L \;\leftarrow\; L \cdot \rho \;\;,\quad \rho \in [0,1]\}$$

où E_{sup} est un seuil arbitrairement fixé à $0,4$. Ce seuil définit la distance t à partir de laquelle une intersection doit être éventuellement réduite. Il permet ainsi d'interdire la réduction d'intersections trop proches du centre du segment.

Afin d'ajuster les segments en fonction de la position des intersections, le coefficient de réduction ρ est lié à l'énergie $E_S(t)$. Sa valeur est par ailleurs

bornée pour éviter des réductions trop importantes.

$$\rho = \begin{cases} (1 - E_S(t)) & si \quad E_S(t) < 0,2 \\ \\ 0.8 & sinon \end{cases} \tag{5.5}$$

L'algorithme 5.1 résume la procédure complète de réduction des extrémités.

Algorithme 5.1 : Rectification des intersections

début
- **répéter**
  - EnergieTotale $\leftarrow 0$
  - **pour** *Chaque segment* S **faire**
    - **pour** *Chaque extrémité de* S **faire**
      - Rechercher le segment S_1 dont l'intersection I est la plus proche de l'extrémité
      - $E_S \leftarrow$ Energie d'intersection entre S_1 et S
      - $E_{S_1} \leftarrow$ Energie d'intersection entre S et S_1
      - %
      - **si** *(*$E_S < E_{sup}$*)* **alors**
        - Rectifier l'extrémité de S proche de I
      - **si** *(*$E_{S_1} < E_{sup}$*)* **alors**
        - Rectifier l'extrémité de S_1 proche de I
      - %
      - EnergieTotale $\leftarrow$ EnergieTotale $+(E_S + E_{S_1})$
    - %
- **jusqu'à** *EnergieTotale suffisamment faible*

fin

Les intersections sont évaluées pour chaque segment jusqu'à ce que l'énergie globale des intersections atteigne un niveau stable. Cette énergie globale est la somme des énergies des intersections répondant aux critères de distances définis ci-dessus.

$$\mathcal{E} = \sum_j E_{S,S_j}$$

Le processus est interrompu lorsque les segments ont atteint un état stable, correspondant à une énergie suffisamment faible pour chaque intersection. Au

fur et à mesure des itérations, les segments dont l'intersection correspond au critère sont rectifiés et leur nombre décroît, ce qui assure la convergence de l'énergie globale $\mathcal{E}$ vers un état stationnaire.

Résultats et perspectives

Le résultat est un ensemble de segments de droites n'autorisant des intersections complètes qu'en cas de portions courbes. Les autres intersections sont restreintes aux extrémités. Comme le montrent les résultats des pages suivantes, les segments ne sont pas nécessairement jointifs.

La complexité de cet algorithme est, dans le pire des cas, de l'ordre de $O((k \cdot N)^2)$ si N est le nombre de groupes issus du réseau de saillance et k le nombre moyen de segments trouvés pour chaque groupe. La complexité réelle est inférieure à cette limite du fait de la réduction du nombre de segments au fur et à mesure des groupements et à cause de l'ordre imposé pour la comparaison des segments.

Les résultats suivants ont été obtenus dans les mêmes conditions de paramètres, pour différentes valeurs de l'écart d'approximation $\epsilon^{\surd}$. Les figures 5.9 et 5.10 donnent des exemples de simplification de segments sur une scène synthétique simple. Malgré la superposition initiale des segments, on peut noter que la localisation des sommets de la partie polygonale est fidèlement respectée. Il en est de même pour la symétrie des points dominants de la partie courbe.

La scène suivante, figures 5.11 à 5.15, illustre l'intérêt du groupement de segments. A partir d'une image de contours relativement bruitée (452 chaînes de contours en gris), l'extraction de 14 groupes dominants permet de représenter l'essentiel de la structure polygonale de la scène en une trentaine de segments seulement. Rappelons que le but recherché n'est pas une représentation de la scène parfaite et sans ambiguïté, mais bien de recouvrer la majeure partie des structures rectilignes présentes dans l'image.

Les exemples des pages 180 à 182 reprennent les scènes utilisées pour le groupement par réseau de saillance. Chaque situation donne deux exemples de détection de segments suivie des résultats du groupement. Ces résultats soulèvent essentiellement deux remarques.

La première concerne le comportement du groupement sur les parties "courbes" de la scène. Selon l'écart d'approximation, ce sont les zones qui présentent le plus d'intersections entre segments non réduites en fin de groupement. A l'inverse, les parties rectilignes restent relativement stables d'une échelle à l'autre.

La seconde remarque concerne la disparition de certains segments. Ces "disparitions" sont dues principalement au critère de proximité entre segments, qui conduit à grouper des segments parallèles trop proches, et à la réduction des

extrémités, qui conduit à tronquer certains segments de manière importante lorsque les intersections sont relativement éloignée des extrémités.

Le dernier exemple, figures 5.22 à 5.25, permet de mieux apprécier l'intérêt du groupement de segments par rapport aux méthodes classiques. La plupart des méthodes de structuration d'image à partir de contours s'appliquent directement sur une approximation polygonale des contours détectés. Le groupement de segments à partir d'un réseau de saillance permet de restituer les principales structures rectilignes, en particulier lorsque la scène est fortement perturbée.

Comme le montrent les exemples précédents, les segments incorrects sont tolérés le long des portions courbes afin d'être comparés aux arcs détectés par ailleurs, et éventuellement éliminés. Le prolongement naturel de ces résultats devrait être une étude de la stabilité des segments détectés pour différents écarts d'approximation afin d'éliminer les segments les plus instables, correspondant à des parties courbes.

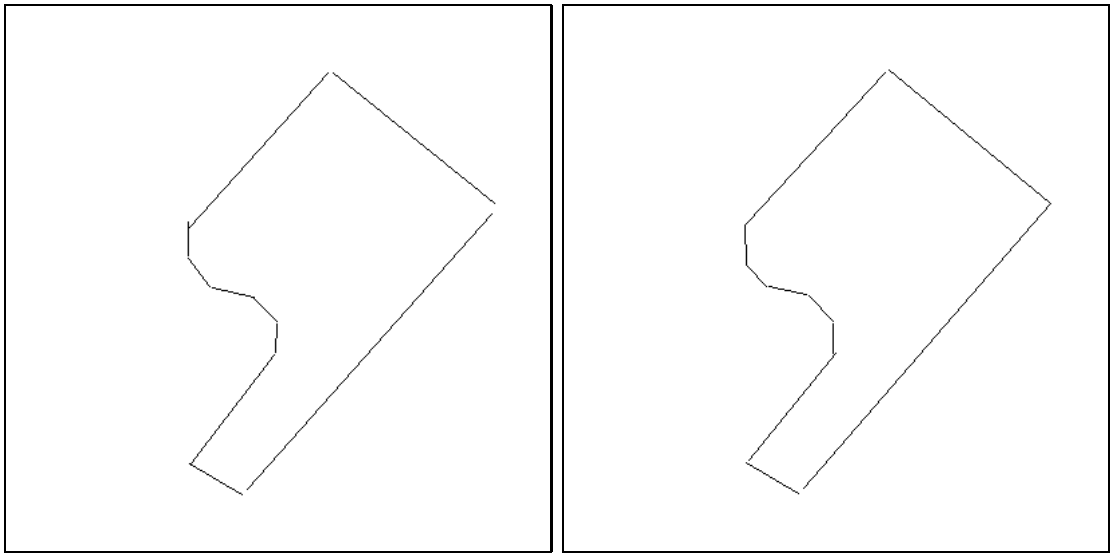

Figure 5.9 - *Groupement de segments pour les écarts* $\epsilon^{\surd} = 1$ *et* $\epsilon^{\surd} = 4$

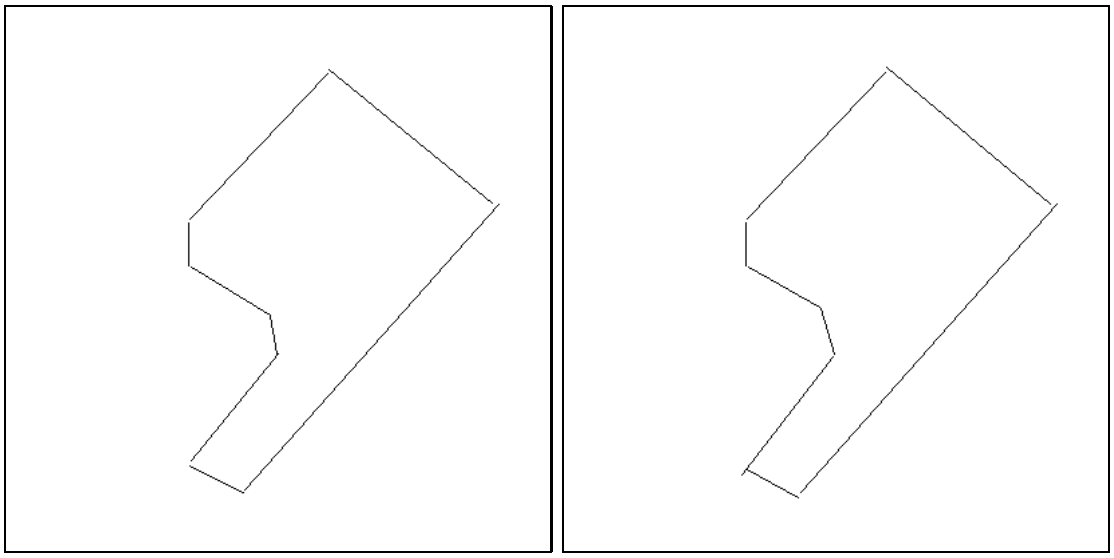

Figure 5.10 - *Groupement de segments pour les écarts* $\epsilon^{\surd} = 8$ *et* $\epsilon^{\surd} = 11$

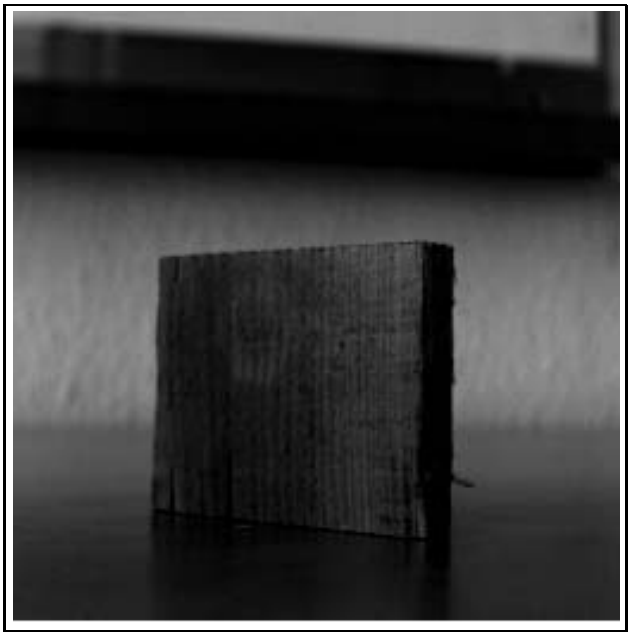

Figure 5.11 - *Pièce en bois*

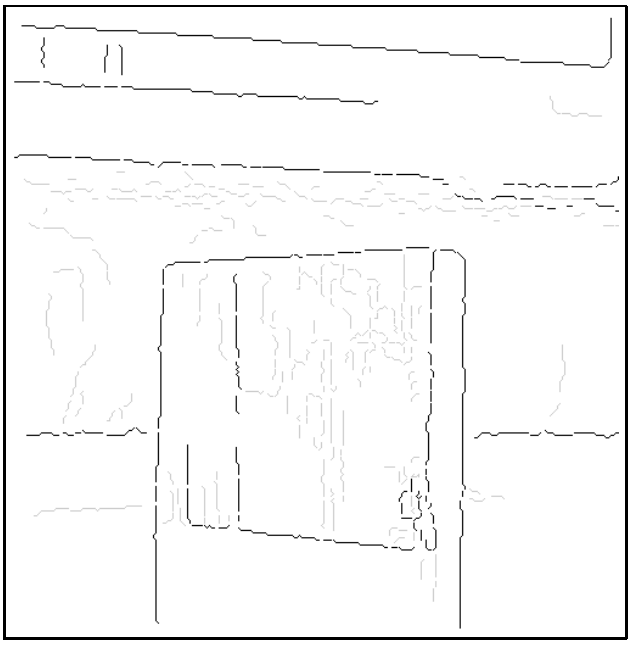

Figure 5.12 - *Détection de contours et sélection des meilleurs groupes - 452 chaînes - 14 groupes*

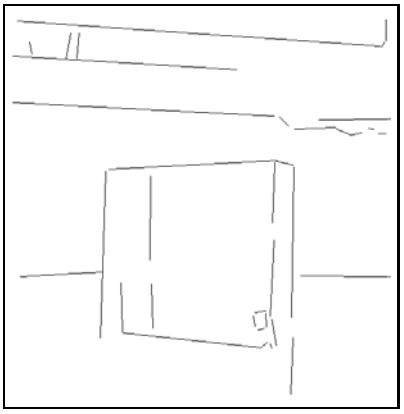

Figure 5.13 - *Groupement - 36 segments* - $\epsilon^{\surd} = 1$

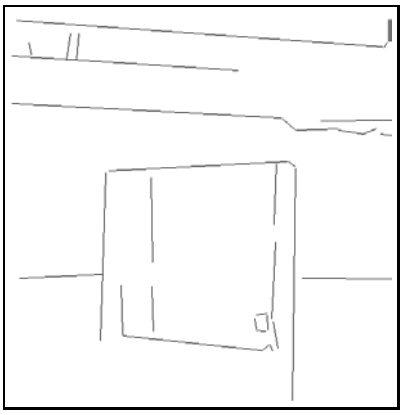

Figure 5.14 - *Groupement - 34 segments* - $\epsilon^{\surd} = 3$

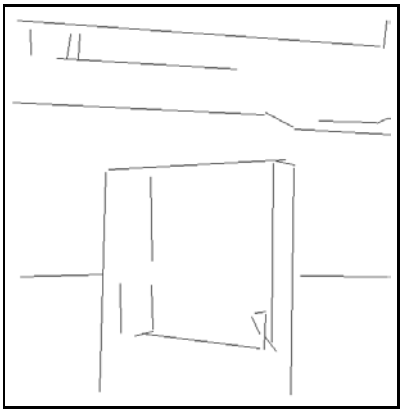

Figure 5.15 - *Groupement - 27 segments* - $\epsilon^{\surd} = 11$

Figure 5.16 - *Téléphone*

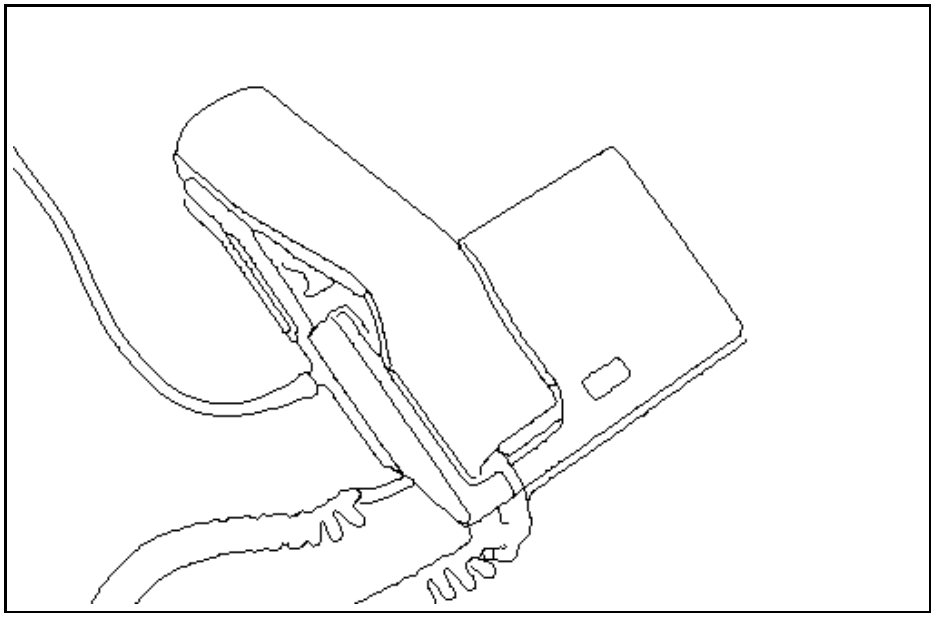

Figure 5.17 - *Détection de contours et sélection des meilleurs groupes - 560 chaînes - 23 groupes*

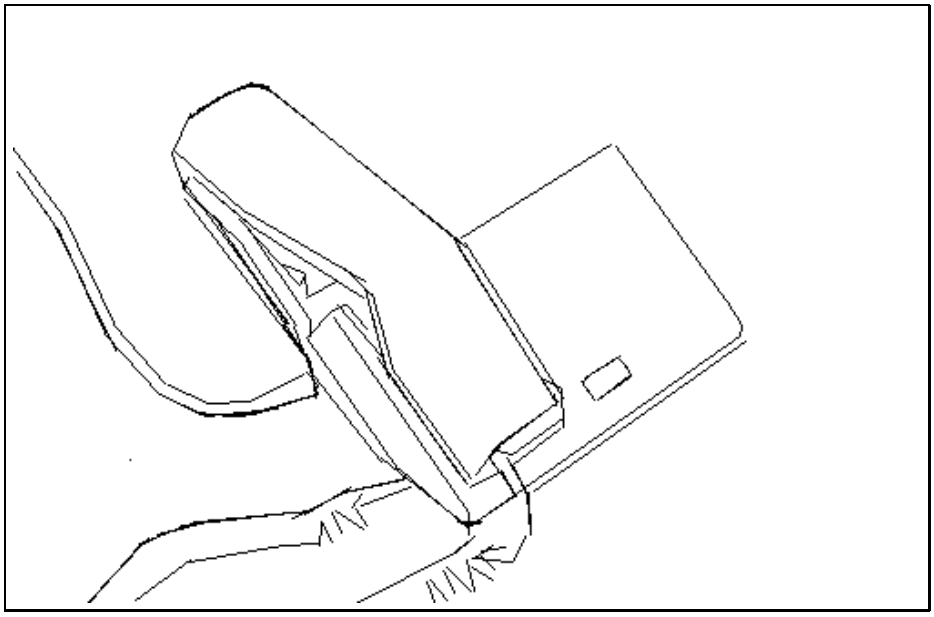

Figure 5.18 - *Avant groupement - 179 segments* - $\epsilon^{\surd} = 3$

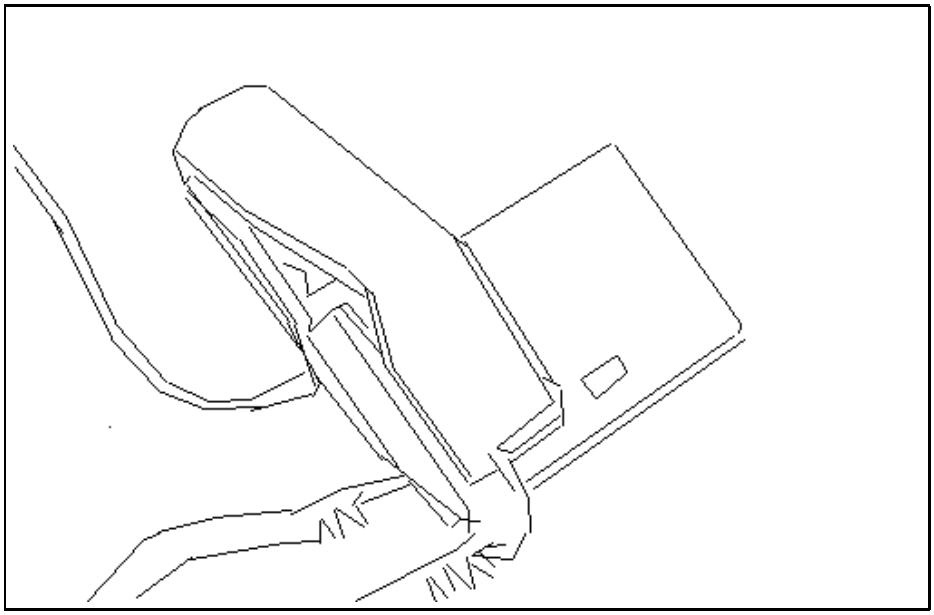

Figure 5.19 - *Aprés groupement - 102 segments*

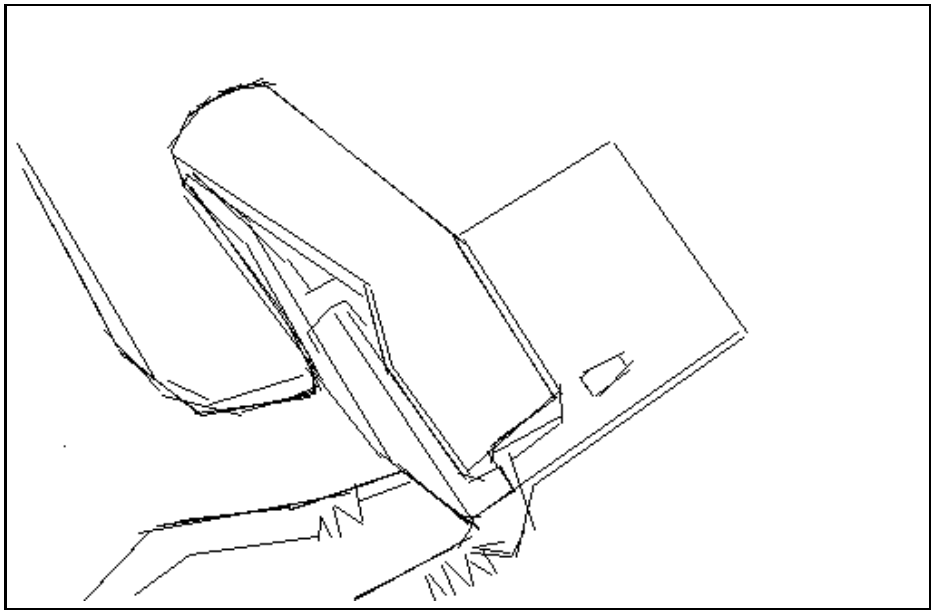

Figure 5.20 - *Avant groupement - 132 segments* - $\epsilon^{\surd} = 6$

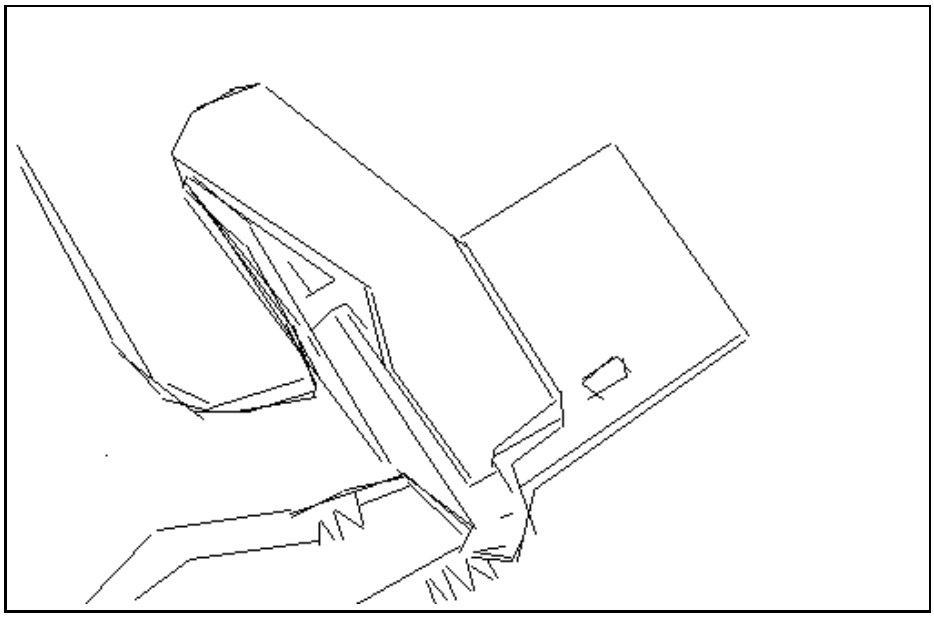

Figure 5.21 - *Aprés groupement - 92 segments*

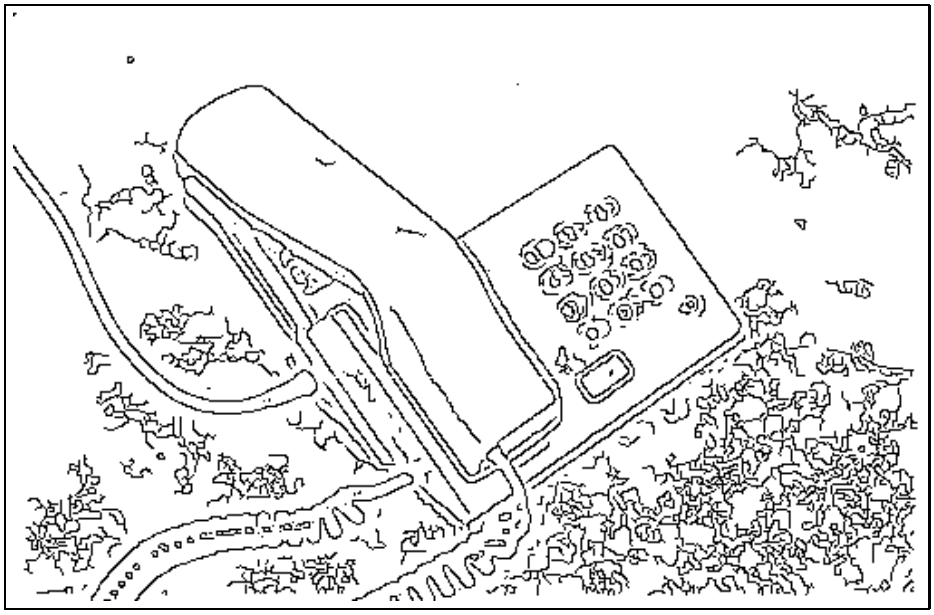

Figure 5.22 - *Téléphone "bruité" - Détection de contours - 2780 Chaînes*

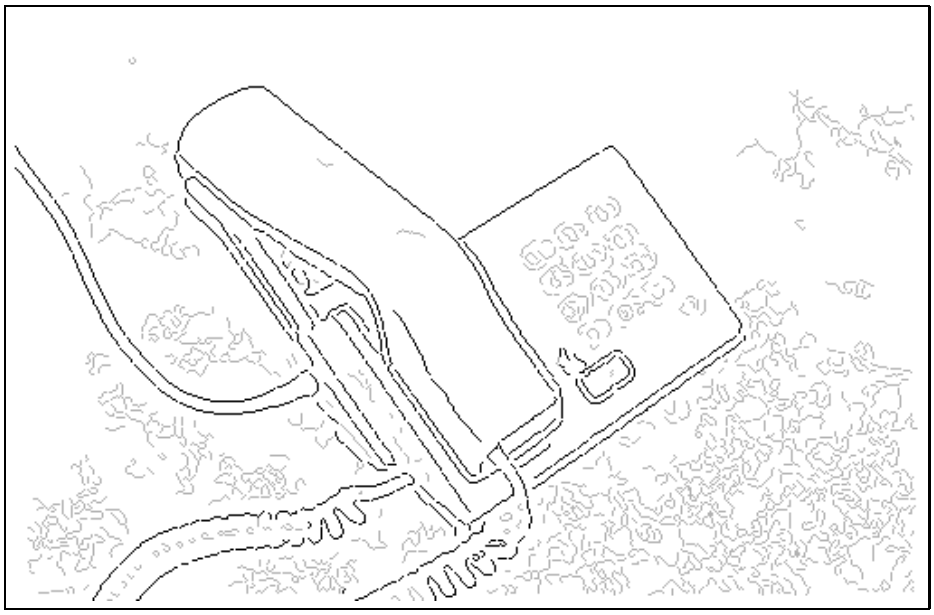

Figure 5.23 - *Sélection des meilleurs groupes - Chaînes couvertes par 29 groupes*

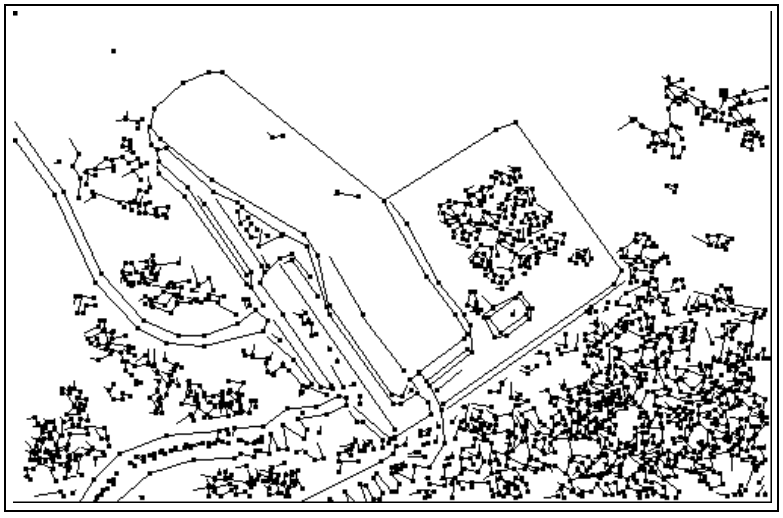

Figure 5.24 - *Approximation polygonale à partir des contours - 1311 segments - $\epsilon^{\surd} = 3$*

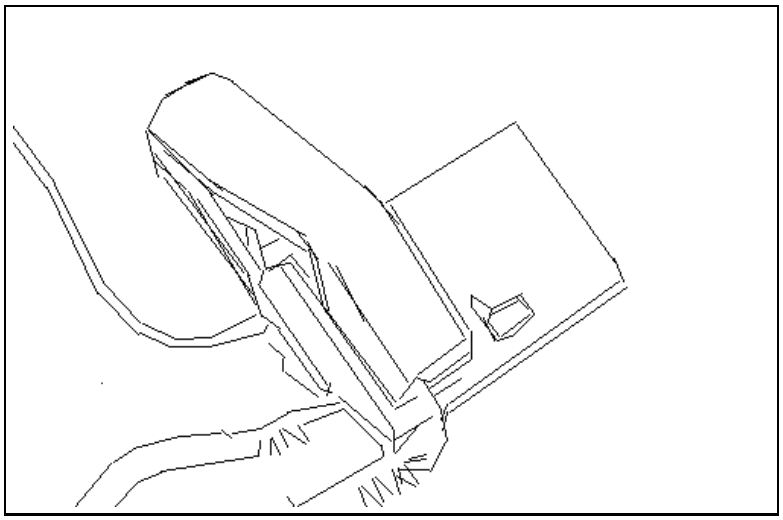

Figure 5.25 - *130 segments aprés détection et groupement - $\epsilon^{\surd} = 3$. Cet exemple permet de comparer le résultat d'une détection de segments classique (par approximation polygonale des contours) avec les hypothèses de segments issues des structures saillantes.*

5.3 Hypothèses "arcs"

En raison de la discrétisation des contours, la discrimination entre segments et courbes est particulièrement délicate. Une courbe sous forme d'une chaîne de points n'est rien d'autre qu'une succession de segments élémentaires reliant les points entre eux. La détection de courbes pose donc les problèmes du choix de l'échelle de lissage avec laquelle la détection des points dominants aura lieu, et du choix du modèle de représentation de la courbe une fois segmentée.

Plusieurs méthodes existent pour extraire les points dominants d'une courbe, la plupart reposent sur une estimation de la courbure. Le choix des points de coupure est ensuite établi à partir de fusion de points de courbure semblable [Wuescher et Boyer, 1991] ou bien, plus généralement, à partir des discontinuités de la mesure de courbure [Wu et Wang, 1993] [Tsang *et al.*, 1994] . Cette dernière approche souligne en général la nécessité de comparer la détection de points dominants selon différentes échelles de lissage, afin d'en extraire une représentation hiérarchique [Fermüller et Kropatsch, 1992] [Rattarangsi et Chin, 1992] .

Concernant le choix d'un modèle de courbe, une première démarche consiste à approcher localement des portions de chaînes par des modèles paramétriques de coniques, comme des arcs de cercles ou d'ellipses [Joseph, 1994] [Ellis *et al.*, 1991] [Cabrera et Meer, 1996] . Le problème est alors de définir des critères correspondant à des arcs visuellement importants sur une chaîne. En tenant compte d'échelles multiples, un découpage est alors possible en fonction de critères d'erreur entre chaîne et modèle d'arc, d'amplitude des discontinuités et d'unicité de détection [Saund, 1991] .

Lorsqu'il n'est pas possible, ou souhaitable, d'utiliser des coniques, des modèles plus généraux sont disponibles. Ces modèles, empruntés à la CAO, sont définis à l'aide de fonctions polynômiales. Ainsi, les modèles d'interpolation de chaîne de points (*C-splines*), ou d'approximation par points de contrôles (*B-splines*) sont disponibles selon l'ordre de continuité désiré le long de la courbe ($\mathcal{C}^1$ pour des *C-splines* ou des *B-splines* quadratiques, $\mathcal{C}^2$ pour des *B-splines* cubiques) [Arbogast, 1990] [Goshtasby, 1993] .

5.3.1 Détection d'arcs élémentaires

La plupart de ces méthodes demandent des calculs intensifs et sont exposées à des erreurs éventuellement importantes étant donné qu'elles reposent principalement sur une estimation précise des modèles de courbes ou de la courbure le long de la chaîne à segmenter. Afin d'apporter une réponse plus qualitative à ce problème, Fischler et Bolles [Fischler et Bolles, 1986] [Fischler et Wolf, 1994] tirent deux principes pour la segmentation de courbes selon des critères perceptuels : le partitionnement doit être stable en cas de faibles perturbations et doit répondre aux principes Gestaltistes de simplicité et globalité. Le principe de continuité est aussi souvent utilisé, pour assurer une régularité de courbure et de co-circularité.

En reprenant ces principes, Gao et Wong [Gao et Wong, 1993] proposent un système de segmentation de courbes en fragments perceptuels inspiré d'expériences psycho-visuelles décrites par Rock (1975). D'après ces expériences, les sujets tendent à découper une courbe selon trois critères. Les points de coupure marquent un changement de propriété visuelle de la courbe. Les fragments d'une courbe sont choisis de manière à faciliter au maximum la reconstruction de celle-ci. Enfin, ces fragments permettent de faciliter la distinction entre différentes courbes.

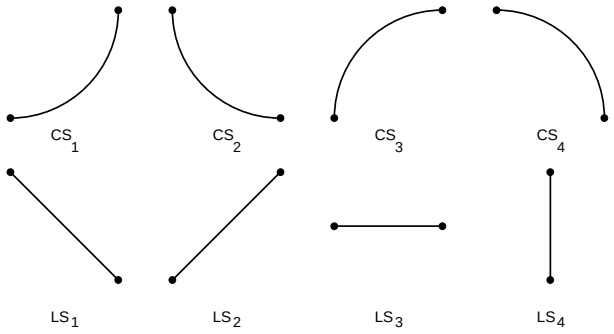

Figure 5.26 - *Huit arcs élémentaires*

Ces observations conduisent à choisir comme propriété visuelle d'une courbe, les variations des dérivées directionnelles le long de son tracé. Une courbe 2D peut être exprimée localement à l'aide de deux fonctions :

$$y = f(x) \quad et \quad x = \phi(y)$$

Les changements de signe des taux de variations de ces deux fonctions permettent de définir huit classes d'arcs élémentaires, représentées par la figure 5.26. Ces arcs correspondent aux propriétés décrites dans le tableau 5.1.

La méthode originalement proposée par Gao et Wong s'applique directement sur l'image d'intensité. Les arcs élémentaires sont extraits à l'aide d'un suivi des contours, en estimant les dérivées directionnelles à l'aide du gradient de la fonction intensité $I(x, y)$:

$$f'(x) = \arctan(\frac{\partial I}{\partial y} / \frac{\partial I}{\partial x})$$

Cette approche directe donne des résultats satisfaisants sur des images bien contrastées. Cependant, un grand nombre de difficultés apparaissent lorsque l'image devient bruitée. En particulier, l'absence de considérations d'échelle accentue les problèmes de discrétisation et du choix du pas de déplacement le long des valeurs importantes du gradient.

Ce formalisme pour la segmentation et le groupement de courbes nous permet cependant d'adopter une démarche semblable au groupement de segments. En effet,

Arc	$y = f(x)$	$x = \phi(y)$	$\dot{y} = f'(x)$	$\dot{x} = \phi'(y)$
CS_1	M^+	M^+	M^+	M^-
CS_2	M^-	M^-	M^+	M^-
CS_3	M^+	M^+	M^-	M^+
CS_4	M^-	M^-	M^-	M^+
LS_1	M^-	M^-	c	c
LS_2	M^+	M^+	c	c
LS_3	c	N/A	∞	0
LS_4	N/A	c	0	∞

Table 5.1 - *Propriétés de variations des fonctions d'une courbe. M^+ et M^- définissent respectivement une croissance et une décroissance monotone. On note c une valeur constante et N/A une valeur non définie.*

une classification des différents arcs possibles est nécessaire afin de pouvoir les comparer, et éventuellement les grouper. Dans un premier temps, des arcs élémentaires sont extraits le long de chaque groupement, pour une échelle de lissage donnée. Ces hypothèses sont ensuite simplifiées à l'aide de règles de composition et de groupement d'arcs élémentaires.

Partitionnement de courbes selon des critères perceptuels

Nous ne gardons que les principes de découpage perceptuel des courbes pour les transposer à notre approche. A la différence de l'approche directe, le parcours des courbes est réalisé au préalable par l'optimisation du réseau de saillance. Les fonctions dont nous étudions les variations sont donc extraites à partir des dérivées première et seconde du tracé de chaque groupement.

Dans une étude comparée de différentes méthodes d'estimation de la courbure d'une chaîne de pixels, Worring et Smeulders [Worring et Smeulders, 1993] soulignent l'importance du choix de l'estimation des dérivées sur l'accumulation des erreurs de calcul. La méthode qu'ils recommandent repose sur l'application d'un filtre gaussien de lissage et de dérivation le long de la chaîne. Ils démontrent en outre que, dans le cas d'un filtre à réponse impulsionnelle finie, les erreurs dues à la largeur du filtre choisi deviennent négligeables à partir d'une largeur maximale, fonction de l'échelle de lissage utilisée pour le filtre.

Nous proposons une estimation de ces dérivées par application d'un filtre à réponse impulsionnelle infinie. En particulier, le filtre de Deriche [1] [Deriche, 1990] permet une implémentation récursive, particulièrement efficace dans le cas d'une chaîne de pixels.

1. Voir le §2.3.1.3, page 55, sur les détecteurs de contours optimaux.

Il est enfin indispensable de définir un sens de parcours le long des chaînes afin de pouvoir détecter les différentes classes d'arcs plus facilement. Nous choisissons pour cela le signe de la courbure de la chaîne, définie à l'aide des dérivées première et seconde par :

$$\kappa(t) = \frac{\dot{x}(t)\ddot{y}(t) - \dot{y}(t)\ddot{x}(t)}{[\dot{x}(t)^2 + \dot{y}(t)^2]^{3/2}}$$

On peut noter dès à présent que, en ne reposant que sur des changements de signes, ces critères de découpage sont moins exposés aux erreurs de calculs des valeurs des dérivées, comme le montre la figure 5.27.

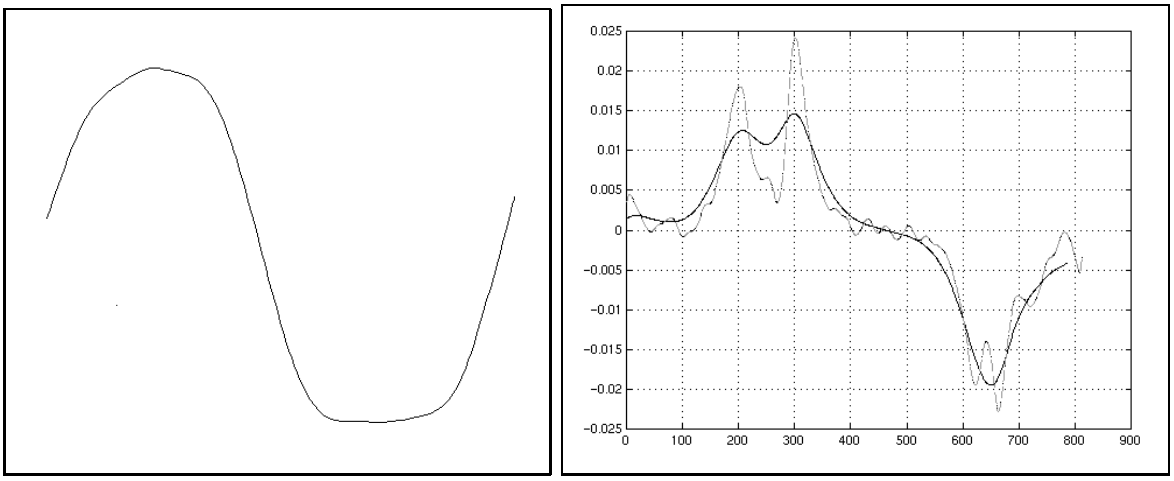

Figure 5.27 - *La figure de droite représente l'estimation de la courbure d'une chaîne (figure de gauche) pour les échelles $\alpha = 0.2$ (en gris) et $\alpha = 0.09$ (en noir).*

Méthode d'extraction des hypothèses "arcs"

Les différents arcs sont donc définis, pour chaque composante de γ, à l'aide du sens de croissance et des changements de signe de la dérivée première, ainsi que du signe de la dérivée seconde.

Le tableau 5.2 résume l'application des règles du tableau 5.1 pour un sens de parcours ($\kappa(t) < 0$). Seules les classes d'arcs CS_i sont utiles à ce niveau de groupement de courbes. Les classes d'arcs LS_i correspondent aux parties "rectilignes" des courbes, détectées de manière plus efficace lors du groupement de segments. Il n'est donc pas nécessaire de faire la différence entre les quatre classes de fragments rectilignes. Elle sont remplacées par une classe générique de fragments rectilignes LS.

L'algorithme de détection proprement dit consiste à étiqueter, dans un premier temps, chaque point de la chaîne selon l'une des huit classes d'arcs. Dans un

Arc	$\dot{x}(t)$	$\dot{y}(t)$	$\ddot{x}(t)$	$\ddot{y}(t)$
CS_1	$Max^+ \searrow 0$	$0 \searrow Min^-$	< 0	< 0
CS_2	$0 \nearrow Max^+$	$Max^+ \searrow 0$	> 0	< 0
CS_3	$Min^- \nearrow 0$	$0 \nearrow Max^+$	< 0	> 0
CS_4	$0 \searrow Min^-$	$Min^- \nearrow 0$	> 0	> 0
LS	$\cdots$	$\cdots$	$= 0$	$= 0$

Table 5.2 - *Propriétés de variations des fonctions d'une courbe. Max^+ et Min^- définissent respectivement un extremum positif et un extremum négatif.*

second temps, les points consécutifs de la chaîne appartenant à une même classe sont agglomérés sous forme d'arcs élémentaires.

Cette approche en deux passes permet de rectifier d'éventuelles erreurs d'étiquetage afin de reconstituer des arcs aussi complets que possible. Les erreurs les plus fréquentes sont de courtes séquences de points étiquetés $[LS_i]$ au milieu d'une série de points appartenant à la même classe $[CS_j]$. Ce type de situation apparaît en particulier lorsque le tracé d'une courbe correspond localement à une droite par rapport à l'échelle envisagée.

Les règles de groupement suivantes permettent de réparer ces erreurs :

$$\begin{cases} [CS_j]^n \cdot [LS]^p \cdot [CS_j]^m \implies [CS_j]^{n+p+m} \\ [CS_j]^n \cdot [LS]^p \cdot [CS_k]^m \implies [CS_j]^{n+\frac{p}{2}} \cdot [CS_k]^{m+\frac{p}{2}} \\ \qquad\qquad\qquad\qquad avec \;\; j \neq k \;\; et \;\; p < Min(\frac{m}{2}, \frac{n}{2}) \end{cases} \tag{5.6}$$

en notant $[CS_j]^n$ une séquence de n points de classe CS_j.

Enfin, comme pour la détection de segments, les points dominants entre arcs élémentaires ainsi que les extrema de courbure sont ajoutés à une liste de points d'intérêt afin de compléter la détection de jonctions.

Analyse des résultats

Le résultat de cette première étape est donc la détection d'arcs élémentaires et de leurs points dominants associés. Les figures suivantes illustrent le résultat du découpage en arcs, dans un premier temps sur des courbes synthétiques, puis sur des courbes extraites de scènes réelles.

Les cercles de la figure 5.28 et les ellipses des figures 5.29 et 5.30 permettent de vérifier la segmentation de courbes simples en arcs perceptuellement importants. Dans chacune de ces figures, les arcs de classes CS_1 et CS_3 sont tracés en noir et les arcs de classes CS_2 et CS_4 en gris. La figure 5.30 permet en particulier de noter la stabilité de détection des transitions entre arcs sur des structures de tailles différentes, et ce, malgré un lissage relativement important. Les erreurs de positionnement qui apparaissent sur les figures 5.29 et 5.30 sont dues au point de départ de la chaîne de pixels lorsque celle-ci forme une courbe fermée.

La figure 5.31 illustre la distinction entre arcs de classe CS_i (en noir) et fragments rectilignes de classe LS (en gris). Cette distinction tend naturellement à s'estomper pour des échelles de lissage plus importantes.

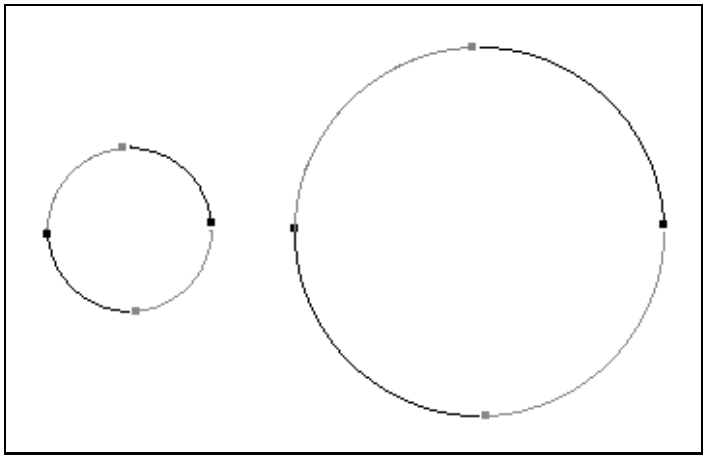

Figure 5.28 - *Segmentation de cercles - rayons 40 et 100 pixels* - $\alpha = 0,125$.

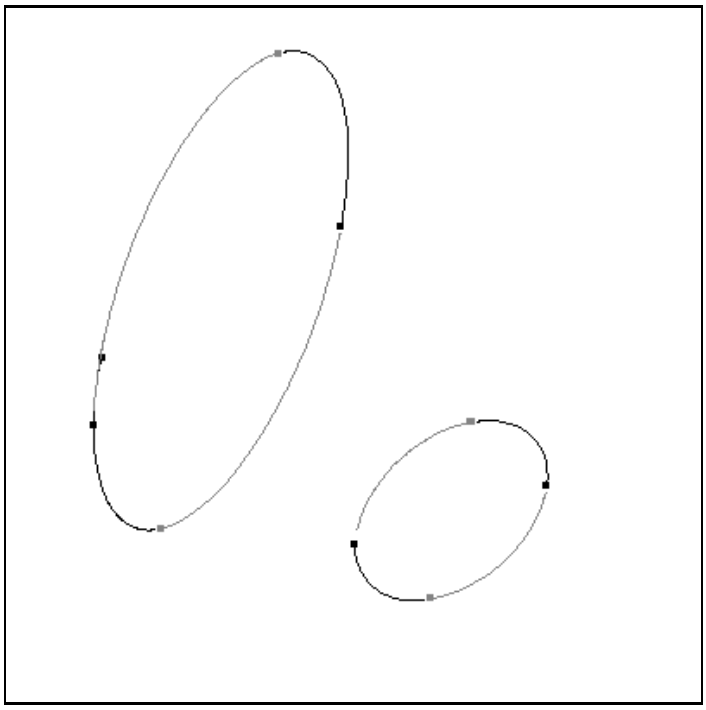

Figure 5.29 - *Segmentation d'ellipse inclinées* - $\alpha = 0,125$. *Le point supplémentaire sur l'ellipse de gauche vient du point de départ de la chaîne de contour.*

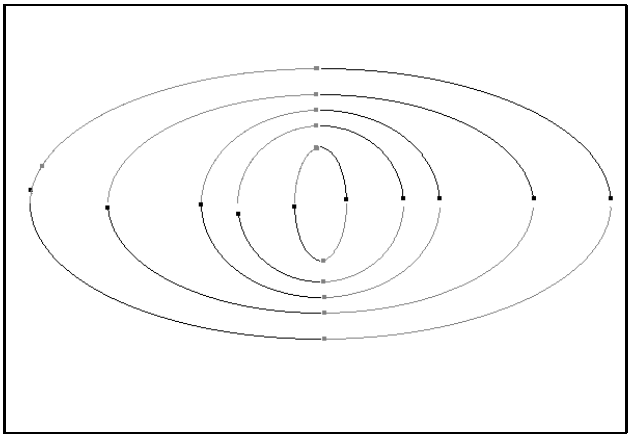

Figure 5.30 - *Ellipses droites de tailles variables* - $\alpha = 0,125$. *Comme pour la figure précédente, les erreurs de localisation des points de la partie gauche des ellipses viennent du choix de point de départ sur chaque chaîne de contour.*

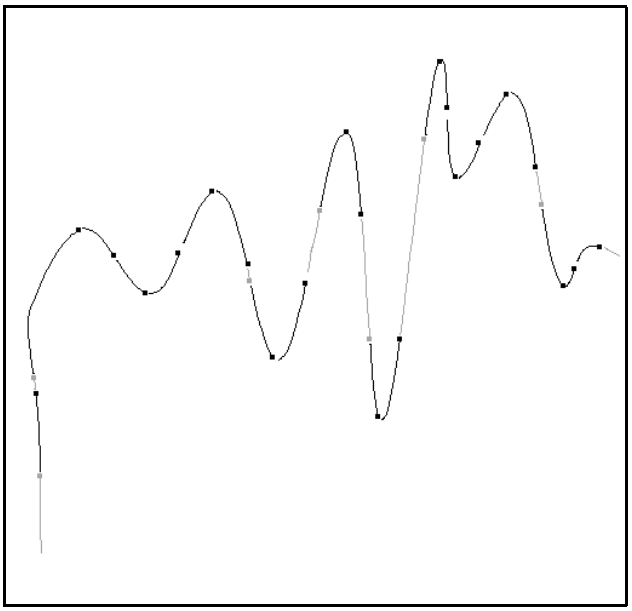

Figure 5.31 - *Courbe quelconque* - $\alpha = 0,2$. *Les arcs en "gris" ont été classifiés en tant que segments rectilignes (classe LS).*

5.3.2 Organisation perceptuelle des arcs

Le groupement des arcs élémentaires répond aux mêmes impératifs que le groupement des segments : constituer un ensemble d'hypothèses aussi réduit que possible. Cette étape est organisée selon le même principe de groupement hiérarchique afin de réduire au maximum les redondances d'hypothèses.

Principes du groupement d'arcs

La comparaison entre arcs est en grande partie facilitée par la classification préalable. Cette classification apporte en effet des contraintes importantes sur les groupements possibles entre arcs élémentaires. Ici encore, les arcs sont comparés par ordre de longueur croissante.

Soient A_1 et A_2 deux arcs élémentaires, de longueurs respectives L_{A_1} et L_{A_2}. On suppose $LA_1 > L_{A_2}$. L'arc A_1 est composé des pixels $\{P_1^1, \cdots, P_n^1\}$, et l'arc A_2 des pixels $\{P_1^2, \cdots, P_m^2\}$. On note de plus $\mathcal{C}_1$ et $\mathcal{C}_2$ la classe de chaque arc.

- *Groupement par similarité et proximité.*

  Deux arcs sont considérés comme similaires s'ils appartiennent à la même classe et s'ils partagent une majorité de pixels en commun. L'appartenance d'un point de A_2 à l'arc A_1 est définie par rapport à une distance maximale $\epsilon^{\approx}$. Ce critère qualitatif donne en pratique de meilleurs résultats que des méthodes de comparaison plus classiques telles que la corrélation. En effet, il est beaucoup plus rapide qu'une comparaison point à point entre arcs qu'impliquerait un calcul de corrélation, et moins exposé aux erreurs de tracé des arcs. Il permet en particulier plus de souplesse face aux variations du tracé des arcs pour une échelle donnée.

  On note $\Gamma_{1,2}$ l'ensemble des pixels communs aux deux arcs et $N_{1,2}$ le nombre de pixels de cet ensemble.

$$\forall P \in \Gamma_{1,2}, \quad P \in A_2 \quad et \quad \exists Q \in A_1 / \|\overrightarrow{PQ}\| < \epsilon^{\approx}$$

  Le taux de recouvrement de A_2 par A_1 est simplement défini par le rapport :

$$T_{1,2} \; = \; \frac{N_{1,2}}{m}$$

  avec m, nombre de pixels de l'arc A_2.

  Ce taux prend la valeur 1 lorsque les deux arcs sont confondus et 0 lorsqu'ils sont suffisamment éloignés. Le critère de groupement revient donc à décider d'un seuil au dessus duquel deux arcs doivent être fusionnés. Ce seuil est arbitrairement fixé à $0,8$, soit 80% de la longueur de A_2.

Afin de réduire la complexité de la comparaison point à point des distances entre deux arcs, de l'ordre de $\mathcal{O}(n \cdot m)$, nous adoptons l'heuristique suivante : lorsque plusieurs points consécutifs de A_2 remplissent le critère de distance avec A_1, l'hypothèse d'un recouvrement possible est établie. Il suffit ensuite de comparer les points restant de A_2 avec les successeurs du dernier point trouvé sur A_1.

Avec un seuil de recouvrement supérieur à 0.9, le groupement d'arcs par proximité revient à éliminer les arcs redondants tout en gardant les arcs les plus longs.

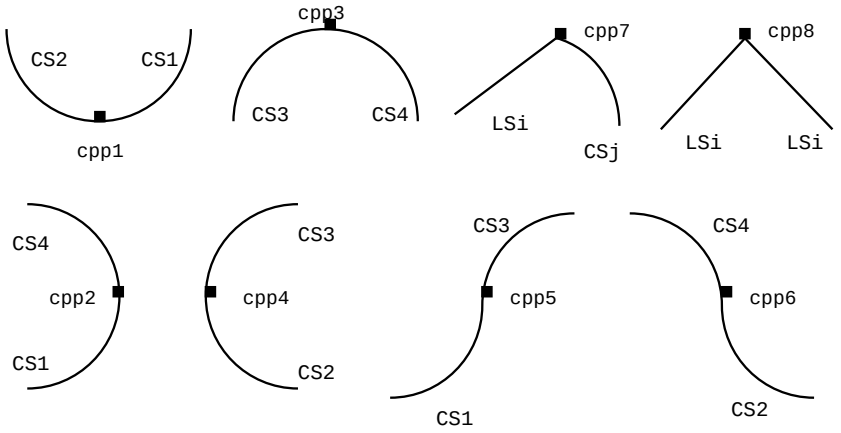

Figure 5.32 - *Huit groupements élémentaires*

– *Groupements élémentaires par co-circularité.*

Les arcs élémentaires de classes CS_i ne sont pas les seules hypothèses formulées lors de la détection des fragments de courbes. Deux autres types d'hypothèses sont générées durant cette étape :

1. Groupements élémentaires entre arcs co-circulaires.
   Les points de coupure correspondant au passage d'un type d'arc à l'autre le long d'une courbe, les triplets constitués de deux arcs consécutifs et d'un point de coupure constituent des groupements élémentaires, notés $\mathcal{G}_i$. En reprenant la nomenclature définie par Gao et Wong, huit types de groupements sont ainsi définis, chacun associé à un type de point de coupure cpp_i particulier. Les groupements élémentaires sont résumés dans la figure 5.32.
   Les hypothèses d'arcs co-circulaires correspondent aux groupements des points dominants $cpp_1, \cdots, cpp_4$.
2. Compositions d'arcs co-circulaires
   Cette hypothèse d'arc étend la précédente à un nombre indéterminé d'arcs co-circulaires. Des séquences d'arcs telles que $cpp_1 \rightarrow cpp_2$ ou

bien $cpp_2 \rightarrow cpp_3 \rightarrow cpp_4$ constituent autant d'hypothèses de courbes fermées.

Ces hypothèses d'arcs particulières sont ensuite simplifiées à l'aide du groupement par proximité et similarité décrit précédemment.

Résultats et discussion

Les résultats de la détection et du groupement d'arcs démontrent un comportement complémentaire du groupement de segments, c'est à dire, une détection des parties courbes satisfaisante et des ambiguïtés le long des parties rectilignes.

Comme le montrent les ellipses de la figure 5.34, les paires d'arcs co-circulaires sont correctement regroupées, jusqu'à reconstituer, en l'absence de point d'inflexion, la courbe d'origine. La courbe de la figure 5.36 donne un autre exemple de groupement d'arcs en présence d'inflexions. Les fragments de courbes absents ont été identifiés comme des segments rectilignes.

Les figures des pages 199 à 199 reprennent la scène de téléphone précédemment utilisée. Dans chaque cas, un exemple de groupement d'arcs élémentaires est présenté pour deux échelles de lissage. Les arcs détectés y sont représentés en noir et les fragments rectilignes en gris. Dans les résultats complémentaires donnés en annexe B, les figures B.9 ,B.11 et B.13 donnent des exemples de paires d'arcs co-circulaires.

Ces résultats sont représentatifs des effets de la présence de structures de tailles différentes le long des chaînes. En particulier, des structures d'échelle trop petite introduisent des inflexions locales qui interrompent des séquence d'arcs co-circulaires. A l'inverse, des structures d'échelle trop grande admettent des inflexions trop importantes. Enfin, la détection de fragments rectilignes devient rapidement ambiguë en cas de prolongement d'un arc par un segment.

Une méthode pour départager les courbes correspondant à des "coins arrondis" des courbes correspondant à de véritables arcs consiste à comparer chaque arc élémentaire avec un modèle d'arc circulaire et ne conserver que ceux qui s'approchent le plus du modèle. Les figures B.14 et B.15 donnent des exemples de détection d'arcs d'ellipses à partir de la figure B.10. La méthode d'approximation utilisée à cette fin est inspirée de [Roth et Levine, 1993] . A l'aide d'une procédure de recuit simulé, une série de tirages aléatoires de points le long d'un arc permet de faire converger un modèle d'ellipse vers une représentation fidèle de cet arc. D'une manière plus générale, l'identification d'arcs particuliers pourrait être étendue à d'autre modèles de courbes afin de simplifier la représentation des hypothèses autant que possible.

Enfin, les résultats obtenus pour la détection des arcs élémentaires imposent un remarque concernant l'invariance de la détection. Les critères de classification des arcs sont effet liés au choix d'un repère dans l'image, indispensable

pour les calculs de dérivées. Dans notre cas, le repère choisi correspond aux directions horizontales et verticales de l'image. Cette remarque est importante car une même figure aura une décomposition en arcs différente si elle subit une rotation.

Les critères de détection restent cependant cohérents avec le comportement de la vision humaine. De nombreuses expériences sur des figures relativement simples, montrent en effet des difficultés d'identification et de perception de symétries en cas de forte rotation. Il suffit d'essayer d'identifier un visage présenté la tête en bas pour s'en persuader.

Pourtant, si le découpage en arcs élémentaires est susceptible de varier, le groupement d'arcs, lui, rétablit l'invariance par rotation. En effet, les propriétés visuelles de co-circularité et d'inflexions restent invariantes par rotation. Par exemple, quel que soit le repère choisi pour découper en arcs une ellipse telle que celle de la figure 5.34, la composition des arcs co-circulaires donnera toujours au final une ellipse complète.

Cependant, la classification d'arcs élémentaires peut jouer un rôle important lorsque la différence entre deux images est minime, comme dans le cas de paires stéréoscopique en vision binoculaire ou bien pour une séquence d'images.

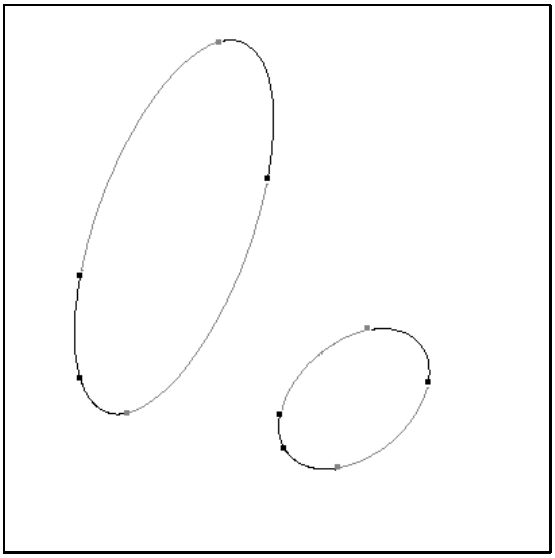

Figure 5.33 - *Arcs élémentaires* - $\alpha = 0,125$

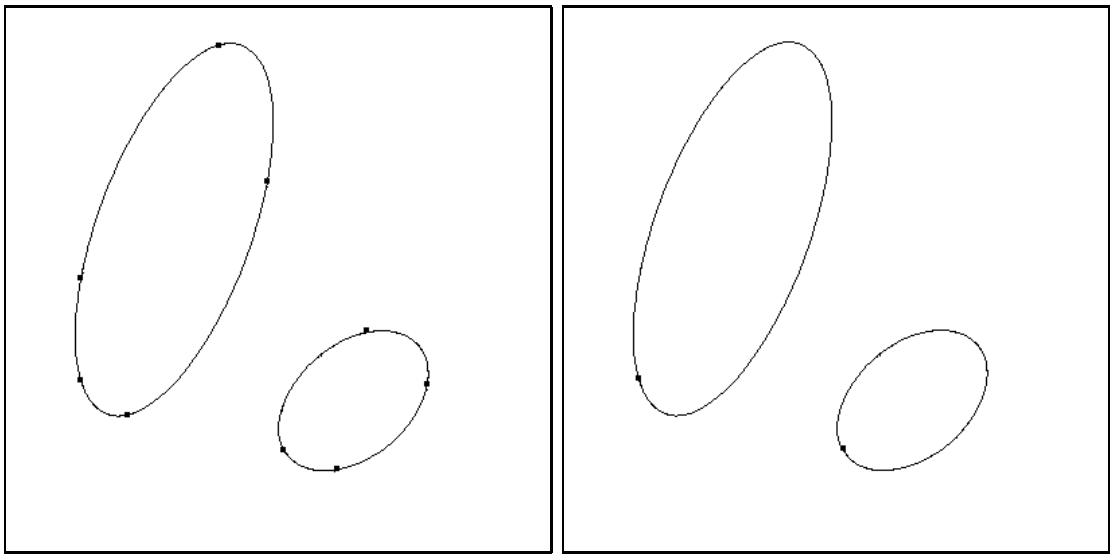

Figure 5.34 - *Paires d'arcs co-circulaires et groupements.*

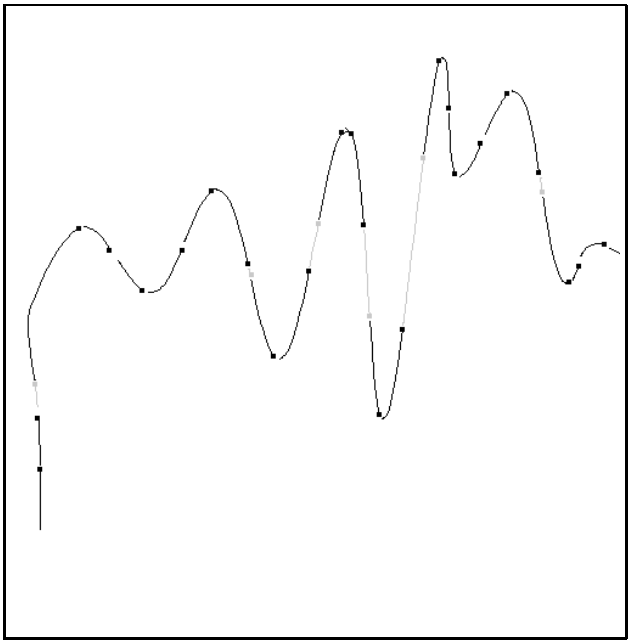

Figure 5.35 - *Arcs élémentaires sur une courbe quelconque* - $\alpha = 0,2$

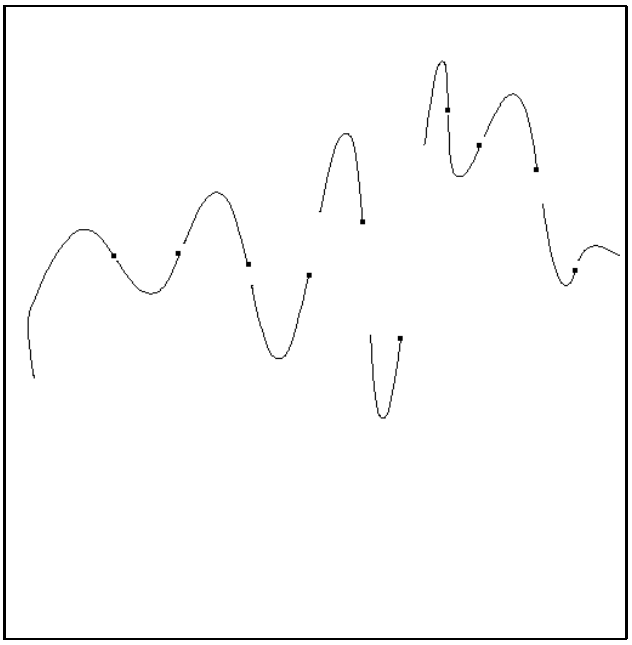

Figure 5.36 - *Paires d'arcs co-circulaires*

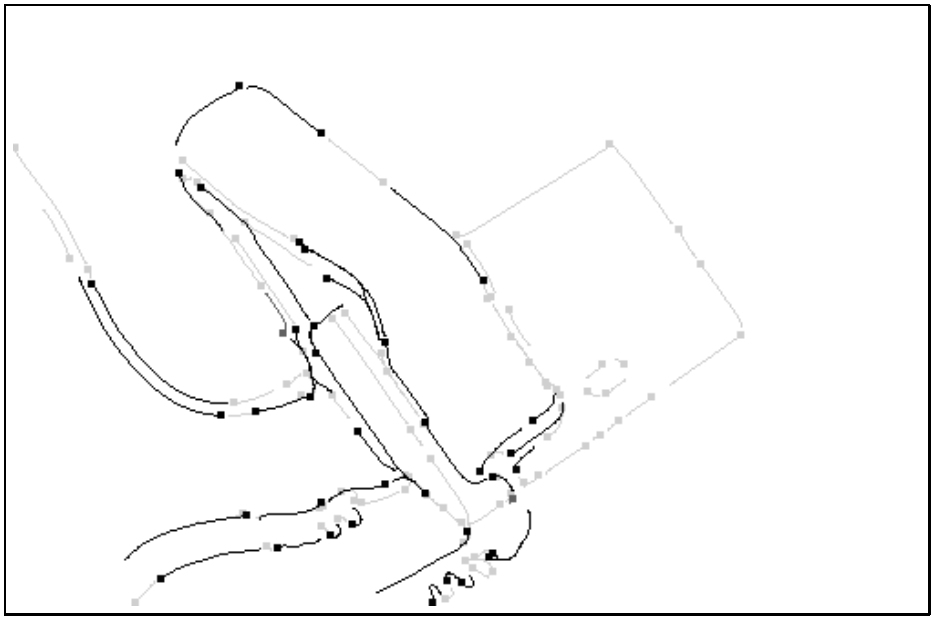

Figure 5.37 - *Téléphone - 105 arcs élémentaires* - $\alpha = 0,15$

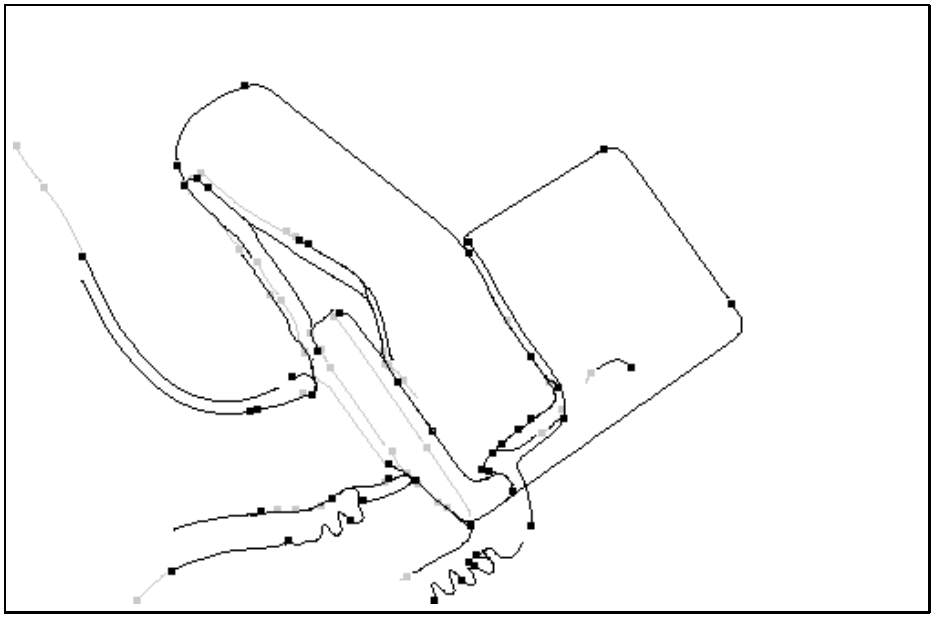

Figure 5.38 - *Téléphone - 90 arcs élémentaires* - $\alpha = 0,07$

5.4 Conclusions et perspectives

Nous avons présenté dans ce chapitre des principes d'extraction d'éléments visuels à partir des groupements résultant d'un réseau de saillance. Ces principes ont été appliqués à la perception de segments, d'arcs et de jonctions. Dans chaque cas, un ensemble d'hypothèses élémentaires est d'abord détecté à partir de chaque groupement individuel. Ces hypothèses sont ensuite mises en commun afin d'être groupées de manière hiérarchique. Le résultat est un ensemble simplifié d'éléments visuels représentatifs des structures curvilinéaires de la scène.

Le groupement de chaque type d'élément visuel a été illustré à l'aide d'exemples de scènes artificielles et d'images réelles. Les résultats obtenus confirment la détection correcte de la majeure partie des structures d'intérêt de la scène, mais ils présentent encore un grand nombre d'ambiguités qui sont autant d'obstacles à une application réelle. L'amélioration de cette méthode selon deux axes pourrait apporter à court terme une plus grande stabilité aux résultats obtenus.

- *Optimisation de ressources*

  L'une des caractéristiques de cette méthode de groupement est la manipulation de grandes quantités d'hypothèses lors de l'organisation des éléments visuels. Le nombre de segments ou d'arcs peut temporairement atteindre quelques milliers pour aboutir au final à quelques centaines d'hypothèses optimisées. Les temps de traitement, de l'ordre d'une dizaine de minutes pour l'organisation de 3000 à 4000 hypothèses, sont essentiellement dus à l'implémentation sous forme de listes chaînées. D'autres types de structures de données pourraient accélérer les calculs tout en économisant les ressources mémoires.

  En particulier, l'utilisation de méthodes d'indexation pourrait accélérer la recherche et la comparaison entre hypothèses lors de la détection d'un nouvel élément. Un exemple d'utilisation d'index pour optimiser un groupement perceptuel peut être trouvé dans [Havaldar *et al.*, 1996] .

- *Détection multi-échelles*

  La notion d'échelle est une autre caractéristique importante de cette approche. Elle est présente tant au niveau de la détection que celui du groupement mais reste, pour le moment, à la discrétion de l'utilisateur. Un grand nombre d'ambiguités pourrait être levées en généralisant la détection à une étude multi-échelle. La sélection des segments et des arcs les plus stables selon plusieurs échelles de détection permettrait ainsi de ne conserver, dans chaque cas, que les éléments véritablement rectilignes ou courbes.

  [Jacot-Descombes et Pun, 1997] donnent un exemple de groupement à partir d'une détection de contours dans un espace échelle. A l'aide d'un filtre gaussien de lissage et dérivation, les contours sont extraits selon différentes échelles.

Pour une échelle donnée, ceux-ci sont découpés selon des intervalles pour lesquels la variation de courbure reste inférieure à un certain seuil. A chaque intervalle est associé une mesure de saillance, en fonction de sa longueur et de l'échelle de détection. Cette mesure favorise les intervalles longs pour une échelle faible et des intervalles de plus en plus courts lorsque l'échelle croît. Les intervalles sélectionnés finalement sont ceux présentant la meilleure saillance sur l'ensemble des échelles.

Ces éléments visuels ne forment en aucun cas une représentation complète de la scène. Ils doivent être considérés comme un ensemble d'indicateurs probables de la structure de l'image, admettant une certaine part d'ambiguités. A plus long terme, deux prolongements de cette méthode devraient permettre une réduction significative du nombre hypothèses ambigües.

– *Discrimination entre hypothèses.*

Les problèmes d'échelle et de discrétisation se font particulièrement sentir lorsque les chaînes ne sont ni franchement rectilignes, ni particulièrement courbes. Il y a alors ambiguïté entre plusieurs modèles pour une même portion de chaîne, avec éventuellement des recouvrements entre modèles. En appliquant le principe de simplicité commun aux règles de groupement, le modèle le plus simple devrait être choisi en cas d'ambiguité.

La complexité d'un modèle peut être évaluée numériquement par une mesure de vraisemblance empruntée à la théorie de l'information. Pour simplifier, ce type d'approche mesure la complexité d'un modèle à partir du nombre de paramètres nécessaires à sa description et d'écarts entre le modèle et la portion de chaîne d'origine [Lindeberg et Li, 1997] . Plus simplement, il est souvent suffisant d'appliquer différents modèles, par ordre de complexité croissante, à une portion de chaîne ambigüe et de retenir le modèle présentant le meilleur écart quadratique avec la chaîne [Leonardis et Bajcsy, 1992] .

– *Intégration d'autres sources de primitives*

Le résultat du groupement est un ensemble d'éléments de représentation probables. Il est donc possible de rectifier ces éléments de manière à ce qu'ils représentent les structures rectilignes de la scène aussi fidèlement que possible. Ce type de démarche consiste à lever les ambiguïtés en utilisant le résultat du groupement comme centres d'attention.

L'ajustement des hypothèses pourrait être réalisé selon un procédé de prédiction et vérification, en comparant chaque hypothèse de segment, d'arc ou de point d'intérêt soit avec l'image d'origine, soit avec le résultat d'autres méthodes de détection plus directes dans le voisinage immédiat des hypothèses. La localisation des points d'intérêt peut être ainsi rectifiée en appliquant un détecteur de coin spécialisé autour de chaque hypothèse de point. Ou encore,

les hypothèses d'arcs peuvent être ajustées de cette manière à l'aide de modèles paramétriques déformables [Blaszka et Deriche, 1994b] .

Une dernière caractéristique des groupements intermédiaires, parmi les plus importantes, est de privilégier des critères qualitatifs de détection et de groupement. Les techniques quantitatives sont appliquées le plus tard possible dans la chaîne de traitement afin d'éviter l'accumulation d'erreurs de calculs d'une étape à l'autre. Définir à partir de quel moment les ambiguïtés deviennent tolérables reste une tâche très délicate sachant qu'un certain nombre d'entre elles sont impossibles à départager sans l'apport d'informations contextuelles ou de connaissances liées à l'environnement ou à la tâche visuelle recherchée. C'est précisément le rôle des niveaux supérieurs de groupements présentés dans le chapitre suivant.

Chapitre 6

Groupements de haut niveau et mise en correspondance

Les éléments visuels extraits à partir des groupements précédents peuvent être comparés à un croquis sommaire des formes saillantes de la scène. Nous montrons dans ce chapitre comment établir des relations structurelles plus complexes à partir de ces éléments visuels afin de mettre en correspondance des structures issues de deux scènes. Nous illustrons enfin ce dernier niveau de groupement par une méthode de mise en correspondance de jonctions.

6.1 Mise en correspondance structurelle

La *mise en correspondance*, ou problème d'*appariement*, est l'une des principales tâches que doit accomplir un système visuel. D'une manière générale, nous entendons par "correspondance" l'identification d'attributs soumis à une relation commune. Ces attributs peuvent être extraits à partir d'images ou bien de modèles d'objets à identifier.

Ce problème, parmi les plus difficiles de la vision par ordinateur, se retrouve sous différentes formes en fonction de la nature des attributs (points, segments, régions ou structures plus élaborées), de leur origine et des applications envisagées.

- *Reconnaissance de formes*

  Etant donné un ensemble de modèles d'objets connus, le problème de la reconnaissance de formes consiste à mettre en correspondance des attributs extraits d'une image avec ceux des modèles afin de détecter la présence d'objets connus dans la scène observée.

- *Stéréo-vision*

  La stéréo-vision, ou problème d'appariement stéréoscopique, consiste à identifier les attributs communs à deux scènes afin de déduire le relief à partir de la

disparité entre les deux scènes. Selon les systèmes, la comparaison des images fournies par deux ou trois caméras permet de reconstituer la profondeur et finalement de modéliser la scène observée en trois dimensions.

– *Analyse de séquence d'images.*

Cette dernière variante est une généralisation de la vision stéréoscopique de par l'absence d'information sur le mouvement des objets de la scène et des changements de points de vues de la caméra. La mise en correspondance entre images d'une séquence se heurte aux problèmes posés par la présence d'objets multiples dont les mouvements peuvent être différents. Parmi les applications de ce type de mise en correspondance, on peut citer le suivi automatique d'objets ou encore la déduction de vues intermédiaires entre deux points d'observation clefs.

Pour plus de détails sur les variantes de la mise en correspondance et les méthodes utilisées pour résoudre ce problème, le lecteur pourra se référer aux états de l'art établis par [Zhang, 1993] et [Jones, 1997] . Il ressort de ces deux études une classification des méthodes de mise en correspondance selon trois critères : le choix des attributs à apparier, le type de contraintes utilisées pour comparer ces attributs et enfin, une méthode d'optimisation pour établir les correspondances entre attributs.

Une vue d'ensemble des méthodes existantes soulève trois remarques. Il existe d'une part quelques approches génériques au problème de la mise en correspondance, comme par exemple, l'isomorphisme de sous graphes, directement issues de la théorie des graphes. Ces approches sont cependant de complexité exponentielles et par conséquent difficilement appliquables à tout type d'attributs. D'autre part, de nombreux algorithmes existent mais restent adaptés à des applications dans des conditions bien précises. Il en est ainsi, par exemple, de l'utilisation de contraintes épipolaires en vision stéréoscopique.

Enfin, ces études soulignent l'importance des méthodes hiérarchiques, reposant sur la mise en correspondance d'attributs complexes. Ces méthodes sont plus rapides et plus robustes car des structures complexes sont peu nombreuses par comparaison avec des primitives tels que des points d'intérêt ou des segments. La richesse des structures hiérarchiques assure également moins d'ambiguités. Enfin, et c'est là le principal avantage de ces méthodes, un appariement entre deux structures complexes peut être propagé facilement à chaque élément de celles-ci.

Pour ces raisons, la mise en correspondance structurelle est un prolongement fréquent des méthodes d'organisation perceptuelle. L'application de règles de groupement complexes permet d'établir des relations hiérarchiques fortes entre éléments visuels. Nous présentons à présent les principales relations structurelles utilisées par ces méthodes. Ces relations, ainsi que des exemples significatifs de mise en correspondance à partir de ces groupements, nous permettront de définir une approche adaptée aux éléments de représentation définis dans le chapitre précédent.

6.1.1 Relations structurelles

En pratique, le choix des règles de groupement de haut niveau dépend essentiellement du type de scène observée ou de l'application recherchée. Ces règles ont toutes en commun de produire des structures dont la probabilité d'apparition accidentelle est particulièrement faible. Chacune de ces relations peut être directement appliquée aux hypothèses de segments, d'arcs et de points d'intérêt extraites à partir du réseau de saillance. Parmi les plus utilisées, on peut citer les relations suivantes.

- *Symétrie et parallélisme*

  Des groupes de segments localement parallèles [Ylä-Jääski et Ade, 1992] ou constituant des rubans symétriques [Cham et Cipolla, 1995] permettent souvent de prévoir un grand nombre de structures géométriques 2D. Par exemple, [Mohan et Nevatia, 1992] exploitent les configurations entre rubans pour retrouver des axes de symétrie et former des structures convexes telles que des quadrilatères. Ces résultats ont été étendus ensuite à des groupements courbes. Selon les mêmes principes, [Ip et Wong, 1997] utilisent des groupements de courbes parallèles afin de faciliter le suivi de routes sur des séquences vidéo.

- *Convergence et Proximité*

  A un niveau local, la relation de convergence et de proximité définit la présence de jonctions entre extrémités de segments ou de courbes. D'une manière plus globale, cette relation est utile pour la détection de points de fuite. Ceux-ci sont détectés en projetant les segments de la scène dans un espace de paramètres exprimés en fonction de l'orientation et de l'équation des droites porteuses de chaque segment. Les segments qui convergent vers un même point correspondent à des points similaires dans cet espace [Straforini *et al.*, 1993] [Tai *et al.*, 1993]

- *Convexité et cycles*

  La détection de groupements circulaires entre éléments de représentation est également une structure hiérarchique importante pour la mise en correspondance. A plus forte raison lorsque ces arrangements circulaires peuvent être identifiés à l'aide de modèles tels que des formes cycliques simples (quadrilatères, cercles, ellipses) ou bien à l'aide de propriétés particulières telles que la convexité.

  [Jacobs, 1996] propose d'évaluer directement une mesure de convexité et de fermeture en termes de probabilité d'apparition par accident. Il démontre en particulier que le choix de bons critères de groupements permet de réduire la complexité d'une recherche quasi-exhaustive, précisément parce que les structures recherchées ont peu de chances de remplir ces critères par accident.

Pour chaque exemple, des méthodes spécialisées permettent d'extraire efficacement chaque type de groupement. Il en est ainsi de la détection de points de fuite ou de la recherche de cycles et d'ensembles convexes par des méthodes de parcours de graphes.

Il existe également une approche plus générale, qui consiste à prédire des hypothèses à partir de configurations locales entre primitives et de vérifier ensuite ces hypothèses à partir des primitives restantes [Denasi *et al.*, 1992] . Des groupements de haut niveau sont ainsi déduits de manière hiérarchique à partir de groupements plus simples. Par exemple, une succession d'arcs élémentaires co-circulaires constitue une bonne hypothèse pour la présence de cercles ou d'ellipses, qu'il suffit de vérifier ensuite par application d'un modèle.

Cette approche n'utilise pas de connaissance du type de scène observé. La seule connaissance utilisée porte sur les critères d'apparition d'une relation structurelle à partir de groupes plus simples. [Sarkar et Boyer, 1993a] proposent une méthode générique pour modéliser ce type de connaissance et l'exploiter afin de déduire automatiquement l'existence de structures plus complexes.

Leur démarche est divisée en trois parties. A partir de portions de contours de courbure continue, une première partie groupe des fragments compatibles en fonction de règles simples. Pour chaque type de groupement recherché, l'espace des paramètres est discrétisé de manière à ce que chaque segment vote pour les points de cet espace satisfaisant cette relation. Ils constituent ainsi ce qu'ils nomment des "Gestalt graphs" reflétant des propriétés de proximité, continuité, fermeture et région commune. On pourra se reporter à [Sarkar et Boyer, 1992] pour une comparaison plus détaillée de cette méthode avec la transformée de Hough.

La deuxième étape combine les graphes entre eux et utilise des techniques de parcours de graphe et recherche de cliques pour établir des hypothèses. Par exemple, les graphes de "région communes" ET de "proximité" donnent des segments parallèles. Ou encore, la recherche de cycles sur les graphes de jonctions donne des polygones.

Finalement, des hypothèses plus complexes sont déduites dans une troisième partie à l'aide d'un réseau d'inférence Bayesien (*PIN - Perceptual Inference Network*). Ce réseau modélise une connaissance *a priori* de règles de groupement sous la forme d'un graphe de probabilité. La probabilité de détection de rubans, cercles ou rectangles dépend alors des probabilités d'apparition de structures génératrices de ces hypothèses [Sarkar et Boyer, 1994] . Cette démarche constitue l'une des rares tentatives d'approche globale du groupement perceptuel à l'aide d'un seul formalisme.

6.1.2 Organisation perceptuelle et mise en correspondance

Les différentes applications du groupement perceptuel à la mise en correspondance structurelle partagent deux principes.

D'une part, la formation de groupement complexes permet un appariement sommaire de structures globales, qui sert ensuite de centre d'attention pour une mise en correspondance plus fine. [Mohan et Nevatia, 1989] démontrent l'intérêt de cette ap-

proche pour l'extraction de structures tridimensionnelles, en particulier pour faciliter la détection de bâtiments en imagerie aérienne. Des paires de segments parallèles sont d'abord extraites à partir de la détection de contours. Des hypothèses intermédiaires de structures en U sont ensuite élaborées à partir de ces paires et complétées sous forme de rectangles. Dans un premier temps, une correspondance sommaire est établie entre les rectangles correspondant aux toits des bâtiments. Les arêtes de chaque rectangle sont ensuite appariées de façon plus précise. Les rectangles donnent ainsi une contrainte forte sur les candidats possibles pour la mise en correspondance précise des contours des toits.

D'autre part, la mise en correspondance est souvent considérée comme un groupement perceptuel particulier. On retrouve en effet dans l'opération d'appariement l'idée d'association entre structures présentant des mouvements similaires. C'est pourquoi ces deux opérations sont souvent présentées avec le même formalisme. [Horaud et Skordas, 1989] ou encore [Sarkar, 1994] représentent les groupements effectués au sein d'une même image sous la forme de graphes relationnels entre primitives. Les appariements possibles sont présentés de la même manière sous la forme de graphes de correspondances. Ce second graphe représente les appariements possibles pour chaque segment de l'image de départ. L'appariement revient ainsi à parcourir ce graphe pour extraire les ensembles de noeuds mutuellement compatibles. Sarkar et Boyer poussent plus loin ce parallèle en étudiant des séquences d'images par superposition de groupements 2D issus d'images consécutives. Cette représentation composite permet d'utiliser des techniques de groupement perceptuel afin de reconstituer les trajectoires des structures de la scène. La mise en correspondance n'est ici rien d'autre qu'un groupement "temporel" entre structures similaires.

Enfin, les approches hiérarchiques élaborées par [Havaldar *et al.*, 1996] ou bien [Venkateswar et Chellappa, 1995] sont autant d'exemples de synthèse de ces deux principes. Dans ce dernier cas, par exemple, les primitives 2D sont organisées selon une hiérarchie de structures de complexité croissante : "segments", "coins", "arêtes" (groupement de segments colinéaires entre deux coins) et enfin "facettes" (succession de contours et de coins consécutifs). La mise en correspondance de structures de haut niveau apporte des contraintes de localisation fortes sur les structures de niveaux inférieurs. Les opérations de groupement et d'appariement sont ici encore présentées selon le même formalisme de graphe.

Nous proposons d'utiliser les groupements définis par les chapitres précédents à la mise en correspondance de jonctions. Dans un premier temps, nous montrons comment détecter et grouper les jonctions à partir des points d'intérêt et des segments. Ces jonctions sont ensuite appariées à l'aide d'une méthode de relaxation stochastique reprenant les deux principes que nous venons d'évoquer.

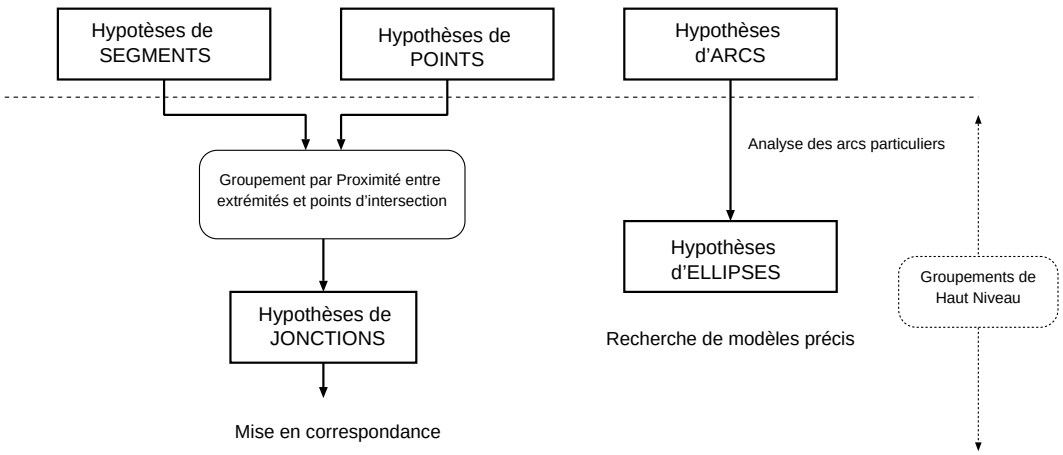

Figure 6.1 - *Principes des niveaux supérieurs de groupement. Les éléments visuels extraits par les niveaux inférieurs sont soit manipulés directement sous la forme de structures plus complexes (mise en correspondance structurelle), soit utilisés comme centre d'attention pour valider des hypothèses de manière plus précise (prédiction et vérification d'hypothèses).*

6.2 Extraction et groupement de jonctions

Les jonctions entre primitives géométriques correspondent en général à des sommets ou des occlusions entre arêtes des objets de la scène. Elles peuvent être envisagées de deux manières, soit comme une relation entre segments et/ou arcs, soit comme des entités à part entière, localisées à l'aide de détecteurs spécialisés.

Si les jonctions sont considérées comme des primitives particulières, leur extraction passe d'abord par une localisation de points d'intérêt. En tant qu'intersections entre extrémités de primitives linéaires, les jonctions correspondent en effet à des "coins" (deux branches) ou des "sommets" (trois branches et plus). Les jonctions proprement dites peuvent alors être définies à partir du modèle de coin détecté, ou à l'aide des extrémités de segments ou courbes présentes dans leur voisinage [Lindeberg et Li, 1997] .

Si au contraire, elles sont considérées comme des relations entre segments ou courbes (relation de connectivité par exemple), les jonctions peuvent être extraites à partir de l'ensemble des intersections possibles entre primitives linéaires, moyennant une certaine zone de recherche autour des extrémités de ces primitives. L'extraction de jonctions cohérentes peut alors se rapporter à un problème d'étiquetage et peut être résolu par un procédé de relaxation sur les probabilités de connexions entre extrémités [Regier, 1991] .

Nous conservons, pour l'extraction des jonctions, des principes comparables à ceux appliqués pour le groupement de segments et d'arcs. Dans un premier temps,

des jonctions élémentaires sont détectées à partir des intersections entre hypothèses issues des groupements précédents. Elles sont ensuite groupées entre elles afin d'éliminer les jonctions redondantes et constituer des jonctions multiples. A la différence des groupements précédents, les jonctions sont détectées à partir des primitives géométriques et non plus directement à partir des chaînes sélectionnées par le réseau de saillance.

6.2.1 Détection des jonctions élémentaires

Les hypothèses de jonctions élémentaires sont établies à partir de paires de segments non colinéaires répondant à un certain critère de voisinage. A l'image des points dominants détectés lors du groupement d'arcs élémentaires, chaque jonction est associée aux segments qui la génèrent afin de former un triplet $(I_{1,2}, S_1, S_2)$.

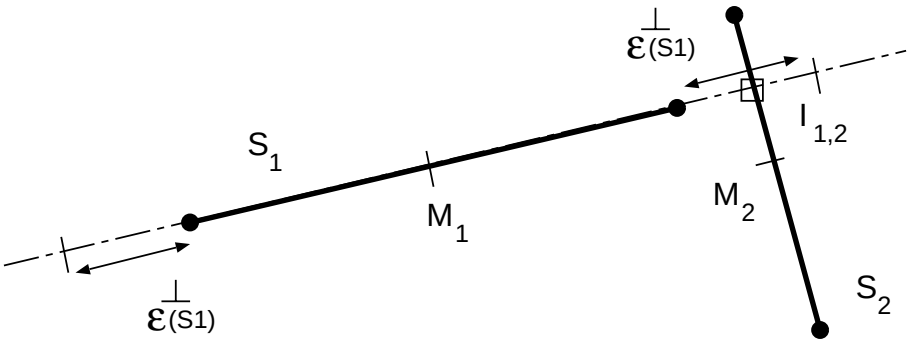

Figure 6.2 - *Notations utilisées pour une intersection entre deux segments. Une marge d'erreur permet de définir des jonctions "réelles" et "virtuelles". Ici, la jonction entre S_1 et S_2 est virtuelle.*

Soient S_1 et S_2 deux segments non colinéaires et $I_{1,2}$ leur intersection. On note de plus M_1 et M_2 les milieux respectifs de chaque segment.

Malgré l'ajustement des extrémités des segments réalisé en fin de groupements, la détection d'intersections entre segments doit tenir compte d'une marge d'erreur liée à la longueur de chaque segment. Cette distance autour des extrémités permet de rattraper la mauvaise localisation des coins lors de l'extraction des contours.

Notons $\epsilon^{\perp}(S_1)$ la marge d'erreur du segment S_1. En pratique, cet écart est défini par :

$$\epsilon^{\perp}(S_1) \;=\; 0.3 \cdot \|\overrightarrow{S_1}\|$$

Cette définition entraîne la détection de trois types de jonctions, résumées par la figure 6.3.

– *Jonctions réelles*

Elles sont caractérisées par les relations suivantes :

$$\left.\begin{array}{rcl} \|\overrightarrow{M_1, I_{1,2}}\| & < & \frac{\|\overrightarrow{S_1}\|}{2} \\ et & & \\ \|\overrightarrow{M_2, I_{1,2}}\| & < & \frac{\|\overrightarrow{S_2}\|}{2} \end{array}\right\} \Longrightarrow (I_{1,2}, S_1, S_2) \text{ Réelle} \qquad (6.1)$$

Ce sont les jonctions en L, T ou X selon les positions respectives du point d'intersection sur chaque segment.

– *Jonctions virtuelles*

Ces jonctions en V correspondent au cas inverse où le point d'intersection n'appartient à aucun des segments.

$$\left.\begin{array}{rcl} |\ \|\overrightarrow{M_1, I_{1,2}}\| - \frac{\|\overrightarrow{S_1}\|}{2}\ | & < & \epsilon^{\perp}(S_1) \\ et & & \\ |\ \|\overrightarrow{M_2, I_{1,2}}\| - \frac{\|\overrightarrow{S_2}\|}{2}\ | & < & \epsilon^{\perp}(S_2) \end{array}\right\} \Longrightarrow (I_{1,2}, S_1, S_2) \text{ Virtuelle} \quad (6.2)$$

– *Jonctions λ*

Cette dernière classe de jonction correspond au cas intermédiaire où le point d'intersection n'appartient qu'à l'un des deux segments.

La recherche des jonctions consiste simplement à comparer les segments deux à deux. Chaque intersection étant symétrique ($I_{1,2} = I_{2,1}$), cette recherche revient à comparer $\frac{n(n-1)}{2}$ segments, si n est le nombre de segments.

6.2.2 Groupement en jonctions complexes

Comme pour la détection de segments et d'arcs, les erreurs de localisation des extrémités de segments entraînent des redondances de jonctions. Les jonctions ainsi extraites sont superposées ou bien très proches les unes des autres, et nécessitent donc une étape de groupement. Le groupement de jonctions consiste donc à produire un ensemble simplifié de *n-uplets* $(I_1^j, S_1^1, \cdots, S_j^1)$ constitués d'un point central et d'un ensemble de segments ou "branches".

– Simplification des jonctions superposées

Les règles de groupement entre jonctions concernent à la fois leur position et leurs branches. Deux jonctions $(I_j^1, S_1^1, \cdots, S_j^1)$ et $(I_k^2, S_1^2, \cdots, S_k^2)$ sont fusionnées si leurs centres respectifs sont suffisamment proches. Si, de plus, elles ont

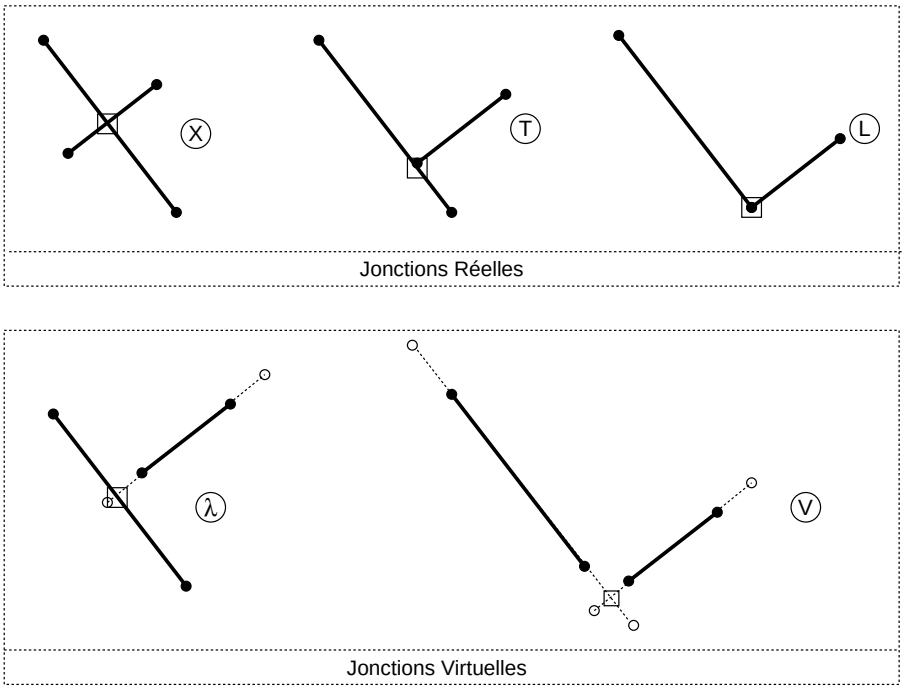

Figure 6.3 - *Catalogue des différentes classes de jonctions élémentaires entre deux segments.*

des branches en commun, celles-ci sont fusionnées à leur tour.

$$\begin{cases} \|\overrightarrow{I_j^1, I_k^2}\| & < \epsilon^{\vdash} \implies I_j^1 \equiv I_k^2 \\ \exists m < j, \exists n < k, \quad |\widehat{\overrightarrow{S_m^1}, \overrightarrow{S_n^2}}| & < \epsilon^{\theta} \implies S_m^1 \equiv S_m^2 \end{cases} \tag{6.3}$$

Le centre de la jonction issue du groupement est défini par le milieu du segment $\overline{I_j^1, I_k^2}$. Les branches fusionnées sont, quand à elles, remplacées par un segment de longueur :

$$L = Max(\|\overrightarrow{S_m^1}\|, \|\overrightarrow{S_n^2}\|)$$

et d'orientation :

$$\Theta = \frac{\widehat{S_m^1} + \widehat{S_m^2}}{2}$$

L'algorithme de fusion des branches entre deux jonctions consiste simplement à mettre en commun les branches des deux voisinages et à les trier par ordre croissant d'orientation. Il suffit ensuite de les comparer deux à deux et de fusionner les branches présentant un faible écart angulaire (inférieur à 5 degrés). Enfin, les écarts $\epsilon^{\vdash} = 5$ pixels et $\epsilon^{\theta} = 10$ degrés sont définis empiriquement.

– Mesure de saillance à partir des points d'intérêt

Afin de tenir compte des points d'intérêts détectés lors des groupements de segments et d'arcs, une mesure de saillance définie à partir de ces points est associée à chaque jonction. Cette mesure récompense les jonctions situées à proximité de coins ou d'extrema de courbures.

Soit une jonction I_j. On note $\{P_1, \cdots, P_n\}$ l'ensemble des points d'intérêt situés dans un voisinage de $\epsilon^{\vdash}$ pixels du centre de I_j. La mesure de saillance de la jonction est simplement définie par :

$$\mathcal{S}(I_j) \;=\; \sum_{i=1}^{n} e^{-\frac{d_i}{\sigma^{\vdash}}}, \qquad avec \;\; d_i \;=\; \|\overrightarrow{I_j, P_i}\| \;\; et \;\; \sigma^{\vdash} \;=\; \frac{\epsilon^{\vdash}}{2} \tag{6.4}$$

où $\sigma^{\vdash}$ est une constante qui permet d'ajuster l'étendue de l'exponentielle.

Ainsi, la saillance d'une jonction est d'autant plus forte que la localisation du centre est confirmée par des points d'intérêts détectés par ailleurs.

Le résultat de cette opération de groupement est un ensemble simplifié de jonctions, définissant une relation de proximité entre un point d'intérêt et un ensemble de segments. Les branches de ces jonctions sont une liste de segments disjoints. Leur point central est le barycentre des jonctions élémentaires impliquées dans le groupement, auquel est associé une mesure de saillance.

6.2.3 Résultats sur les jonctions de segments

Comme dans le cas des groupements de segments et d'arcs, nous commençons par présenter les résultats du groupement de jonctions dans une scène artificielle de test. Les figures 6.4 et 6.5 illustrent la stabilité en rotation du groupement de jonctions. Pour des raisons de clarté, nous ne représentons que les centres et les directions des branches de chaque jonction. Malgré des irrégularités importantes le long des arêtes des polygones, les jonctions sont correctement groupées. Les jonctions superflues présentes dans les deux images de la figure 6.5 donnent une idée des jonctions résiduelles qui échappent au groupement. Ce type de résidu ne remet pas en cause la détection des autres jonctions, plus stables. En effet, elles seront considérées par l'algorithme de mise en correspondance comme des éléments incohérents, et seront alors rejetées du voisinage des jonctions stables.

Les figures des pages 215 à 216 représentent une scène de bureau. La détection de contours et la sélection des groupements les plus saillants permettent de passer de 440 chaînes de contours à seulement 54 groupements, dont sont extraits les hypothèses de segments (Figure 6.7). Les figures suivantes montrent successivement la détection de 718 jonctions doubles (Figure 6.8) et la simplification de ces hypothèses en 229 jonctions groupées (Figure 6.9). Les jonctions virtuelles sont

représentées ici en noir. Bien qu'elles n'aient, dans la plupart des cas, aucune signification physique, ces jonctions virtuelles apportent néanmoins des informations sur les positions relatives de segments importants de la scène.

Notre méthode est une approche structurelle au problème de la détection de coins. Par comparaison, les méthodes de détection de coins évoquées dans le sous-chapitre 2.3.3 (page 59), abordent le problème à partir des propriétés photométriques de l'intensité lumineuse de l'image d'origine. Celle-ci pourraient être utilisées en complément au groupement de jonctions afin d'obtenir une localisation précise de leur centre. En particulier, le groupement de jonctions pourrait servir d'initialisation au détecteur de coins de [Blaszka et Deriche, 1994b] à l'aide de modèles déformables.

Enfin, les figures des pages 217 et 218 reprennent les scènes utilisées dans le chapitre précédent en présentant, dans chaque cas, la détection et le groupement des jonctions. Malgré une diminution significative du nombre de jonctions superposées, on peut cependant remarquer que les règles de groupement entre jonctions sont insuffisantes pour simplifier toutes les situations. C'est, pour l'instant, la principale limitation de cette méthode. L'un des prolongements immédiats est la définition de règles plus complètes, comme nous l'avons fait avec les groupements de segments. Ces règles pourraient s'inspirer de méthodes de groupement de jonctions par relaxation d'un critère de proximité [Matas et Kittler, 1993] ou encore d'une construction progressive d'un graphe de voisinage entre segments saillants de la scène [Jacot-Descombes et Pun, 1997] .

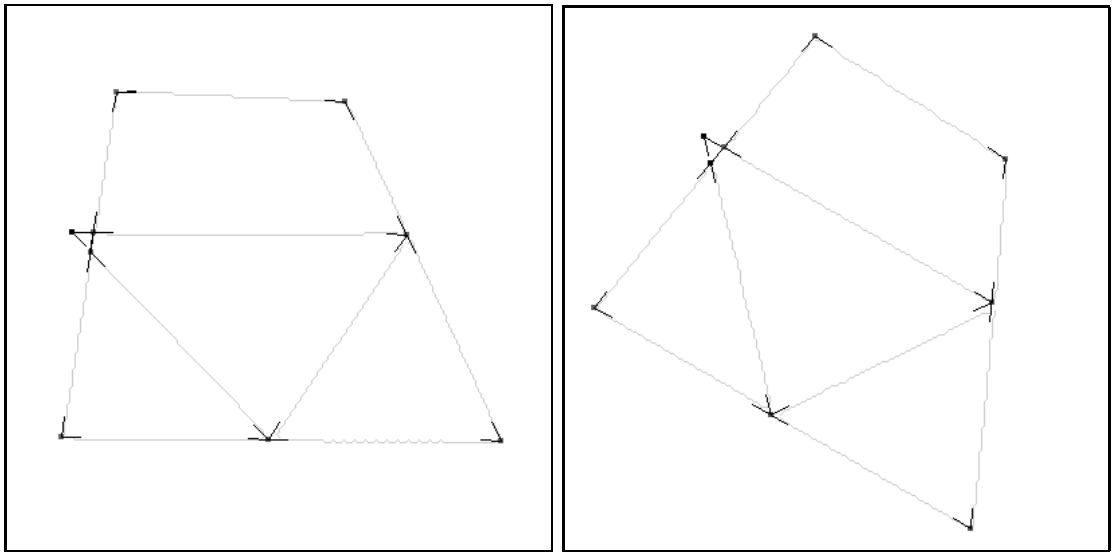

Figure 6.4 - *Stabilité du groupement de jonctions en rotation - $\theta = 0$ et $\theta = \frac{\Pi}{6}$*

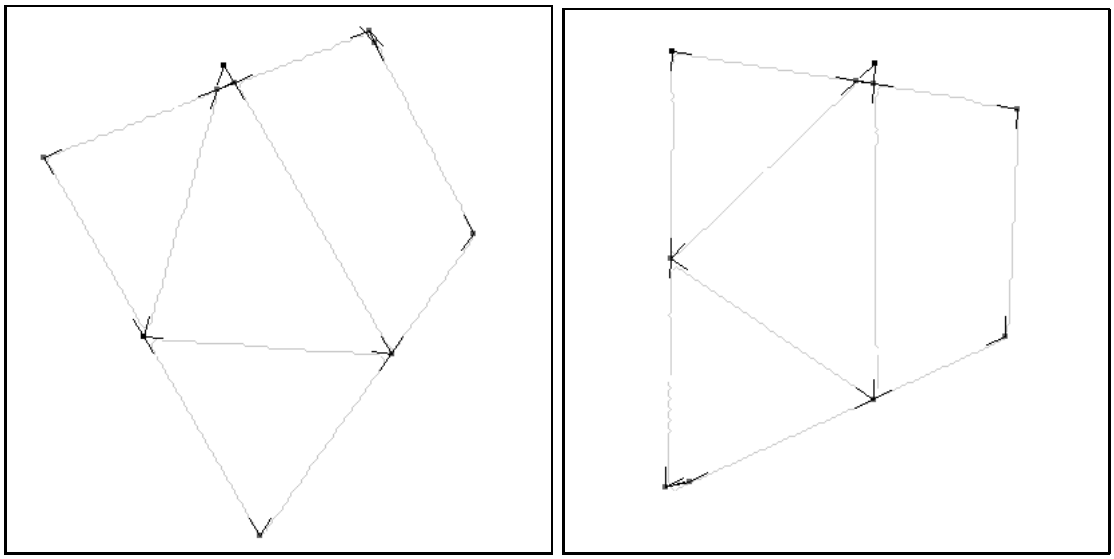

Figure 6.5 - *Stabilité du groupement de jonctions en rotation - $\theta = \frac{2\Pi}{6}$ et $\theta = \frac{\Pi}{2}$*

Figure 6.6 - *Scène de bureau*

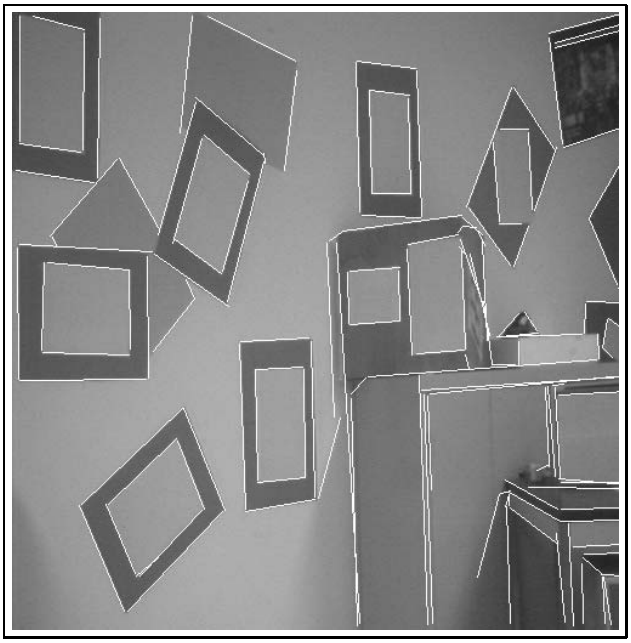

Figure 6.7 - *Scène de bureau - Détection et groupement de segments - 144 segments extraits à partir de 54 groupements sur 440 chaines (note : les discontinuités des segments en blanc sont dues à un défaut d'impression).*

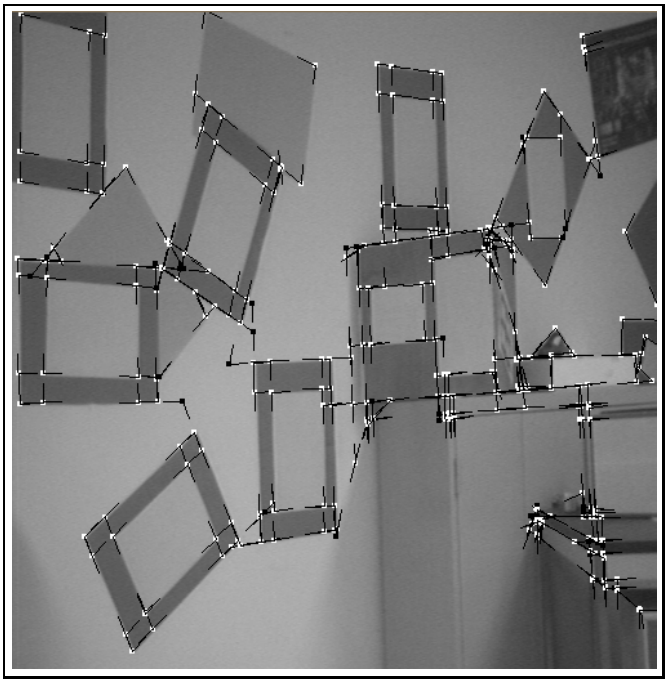

Figure 6.8 - *Scène de bureau - Détection de 718 jonctions doubles*

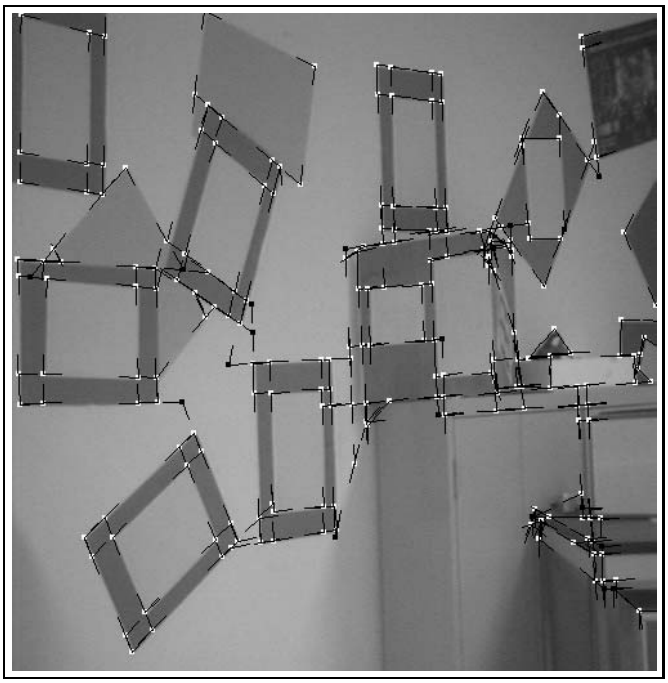

Figure 6.9 - *Scène de bureau - Groupement de jonctions - restent 229 jonctions groupées*

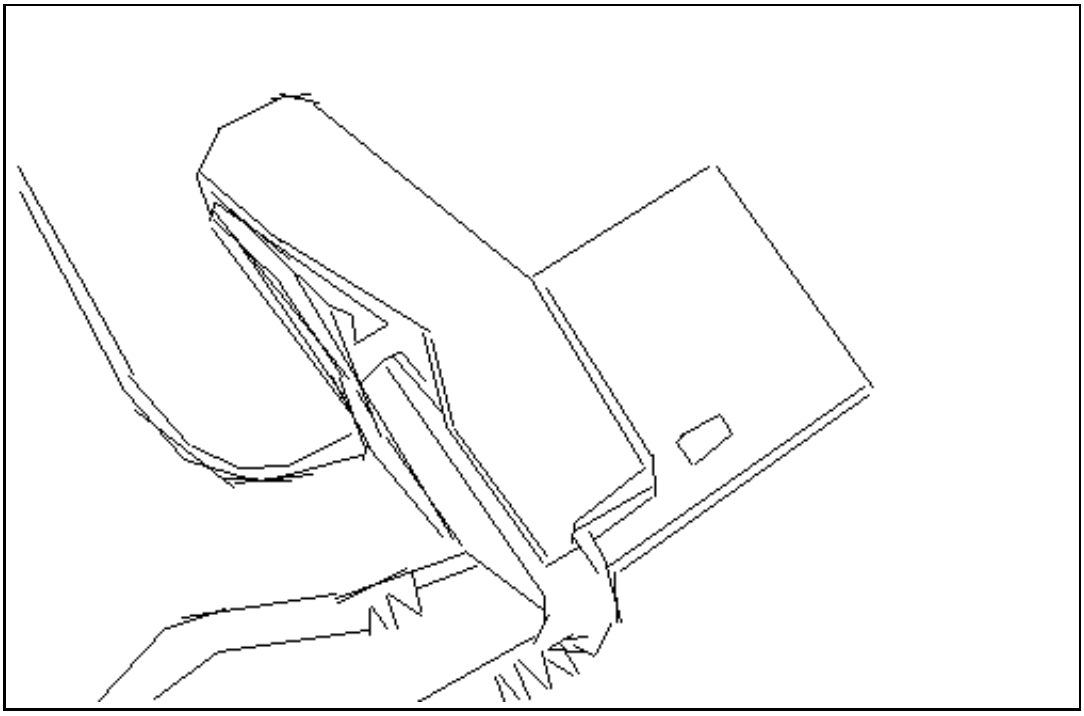

Figure 6.10 - *Téléphone - Détection et groupement de segments - 101 segments extraits à partir de 22 groupements sur 550 chaines*

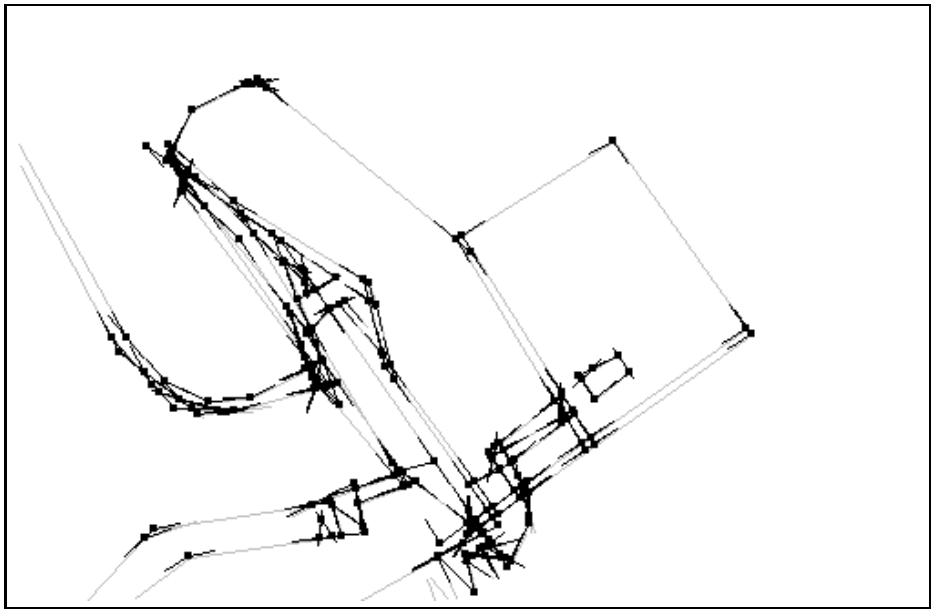

Figure 6.11 - *Téléphone - Détection de 482 jonctions doubles*

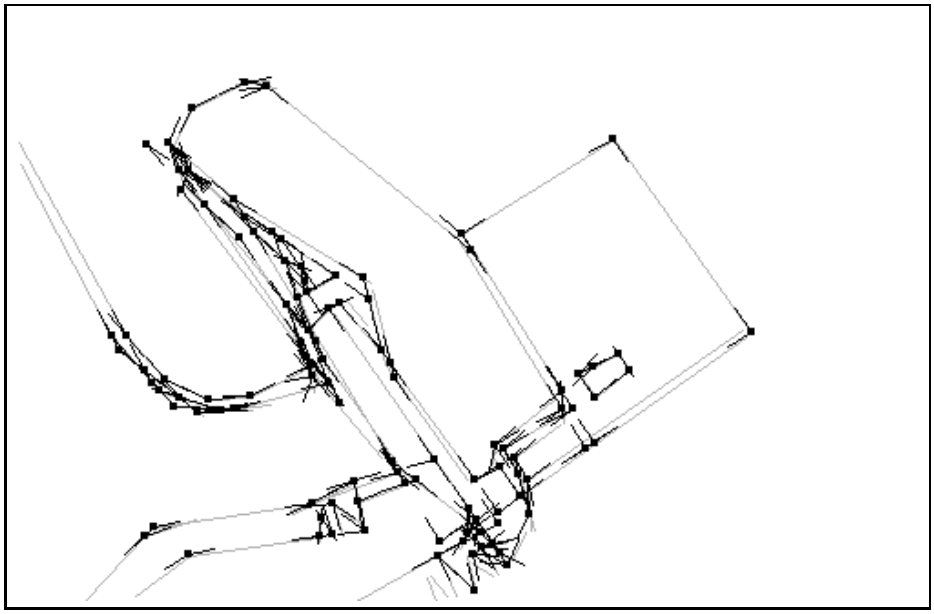

Figure 6.12 - *Téléphone - Groupement de jonctions - restent 101 jonctions groupées*

6.3 Mise en correspondance de jonctions

Nous nous inspirons, pour l'appariement de jonctions, d'un algorithme de coopération entre mise en correspondance et groupement perceptuel proposé par [Chang et Aggarwal, 1997] . Cette méthode tient compte en particulier des principes d'application du groupement perceptuel à la mise en correspondance. Originellement appliquée à la mise en correspondance de segments, elle suit une approche suffisamment générique pour être adaptée simplement à tout type de structure, quel que soit son niveau de hiérarchie ou de complexité. Elle peut donc être utilisée pour guider une mise en correspondance précise à partir d'appariements de structures complexes. Enfin, elle considère la mise en correspondance comme une étape de groupement temporel d'une image vers l'autre, partageant le même formalisme que le groupement perceptuel.

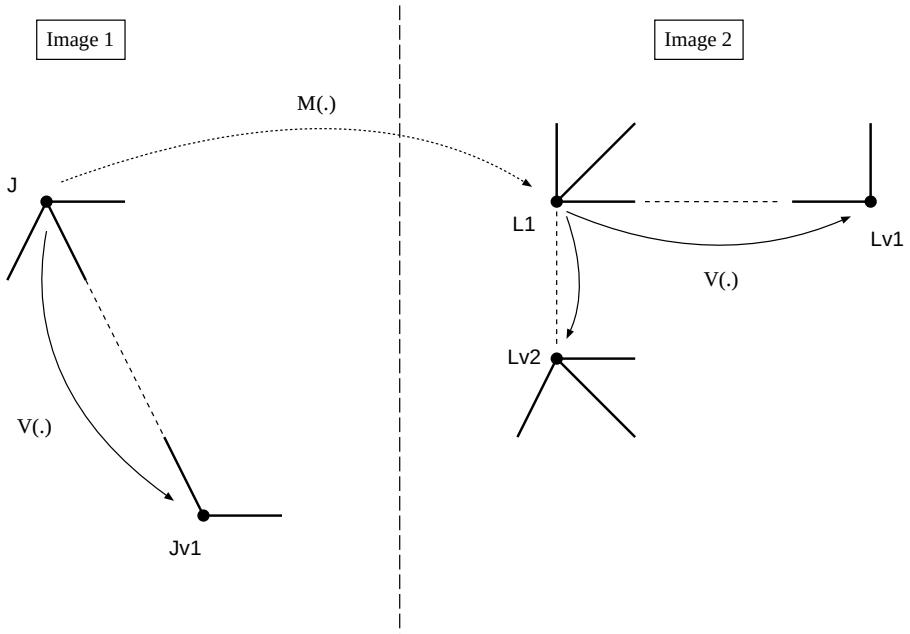

Figure 6.13 - *Voisinages Temporels et Perceptuels pour une jonction J.*

Cet algorithme part de l'hypothèse d'un mouvement rigide des objets d'une image à l'autre de la scène. Ainsi, les éléments d'un même groupement perceptuel établi dans une image doivent avoir le même mouvement dans l'autre image. Les relations de structures d'un groupement sont donc utilisées comme contraintes sur les appariements possibles.

En retour, la mise en correspondance doit également préserver les relations structurelles. Il est donc possible d'utiliser les hypothèses d'appariements de chaque membre d'un groupement perceptuel afin de rejeter ceux dont le mouvement est incohérent avec celui du groupe.

L'algorithme de mise en correspondance suit une procédure de relaxation stochastique afin d'apparier de manière cohérente chaque élément visuel, tour à tour avec ses correspondants possibles, et ses voisins immédiats.

Nous ne retenons de cet algorithme que ses principes de coopération entre deux types de contraintes pour l'appliquer directement à la mise en correspondance de jonctions. Celles-ci sont en effet moins nombreuses que des segments et surtout, moins ambigües lorsqu'il s'agit de les comparer.

6.3.1 Coopération entre appariement et groupement

Le problème de l'appariement de jonctions consiste donc à établir une correspondance entre un ensemble $\mathcal{J}_1$ de jonctions détectées dans une image I_1 et un autre ensemble de jonctions $\mathcal{J}_2$ extraites d'une image I_2.

La coopération entre mise en correspondance et groupement perceptuel suppose la définition de deux types de voisinages pour chaque jonction $J \in \mathcal{J}_1$.

- Un voisinage *temporel*, noté $\mathcal{M}(J)$, qui représente l'ensemble des candidats possibles dans $\mathcal{J}_2$ pour l'appariement avec J. Chaque élément de ce voisinage vérifie la relation :

$$\begin{array}{rccl} \mathcal{M} : & \mathcal{J}_1 & \longmapsto & \mathcal{J}_2 \\ & J & \longrightarrow & \mathcal{M}(J) = \{L_1, \cdots, L_k\}, \quad \text{correspondants possibles pour J} \end{array} \tag{6.5}$$

  Chaque élément L_i de $\mathcal{M}(J)$ définit une probabilité d'appariement avec la jonction J, notée : $S_M(J, L_i)$.

- Un voisinage *perceptuel*, noté $\mathcal{V}(J)$, qui représente une relation structurelle entre la jonction J et un certain nombre de jonctions de $\mathcal{J}_1$.

$$\begin{array}{rccl} \mathcal{V} : & \mathcal{J}_1 & \longmapsto & \mathcal{J}_1 \\ & J & \longrightarrow & \mathcal{V}(J) = \{Jv_1, \cdots, Jv_n\}, \quad \text{voisins de J pour la relation } \mathcal{V} \end{array} \tag{6.6}$$

  Le voisinage défini par $\mathcal{V}$ peut être construit, par exemple, à l'aide de l'une des relations définies à la section 6.1.1, ou encore la somme de relations de ce type. Dans le cas présent, le voisinage perceptuel d'une jonction $J \in \mathcal{J}_1$ est constitué des jonctions de $\mathcal{J}_1$ dont le centre est aligné avec l'une des branches de J, comme le montre la figure 6.14. On définit un voisinage similaire pour les jonctions appartenant à $\mathcal{J}_2$.

  Enfin, chaque élément Jv_i de $\mathcal{V}(J)$ définit également un score de groupement avec la jonction J, noté : $S_G(J, Jv_i)$. Contrairement aux probabilités d'appariement, ce score ne représente pas une probabilité mais plutôt un coefficient de compatibilité entre une jonction et ses voisins.

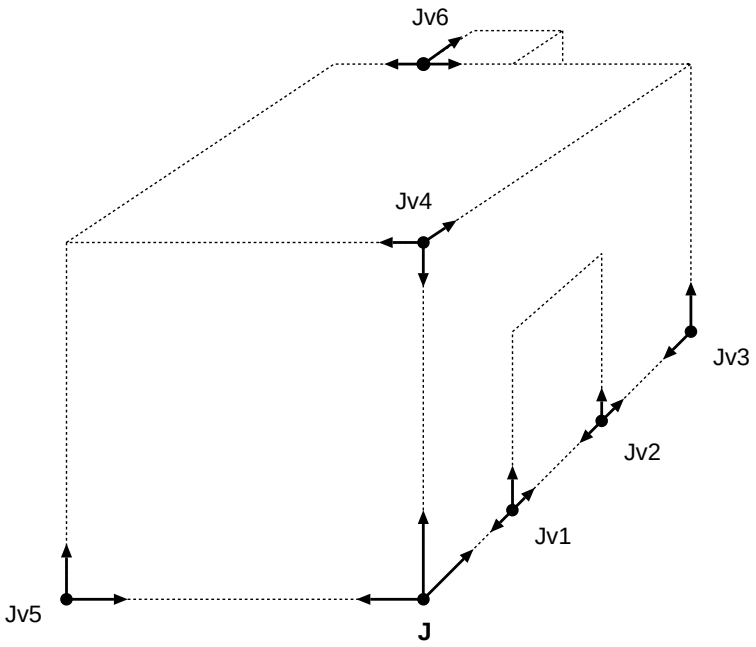

Figure 6.14 - *Le voisinage perceptuel d'une jonction J est constitué des jonctions dont le centre se trouve aligné avec l'une des branches de J.*

Algorithme 6.1 : Coopération entre Groupement Perceptuel et Mise en Correspondance - algorithme principal

début
  pour *Chaque jonction J de $\mathcal{J}_1$* **faire**
    Initialiser le voisinage perceptuel $\mathcal{V}(J)$
    Initialiser le voisinage temporel $\mathcal{M}(J)$
  pour *Chaque jonction L de $\mathcal{J}_2$* **faire**
    Initialiser le voisinage perceptuel $\mathcal{V}(L)$
  répéter
    % Relaxation temporelle - deux passes
    Mise à jour des probabilites d'appariement
      sur les voisinages temporels des jonctions de $\mathcal{J}_1$
    %
    % Relaxation perceptuelle
    Mise à jour des scores de groupement
      sur les voisinages perceptuels des jonctions de $\mathcal{J}_1$
    %
    pour *Chaque jonction J de $\mathcal{J}_1$* **faire**
      Eliminer les voisins de $\mathcal{V}(J)$ et $\mathcal{M}(J)$ dont le score est trop faible
  jusqu'à *stabilité des voisinages temporels des jonctions de $\mathcal{J}_1$*
fin

L'algorithme proprement dit se décompose en deux étapes. Dans un premier temps, un processus de relaxation stochastique est appliqué au voisinage temporel de chaque jonction J afin d'attribuer à chaque hypothèse une probabilité d'appariement. Cette étape compare les voisinages perceptuels de J et de ses correspondants afin de renforcer itérativement les hypothèses dont le voisinage est cohérent avec celui de J.

Les probabilités d'appariements sont ensuite utilisées dans la seconde étape de relaxation stochastique afin de renforcer, pour chaque jonction J, les voisins dont les hypothèses d'appariement sont compatibles avec celles de J.

Cette seconde relaxation permet de ne conserver que les éléments du voisinage perceptuel de J dont le mouvement est cohérent avec celui de J. Les voisins incohérents correspondent en pratique à des groupements accidentels comme ce peut être le cas pour des jonctions redondantes, et sont finalement retirés du voisinage de la jonction. Ainsi, lors de la phase suivante de relaxation "temporelle" seuls les voisins dont le mouvement est cohérent contribueront à renforcer les hypothèses d'appariements, comme le montre la figure 6.15.

Si une jonction d'un objet statique est initialement appariée avec une jonction d'un objet en mouvement à cause de la ressemblance accidentelle de leurs voisinages respectifs, cette hypothèse sera invalidée lorsqu'il s'agira d'estimer la cohérence des mouvements des voisins de cette jonction en fonction de cette hypothèse.

Cette alternance de deux types de relaxation afin de renforcer mutuellement chaque étiquetage permet d'accélérer la convergence et rend la mise en correspondance plus robuste en interdisant à des voisins incohérents de renforcer les hypothèses d'appariement.

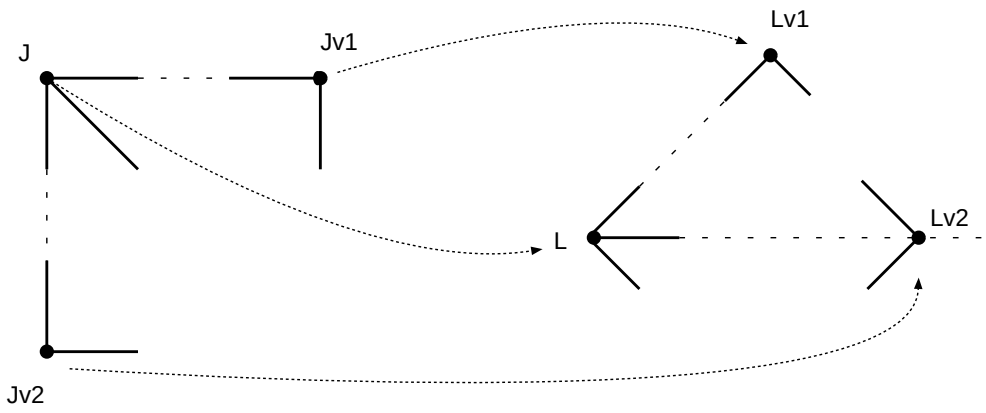

Figure 6.15 - *Incohérence dans les voisinages perceptuels d'une jonction J et d'un correspondant possible L. Dans cette situation, le déplacement de Lv_1 par rapport à L est cohérent avec celui de Jv_1 et de J. La jonction Lv_2 se comporte de manière incohérente - elle doit donc être retirée du voisinage de L.*

En pratique, chaque itération est composée de deux mises à jour des probabilités d'appariement pour une mise à jour des scores de groupement. Les voisinages perceptuels sont en effet jugés plus fiables que les hypothèses d'appariement car ils proviennent de l'environnement direct de chaque jonction. Enfin, toutes les deux itérations, les voisinages temporels et perceptuels sont étudiés afin d'éliminer les hypothèses dont la probabilité est trop faible. Cette dernière étape permet de réduire graduellement la complexité algorithmique des comparaisons entre voisinages.

Nous abordons à présent le détail des algorithmes de mise à jour des scores de groupement et des probabilités d'appariement, en commençant par les différentes mesures utilisées pour comparer les configurations entre jonctions.

6.3.2 Mesures de distances entre jonctions

Deux types de mesures sont utiles à la mise en correspondance de jonctions. D'une part, il est nécessaire d'évaluer le déplacement entre une jonction et les éléments de ses voisinages temporels et perceptuels. D'autre part, une mesure de similarité doit permettre, dans chaque cas, de comparer les déplacements de deux couples de jonctions. Enfin, chacune de ces mesures suppose la comparaison préalable d'une jonction avec une autre.

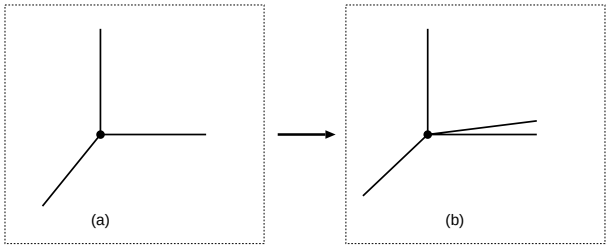

Figure 6.16 - *Exemple de différences de groupement d'une même jonction dans deux images différentes. La mesure de similarité doit être suffisamment tolérante pour accepter ce genre de distorsion.*

– Coefficient de similarité entre deux jonctions - $d_s(J_1, J_2)$

Comparer deux jonctions consiste à déterminer dans quelle mesure leurs branches sont superposables. Le coefficient de similarité doit tenir compte, en particulier, du nombre de branches communes aux deux jonctions.

Comme le montre la figure 6.16, le résultat du groupement d'une même jonction peut être légèrement différent entre deux images. L'apparition, ou la disparition, de branches d'une jonction à l'autre est généralement due à des configurations entre branches à la limite des critères de groupement. La mesure

de similarité doit être suffisamment tolérante pour accepter de légères variations entre jonctions. Son principe consiste à évaluer le nombre de branches *consécutives* communes aux deux jonctions.

Chacune des branches est un segment caractérisé par sa longueur L et son orientation λ, exprimée par rapport à une direction de référence. Comme le montre la figure 6.17, cette direction de référence est, initialement, l'axe horizontal du repère de l'image. Afin de pouvoir prétendre à superposer les deux jonctions, il est nécessaire de les "aligner" de manière à faciliter la comparaison des branches. Une branche de chaque jonction est arbitrairement choisie comme référence par rapport à laquelle sont calculés les angles des branches restantes, noté Θ.

On désigne par "zone de comparaison" l'écart angulaire maximum couvert entre la branche de référence et les branches restantes. Afin de faciliter la comparaison, les branches des deux jonctions sont triées par ordre croissant de leur angle Θ.

Soit $\mathcal{B}_1 = \{S_1^1, \cdots, S_n^1\}$ et $\mathcal{B}_2 = \{S_1^2, \cdots, S_M^2\}$ les ensembles de branches de J_1 et J_2, triées par rapport à leur angle. Les branches de référence sont ici S_1^1 et S_1^2.

L'algorithme de comparaison des deux jonctions consiste, pour chaque branche S_i^1 de J_1, à rechercher la branche S_j^2 de J_2 qui maximise la similarité entre les angles et longueurs de chaque branche.

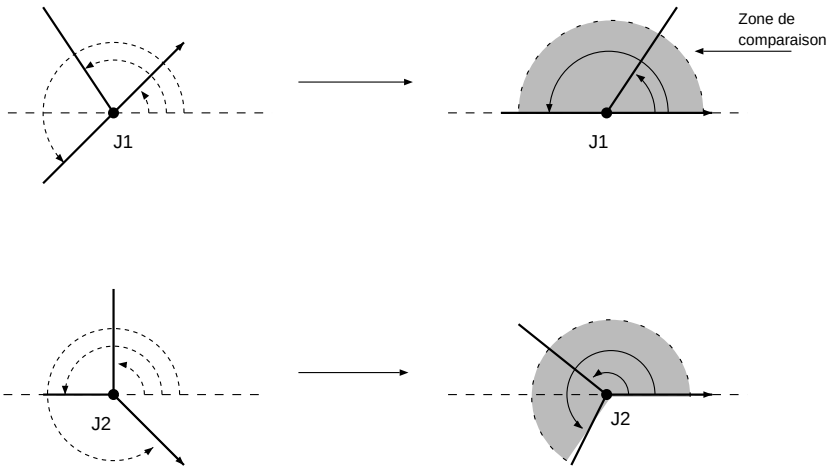

Figure 6.17 - *Alignement de deux jonctions par rapport à une direction de référence commune. Les jonctions sont initialement exprimées par rapport à l'axe horizontal (repère de l'image). La partie en gris signale la zone de comparaison entre les deux jonctions.*

La similarité entre deux branches est déterminée de la manière suivante :

$$\sigma_{i,j} \;=\; \frac{\sigma_\Theta(S_i^1, S_j^2) + \sigma_L(S_i^1, S_j^2)}{2.0}$$

avec :

$$\sigma_\Theta(S_i^1, S_j^2) \;=\; \begin{cases} 1 & si \;\; \Theta_i^1 \;=\; \Theta_j^2 \\ f_\alpha(\Theta_i^1, \Theta_j^2) & sinon \end{cases} \tag{6.7}$$

et :

$$\sigma_L(S_i^1, S_j^2) \;=\; \begin{cases} 1 & si \;\; L_i^1 \;=\; L_j^2 \\ f_\alpha(L_i^1, L_j^2) & sinon \end{cases} \tag{6.8}$$

La fonction de similarité $f(x, y)$ utilisée est la suivante :

$$f_\alpha(x, y) \;=\; \exp(-\alpha \cdot (1.0 - \frac{\mathbf{Min}(x, y)}{\mathbf{Max}(x, y)})^2)$$

Cette fonction permet de ramener l'écart entre deux valeurs x et y à une mesure qui vaut 1 si x et y sont similaires et qui tend vers $e^{-\alpha}$ sinon. Le choix de α permet de régler la vitesse de l'atténuation. Sa valeur est fixée en pratique à $\alpha = 2.0$ pour les deux similarités.

La comparaison se poursuit ensuite entre la branche S_{i+1}^1 et les branches restantes $\{S_{j+1}^2, \cdots, S_M^2\}$ jusqu'à ce qu'il ne reste plus de branche à comparer.

On note alors $in_{1,2}$ le nombre de branches similaires pour les deux jonctions, out_2 le nombre de branches de J_2 restées à l'intérieur de la zone de comparaison et qui n'ont pas trouvé de correspondant dans J_1, et enfin, out_1 le nombre de branches de J_1 qui n'ont pas pu être comparées avec celles de J_2.

La mesure de similarité finale est la somme normalisée des similarités des branches communes aux deux jonctions, pondérée par le nombre de branches délaissées par la comparaison.

$$\sigma(J_1, J_2) \;=\; (\frac{\sum_{similaires} \sigma_{i,j}}{in_{1,2}}) \cdot e^{-out_2} \cdot e^{-out_1}$$

L'idée principale pour cette mesure est d'accepter l'existence de branches superflues entre deux jonctions à comparer tant que ces branches se trouvent hors de la zone de comparaison. De plus, cette mesure établit une relation d'ordre entre jonctions en encourageant les comparaisons entre "petites" et "grandes" jonctions (par rapport au nombre de branches).

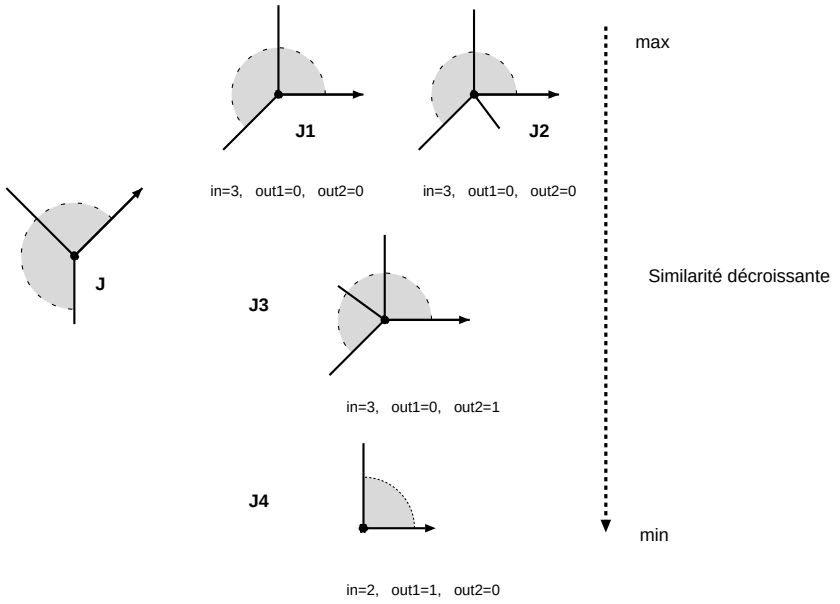

Figure 6.18 - *Similarité entre une jonction J et quatre candidats J_1 à J_4. Les directions de référence sont les branches marquées d'une flèche. Dans chaque cas, la zone de comparaison est signalée en gris. Les jonctions sont classées par ordre décroissant de similarité.*

Les exemples de la figure 6.18 montrent l'influence des coefficients e^{-out_1} et e^{-out_2}. Les similarités $\sigma(J, J_1)$ et $\sigma(J, J_2)$ sont ici identiques. La branche supplémentaire de J_2 ne gène en rien la superposition de J_2 par J. Elle n'est donc pas prise en compte dans le calcul de similarité.

Par contre, la jonction J_3 pourrait être superposable avec J mais sa branche superflue se trouve dans la zone de comparaison et gène donc la superposition. La mesure de similarité $\sigma(J, J_3)$ est ainsi plus faible.

Enfin, la jonction J_4 n'a pas assez de branches pour être correctement comparée à J. La mesure de similarité $\sigma(J, J_4)$ est ici inférieure à $\sigma(J_4, J)$. Cet exemple illustre la notion d'ordre imposée sur la comparaison de jonctions.

Cette mesure de similarité n'est définie qu'en fonction d'une direction de référence pour chaque jonction. L'algorithme complet de comparaison consiste finalement à prendre successivement comme référence chaque branche de chaque jonction et de conserver la configuration pour laquelle la similarité est maximale.

– Déplacement entre une jonction et un voisin perceptuel - $d_G(J, Jv)$

Cette mesure permet, lors de la relaxation, de comparer les configurations de couples de jonctions. Une configuration entre deux jonctions est définie selon les même principes que la comparaison de jonctions.

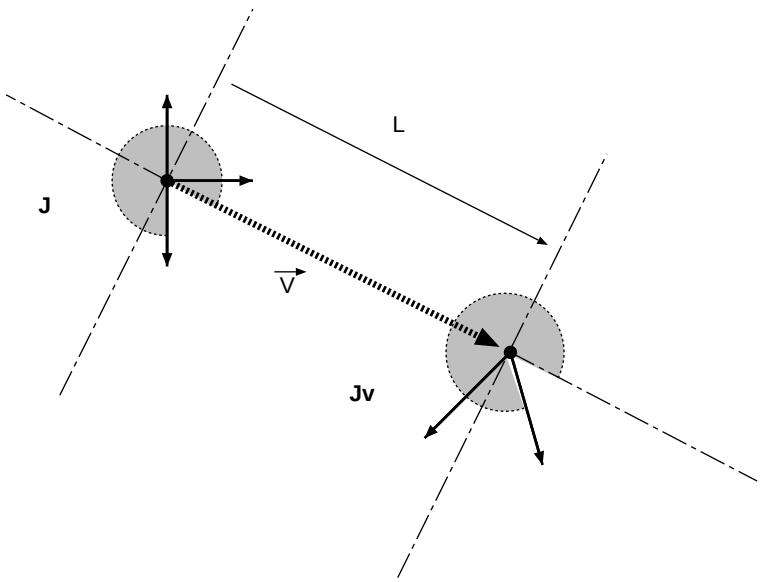

Figure 6.19 - *Configuration entre deux jonctions J_1 et J_2. Chaque jonction est alignée avec le vecteur $J_1 \rightarrow J_2$. Les zones de comparaison sont ici encore signalées en gris.*

Soient deux jonctions $J \in \mathcal{J}_1$ et $Jv \in \mathcal{J}_1$. La relation de groupement qui lie ces deux jonctions est notée $d_G(J, Jv)$. Cette relation est associée au score de groupement $S_G(J, Jv)$.

Comme le montre la figure 6.19, les angles des branches de chaque jonction sont exprimés par rapport au vecteur $\overrightarrow{V} = \overrightarrow{J, Jv}$. Afin de tenir compte des proportions de la configuration, les longueurs des branches de chaque jonction sont rapportées à la distance $L = \|\overrightarrow{J, Jv}\|$.

Exprimés ainsi, les angles et les longueurs des branches sont autant de paramètres propres au couple de jonctions. Les valeurs de ces paramètres sont de plus invariants par transformation du couple de jonctions en rotation, translation et changement d'échelle.

Soient (J_1, Jv_1) et (J_2, Jv_2) deux couples de jonctions. La similarité entre ces deux couples est alors simplement définie par le produit des similarités des jonctions prises deux à deux.

$$\sigma_G((J_1, Jv_1), (J_2, Jv_2)) \;=\; \sigma(J_1, J_2) \cdot \sigma(Jv_1, Jv_2)$$

Ce produit est comparable à une probabilité conditionnelle. Il représente la compatibilité entre la jonction J_1 et J_2 en supposant que Jv_1 correspond à Jv_2. Sa valeur est maximale lorsque les jonctions sont effectivement similaires deux à deux.

– Déplacement entre une jonction et un voisin temporel - $d_M(J, L)$

On désigne par $d_M(J, L)$ la transformation d'une jonction $J \in \mathcal{J}_1$ en son correspondant $L \in \mathcal{J}_2$. Cette relation est associée à la probabilité d'appariement $S_M(J, L)$.

La complexité de la transformation $d_M(\cdot, \cdot)$ est liée au choix d'hypothèses sur les conditions de mise en correspondance.

Le cas le plus simple correspond à une vision stéréoscopique, avec un faible écart angulaire entre les points de vues et des objets relativement éloignés de l'observateur. La seule transformation envisageable est alors réduite à une simple translation.

Dans un cas plus général où la scène est toujours éloignée de l'observateur mais où les objets peuvent éventuellement bouger librement, la transformation d_M est la composition d'une translation, d'une rotation et éventuellement d'un changement d'échelle. Il faudrait alors rechercher les paramètres de la plus petite transformation permettant de changer les branches de J en branches de L.

Enfin, le cas le plus général est celui d'une transformation projective. Dans ce dernier cas, les angles ne sont plus garantis et la mesure de similarité entre jonctions devrait être redéfinie.

Dans le cadre de notre application, nous nous plaçons dans l'hypothèse d'une paire d'images stéréoscopiques sans mouvement de rotation. Dans ces conditions, résumées par la figure 6.20, la comparaison entre deux appariements possibles est définie de la manière suivante.

Soit $\overrightarrow{V}_{J,L}(\ell_V, \Theta_V)$ la translation qui transforme J en L, et $\overrightarrow{U}_{Jv_1,Lv_1}(\ell_U, \Theta_U)$ qui transforme Jv_1 en Lv_1. La compatibilité entre les deux appariements est donnée par :

$$\sigma_M((J, L), (Jv_1, Lv_1)) \;=\; \frac{1}{3}(f_\alpha(\ell_V, \ell_U) + f_\alpha(\Theta_V, \Theta_U) + \sigma_G((J, Jv_1), (L, Lv_1)))$$

Comme dans le cas précédent, cette mesure est comparable à la probabilité conditionnelle que J soit apparié avec L en supposant que Jv_1 le soit avec Lv_1. On retrouve ici le même formalisme entre groupement perceptuel (mesure de similarité σ_G) et mise en correspondance (mesure de similarité σ_M).

6.3.3 Relaxation "temporelle"

Le principe de la relaxation "temporelle" consiste à renforcer la probabilité d'un appariement entre deux jonctions $J \in \mathcal{J}_1$ et $L \in \mathcal{J}_2$ si celui-ci est cohérent avec les appariements des voisins de J. La probabilité d'appariement entre J et L est renforcée par les probabilités d'appariement des voisins de J dont le déplacement est similaire à $d_M(J,L)$.

Soient $Jv_i \in \mathcal{V}(J)$ un élément du voisinage perceptuel de J et $L_j \in \mathcal{M}(Jv_i)$ l'un des correspondants de Jv_i.

La mise à jour des probabilités d'appariement est effectuée de la manière suivante :

$$S_M(J,L) \;=\; S_M(J,L) \cdot (1.0 + \sum_{(Jv_i,L_j)} \{S_M(Jv_i,L_j) \;\; / \;\; d_M(J,L) \equiv d_M(Jv_i,L_j)\})$$

Ainsi, chaque voisin Jv_i de J apporte une contribution à la probabilité de l'appariement entre J et L si et seulement si il existe un appariement semblable parmi les voisins temporels de Jv_i.

Le voisinage temporel de chaque jonction $J \in \mathcal{J}_1$ contient initialement toutes les jonctions de $\mathcal{J}_2$. Les appariements initiaux entre jonctions sont initialisés par la mesure de similarité entre jonctions notée $d_s(J,L)$.

$$\forall J \in \mathcal{J}_1, \;\; \forall L \in \mathcal{J}_2, \;\; \longrightarrow \;\; L \in \mathcal{V}(J) \;\; et \;\; S_M(J,L) \;=\; d_s(J,L)$$

Après avoir reçu les contributions des voisinages des jonctions, les scores d'appariement sont normalisés de manière à ce que leur somme soit égale à 1 pour une jonction donnée.

Le résultat de cet algorithme est, pour chaque jonction $J \in \mathcal{J}_1$, un ensemble de correspondants les plus probables et une probabilité d'appariement pour chacun de ces correspondants.

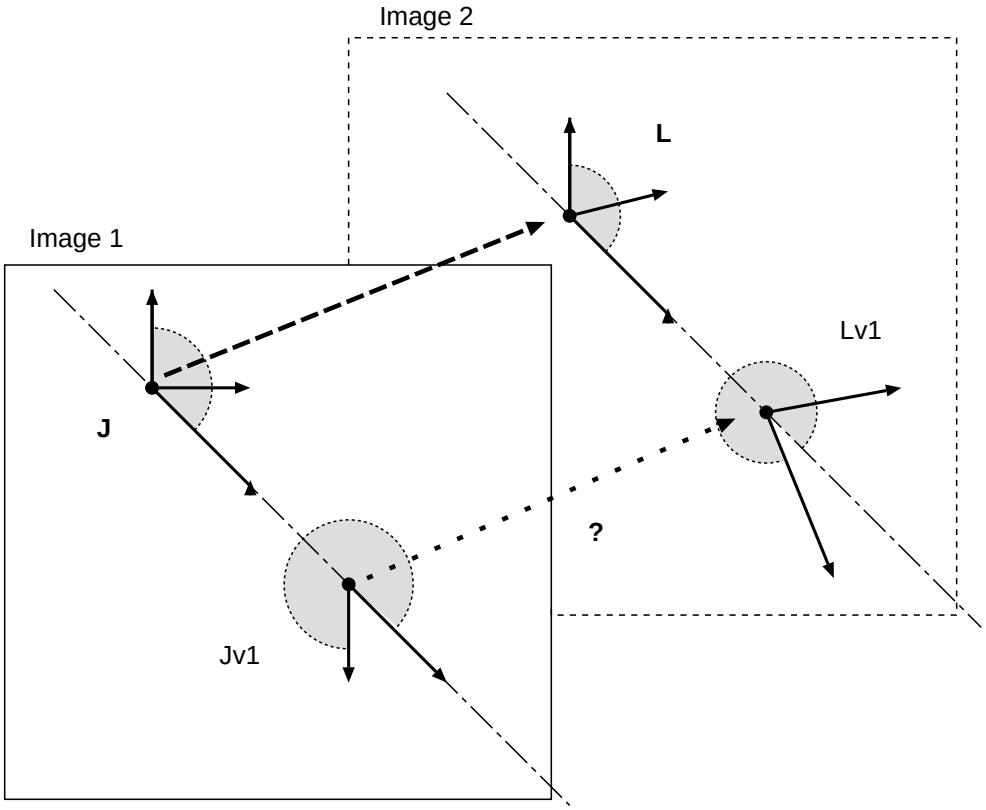

Figure 6.20 - *Comparaison entre hypothèses d'appariements. En supposant que Lv_1 est apparié avec Jv_1, dans quelle mesure peut on considérer que J est associé à L ?*

Algorithme 6.2 : Relaxation temporelle

début
  pour *Chaque jonction* $J \in \mathcal{J}_1$ **faire**
    pour *Chaque hypothèse d'appariement* $L \in \mathcal{M}(J)$ **faire**
      $Contributions \leftarrow 0$
      pour *Chaque voisin* $Jv_i \in \mathcal{V}(J)$ **faire**
        pour *Chaque hypothèse d'appariement* $L_j \in \mathcal{M}(Jv_i)$ **faire**
          si *le déplacement* $d_M(J, L)$ *similaire au déplacement* $d_M(Jv_i, L_j)$ **alors**
            $Contributions \leftarrow Contributions + S_M(Jv_i, L_j)$
      si $Contributions \neq 0$ **alors**
        % Renforcement de la probabilité $S_M(J, L)$
        $S_M(J, L) = S_M(J, L) \cdot (1.0 + Contributions)$
fin

6.3.4 Relaxation "perceptuelle"

Cette relaxation sur le voisinage des jonctions est semblable à la phase de relaxation "temporelle". Pour chaque jonction $J \in \mathcal{J}_1$, elle consiste à rechercher parmi les voisins de J ceux dont les correspondants dans $\mathcal{J}_2$ sont les plus cohérents avec les correspondants de J.

Les mécanismes de comparaison des configurations entre jonctions et de mise à jour des scores de groupement sont identiques à ceux de la relaxation "temporelle". La différence majeure vient du choix des jonctions utiles au renforcement de chaque hypothèse. Pour la relaxation temporelle, il s'agissait des jonctions $Lv_1 \in \mathcal{V}(L)$, éléments du voisinage perceptuel des jonctions L. Dans le cas de la relaxation spatiale, ces jonctions sont $L_1 \in \mathcal{M}(Jv_1)$, appariements possibles pour les voisins des jonctions J.

Le voisinage perceptuel de chaque jonction est défini par une relation de colinéarité avec les branches de chaque jonction. Les groupements sont ici supposés équiprobables. Ils sont donc initialisés par la quantité $(\frac{1}{n})$, 'n' étant le nombre de voisins du voisinage initial.

Le renforcement du score de groupement est, quant à lui, différent de celui utilisé pour la relaxation temporelle.

$$S_G(J, Jv_i) \;=\; S_G(J, Jv_i) \cdot (0.1 + 2.0 \cdot \sum_{(L,L_j)} \{ S_G(L, L_j) \;\; / \;\; d_G(J, Jv_i) \equiv d_G(L, L_j) \})$$

Un coefficient inférieur à 1 permet ici d'atténuer les scores des groupements qui n'ont reçu aucun soutien des jonctions de leur voisinage temporel. Ces scores ne sont pas normalisés de manière à éviter d'éliminer trop vite des voisins dont la contribution serait faible.

La figure 6.14 montre un exemple d'initialisation de ce type de voisinage. Dans cet exemple, le centre de la jonction J_6 est aligné avec une branche de J. Elle fait donc partie de son voisinage initial. Mais comme il y a peu de chances pour que cet alignement soit présent dans une autre vue de la même scène (même en conservant un écart faible entre les points de vue), la relaxation perceptuelle aura pour conséquence d'éliminer J_6 du voisinage de J.

Le résultat de cette phase de relaxation est donc, pour chaque jonction J, un ensemble de voisins perceptuels appartenant probablement au même objet que J (jonctions cohérentes avec un déplacement rigide). On retrouve ici le principe Gestaltiste de groupement perceptuel par "comportement commun" (*common fate*). Les scores de groupement traduisent la qualité de chaque association.

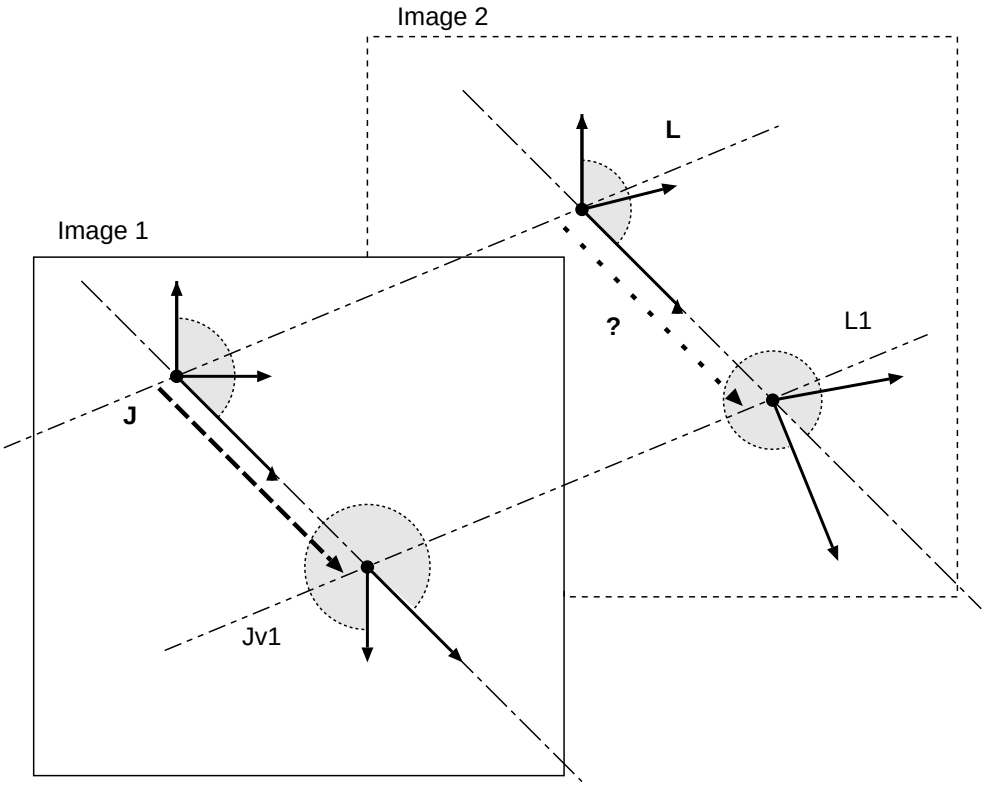

Figure 6.21 - *Comparaison entre hypothèses de groupements. En supposant que L est groupée avec L_1, dans quelle mesure peut on considérer que J est groupée à Jv_1 ?*

Algorithme 6.3 : Relaxation perceptuelle

début
  pour *Chaque jonction* $J \in \mathcal{J}_1$ **faire**
    pour *Chaque voisin* $Jv_i \in \mathcal{V}(J)$ **faire**
      $Contributions \leftarrow 0$
      pour *Chaque hypothèse d'appariement* $L \in \mathcal{M}(J)$ **faire**
        pour *Chaque voisin* $L_j \in \mathcal{V}(L)$ **faire**
          si *la configuration* $d_G(J, Jv_i)$ *similaire à la configuration* $d_G(L, L_j)$ **alors**
            $Contributions \leftarrow Contributions + S_G(L, L_j)$
      si $Contributions \neq 0$ **alors**
        % Renforcement du score de groupement $S_G(J, Jv_i)$
        $S_G(J, Jv_i) = S_G(J, Jv_i) \cdot (0.1 + 2.0 * Contributions)$
fin

6.3.5 Résultats de mise en correspondance

Comme c'était le cas pour les précédentes étapes de groupement, nous avons d'abord appliqué cet algorithme de mise en correspondance de jonctions sur des scènes artificielles puis sur des images réelles. Dans chaque cas, la limite du nombre d'itérations a été fixé à dix cycles de relaxation temporelle et spatiale. Chang et Aggarwal ont montré en effet que la conjugaison des deux relaxations et l'élimination progressive des appariements les plus mauvais renforcent la convergence. En plus de cette condition limite, l'analyse des appariements permet d'arrêter les cycles de relaxation lorsque les voisinages temporels des jonctions sont réduits à un voisin dominant.

Afin d'éliminer une grande partie des ambiguïtés d'appariements, l'algorithme est appliqué successivement de l'image I_1 vers l'image I_2, puis de I_2 vers I_1. Seules les jonctions pour lesquelles il existe un appariement réversible entre les deux images sont conservées. Les meilleures hypothèses d'appariements sont enfin extraites selon leur score.

Les figures de la page 234 visualisent les résultats d'appariement pour une scène simple. Les deux rectangles ont ici des déplacement très différents d'une image à l'autre, illustrés par la figure 6.24 des vecteurs de disparité entre les deux ensembles de jonctions. Cet algorithme permet donc de s'affranchir d'une contrainte de déplacement cohérent pour l'ensemble des jonctions grâce au groupement perceptuel des jonctions présentant des déplacements semblables.

Les images des pages 235 et 236 illustrent bien la robustesse de l'appariement malgré des variations importantes dans l'orientation des branches des jonctions. En effet, la reconstitution des arêtes à partir de l'image de contours subit des perturbations importantes dues essentiellement à la discrétisation des contours. C'est le cas, en particulier, pour les arêtes des fenêtres (jonctions 13, 14 et 15).

Cet exemple illustre également l'utilité des jonctions virtuelles. Même si elles sont issues de groupements accidentels, des jonctions apparaissant sur des points de vues différents constituent des points de repères utiles à la mise en correspondance. Les jonctions 9 et 11 sont ainsi appariées bien qu'elles n'aient pas de signification physique particulière.

Les figures des pages 237 et 238 représentent un exemple d'appariement dans des conditions réelles. Malgré l'apparente simplicité de la scène, les différences de contraste et le bruit de chaque image introduisent des différences importantes entre les contours des deux scènes. Comme le montre la figure 6.31, le groupement de segments permet de compenser une majeure partie des discontinuités le long des contours. La principale difficulté de l'appariement dans le cas de scènes réelles vient des jonctions sans correspondant, comme la jonction 3 dans cet exemple. La souplesse de l'algorithme d'appariement attribue tout de même un correspondant à cette jonction. Ce résultat est toutefois intéressant dans la mesure où les jonctions correctement appariées présentent des différences importantes de localisation (représentées par la figure 6.33) et d'orientation de leurs branches.

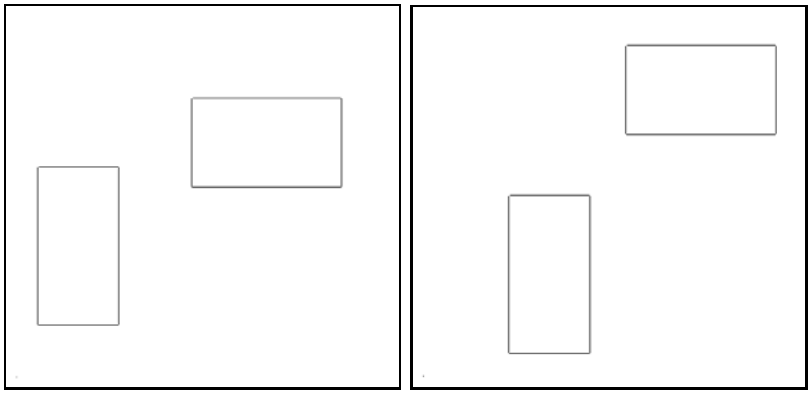

Figure 6.22 - *Appariement simple - rectangles*

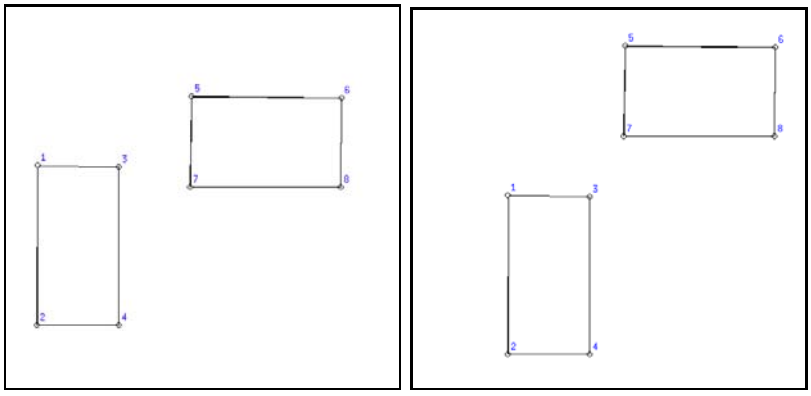

Figure 6.23 - *Appariement simple - rectangles - appariements*

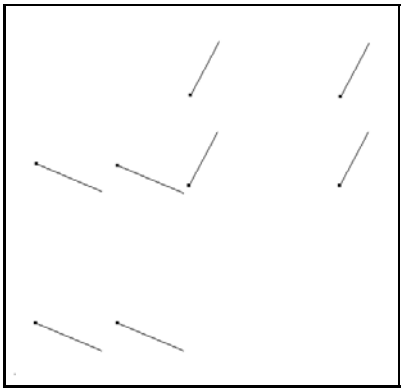

Figure 6.24 - *Appariement simple - rectangles - vecteurs de déplacement*

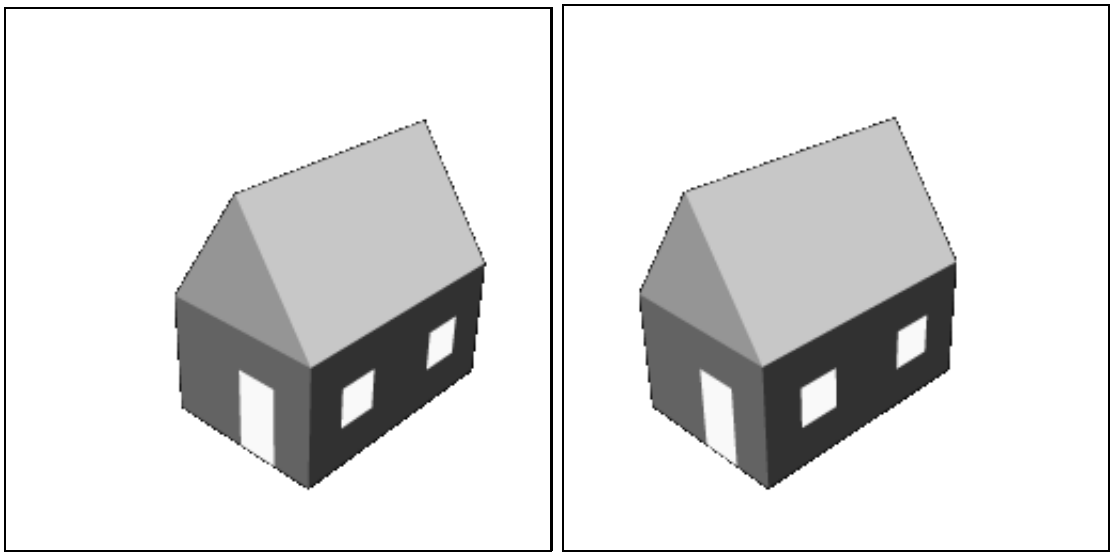

Figure 6.25 - *Appariement complexe - maison*

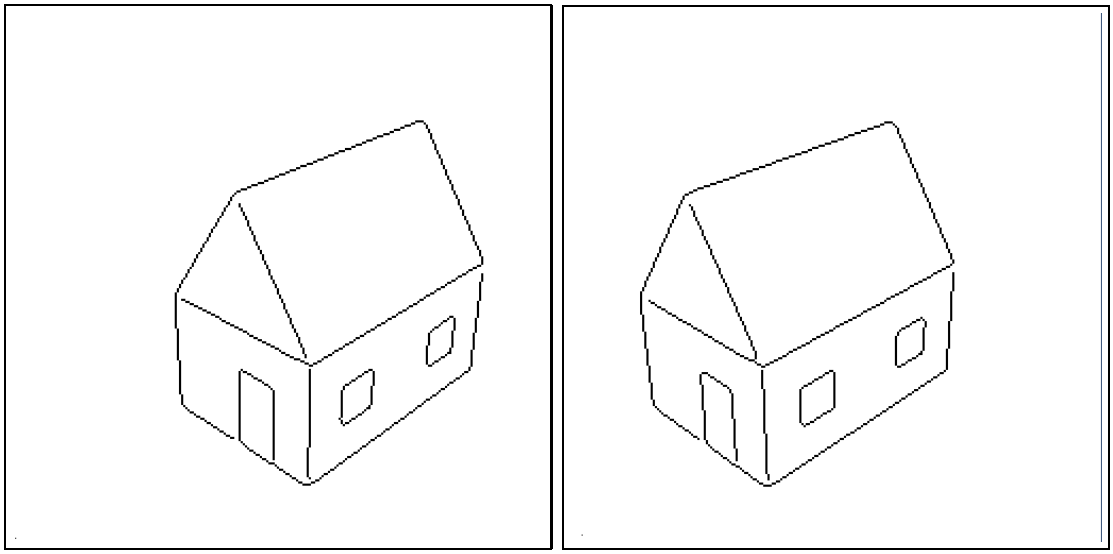

Figure 6.26 - *Appariement complexe - maison - détection de contours*

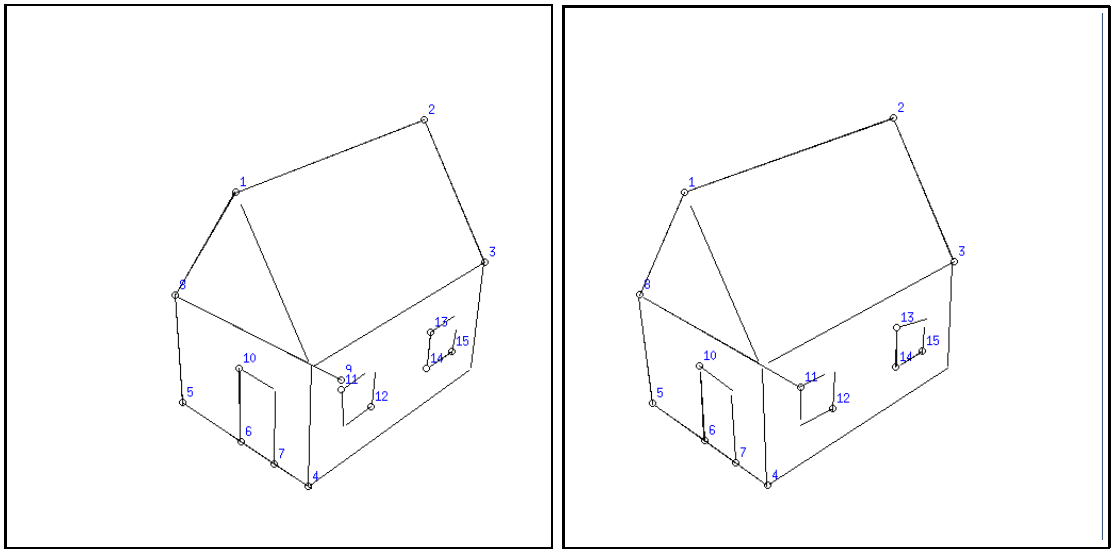

Figure 6.27 - *Appariement complexe - maison. On peut noter l'appariement correct de la jonction 13 dans les deux images, malgré les différences d'orientation des branches. Les jonctions virtuelles 9 et 11 sont superposées dans l'image de droite.*

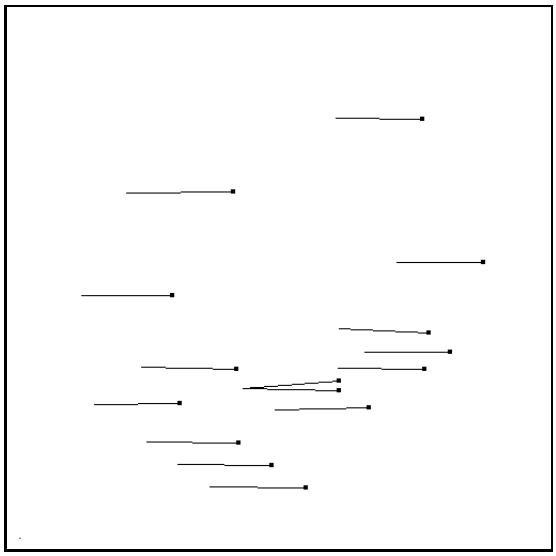

Figure 6.28 - *Appariement complexe - maison - vecteurs de déplacement*

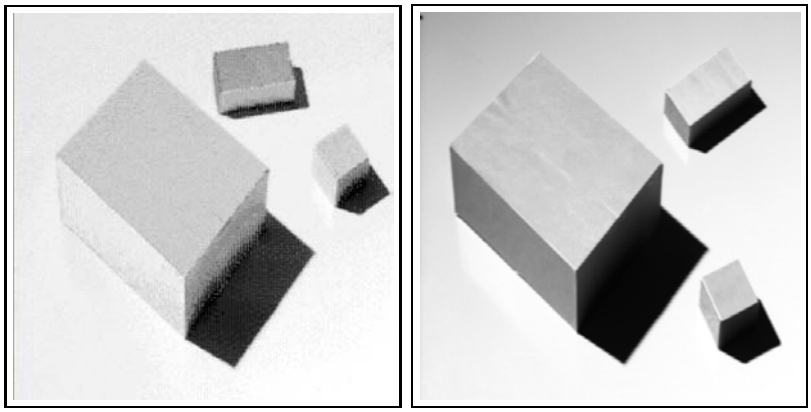

Figure 6.29 - *Appariement de jonctions - cube*

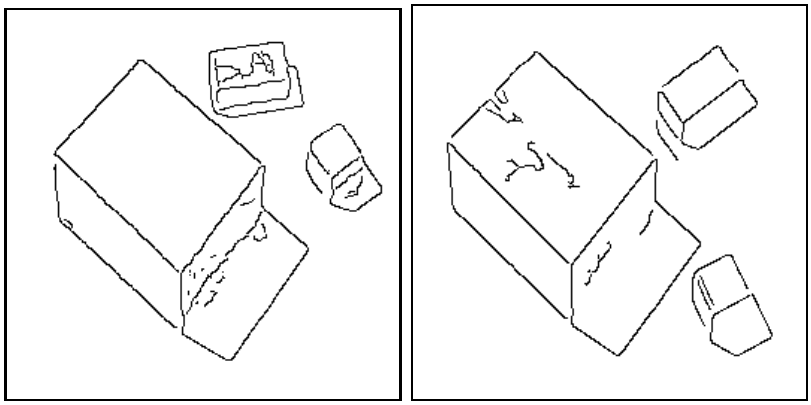

Figure 6.30 - *Appariement de jonctions - cube - détection de contours*

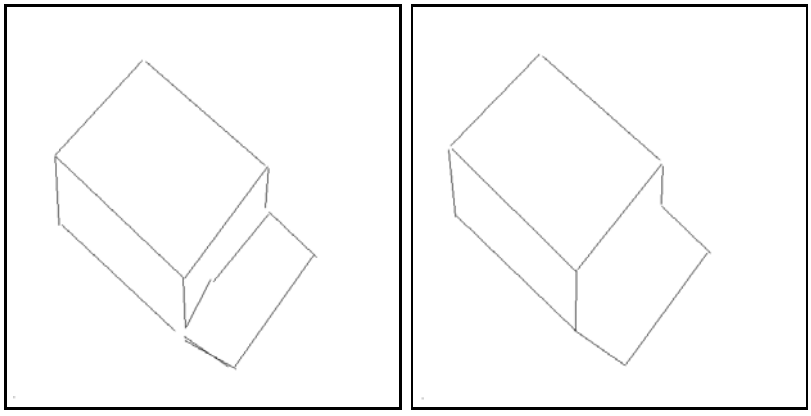

Figure 6.31 - *Appariement de jonctions - cube - hypothèses de segments aprés groupement*

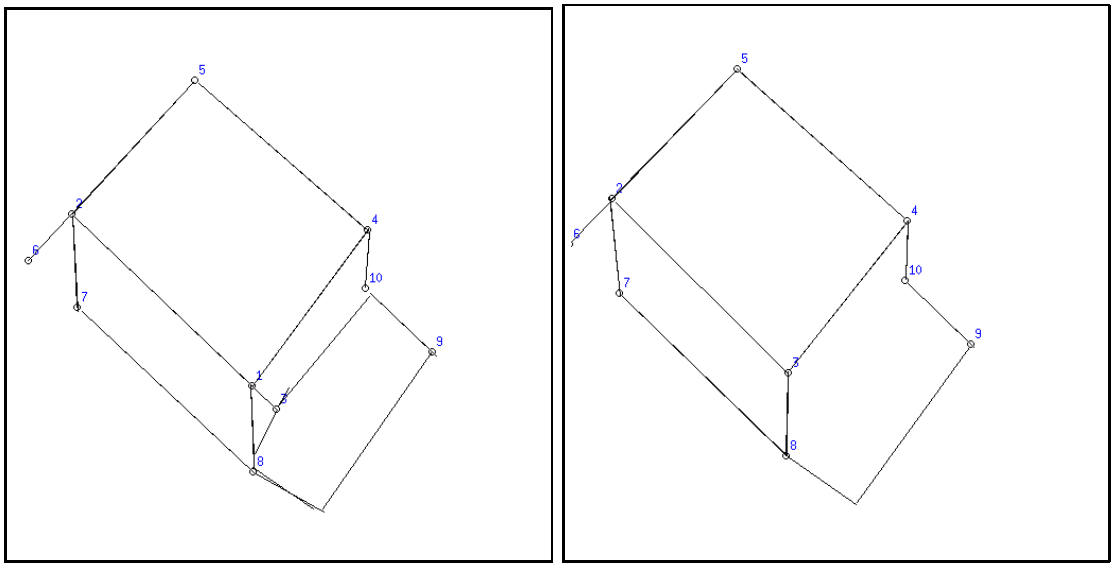

Figure 6.32 - *Appariement de jonctions - cube. On peut noter l'appariement correct de la jonction 10 malgré le passage de 3 à 2 branches d'une scène à l'autre. Les jonctions 1 et 3 sont superposées dans l'image de droite. La jonction 6 est une jonction virtuelle.*

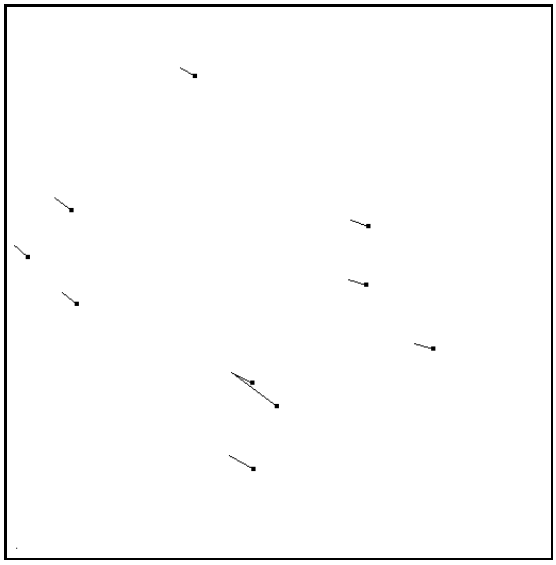

Figure 6.33 - *Appariement de jonctions - cube - vecteurs de déplacement*

6.4 Conclusion

Nous avons montré dans ce chapitre comment utiliser les éléments de représentations extraits à partir des chaînes saillantes d'une image pour constituer des structures plus globales et les manipuler. En particulier, nous avons présenté une méthode de groupement de segments et de points d'intérêt selon un ensemble de jonctions. L'utilité de ce type de groupement de haut niveau est finalement illustré à l'aide d'un algorithme de mise en correspondance de jonctions.

Le principal avantage de cette approche est sa flexibilité et sa robustesse face au bruit grâce à la généricité des méthodes de groupement. Il serait intéressant de la développer selon deux axes privilégiés.

D'une part, la construction d'hypothèses de groupement selon des règles plus complexes. Des propriétés telles que la symétrie ou le parallélisme apportent en effet des contraintes plus fortes pour la mise en correspondance de structures d'intérêt. Leur complexité élevée assure un faible nombre de structures et une recherche d'autant plus rapide.

D'autre part, la mise en correspondance pourrait bénéficier de la hiérarchie de groupements. En effet, l'appariement de jonctions ne peut qu'être le début d'une mise en correspondance hiérarchique. En servant de centre d'intérêt, les jonctions appariées peuvent être utilisées pour mettre en correspondance les jonctions de leur voisinage. Ces appariements seraient ensuite propagés à chaque branche, jusqu'aux pixels de l'image correspondant à ces branches si l'application exige une grande précision. Cette démarche, semblable à celle de [Venkateswar et Chellappa, 1995] , consiste en une succession de prédictions et vérifications d'appariements entre structures, les appariements de haut niveau servant de centre d'attention pour niveaux inférieurs.

Enfin, une amélioration plus matérielle est nécessaire à court terme. En tant que structures complexes, les jonctions sont moins nombreuses que de pixels ou des segments, mais leur représentation en termes de mémoire est beaucoup plus volumineuse. Comme pour la manipulation de grands nombres d'hypothèses évoquée en fin du chapitre précédent, il serait utile ici aussi d'adopter des représentations internes plus adaptées à l'utilisation de structures complexes.

Chapitre 7

Conclusion

Au cours de cette thèse, nous avons abordé le problème de la perception de structures régulières à partir d'une détection de contours dans des images d'intensité lumineuse.

A partir d'une étude bibliographique de la perception visuelle, nous avons souligné les différentes sources d'ambiguités qui font de la vision par ordinateur un problème d'une extraordinaire complexité. En particulier, nous nous sommes intéressés aux problèmes que pose l'interprétation de scènes de contours ainsi qu'à différentes approches proposées en vision artificielle pour traiter ce type de scènes.

Le choix des contours comme support de notre travail est délibéré. Le propos n'est pas de réduire l'analyse d'images aux seuls contours. Ce choix doit être placé dans un contexte plus général d'un système de vision par ordinateur qui mettrait en commun les résultats d'analyse d'images selon divers indices visuels, dont les contours.

Enfin, nous avons étudié le rôle du groupement perceptuel pour réduire la complexité de cette tâche et nous avons conclu par la proposition d'une méthodologie de structuration progressive des contours. Afin de faire face aux nombreuses sources d'ambiguité posées par la détection de contours avec un minimum d'hypothèses sur le type de scène observée, nous proposons une analyse qualitative des contours de l'image. L'utilisation de principes de groupement perceptuel permet de repousser le plus loin possible dans la chaîne de traitements l'intervention de méthodes d'analyse précises, plus sensibles aux erreurs de détection.

Une première phase détecte les structures curvilignes les plus régulières à l'aide de réseaux de saillances. Il s'agit de définir des critères de régularité pour des groupements possibles entre éléments de contour, de mettre en valeur les structures les plus régulières afin d'en extraire les principaux groupements. En plus d'apporter une approche générique pour ce type d'optimisation, notre contribution à ce niveau de traitement inclus un nouveau formalisme pour évaluer la régularité d'un groupement, un algorithme différent pour assurer la convergence du réseau vers des structures stables et un ensemble d'heuristiques pour l'extraction des meilleurs groupements après optimisation.

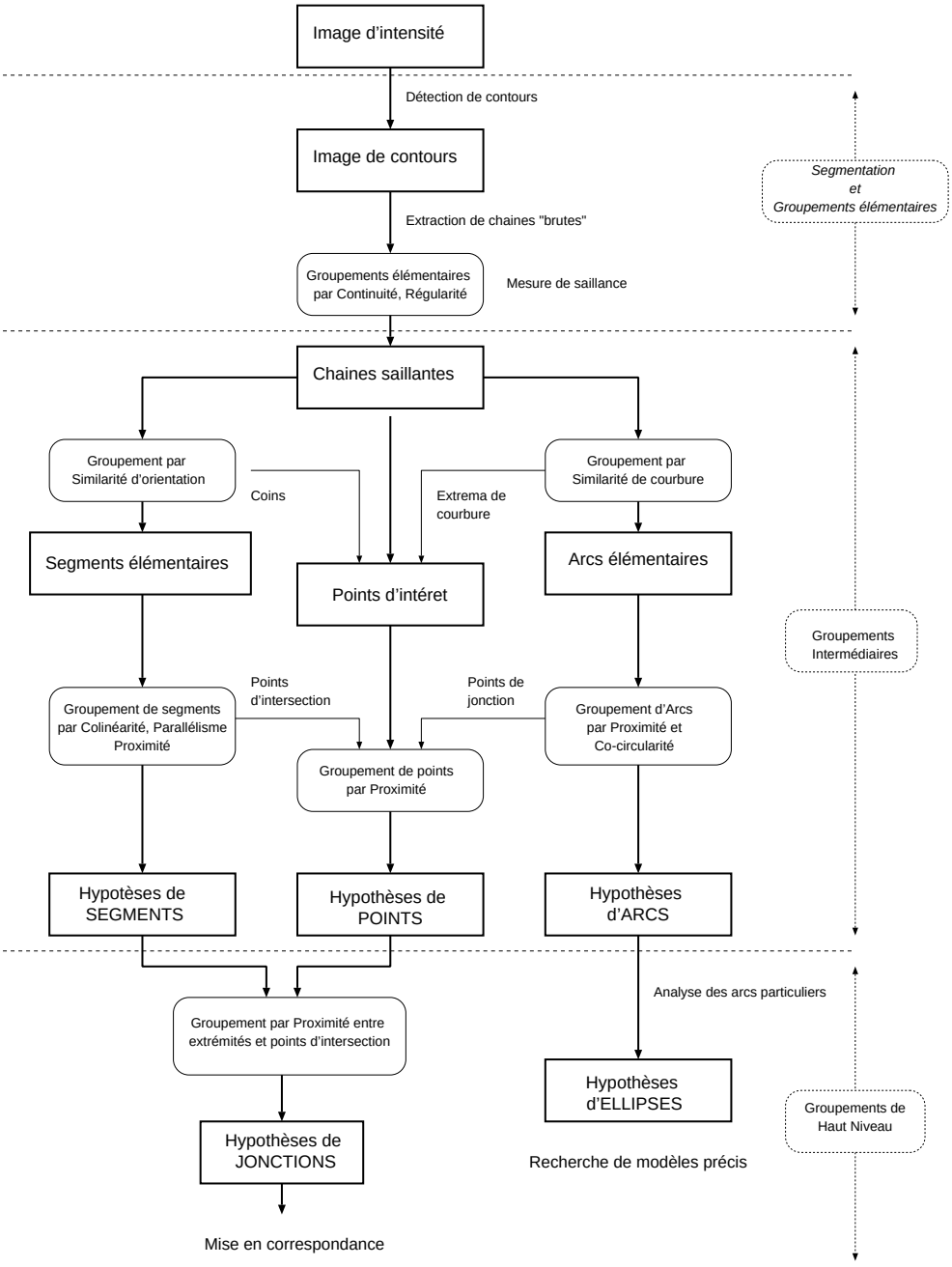

Figure 7.1 - *Récapitulatif des trois niveaux d'organisation perceptuelle.*

Les groupements saillants ainsi détectés jouent un rôle de centre d'attention pour l'extraction d'éléments visuels représentatifs des contours de l'image. Ils permettent ainsi de réduire la complexité en ne conservant que les hypothèses les plus régulières. Notre approche est hiérarchique et modulaire. Les éléments visuels sont extraits progressivement, sous forme d'hypothèses de groupements de plus en plus complexes.

Chaque niveau est ouvert aux contributions d'autres détections. Par exemple, le groupement par réseau de saillance peut être enrichi par la détection de frontières entre régions, dans la définition de la fonction de qualité ou encore lors de l'extraction des meilleurs groupements. De même, les hypothèses de points d'intérêts peuvent être complétées par une détection directe de coins si l'application le demande.

Le résultat de cette méthodologie est un ensemble d'éléments visuels représentatifs des structures curvilignes de l'image. L'utilisation de groupements selon des hypothèses de plus en plus complexes offre de nombreux avantages. En tolérant une certaine part de redondance, les hypothèses permettent l'interprétation de contours selon des échelles différentes. Elles apportent à la méthode une plus grande robustesse devant les discontinuités et les fausses détections. Elles peuvent être enfin utilisées directement au sein de structures plus complexes, comme nous l'avons montré avec la mise en correspondance de jonctions, ou bien servir de point de départ à une analyse plus précise par l'application de modèles déformables.

Nous avons enfin implanté chaque niveau de groupement en insistant sur l'application de notre méthode à de nombreuses images complexes (scènes artificielles, urbaines et naturelles), sur des plateformes usuelles [1]. Par comparaison, la majorité des travaux antérieurs portent sur des types d'images particuliers ou bien nécessitent une mise en oeuvre sur des machines parallèles. En ce sens, notre approche est comparable, par sa généricité, aux démarches de [Mohan et Nevatia, 1992] et de [Sarkar et Boyer, 1994] .

Malgré la qualité des résultats obtenus, un certain nombre de problèmes persistent. Un grand nombre de paramètres, propres à chaque phase de détection et de groupement, restent encore définis de manière empirique. En pratique, ces paramètres sont réutilisables pour une même classe d'images en produisant des résultats similaires. Nous avons privilégié l'étude qualitative d'ensemble du système à une étude quantitative en profondeur de chacune de ses composantes. Bien qu'il soit envisageable de confronter les résultats de manière automatisée avec des images de référence, ce dernier aspect pose en particulier le délicat problème d'une estimation quantitative de la qualité visuelle d'un groupement.

La sélection semi-automatique des structures saillantes offre un meilleur confort d'utilisation, en évitant à l'utilisateur de choisir manuellement les points de départ des chaînes saillantes. Elle suppose cependant une recherche manuelle des seuils de sélection. L'analyse des groupes saillants ne tient compte que d'un seul paramètre d'échelle, fixé par l'utilisateur. Il serait utile d'automatiser l'analyse à différentes

1. Les résultats d'analyse sont de l'ordre de 5 à 20 minutes, tous traitements compris, sur stations de travail SUN Sparc-10 et PC, pour des images de tailles inférieures à 800×600 pixels.

échelles, en recherchant les structures les plus stables et en définissant des critères de qualité pour les hypothèses.

Les ambiguïtés de la mise en correspondance de jonctions, encore trop nombreuses, pourraient être réduites en validant les hypothèses finales d'appariement à l'aide des voisinages de chaque jonction.

Enfin, la limitation la plus récurrente vient de l'implantation elle-même. Le grand nombre d'hypothèses entraîne des besoins en ressources mémoires encore trop importantes et limite les performances en temps de calculs. Une implantation plus efficace à l'aide de techniques adaptées à de grands nombres d'hypothèses, comme des méthodes d'indexation par exemple, pourrait résoudre ce problème.

Ces observations permettent de dégager deux axes de recherche principaux pour prolonger notre travail. Les hypothèses d'éléments visuels peuvent être validées et ajustées précisément en servant de point de départ à des méthodes précises selon un cycle de prédiction-vérification. A l'inverse, elles peuvent être directement groupées selon des règles plus complexes (parallélisme, symétrie, similarité, convexité) et formuler ainsi des hypothèses de plus en plus structurées.

A plus long terme, l'objectif d'un tel système est bien entendu une coopération avec d'autres processus d'analyse au sein d'une application plus vaste. Citons, à titre d'exemple, l'aide à l'indexation automatique de modèles. Les hypothèses géométriques produites par ce type d'approche pourrait s'insérer naturellement dans la construction de "modèles d'apparance" tels que les ont définis A. Pope et D. Lowe[2].

2. Cf. [Pope et Lowe, 1993] , page 74.

Annexes

Annexe A

Réseaux de saillance de Shashua et Ullman

Malgré des définitions très différentes, les mesures de saillance proposées dans la littérature ont en commun l'application des principes Gestaltistes de continuité et de "bonne forme". Elles tiennent compte, selon les cas, de mesures de courbure, de proximité, d'orientations ou encore de co-circularité. Elles se distinguent surtout par leur mécanisme de calcul, qui peut aller de l'optimisation combinatoire sous toutes ses formes à l'application directe d'un filtre adapté par convolution.

A.1 Mesures de saillance structurelle

On pourra également trouver dans [Williams et Thornber, 1997] une étude comparée de différentes mesures de saillance. Notons qu'une conséquence intéressante des mesures de saillance est la fermeture des discontinuités. Elles sont souvent appliquées à l'élaboration d'hypothèses sur la perception de contours fictifs.

A.1.1 Optimisation combinatoire

En théorie, une fois définie une fonction de qualité entre éléments de contours à grouper, toute méthode d'optimisation combinatoire adaptée aux problèmes NP-complets peut être envisagée.

- Recuit simulé et variantes - [Hérault, 1991]

  Cette démarche constitue un exemple significatif de ce type d'approche. Le problème posé est le suivant :

  "*Etant donnés des points de contours dans l'image et connaissant le gradient en chaque point, quels sont les points par lesquels passent des courbes saillantes dans l'image ?*"

Il s'agit donc de mesurer la saillance de points de contours sous forme d'une classification binaire entre les points des structures linéaires et les points de perturbations. Deux critères de sélection sont choisis pour évaluer l'appartenance possible de deux points de contours à une structure courbe.

- **La co-circularité** qui mesure la probabilité pour qu'un cercle passe par deux points et leurs tangentes associées. Ce critère tient compte de l'erreur de quantification sur la localisation des points et l'orientation des tangentes.

  Formellement, deux points de contours i et j avec leurs tangentes associées T_i et T_j, sont sur un même cercle, si et seulement si :

  $$\lambda_i + \lambda_j \ = \ \pi$$

  où λ_i et λ_j désignent respectivement les angles formés par le segment S_{ij} et les tangentes T_i et T_j. Si on note : $\Delta_{ij} \ = \ |\lambda_i + \lambda_j - \pi|$, le coefficient de co-circularité est donné par :

  $$c_{ij}^{cocirc} \ = \ (1 - \frac{\Delta_{ij}^2}{\Pi^2}) \cdot exp(-\frac{\Delta_{ij}^2}{k})$$

  Le paramètre k est choisi de façon à ce que le coefficient de co-circularité décroisse rapidement pour des configurations non co-circulaires.

- **La proximité** qui permet de favoriser les interactions locales entre éléments de contours. En notant d_{ij} la distance entre les deux points, ce coefficient donne :

  $$c_{ij}^{prox} \ = \ exp(-\frac{d_{ij}^2}{2\sigma_d^2})$$

  où σ_d est l'écart type de la distribution des distances entre les tangentes.

Ces deux critères permettent de définir une énergie de co-circularité de la manière suivante. A chaque point i est associée une variable binaire x_i, qui vaut 1 si le point fait partie d'une structure courbe, et 0 s'il s'agit d'un point de bruit. Une structure aura une forte co-circularité globale si elle maximise sur les variables x_i la valeur suivante :

$$E_{cocirc} \ = \ \sum_{i=1}^{N} s_i \cdot x_i \ = \ \sum_{i=1}^{N} (\sum_{j=1}^{N} c_{ij}^{cocirc} \cdot c_{ij}^{prox} \cdot x_j) \cdot x_i$$

Le terme s_i mesure le degré de saillance de la tangente en i.

Afin d'éviter une solution triviale pour laquelle tous les x_i seraient égaux à 1, cette énergie est associée à une énergie de contrainte sur la taille de l'ensemble de points recherché :

$$E_{max} = (\sum_{i=1}^{N} x_i)^2$$

Le problème d'optimisation revient donc à minimiser, pour les variables x_i la quantité : $E = -(E_{cocirc} + \lambda \cdot E_{max})$, λ étant un paramètre positif ajusté expérimentalement.

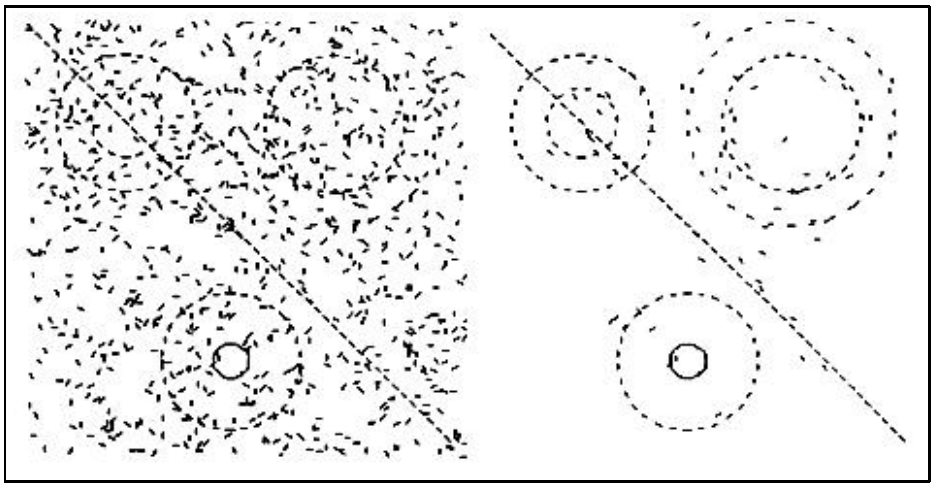

Figure A.1 - *Exemple de groupement perceptuel par recuit simulé. Sur les 1000 segments de départ, 352 ont été sélectionnés parmi les plus saillants. Exemple tiré de [Hérault et Horaud, 1992].*

Ce problème est finalement codé sous la forme d'un réseau de neurones de Hopfield, puis optimisé en comparant différentes méthodes telles que l'approximation du champ moyen (inspirée d'un modèle physique de ferromagnétisme), le recuit simulé simple (inspiré d'un modèle thermique), et du recuit en champ moyen [Hérault et Horaud, 1992] .

Cette approche est bien représentative de la construction d'une mesure de qualité, puis du codage et de l'optimisation de cette mesure. Elle donne des résultats intéressants sur images synthétiques et a été appliquée à des images réelles simples. Les structures saillantes sont bien conservées et une certaine quantité de bruit est éliminée. Le temps de calcul est, quand à lui, relativement long [1].

1. De l'ordre de 30 minutes à 3 heures sur un Cray pour un traitement sur 1000 segments, selon les méthodes d'optimisation.

Enfin, cette méthode se concentre uniquement sur une classification des éléments de contour en 'bruit' et 'forme', sans reconstituer les structures après détection. Hérault propose d'utiliser des processus de classification pour séparer les structures entre elles.

– Relaxation - [Parent et Zucker, 1989]

Par une méthode en deux étape, Parent et Zucker soulignent l'importance d'une séparation entre des mesures locales nécessairement erronées et une détection des courbes par optimisation globale de critères visuels.

Un premier niveau de traitement établit une estimation grossière des tangentes présentes dans l'image. Cette estimation est réalisée à l'aide d'une convolution de l'image avec un ensemble de filtres linéaires selon des orientations prédéfinies. Ces opérateurs, semblables à des détecteurs de contours, mettent en valeur la présence de tangentes dans l'image.

Ces mesures étant dépendantes de mesures limitées à un voisinage relativement réduit, elles sont sensibles au bruit présent dans l'image de contours. Une seconde étape d'optimisation est donc nécessaire pour éliminer les fausses détections de tangentes et renforcer les orientations appartenant à courbes communes. Cette étape optimise, à l'aide d'un processus de relaxation, une mesure de saillance des tangente.

Comme pour l'exemple précédent, cette mesure intègre des contraintes de proximité, de co-circularité et continuité de courbure. Ces contraintes sont définies dans un voisinage de chaque point de contour. Pour des raisons d'éfficacité, ce voisinage décrit sept classes de courbures, correspondant à une discrétisation d'arcs de cercles tangents au contour.

Cette méthode ne se contente pas de séparer les points de bruit de ceux des courbes. En plus de mettre en valeur les structures courbes de l'image, elle produit un champ de tangentes et de courbures optimisées. Afin d'extraire les structures courbes de ces champs de tangentes, une méthode de chaînage est également proposée par les mêmes auteurs. En définissant un champ de potentiels à partir des tangentes, ce groupement de plus haut niveau procède par optimisation de contours actifs. Ces contours, initialisés le long des tangentes, convergent dans le champ de potentiel vers un ensemble de courbes saillantes. [Zucker *et al.*, 1989]

Les résultats sur des images artificielles et réelles sont nombreux et démontrent une robustesse en s'appliquant à différentes situations (images satellitaires, médicales et empreintes digitales). L'algorithme de détection des courbes est entièrement parallèle mais présente des temps de calculs très longs [2].

2. Jusqu'à 12 heures sur un DEC VAX 11/780

A.1.2 Mesures 'directes'

L'un des principaux reproches faits aux méthodes précédentes est leur incompatibilité avec l'expérience biologique de la perception visuelle. En effet, la convergence itérative vers une solution approchée est un modèle peu satisfaisant pour un mécanisme aussi immédiat que la perception de contours saillants dont les temps de réponse sont de l'ordre de la centaine de milli-secondes. Ces observations suggèrent des méthodes plus directes dont les exemples suivants sont représentatifs.

- Champs d'extensions - [Guy et Medioni, 1996]

  Cette mesure de saillance est définie comme la somme de compromis entre les réponses de filtres directionnels appliqués aux éléments de contours de l'image. Ces filtres, qu'ils nomment "champs d'extensions" (*extension fields*) représentent la probabilité de relier l'extrémité d'une courbe incomplète à partir du point d'application. Ils tiennent compte à la fois de la tangente au point de contour et de la forme globale de la courbe, en définissant la contribution de ce point pour ses voisins en termes de longueur et d'orientation.

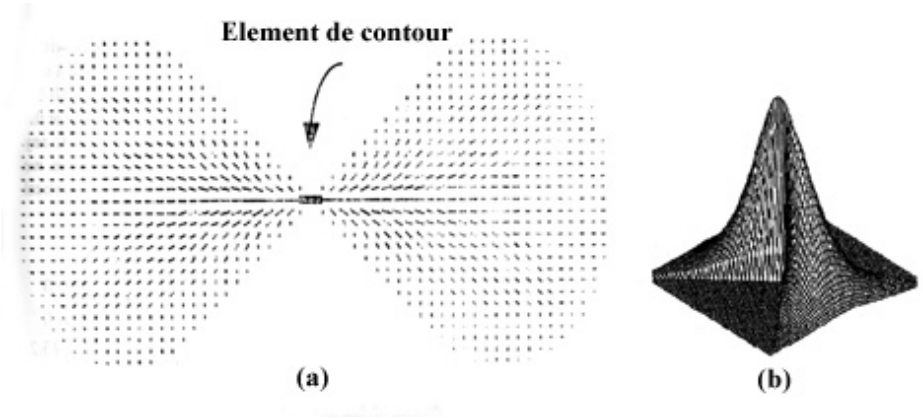

Figure A.2 - *Aspect d'un champ d'extension. La figure (a) représente la distribution des orientations autour d'un élément de contour. La figure (b) représente la variation de l'amplitude du champ en fonction de la distance et de l'orientation. Exemple tiré de [Guy et Medioni, 1996].*

Une forme analytique de ce type de champ peut être définie par :

$$\overline{EF}(x,y) \;=\; \begin{cases} e^{-Ax^2}(1,0)^T & ,y \;=\; 0 \\ (e^{(-Ax^2 \;+\; B\cdot\arctan(|y|,|x|)^2)}(\frac{x}{R},\frac{y}{R}-1)^T & ,y \;\neq\; 0 \end{cases}$$

avec :

$$R(x,y) \;=\; \frac{(x^2+y^2)}{2y}$$

Le paramètre A contrôle l'atténuation due à la distance et B contrôle l'atténuation due au changement de courbure. Ces paramètres permettent d'influer sur la forme du champ. Celle-ci est donc une fonction qui décroît exponentiellement selon la distance et le rayon de l'arc circulaire défini à partir de l'origine. Le choix de la forme du champ permet de détecter différents types de structures (jonctions ou courbes). Par construction, les champs d'extension intègrent des contraintes de co-circularité, de continuité de courbure, et de proximité.

L'optimisation sur l'ensemble des points de contours de l'image est mise en oeuvre par une technique semblable à la transformée de Hough. Chaque point de contour p reçoit des votes de la part de tous les autres points dont le champ d'extension traverse p. Chaque vote est un vecteur, défini par une intensité et une orientation. Appliqué à tous les autres points, ce processus est comparable à une convolution entre les points de l'image et un masque défini par le champs d'extension. A la seule différence que le résultat de la convolution n'est pas un scalaire mais un vecteur.

Après application de tous les votes, chaque point de contour est le site d'une accumulation de vecteurs dont il suffit d'extraire les directions privilégiées. L'analyse statistique des moments de ce système de vecteurs permet de définir en chaque point une ellipse dont les axes correspondent aux directions des moments principaux. Si on note λ_{min} et λ_{max} les valeurs propres de la matrice de co-variance correspondant à ces moments, une mesure de la saillance d'un point de contour peut être définie par la simple différence $(\lambda_{max} - \lambda_{min})$. Sans entrer dans les détails, λ_{max} est fonction croissante du nombre de votes accumulés sur un site et λ_{min} est d'autant plus faible que les vecteurs accumulés sur ce site ont une direction proche.

Le résultat de cette méthode est à la fois une carte d'éléments orientés, et une mesure de la saillance des points de contours. Cette approche permet ainsi de définir une mesure de saillance globale, chaque point recevant une contribution de la part de tous les autres. Par opposition, les méthodes d'optimisation combinatoire cherchent à optimiser globalement des mesures locales.

Les résultats sur des images synthétiques et quelques images réelles montrent une bonne mise en valeur des structures globales ainsi que des applications possibles pour la détection de jonctions et la fermeture de contours fictifs. Les auteurs suggèrent une extraction possible des structures saillantes en suivant les crêtes de la carte de saillance et ses vecteurs associés. Ils restent cependant peu clairs sur le traitement des intersections entre courbes ainsi que sur les zones de saillance homogène.

– Champs stochastiques de fermeture - [Williams et Jacobs, 1994]

Cet autre exemple de mesure directe de saillance peut être vu comme une approche rigoureuse, d'un point de vue statistique, des idées avancées par

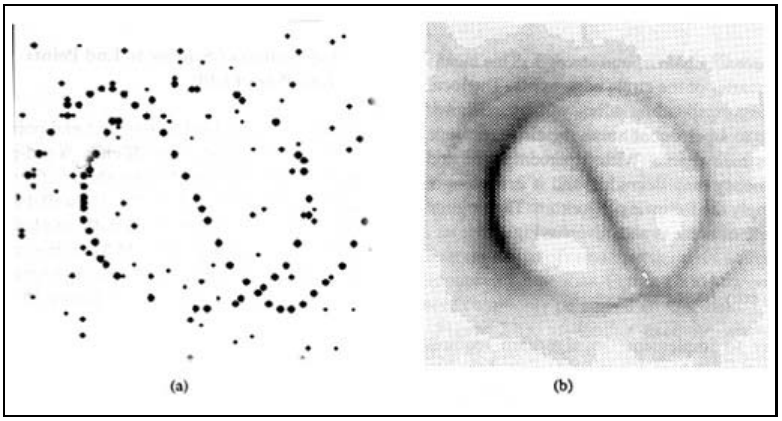

Figure A.3 - *Exemple de carte de saillance obtenue à l'aide de champs d'extension. L'intensité est d'autant plus faible que la saillance des points est plus grande. (a) Image d'origine. (b) Carte de saillance. Exemple tiré de [Guy et Medioni, 1996].*

Guy et Medioni. Tout comme la mesure précédente, celle-ci définit un champ de vecteurs autour de chaque point de contour.

Désigné par "Champ stochastique de fermeture" ou *Stochastic Completion Fields*, ce champ représente une distribution de tous les chemins possibles partant d'un point selon une direction initiale. Cette distribution est modélisée par un mouvement de particules selon certaines contraintes de position, d'orientation et vitesse (mouvements Browniens).

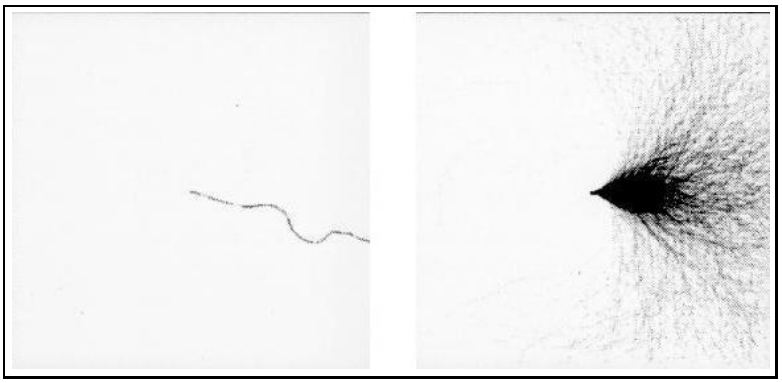

Figure A.4 - *Exemple de champ stochastique de fermeture. La figure de droite représente le mouvement aléatoire d'une particule. La figure de gauche montre la distribution des trajectoires d'un ensemble de particules. Exemple tiré de [Williams et Jacobs, 1994].*

Dans ce cas, la saillance d'un élément de contour p est la probabilité qu'une particule passe par p en suivant un mouvement stochastique reliant deux autres

éléments de contours.

La généralisation de cette mesure sur l'ensemble des éléments de contours est ici aussi un produit de convolution entre champs de vecteurs. Elle revient à chercher les courbes d'énergie minimale parmi les distributions stochastiques de tous les chemins possibles. Pour des raisons d'éfficacité, ces distributions sont pré-calculées sur un ensemble de positions et orientations.

[Thornber et Williams, 1997] ont proposé depuis une variante de cette mesure en considérant une somme de mouvements stochastiques sur des chemins reliant plusieurs éléments de contours entre eux. Par comparaison, la mesure précédente se concentre sur un seul mouvement de particule entre deux éléments de contours. La saillance est définie alors par la fraction de chemins stochastiques fermés passant par un élément de contours donné.

Dans les deux cas, l'application privilégiée est la fermeture de contours, et la perception de contours illusoires. Les résultats sont particulièrement intéressants sur les contours illusoires, de part leur similarité avec les observations psycho-visuelles sur ces mêmes contours. Ils restent cependant limités à des images artificielles.

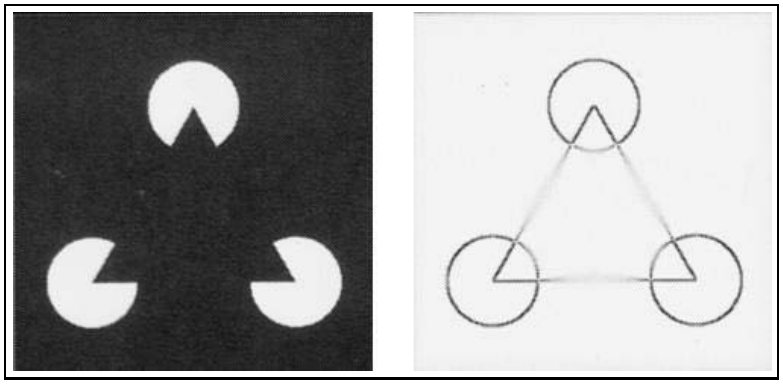

Figure A.5 - *Résultat de fermeture de contours fictifs par application de champs stochastiques. Exemple tiré de [Williams et Jacobs, 1994].*

A.2 Réseaux de Saillance de Shashua et Ullman

Les réseaux de saillances, tels que définis par [Shashua et Ullman, 1988] offrent un formalisme original pour l'optimisation globale d'une mesure de saillance à partir de calculs locaux. Afin de faciliter la comparaison avec notre propre approche, nous présentons les réseaux de saillance de Shashua avec un formalisme semblable à celui du chapitre 4.

A.2.1 Définitions et notations

Comme la plupart des autres mesures de saillance, celle-ci concerne le groupement de pixels de contours selon des courbes visuellement importantes tout en éliminant les pixels de bruit. Elle s'applique donc à une image de détection de contours.

Soit un pixel P de cette image. Son intensité vaut 1 s'il s'agit d'un point de contour et 0 sinon.

L'image est considérée comme un réseau de pixels inter-connectés. Chaque pixel P est relié à k voisins notés $N_i,\ i \in [0,k]$. Ce voisinage, noté $\mathcal{V}(P)$, définit ainsi k éléments d'orientations v_i. Ces éléments sont dits "réels" si le voisin relié est un point de contours, et "virtuels" sinon. Ce réseau peut être vu comme un graphe pour lequel les pixels sont les noeuds et les éléments d'orientation les arcs.

Soit $\Gamma_N(P, v, \overline{v})$ une courbe de $2N$ éléments traversant P. Cette courbe arrive en P par la direction d'un élément $\overline{v}$ et en repart dans la direction de $v \neq \overline{v}$. On note $\gamma_N(P, v)$ la branche de cette courbe composée des éléments de connexion $\{e_1, e_2, \cdots, e_N\}$, et $\gamma_N(P, \overline{v})$ la branche composée des éléments de connexion $\{e_{-1}, e_{-2}, \cdots, e_{-N}\}$.

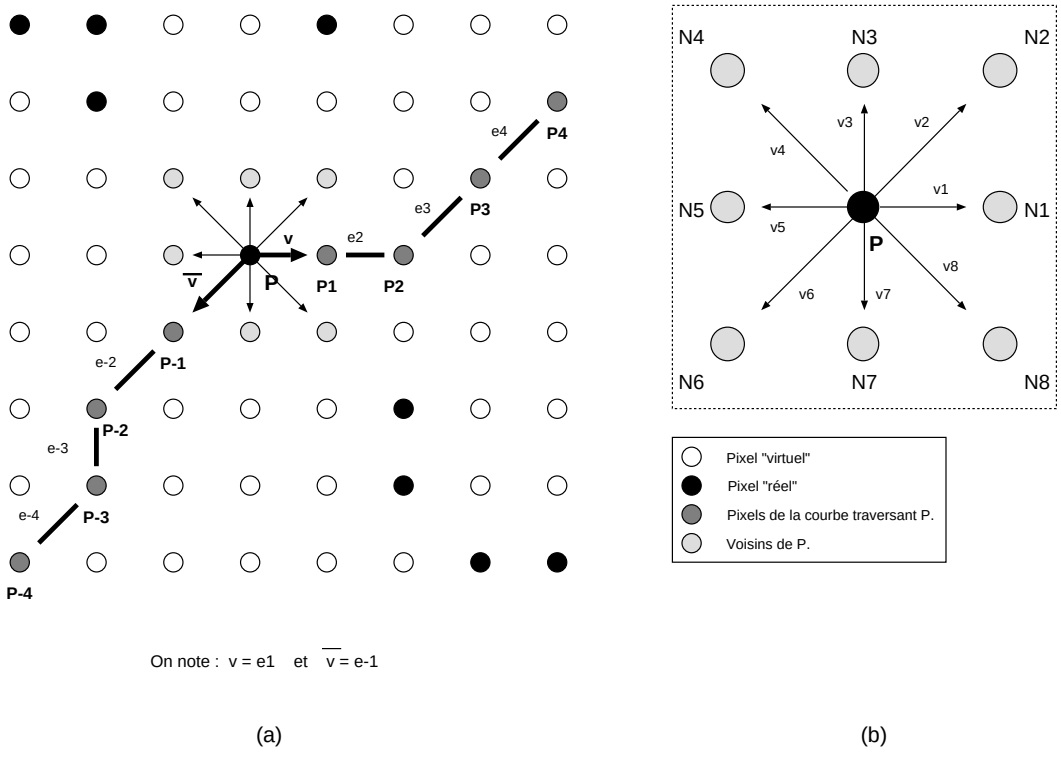

Figure A.6 - (a) *Notations pour une courbe traversant un pixel P* - (b) *Exemple de voisinage en 8 connexité. Dans cet exemple, $e_1 = v_1$ et $e_{-1} = v_6$.*

Comme le montre la figure A.6, les éléments e_1 et e_{-1} sont définis par : $e_1 = v$ et $e_{-1} = \overline{v}$.

On parlera par la suite indifféremment de groupements et de courbes.

A.2.2 Mesure de saillance

Shashua et Ullman définissent une mesure de saillance qui doit favoriser la formation de courbes longues et uniformes, répondant à des critères visuels. Cette mesure doit de plus se conformer à l'algorithme d'optimisation choisi.

Dans un premier temps, ils dérivent une fonction de qualité à partir de critères de courbures et de longueurs. Ces critères doivent récompenser les courbes lisses, présentant peu de discontinuités. Cette fonction de qualité est ensuite exprimée sous une forme adaptée à la méthode d'optimisation.

Ils distinguent de plus deux types de saillances. D'une part, la saillance d'un élément d'orientation v est définie comme la valeur maximale des qualités des courbes partant de P dans la direction de v. L'ensemble des courbes possibles de longueur N partant de P dans la direction de v est noté $\delta^N(v)$. On peut remarquer que $\delta^1(v)$, noté $\delta(v)$, correspond aux éléments voisins du pixel P_v, à l'exception de l'élément v.

D'autre part, la saillance d'un pixel est définie comme la valeur maximale des saillances des courbes traversant ce pixel. C'est cette valeur qui est utilisée pour constituer une carte de saillance de l'image.

A.2.2.1 Fonction de qualité

Cette fonction étant définie pour une courbe $\Gamma_N(P, v, \overline{v})$, chaque terme de saillance est exprimé pour un seul des deux brins, $\gamma_N(P, v)$.

Longueur

Le terme de longueur représente la contribution des éléments $e_j,\ j \in [1, N]$ à la saillance de la courbe. Cette contribution est d'autant plus faible que le nombre d'éléments virtuels entre le départ de la courbe et e_j est grand. Soit $\rho_{1,j}$ la contribution individuelle de chaque e_j.

$$\rho_{1,j} \ = \ \rho^{g_{1,j}} \ , \quad avec \ \ \rho < 1$$

Le nombre d'éléments virtuels entre e_1 et e_j est noté $g_{1,j}$.

Le terme de longueur est défini par la somme :

$$\sum_{j=1}^{N} \sigma_j \cdot \rho_{1,j}$$

avec :

$$\sigma_j \;=\; \begin{cases} 1, & \text{si } e_j \text{ est réel} \\ 0, & \text{si } e_j \text{ est virtuel} \end{cases}$$

Le facteur σ_j assure que seuls les éléments réels apportent une contribution à la saillance de la courbe.

Courbure

Le terme de courbure accumule les contributions de courbure locale entre éléments consécutifs depuis le début de la courbe. Dans sa forme continue, ce terme est défini par :

$$\mathcal{C}_{1,j} \;=\; e^{-\int_{e_1}^{e_j} \kappa^2(s)ds}$$

où $\kappa(s) \;=\; (\frac{d\theta}{ds})$ est la courbure à l'abscisse curviligne s le long de la courbe. Cette mesure est bornée, et inversement liée à la courbure de la courbe. Elle vaut en effet 1 pour une ligne droite et décroît vers 0 lorsque la courbure globale tend vers l'infini.

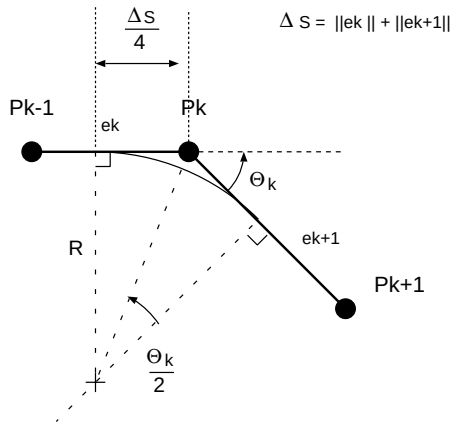

Figure A.7 - *Estimation de la courbure entre deux éléments de connexion consécutifs e_k et e_{k+1}.*

La courbure peut être évaluée de manière discrète entre deux éléments consécutifs e_k et e_{k+1}

$$K_{k,k+1} \;=\; \frac{2\theta_k tan\frac{\theta_k}{2}}{\Delta s}$$

On note $\theta_k \;=\; e_k,\widehat{e_{k+1}}$ la différence d'orientation entre les deux éléments et Δs la longueur totale des deux éléments.

L'approximation de la courbure totale le long des éléments $e_1, e_2, \cdots, e_{j-1}$ est donnée par :

$$C_{1,j} \;=\; \prod_{k=1}^{j-1} f(e_k, e_{k+1}) \;=\; \exp^{-\sum_{k=1}^{j-1} K_{k,k+1}} \qquad \text{(A.1)}$$

avec :

$$f(e_k, e_{k+1}) \;=\; e^{-K_{k,k+1}}$$

La fonction $f(e_k, e_{k+1})$ est, par construction, un ensemble de constantes d'appariement entre deux éléments d'orientation consécutifs. Ces constantes peuvent être évaluées au préalable à partir des combinaisons possibles entre éléments d'orientation.

La fonction de qualité de $\gamma_N(P, v)$ est définie par la somme des contributions locales σ_j de chaque élément, pondérées par la saillance de chaque élément en termes de longueur et de courbure :

$$\mathcal{F}(\gamma_N(P, v)) \;=\; \sum_{j=1}^{N} \sigma_j \cdot \rho_{1,j} \cdot C_{1,j} \qquad \text{(A.2)}$$

Cette définition assure à la fonction de qualité une croissance monotone en fonction de la longueur des groupements et une décroissance monotone en fonction de son énergie (terme de courbure). Elle pénalise la présence de discontinuités (éléments virtuels) ainsi que les courbes trop sinueuses.

A.2.2.2 Forme récursive et fonctions extensibles

L'originalité de la méthode de Shashua et Ullman est d'exprimer la mesure de saillance sous une forme récursive. Ainsi, pour des courbes de longueur N partant dans la direction de l'élément v, la saillance φ_N de cet élément est fonction de la qualité des courbes de longueur $(N-1)$ partant de chacun de ses éléments voisins (figure A.8).

$$\varphi_N(v) \;=\; \underset{e_k \in \delta(v)}{\mathbf{Max}} F(P, v, \varphi_{N-1}(e_k))$$

La fonction F est définie à partir de φ_{N-1} et de constantes représentant la saillance propre des pixels P et P_v. On peut considérer que $\varphi_{N-1}(e_k)$ représente la contribution de l'élément e_k pour la saillance du pixel P.

Afin de calculer cette valeur maximale sans avoir à parcourir exhaustivement l'ensemble des courbes possibles partant de P dans la direction de v, Shashua et Ullman définissent une certaine classe de fonctions, dites *fonctions extensibles*, par la propriété suivante.

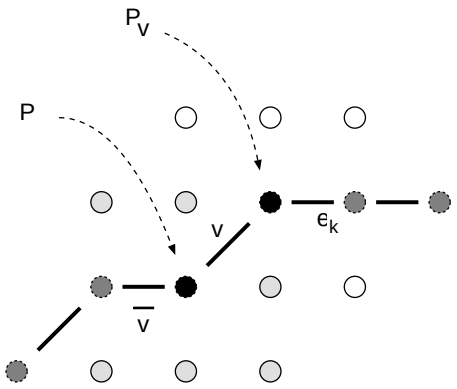

Figure A.8 - *L'élément d'orientation relie les pixels P et P_v. La saillance du meilleur chemin de longueur N partant du pixel P dans la direction de v, est une fonction de la saillance du meilleur chemin de longueur $N-1$ partant du pixel P_v dans la direction de $e_k \in \delta(v)$.*

Définition : Une fonction $\psi_N(\cdot)$, définie sur N valeurs $e_i, e_{i+1}, \cdots, e_{i+N}$ est dite *extensible* si elle vérifie la relation :

$$\underset{\delta^N(e_i)}{\mathbf{Max}}\, \psi_N(e_i, e_{i+1}, \cdots, e_{i+N}) \;=\; \underset{\delta(e_i)}{\mathbf{Max}}\, \psi_1(e_i, \underset{\delta^{N-1}(e_{i+1})}{\mathbf{Max}}\, \psi_{N-1}(e_{i+1}, \cdots, e_{i+N})) \quad (A.3)$$

Cette définition, empruntée à la programmation dynamique [Montanari, 1971] , permet de réduire l'espace de recherche pour une courbe de longueur N partant de v à $(k-1).N$ possibilités au lieu de $(k-1)^N$. Rappelons que $(k-1)$ est le nombre d'orientations possibles autour de l'élément v à partir d'un pixel P. Ce type de fonction assure une construction récursive de courbes optimales autour de chaque pixel.

La fonction F est définie sous la forme suivante :

$$F(P, v, \varphi_{N-1}(e_k)) \;=\; \sigma_v \;+\; \rho_v \cdot C_{1,k} \cdot \varphi_{N-1}(e_k)$$

Ce qui donne, par récursivité :

$$\varphi_N(v) \;=\; \sigma_v \;+\; \rho_v \underset{e_k \in \delta(v)}{\mathbf{Max}} \{ f(v, e_k) \varphi_{N-1}(e_k) \}$$

Définie ainsi, $\varphi_N(v)$ représente bien la meilleure saillance parmi les courbes possibles partant de P dans la direction de v.

On peut définir de même la saillance $\varphi_N(\overline{v})$ du brin opposé. La mesure de saillance à optimiser est donc la somme de ces deux mesures latérales.

$$\Phi N(v, \overline{v}) \;=\; \varphi_N(v) \;+\; \varphi_N(\overline{v})$$

Cette forme récursive ne demande que des calculs locaux à chaque pixel. L'aspect global intervient dans la contribution de chaque voisin au calcul de la valeur de saillance.

A.2.3 Optimisation récursive

Soit un pixel P et les éléments d'orientation qui en dépendent. On associe à chaque élément v une variable d'état $S_v^{(n)}$. Cette variable représente la saillance de la meilleure courbe de longueur n partant de P dans la direction v.

La variable d'état est initialisée par la saillance locale de l'élément v :

$$S_v^{(0)} \;=\; \sigma_v$$

En reprenant l'expression récursive de la mesure de saillance, la valeur de $S_i^{(n)}$ est mise à jour, pour chaque nouvelle itération. Cette mise à jour revient à établir la paire d'éléments $(\ v, e_k\ ),\ e_k \in \delta(v)$ qui maximise l'état de v

$$S_v^{(n+1)} \;=\; \sigma_v \;+\; \rho_v \cdot \mathop{\mathbf{Max}}_{e_k \in \delta(v)} \{ f(v, e_k) S_{e_k}^{(n)} \} \tag{A.4}$$

L'élément e_k choisi est celui qui contribue le plus à l'état de v pour une itération donnée.

Enfin, une courbe traversant P par les directions v et $\overline{v}$, donne à ce pixel la saillance :

$$\Phi^{(n+1)}(\gamma_{n+1}(P, v)) \;=\; S_v^{(n+1)} \;+\; \overline{S}_{\overline{v}}^{(n+1)}$$

Au long de l'optimisation, les éléments présents le long d'une structure courbe reçoivent des contributions fortes dans la direction des tangentes à la courbe. A l'inverse, les éléments isolés reçoivent des contributions d'autant plus faibles que le nombre d'éléments virtuels de leur voisinage est important.

En fin d'optimisation, la saillance de chaque pixel P est la valeur maximale des mesures de saillance des courbes traversant P, soit :

$$\mathcal{S}(P) \;=\; \mathop{\mathbf{Max}}_{v_i \in \mathcal{V}(P)} \Phi(\Gamma(P, v_i, \overline{v_i})) \tag{A.5}$$

Shashua et Ullman apportent la preuve de la convergence de ce type de fonction par un raisonnement sur une courbe fermée. En pratique, le nombre d'itérations dépend de la longueur maximale de discontinuités à remplir. Il faut n itérations pour que deux éléments séparés par n autres puissent contribuer à leurs états mutuels.

A.2.4 Extraction des structures saillantes

De la même manière que pour la mesure directe de Guy et Medioni, cette méthode offre la propriété intéressante d'établir, en plus d'une carte de saillance, une carte de connectivité entre pixels. Les groupements optimaux peuvent être reconstitués en suivant, de proche en proche, les paires d'éléments définies au cours de l'optimisation.

Un pixel, servant de point de départ pour un groupement, définit deux directions privilégiées selon la paire d'éléments de son voisinage qui présente la plus forte

saillance. Le suivi des paires d'éléments, dans chaque direction, permet ainsi de combler les discontinuités. Plusieurs conditions d'arrêt au suivi sont envisageables, comme par exemple la présence d'un cycle ou bien la sortie des limites de l'image.

L'optimisation assure l'existence d'une courbe optimale passant par chaque pixel de l'image. L'ensemble des groupements possibles dans l'image est donc réduit à une seule courbe par pixel. En pratique, les pixels de contours présentant une forte saillance constituent des points de départ suffisants pour la reconstitution des groupements.

Malgré une réduction considérable de l'espace de recherche, l'optimisation par réseau de saillance ne résout pas le problème de l'extraction de groupements significatifs à partir de leur réseau de saillance. En effet, les pixels situés le long d'une structure courbe sont autant de points de départs pour des groupements de saillance semblable. Il en est de même pour tout point de bruit immédiatement voisin d'un contour saillant.

Dans [Shashua et Ullman, 1991] , les auteurs apportent un début de solution en recherchant des groupements optimaux sous la forme d'un ensemble de parcours disjoints dans le réseau d'éléments connectés. Ils démontrent en particulier comment obtenir cette partition en choisissant soigneusement les paires de voisins définies autour de chaque pixel.

En optimisant la saillance sur toutes les courbes possibles traversant chaque point, l'étape précédente définit des paires d'éléments pour des courbes non nécessairement disjointes. Un second niveau d'optimisation est donc nécessaire pour ajuster ces paires de manière à former des groupes disjoints. Ce niveau reprend le même mécanisme d'optimisation en ne changeant que le choix des paires d'éléments et la formule de mise à jour de l'état des éléments.

Une variable d'état $F_v^{(0)}$ est initialisée pour chaque élément d'orientation :

$$F_v^{(0)} \;=\; \sigma_v$$

La valeur de cet état est mise à jour en définissant, localement à chaque pixel, une partition de $\frac{k}{2}$ paires d'éléments dans son voisinage. Les paires sont constituées par ordre décroissant sur les saillances déjà calculées à l'étape n.

Notons $\delta^*(v)$ l'ensemble des voisins de l'élément v qui n'ont pas été appariés. On choisi alors l'élément $e_k \in \delta^*(v)$ tel que :

$$f(v,e_k)F_{e_k}^{(n)} \;=\; \underset{e_j \in \delta^*(v)}{\mathbf{Max}} \{f(v,e_j)F_{e_j}^{(n)}\}$$

La contribution pour la mise à jour est alors :

$$F_v^{(n+1)} \;=\; \sigma_v \;+\; \rho_v \cdot \{f(v,e_k)F_{e_k}^{(n)}\}$$

avec (v,e_k) paire disjointe au voisinage de P.

Ce mécanisme est reproduit jusqu'à ce que les nouvelles saillances $F_v^{(n)}$ soient suffisamment proches des saillances optimales $S_v^{(n)}$ calculées préalablement.

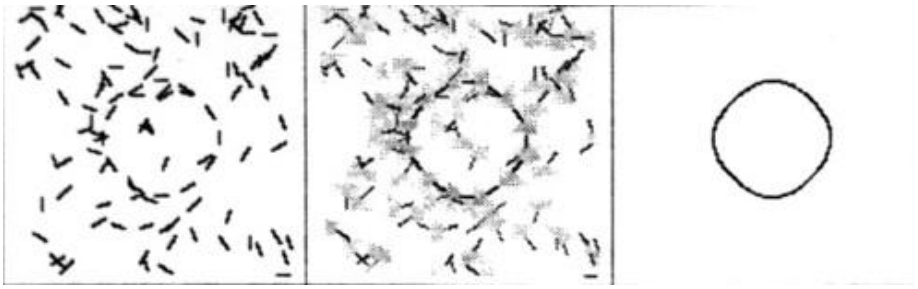

Figure A.9 - *Exemple de carte de saillance et d'extraction du meilleur groupement à partir d'un cercle bruité. Exemple tiré de [Alter et Basri, 1996].*

Cette méthode simple permet de construire dynamiquement les groupes optimaux de l'image. Elle permet également de propager les décisions d'appariement le long de courbes de saillance optimale et d'influencer les décisions d'appariement à l'itération suivante. Si elle donne des résultats satisfaisants en pratique, Shashua et Ullman admettent toutefois qu'elle ne garantit pas une convergence vers une solution optimale.

A.2.5 Discussion

La méthode proposée par Shashua et Ullman est intéressante et originale à plusieurs points de vues.

Le réseau de saillance réduit la complexité du problème de groupement à une optimisation rapide de mesures locales propagées globalement. Sa complexité algorithmique ne dépend que du nombre d'arcs et de sommets constituant le réseau. Dans le cas présent, pour une image de $n \times m$ pixels et un voisinage de k éléments autour de chaque pixel, cette complexité est de l'ordre de $\mathcal{O}(n \cdot m \cdot k^2)$ par itération.

L'optimisation d'une mesure de saillance dense sur l'image, à la fois sur les éléments de contours et les éléments virtuels, permet de fermer les discontinuités des contours. Une conséquence intéressante de cette optimisation est la détection, en chaque point de l'image, de directions privilégiées issues de contributions globales des éléments environnants. Ces directions permettent de guider la reconstitution de groupements par simple suivi de proche en proche.

Il est intéressant de remarquer ici que le résultat de cette optimisation est très proche de celui obtenu avec la mesure globale de Guy et Medioni. En plus d'une valeur de saillance structurelle, chaque point de l'image se retrouve associé à un ensemble de vecteurs, correspondant aux influences des points environnants. La principale différence entre les deux méthodes réside dans la discrétisation des orientations d'influence et le mode de calcul de celles-ci. L'approche de Guy et Medioni demande un vote de la part de tous les pixels, alors que l'optimisation de Shashua et Ullman ne tient compte que de calculs locaux propagés globalement.

A.2.5.1 Applications des réseaux de saillance

Les réseaux de saillance ont fait l'objet de plusieurs extensions et applications, à commencer par Shashua et Ullman. Ceux-ci suggèrent en effet l'extension de la méthode à d'autres fonctions de qualité, en particulier une mesure de saillance reposant sur une différence de courbure.

$$H_v^{(n+1)} \;=\; \sigma_v \;+\; \rho_v \cdot \mathop{\mathbf{Max}}_{e_k \in \delta(v)} \{ \ell(v, e_k, e_l) H_{e_k}^{(n)} \}$$

La fonction $l(v, e_k, e_l)$ est similaire à $f(v, e_k)$ à ceci prés que l'angle θ_k est ici une différence sur trois orientations successives. La comparaison des deux mesures donne des résultats similaires.

Ils proposent également de lisser les groupements en cours d'optimisation en ajustant la position des sommets du réseau. On pourra se reporter à [Shashua, 1988] pour plus de détails sur la méthode et ses applications.

Parmi les applications des réseaux de saillance à d'autres types de problèmes, on peut citer [Subirana-Vilanova et Sung, 1992] avec une extension de la méthode à la définition de squelettes de régions saillantes et [Merlet et Zerubia, 1996] avec une approche semblable adaptée à la recherche de structures courbes sur des images satellites.

A.2.5.2 Problèmes non résolus et remarques

La robustesse des réseaux de saillance et leur simplicité font de cette méthode une bonne approche pour notre premier niveau de groupements. Malgré leurs nombreux avantages, les réseaux de saillance présentent un certain nombre de problèmes qui peuvent en limiter l'utilisation. On pourra trouver dans [Alter et Basri, 1996] une étude quantitative et qualitative des réseaux de saillance. Cette étude concerne en particulier leur stabilité et complexité.

C'est en apportant une réponse à chacun de ces problèmes que nous avons abouti à la méthodologie de groupement par réseau de saillance qui fait l'objet du chapitre 4. Notre contribution à ce type de groupement porte en particulier sur un nouveau formalisme pour la fonction de qualité, une différente procédure d'optimisation et enfin, une heuristique pour l'extraction des groupes les plus importants après optimisation.

Choix du voisinage

Les réseaux de saillance sont définis, a l'origine, à partir de pixels interconnectés. Le choix de la forme du voisinage de ces pixels et le nombre d'éléments d'orientation par voisinage est déterminant pour assurer des groupements suffisamment précis. Ainsi, un voisinage trop petit ne pourra pas garantir une reconstruction fidèle des courbes de l'image. A l'inverse, un voisinage trop grand pèse lourdement sur la complexité de l'optimisation. Chaque pixel

doit en effet garder en mémoire l'état des k éléments d'orientations de son voisinage. Un voisinage trop important entraîne des besoins de mémoire tels qu'il rend la méthode inapplicable pour des images de taille importante.

Ces problèmes de complexité peuvent être réduits de manière significative en généralisant cette méthode d'optimisation à un réseau de chaînes de pixels connectées par leurs extrémités et dotées d'un voisinage dynamique afin de ne conserver que les connexions utiles au groupement.

Mesure de saillance structurelle

La mesure de saillance dépend uniquement de termes de longueur et de courbure, critères géométriques qui se montrent insuffisants dans de nombreuses situations. De nombreuses ambiguïtés apparaissent en particulier lorsque les éléments de contours sont répartis uniformément, ou bien forment des zones de bruit relativement denses par rapport aux structures saillantes. L'ajout de critères de co-circularité ou d'orientation des tangentes permet un contrôle plus important sur le type de courbes détectées.

De plus, la conjugaison des termes de courbure et de longueur au sein d'une même contribution rend difficile l'évaluation de l'influence de chacun de ces termes sur la fonction de qualité. L'un des effets indésirables de ce type de fonction est d'accélérer la convergence sur des structures localement saillantes et de la ralentir de manière trop importante sur des structures plus globales. D'autre part, la multiplication des contributions par le facteur σ_i dans la fonction de qualité interdit toute distinction entre éléments de contours lorsque ceux-ci sont virtuels. Ainsi, deux éléments de connexion virtuels auront la même contribution, nulle, quel que soit leur courbure locale. Multiplier le terme de courbure par σ_i revient à ignorer la courbure des éléments virtuels et accorde une saillance forte à des groupements irréguliers (figure A.10).

L'utilisation d'une fonction de qualité additive, inspirée du formalisme des contours actifs, nous permet de mieux contrôler l'influence de chaque terme de saillance sur la mesure finale.

Reconstitution des groupements

La méthode originale de Shashua et Ullman donne des résultats intéressants dans le cas d'une seule structure saillante dans un environnement bruité. Les auteurs présentent peu de résultats sur le groupement après optimisation, en particulier sur des images réelles.

De nombreux problèmes apparaissent lorsque cette méthode est appliquée à des scènes contenant plusieurs structures d'intérêt. Le suivi des meilleures connexions n'est pas suffisant pour extraire des structures cohérentes depuis les contours de la scène. En particulier, le suivi d'un groupement optimisé peut aisément "sauter" d'une structure à l'autre en cas de jonctions, d'occlusions

ou bien de structures saillantes parallèles (figure A.11). Cette méthode est donc plus adaptée à la détection de "la" meilleure structure.

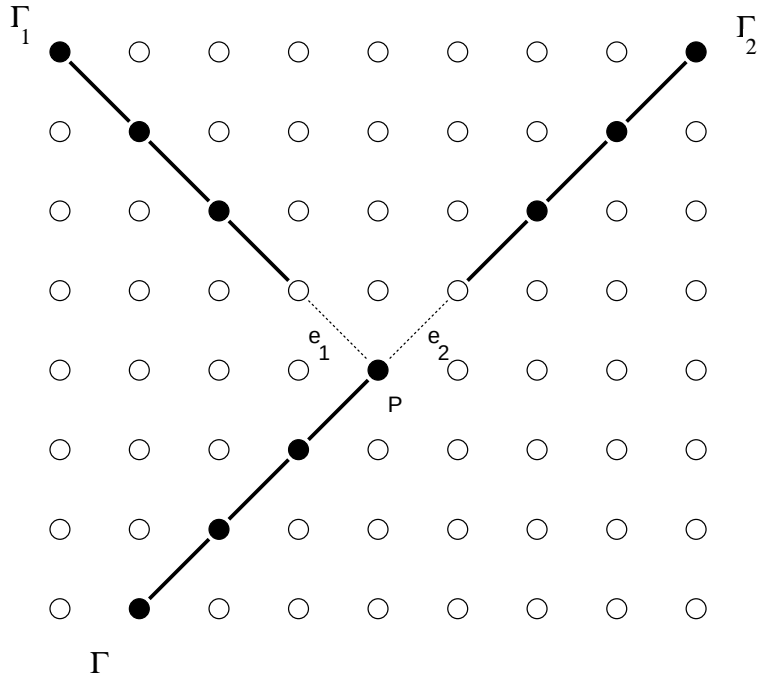

Figure A.10 - *En cas d'éléments de connexion virtuels, la courbure locale ne doit pas être ignorée. En multipliant les termes de saillance, les éléments virtuels* e_1 *et* e_2 *apportent ici une contribution nulle (* $\sigma_1 = \sigma_2 = 0$ *) en* P *alors que pour* Γ *,* Γ_2 *est d'évidence un meilleur groupement que* Γ_1.

L'application d'un second niveau d'optimisation pour extraire un ensemble de groupements disjoints ne répond qu'en partie aux problèmes de suivi. La méthode proposée impose des contraintes trop fortes sur les groupements, surtout autour d'intersections entre courbes. Une sélection automatique des groupes les plus représentatifs tenant compte de ces nombreux problèmes reste encore à définir.

Enfin, cette méthode présente un certain nombre d'incohérences avec des expériences psycho-visuelles qu'il serait bon de rectifier. La mesure de saillance est, par exemple, trop sensible à la taille et la répartition des discontinuités sur une courbe. Un cercle fragmenté régulièrement peut obtenir une saillance inférieure à un incomplet mais continu. Un autre exemple est donné par l'importance trop grande des contributions de voisinage. Des pixels de bruit proches d'un contour peuvent obtenir une saillance importante du fait de la proximité d'une structure linéaire.

Nous proposons enfin des critères de sélection des meilleurs groupements en

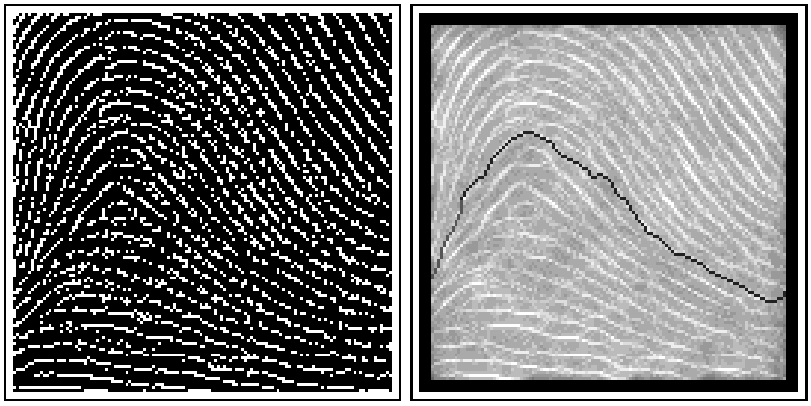

Figure A.11 - *Des structures linéaires rendent difficile le groupement par simple suivi des meilleurs éléments de connexion. Comme le montre cet exemple sur une image d'empreinte digitale, le parcours du graphe de connexions bascule indifféremment d'une structure à l'autre.*

fonction de leur qualité globale et de leur point de départ. Ces critères permettent une plus grande discrimination entre les structures réellement saillantes et les groupements bénéficiant de ces effets indésirables.

Chacun de ces aspects est exposé d'une manière plus détaillée dans le chapitre 4, consacré à la définition d'un formalisme générique pour les réseaux de saillance et son application au groupement de pixels et de chaînes de pixels.

Annexe B

Résultats complémentaires

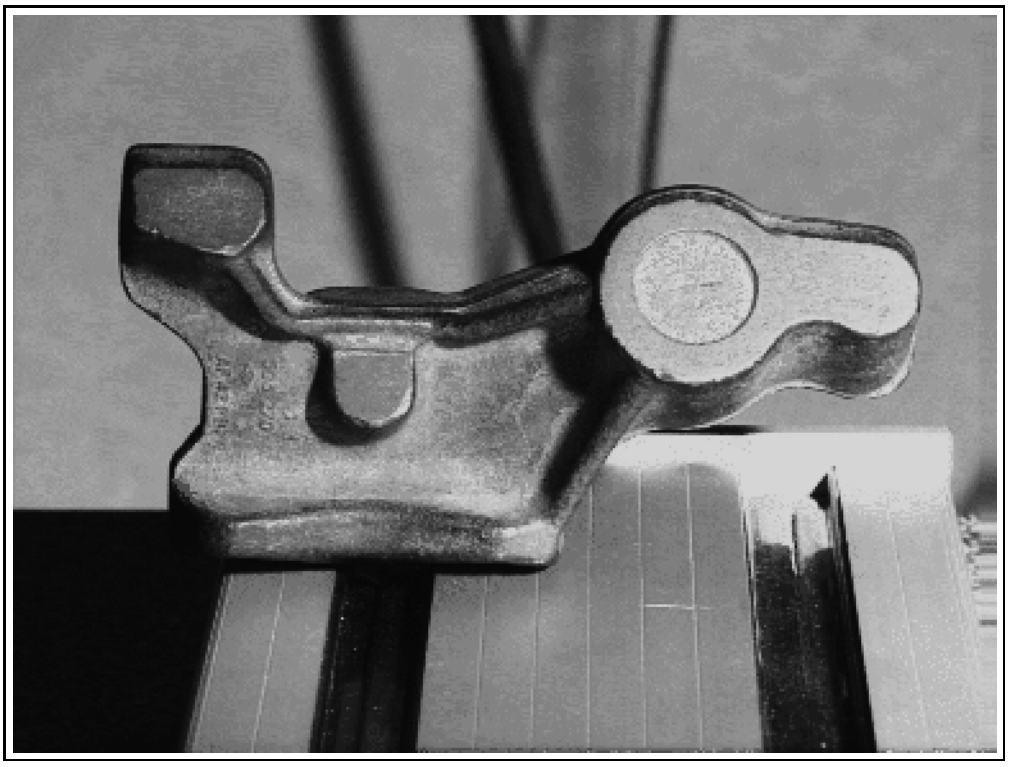

Figure B.1 - *Pièce industrielle - Cette scène est intéressante car elle présente des structures rectilignes et courbes à différentes tailles. La texture de la pièce et l'atténuation de l'arrière plan introduisent de plus de nombreuses perturbations - Photographie © Projet Syntim, INRIA.*

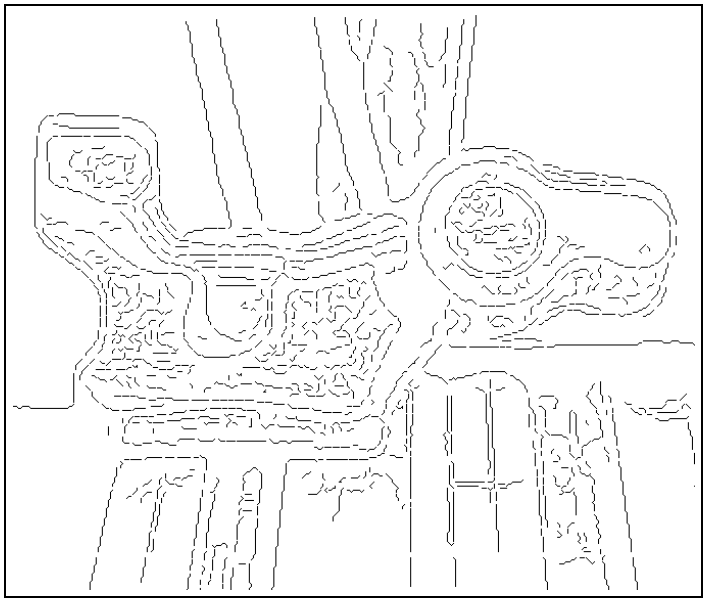

Figure B.2 - *Détection de contours par filtre de Deriche -* $\alpha = 1$ *- 1408 chaines élémentaires.*

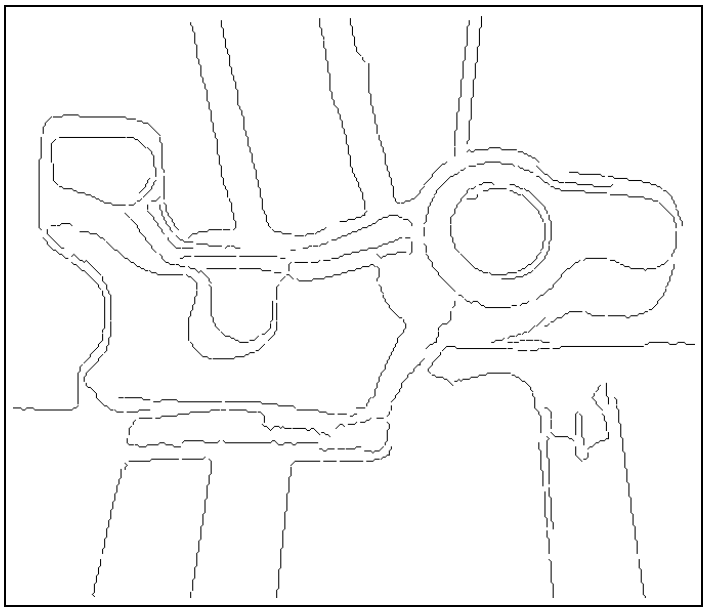

Figure B.3 - *351 chaines sélectionnées automatiquement aprés optimisation du réseau de saillance sur les chaînes de contours. Ces chaînes correspondent à 34 groupes saillants.*

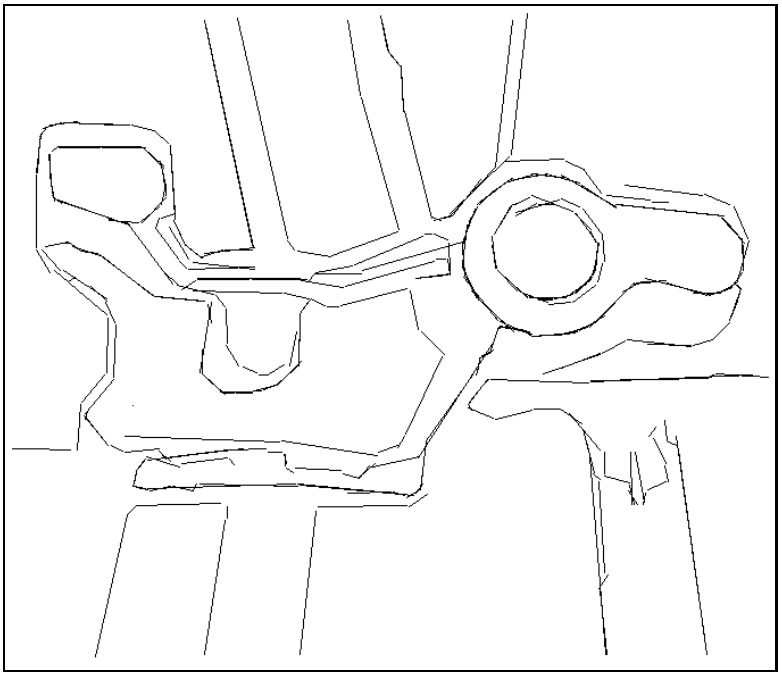

Figure B.4 - *Détection de segments avant groupement - 489 segments - Seuil de découpage récursif* $\epsilon^{\surd} = 3$

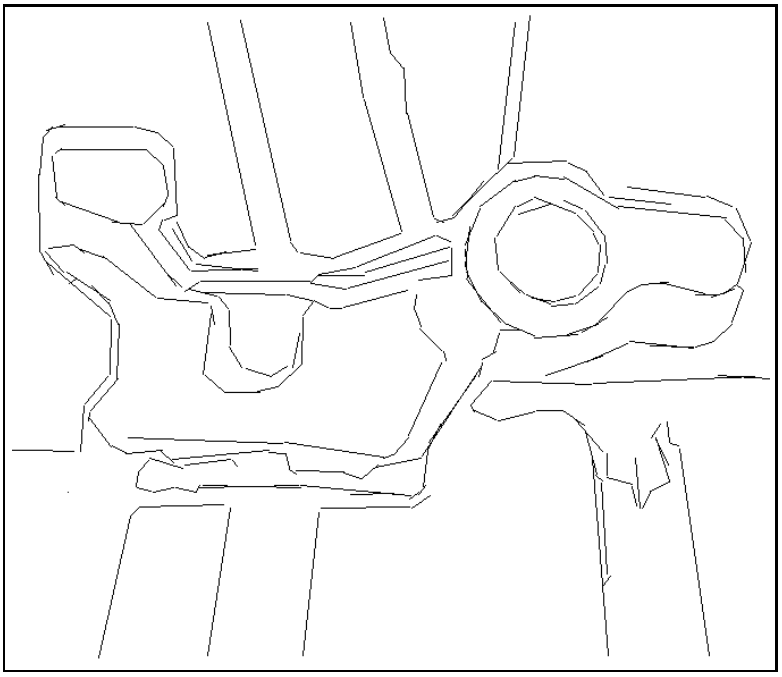

Figure B.5 - *Aprés groupement - 239 segments*

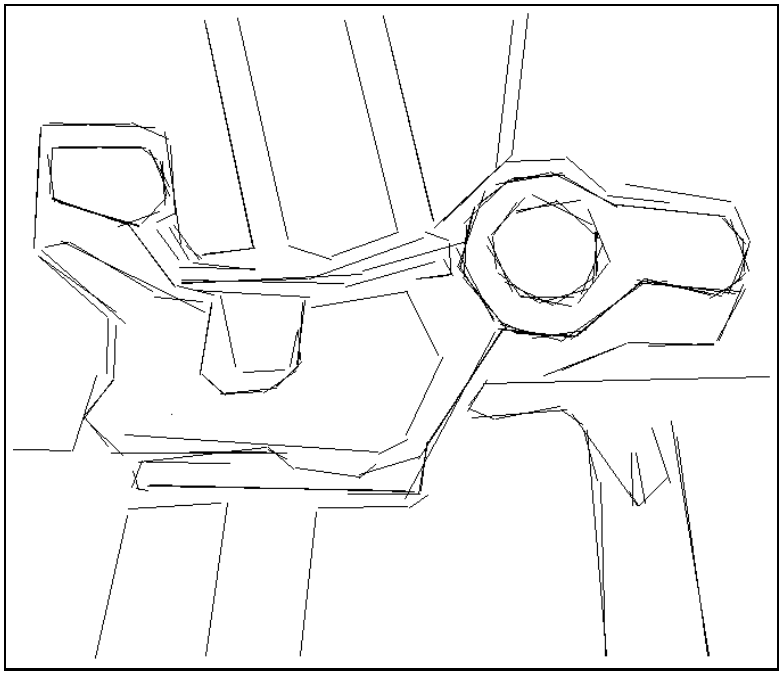

Figure B.6 - *Détection de segments avant groupement - 282 segments - Seuil de découpage récursif* $\epsilon^{\surd} = 11$

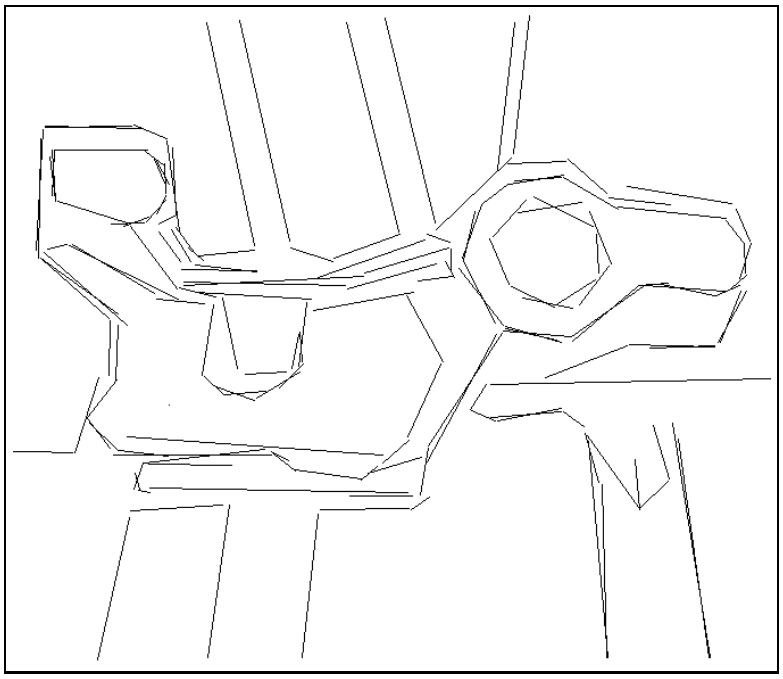

Figure B.7 - *Aprés groupement - 161 segments*

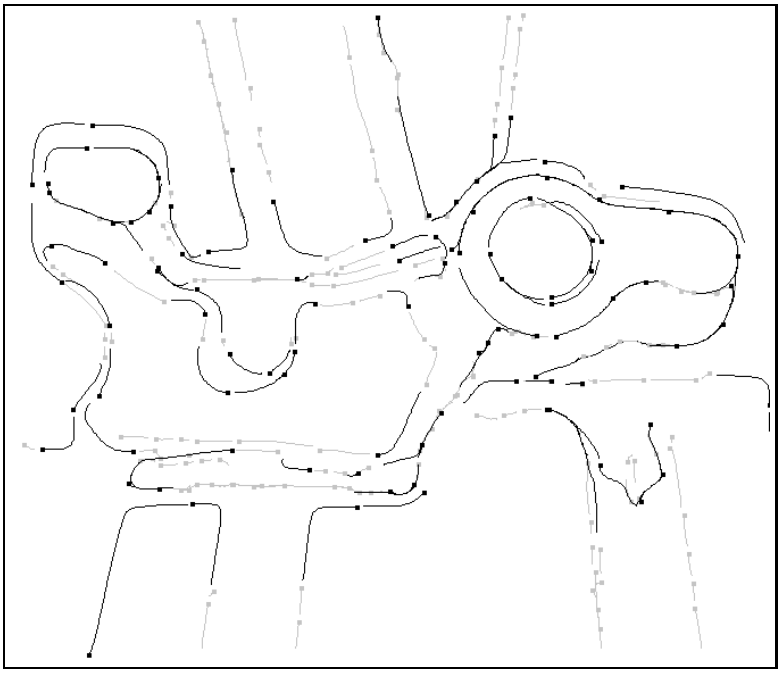

Figure B.8 - *Détection d'arcs - 281 arcs élémentaires - Echelle de lissage* $\alpha = 0,5$

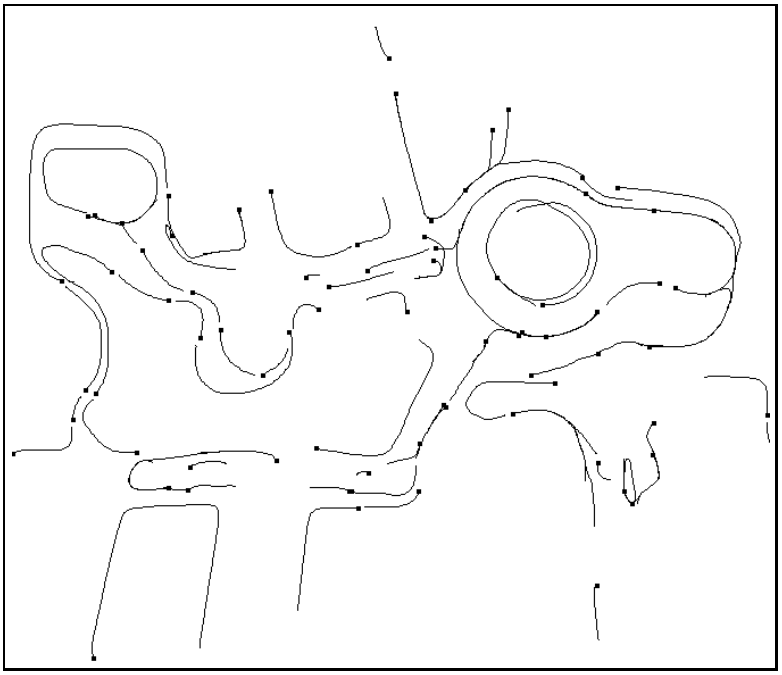

Figure B.9 - *Groupement de 80 paires d'arcs co-circulaires*

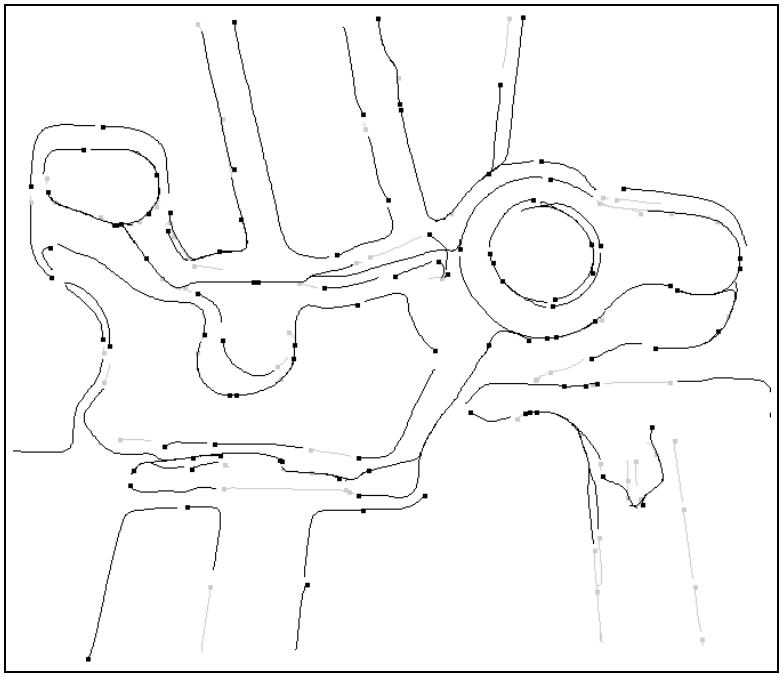

Figure B.10 - *Détection d'arcs - 220 arcs élémentaires - Echelle de lissage* $\alpha = 0,125$

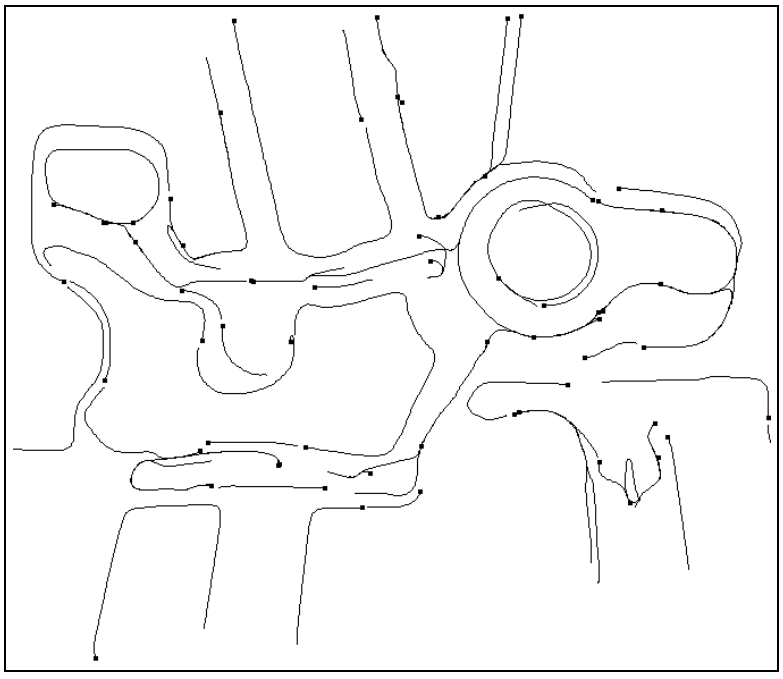

Figure B.11 - *Groupement de 95 paires d'arcs co-circulaires*

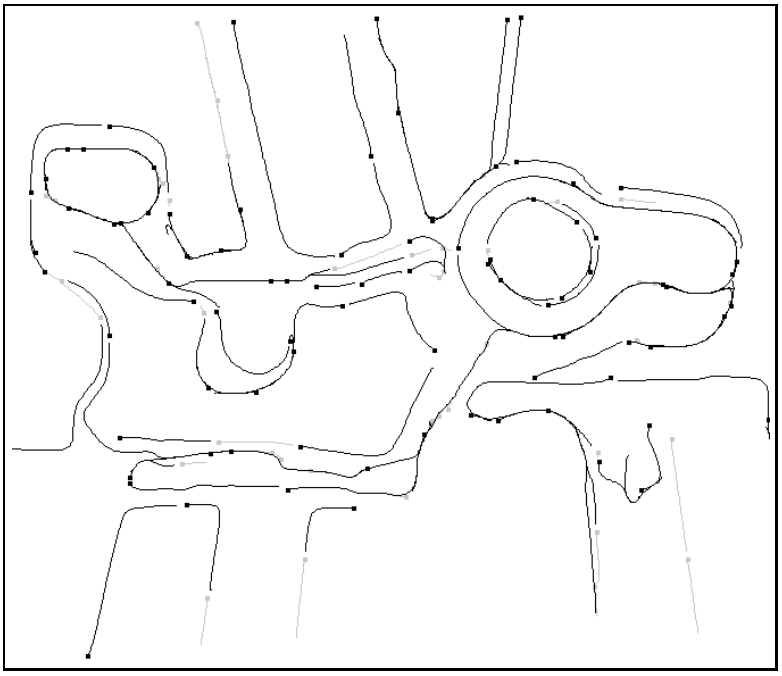

Figure B.12 - *Détection d'arcs - 175 arcs élémentaires - Echelle de lissage* $\alpha = 0,07$

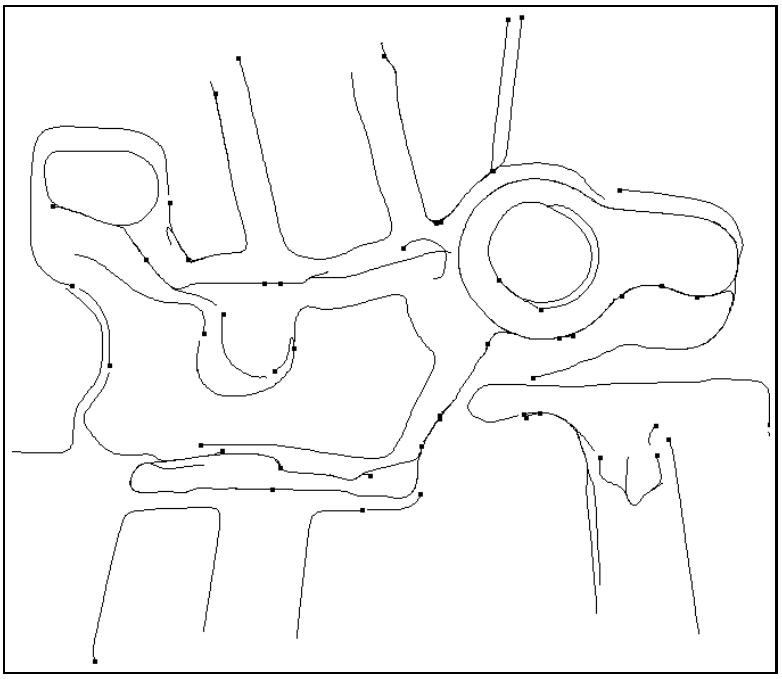

Figure B.13 - *Groupement de 97 paires d'arcs co-circulaires*

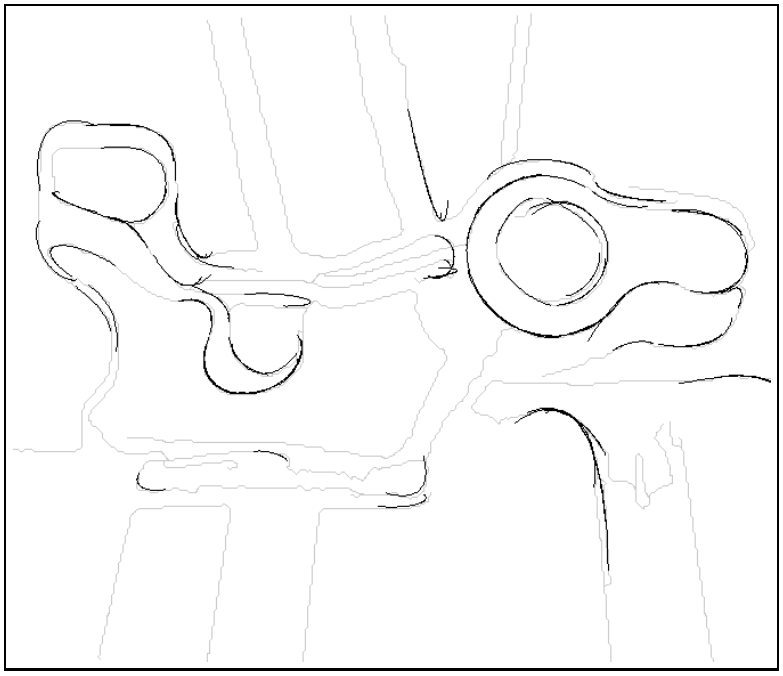

Figure B.14 - *Analyse des arcs élémentaires - détection de 87 arcs d'ellipse - Echelle de lissage* $\alpha = 0,5$

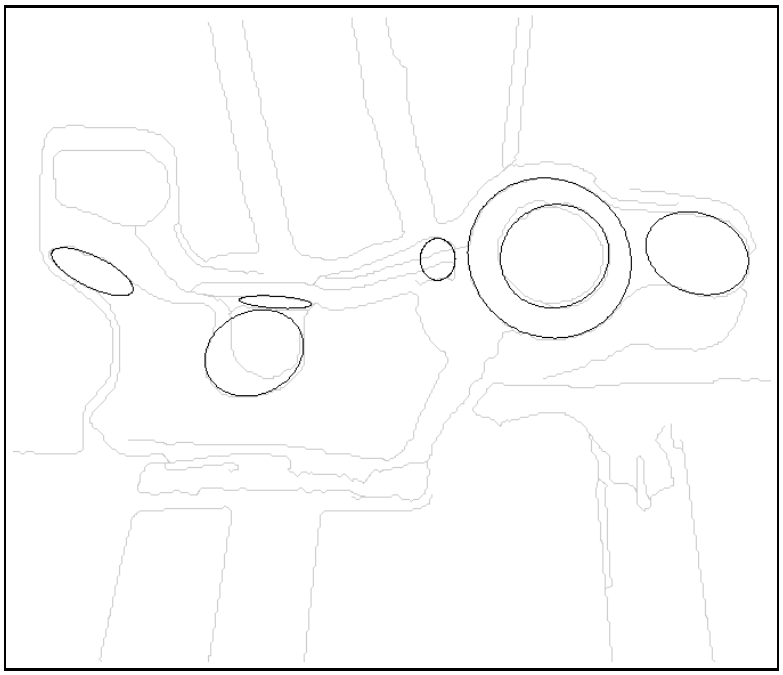

Figure B.15 - *Les arcs les plus longs forment 7 hypothèses d'ellipses.*

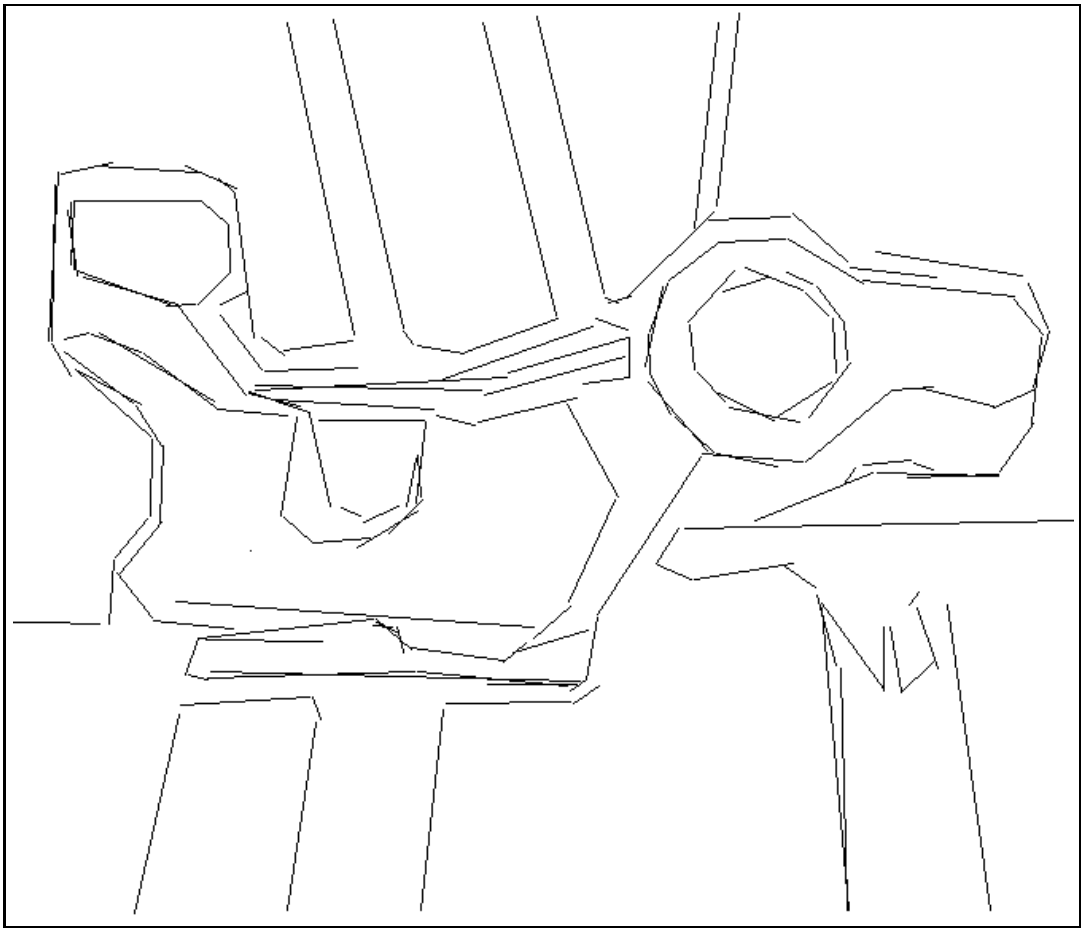

Figure B.16 - *Préliminaire à la construction des hypothèses de jonctions. Détection et groupement de segments - 169 segments extraits à partir de 29 groupements sur 351 chaines - Seuil de découpage récursif* $\epsilon^{\surd} = 5$.

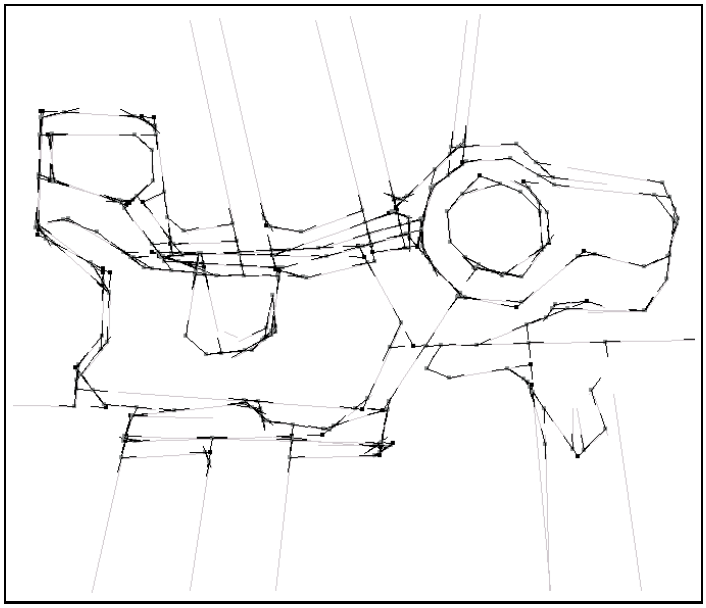

Figure B.17 - *Détection de 744 jonctions élémentaires à partir de 169 segments.*

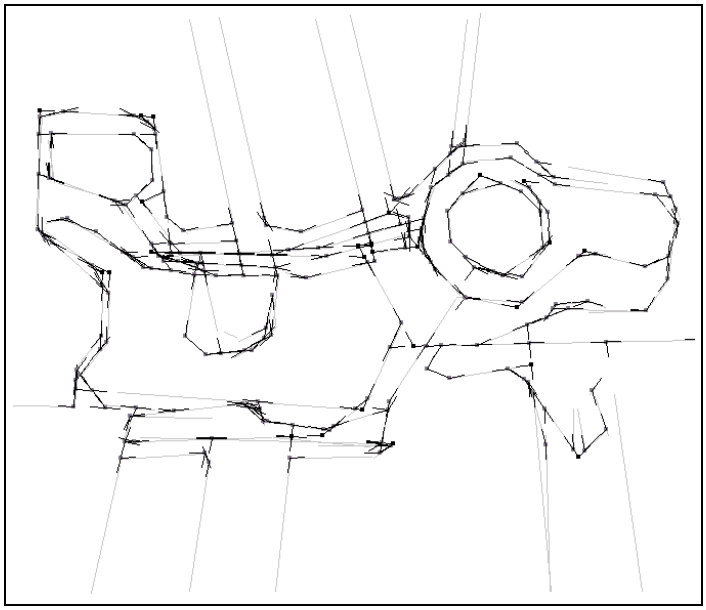

Figure B.18 - *Simplification des hypothèses de jonctions - 224 jonctions aprés groupement.*

Liste des figures

Liste des tableaux

Liste des Algorithmes

Bibliographie

[Aggarwal et Martin, 1994] Aggarwal, J. K. et Martin, W. N. (1994). The role of representation and reconstruction in vision : Is it a matter of definition ? *CVGIP : Image Understanding*, 60(1):100–102. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Allman et Kaas, 1971] Allman, J. M. et Kaas, J. H. (1971). A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey. *Brain Res.*, 81:85–105.

[Aloimonos, 1990] Aloimonos, Y. (1990). Purposive and qualitative active vision. pages 346–360, Atlantic City, NJ. 10th ICPR.

[Aloimonos, 1994] Aloimonos, Y. (1994). What i have learned. *CVGIP : Image Understanding*, 60(1):74–85. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Aloimonos et Rosenfeld, 1992] Aloimonos, Y. et Rosenfeld, A. (1992). *Encyclopedia of Artificial Intelligence*, chapter Visual Recovery. Wiley, New York.

[Aloimonos *et al.*, 1988] Aloimonos, Y., Weiss, I., et Bandopadhay, A. (1988). Active vision. *International Journal of Computer Vision*, 2(1):333–356.

[Alquier, 1994] Alquier, L. (1994). Groupement perceptuel appliqué à la reconnaissance de courbes en vision artificielle. Master's thesis, Université de Montpellier. (in French).

[Alquier et Montesinos, 1997] Alquier, L. et Montesinos, P. (1997). Recursive perceptual grouping for 3d object reconstruction from 2d scenes. Lappeenranta, Finland. 10th SCIA.

[Alter et Basri, 1996] Alter, T. D. et Basri, R. (1996). Extracting salient curves from images : An analysis of the saliency network. pages 13–20, San Francisco, CA. CVPR.

[Aoyama et Kawagoe, 1991] Aoyama, H. et Kawagoe, M. (1991). A piecewise linear polygonal approximation method preserving visual feature points of original figures. *CVGIP: Graphical Models and Image Processing*, 53(5):435–446.

[Arbogast, 1990] Arbogast, E. (1990). La representation des contours et leur segmentation. Rapport de recherche 115, LIFIA.

[Armande *et al.*, 1995] Armande, N., Monga, O., et Montesinos, P. (1995). Thin nets and crest lines: Application to satellite data and medical images. Washington, D.C., USA. 2nd ICIP.

[Attneave, 1954] Attneave, F. (1954). Some informational aspects of visual perception. *Psychology Review*, 61:183–193.

[Ayache et Lustman, 1991] Ayache, N. et Lustman, F. (1991). Trinocular stereo for robotics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):73–85.

[Bajcsy, 1988] Bajcsy, R. (1988). Active perception. *Proc. IEEE (Special issue on computer visoion)*, 76(8):996–1005.

[Bajcsy et Lieberman, 1976] Bajcsy, R. et Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics and Image Processing*, 5(1):52–67.

[Ballard et Brown, 1982] Ballard, D. H. et Brown, C. M. (1982). *Computer Vision*. Prentice Hall.

[Ballard et Brown, 1992] Ballard, D. H. et Brown, C. M. (1992). Principles of animate vision. *CVGIP Image Understanding*, 56(1):3–21.

[Batchelor et Whelan, 1997] Batchelor, B. G. et Whelan, P. F. (1997). *Intelligent Vision Systems for the Industry*. Springer-Verlag.

[Beis et Lowe, 1994] Beis, J. S. et Lowe, D. G. (1994). Learning indexing functions for 3d model-based object recognition. pages 275–280, Seattle, WA. CVPR.

[Beis et Lowe, 1997] Beis, J. S. et Lowe, D. G. (1997). Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. pages 1000–1006, San Juan, PR. CVPR.

[Bengtsson et Eklundh, 1991] Bengtsson, A. et Eklundh, J. O. (1991). Shape representation by multiscale contour approximation. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 13(1):85–93.

[Berger, 1991] Berger, M. O. (1991). *Contours actifs : modélisation, comportement et convergence*. PhD thesis, Institut National Polytechnique de Lorraine, INRIA Lorraine.

[Biederman, 1987] Biederman, I. (1987). Recognition by components. London. 1st ICCV.

[Blake et Zisserman, 1987] Blake, A. et Zisserman, A. (1987). *Visual reconstruction*. MIT Press, Cambridge-MA.

[Blaszka et Deriche, 1994a] Blaszka, T. et Deriche, R. (1994a). A model based method for characterization and location of curved image features. Rapport de recherche 2451, INRIA-Sophia Antipolis.

[Blaszka et Deriche, 1994b] Blaszka, T. et Deriche, R. (1994b). Recovering and characterizing image features using an efficient model based approach. Rapport de recherche 2422, INRIA Sophia Antipolis.

[Bolle et Vemuri, 1991] Bolle, R. M. et Vemuri, B. C. (1991). On three-dimensional surface reconstruction methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):1–13.

[Brillault, 1992] Brillault, B. (1992). High level 3d structures from a single view. *Image and Vision Computing*, 10(7):508–520.

[Brooks, 1987] Brooks, R. (1987). Intelligence without representation. Proceedings. Workshop on the Foundations of Artificial Intelligence.

[Brown, 1994] Brown, C. M. (1994). Toward general vision. *CVGIP : Image Understanding*, 60(1):89–91. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Brunnström *et al.*, 1996] Brunnström, K., Eklundh, J. O., et Uhlin, T. (1996). Active fixation for scene exploration. *International Journal of Computer Vision*, 17(2):137–162.

[Burt, 1988] Burt, P. J. (1988). Smart sensing within a pyramid vision machine. *Proc. IEEE (Special issue on computer visoion)*, 76(8):1006–1015.

[Cabrera et Meer, 1996] Cabrera, J. et Meer, P. (1996). Unbiased estimation of ellipses by bootstraping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):752–756.

[Canny, 1983] Canny, J. F. (1983). Finding edges and lines in images. Rapport de recherche 720, MIT.

[Caselles *et al.*, 1997] Caselles, V., Kimmel, R., et Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79.

[Castan *et al.*, 1990] Castan, S., Zhao, J., et Shen, J. (1990). Optimal filter for edge detection, methods and results. pages 13–17, Antibes, France. 1st ECCV.

[Chabbi, 1993] Chabbi, H. (1993). *Construction de facettes 3D par stéréovision intégrant des principes de géométrie projective*. PhD thesis, Institut National Polytechnique de Lorraine.

[Cham et Cipolla, 1995] Cham, T. J. et Cipolla, R. (1995). Symmetry detection through local skewed symmetries. *Image and Vision Computing*, 13(5):439–450.

[Chang et Aggarwal, 1997] Chang, Y. L. et Aggarwal, J. K. (1997). Line correspondances from cooperating spatial and temporal grouping processes for a sequence of images. *Computer Vision and Image Understanding*, 67(2):186–201.

[Chen *et al.*, 1996] Chen, P. C., Tsai, W. C., et Hwang, S. Y. (1996). A graded approach to shape representation. *Journal of Visual Communication and Image Representation*, 7(2):105–115.

[Christensen et Madsen, 1994] Christensen, H. I. et Madsen, C. B. (1994). Purposive reconstruction. *CVGIP : Image Understanding*, 60(1):103–108. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Cooper, 1993] Cooper, M. C. (1993). Interpretation of line drawings of complex objects. *Image and Vision Computing*, 11(2):82–90.

[Cowie et Perrott, 1993] Cowie, R. et Perrott, R. (1993). From line drawings to impressions of 3d objects : developing a model to account for the shapes that people see. *Image and Vision Computing*, 11(6):342–352.

[Cox *et al.*, 1993] Cox, I. J., Rehg, J. M., et Hingorani, S. (1993). A bayesian multiple-hypothesis approach to edge grouping and contour segmentation. *International Journal of Computer Vision*, 11(1):5–24.

[de Jong et Buurman, 1992] de Jong, J. J. et Buurman, J. (1992). Learning 3d object descriptions from a set of stereo vision observations. volume 1, pages 768–771. 11th ICPR.

[Debray, 1992] Debray, R. (1992). *Vie et mort de l'image*. Gallimard.

[Denasi *et al.*, 1992] Denasi, S., Quaglia, G., et Rinaudi, D. (1992). The use of perceptual organization in the prediction of geometric structures. *Pattern Recognition Letters*, 13:529–539.

[Deriche, 1987] Deriche, R. (1987). Using canny's criteria to derive a recursive implemented optimal edge detector. *International Journal of Computer Vision*, 1:167–187.

[Deriche, 1990] Deriche, R. (1990). Techniques d'extraction de contours. Rapport de recherche, INRIA Sphia-Antipolis.

[Deriche et Faugeras, 1996] Deriche, R. et Faugeras, O. (1996). Les équations aux dérivées partielles en traitement des images et vision par ordinateur. *Reconnaissance des Formes et Intelligence Artificielle - Traitement du Signal*, 13(6):551–577.

[Deriche et Giraudon, 1993] Deriche, R. et Giraudon, G. (1993). A computational approach for corner and vertex detection. *International Journal of Computer Vision*, 10(2):101–124.

[Devernay, 1995] Devernay, F. (1995). A non-maxima suppression method for edge detection with sub-pixel accuracy. Rapport de recherche 2724, INRIA-Sophia Antipolis.

[Dickson, 1991] Dickson, W. (1991). Feature grouping in a hierarchical probabilistic network. *Image and Vision Computing*, 9(1):51–57.

[Dudek et Tsotsos, 1997] Dudek, G. et Tsotsos, J. K. (1997). Shape representation and recognition from multiscale curvature. *Computer Vision and Image Understanding*, 68(2):170–189.

[Duncan et Birkhölzer, 1992] Duncan, J. S. et Birkhölzer, T. (1992). Reinforcement of linear structure using parametrized relaxation labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(5):502–515.

[Dunn *et al.*, 1994] Dunn, D., Higgins, W. E., et Wakeley, J. (1994). Texture segmentation using 2-d gabor elementary functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):130–149.

[Edelman, 1994] Edelman, S. (1994). Representation without reconstruction. *CVGIP : Image Understanding*, 60(1):92–94. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Eggert et Bowyer, 1993] Eggert, D. et Bowyer, K. (1993). Computing the perspective projection aspect graph of solids of revolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2):109–128.

[Ellis *et al.*, 1991] Ellis, T., Abbood, A., et Brillault, B. (1991). Ellipse detection and matching with uncertainty. pages 136–144. BMVC.

[Eric et Grimson, 1993] Eric, W. et Grimson, L. (1993). Why stereo-vision is not always about 3d reconstruction. Rapport de recherche 1435, MIT.

[Fairney et Fairney, 1994] Fairney, D. P. et Fairney, P. T. (1994). On the accuracy of point curvature estimators in a discrete environment. *Image and Vision Computing*, 12(5):259–265.

[Faugeras et Robert, 1994] Faugeras, O. et Robert, L. (1994). What can two images tell us about a third one? Rapport de recherche, INRIA.

[Felleman et Van Essen, 1991] Felleman, D. et Van Essen, D. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47.

[Fermüller et Kropatsch, 1992] Fermüller, C. et Kropatsch, W. (1992). Hierarchical curve representation. volume 3, pages 143–146. IAPR.

[Fiorio, 1995] Fiorio, C. (1995). *Approche interpixel en analyse d'images, une topologie et des algorithmes de segmentation.* PhD thesis, Université de Montpellier II.

[Fischler, 1994] Fischler, M. (1994). The modeling and representation of visual information. *CVGIP : Image Understanding*, 60(1):98–99. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Fischler et Bolles, 1986] Fischler, M. A. et Bolles, R. C. (1986). Perceptual organization and curve partitionning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):100–105.

[Fischler et Wolf, 1994] Fischler, M. A. et Wolf, H. C. (1994). Locating perceptually salient points on planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):113–129.

[Freeman, 1988] Freeman, H. (1988). *Machine vision, algorithms, architectures, and systems*, volume 20 of *Perspectives in computing*. Academic Press.

[Freeman, 1989] Freeman, H. (1989). *Machine vision for inspection and measurement*, volume 24 of *Perspectives in computing*. Academic Press.

[Gao et Wong, 1993] Gao, Q. G. et Wong, A. K. C. (1993). Curve detection based on perceptual organization. *Pattern Recognition*, 26(7):1039–1046.

[Garnesson et Giraudon, 1991] Garnesson, P. et Giraudon, G. (1991). Polygonal approximation : Overview and perspectives. Rapport de recherche 1621, INRIA, Sophia Antipolis.

[Geman *et al.*, 1990] Geman, S., Geman, D., Graffigne, C., et Dong, P. (1990). Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:609–628.

[Gibson, 1950] Gibson, J. J. (1950). *The Perception of the Visual World*. Houghton Mifflin, Boston.

[Gibson, 1979] Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.

[Giraudon, 1987] Giraudon, G. (1987). Chainage efficace de contours. Rapport de recherche 605, INRIA.

[Gordon, 1989] Gordon, I. E. (1989). *Theories of Visual Perception*. John Wiley & Sons.

[Goshtasby, 1993] Goshtasby, A. (1993). Design and recovery of 2-d and 3-d shapes using rational gaussian curves and surfaces. *International Journal of Computer Vision*, 10(3):233–256.

[Gregory, 1974] Gregory, R. L. (1974). *Philosophy of Psychology*, chapter 9 - Perception as hypotheses. Macmillan, London.

[Gros, 1994] Gros, P. (1994). Using quasi-invariants for automatic model building and object recognition : an overview. Workshop NSF/ARPA on 3D object representation in Computer Vision.

[Gros et Mohr, 1992] Gros, P. et Mohr, R. (1992). Automatic object modelization in computer vision. volume 5, pages 385–400, Berne, Suisse. Workshop on Advances in Structural and Syntactic Pattern Recognition.

[Gross et Boult, 1994] Gross, A. D. et Boult, T. E. (1994). Analyzing skewed symmetries. *International Journal of Computer Vision*, 13(1):91–111.

[Grossberg et Mingolla, 1985] Grossberg, S. et Mingolla, E. (1985). Neural dynamics of perceptual grouping : Textures, boundaries and emergent segmentations. *Perception and Psychophysics*, 38(2):141–171.

[Grossmann, 1987] Grossmann, P. (1987). Depth from focus. *Pattern Recognition Letters*, 5:63–69.

[Gunn et Nixon, 1996] Gunn, S. R. et Nixon, M. S. (1996). A robust snake implementation : A dual active contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1):63–68.

[Gupta *et al.*, 1993] Gupta, A., Chaudhury, S., et Parthasarathy, G. (1993). A new approach for aggregating edge points into line segments. *Pattern Recognition*, 26(7):1069–1086.

[Guy et Medioni, 1996] Guy, G. et Medioni, G. (1996). Inferring global perceptual contours from local features. *International Journal of Computer Vision*, 20(1):113–133.

[Harris et Stephen, 1988] Harris, C. et Stephen, M. (1988). A combined corner and edge detector. pages 147–151, Manchester. Proceedings 4th Alvey Conference.

[Hartley et Sturm, 1997] Hartley, R. I. et Sturm, P. (1997). Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157.

[Havaldar *et al.*, 1996] Havaldar, P., Medioni, G., et Stein, F. (1996). Perceptual grouping for generic recognition. *International Journal of Computer Vision*, 20(1):59–80.

[Hérault, 1991] Hérault, L. (1991). *Réseaux de neurones récursifs pour l'optimisation combinatoire : application à la théorie des graphes et à la vision par ordinateur.* PhD thesis, Institut National Polytechnique de Grenoble.

[Hérault et Horaud, 1992] Hérault, L. et Horaud, R. (1992). Figure-ground discrimination by mean field annealing. pages 58–66, Santa Margherita Ligure, Italy. 2nd ECCV.

[Horaud et Skordas, 1989] Horaud, R. et Skordas, T. (1989). Stereo correspondance through feature grouping and maximal cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1168–1180.

[Horaud *et al.*, 1990] Horaud, R., Veillon, F., et Skordas, T. (1990). Finding geometric and relational structures in an image. volume 90, pages 374–384, Antibes, France. 1st ECCV.

[Horn, 1975] Horn, B. K. P. (1975). *Psychology of Computer Vision*, chapter Shape from shading. McGraw-Hill, New York.

[Hubel et Weisel, 1962] Hubel, D. H. et Weisel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Psychology*, 166:106–154.

[Huddleston et Ben-Arie, 1993] Huddleston, J. N. et Ben-Arie, J. (1993). Grouping edgels into structural entities using circular symmetry, the distributed hough transform and probabilistic non-accidentalness. *CVGIP : Image Understanding*, 57(2):227–242.

[Huynh et Owens, 1994] Huynh, D. Q. et Owens, R. A. (1994). Line labeling and region segmentation in stereo image pairs. *Image and Vision Computing*, 12(4):213–225.

[Ip et Wong, 1997] Ip, H. S. et Wong, W. H. (1997). Detecting perceptually parallel curves : criteria and force-driven optimization. *Computer Vision and Image Understanding*, 68(2):190–208.

[Jacobs, 1992] Jacobs, D. W. (1992). *Recognizing 3D objects using 2D images.* PhD thesis, Massachusetts Institute of Technology.

[Jacobs, 1996] Jacobs, D. W. (1996). Robust and efficient detection of salient convex groups. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):23–37.

[Jacot-Descombes et Pun, 1997] Jacot-Descombes, A. J. et Pun, T. (1997). Asynchronous perceptual grouping : From contours to relevant 2d structures. *Computer Vision and Image Understanding*, 66(1):1–24.

[Jolion, 1994] Jolion, J. M. (1994). Computer vision methodologies. *CVGIP : Image Understanding*, 59(1):53–71.

[Jones, 1997] Jones, G. A. (1997). Constraint, optimization, and hierarchy : Reviewing stereoscopic correspondance of complex features. *Computer Vision and Image Understanding*, 65(1):57–78.

[Joseph, 1994] Joseph, S. H. (1994). Unbiaised least square fitting of circular arcs. *CVGIP: Graphical Models and Image Processing*, 56(5):424–432.

[Julesz, 1960] Julesz, B. (1960). Binocular depth perception of computer generated patterns. *Bell. Syst. Tech. J.*, (39):1125–1162.

[Kass *et al.*, 1987] Kass, M., Witkin, A., et Terzopoulos, D. (1987). Snakes : Active contour models. In *Third International Conference on Computer Vision*, pages 259–268.

[Kender, 1978] Kender, J. (1978). Shape from texture : A brief overview and a new aggregation transform. DARPA Image Understanding Workshop.

[Kitchen et Rosenfeld, 1982] Kitchen, L. et Rosenfeld, A. (1982). Grey-level corner detection. *Pattern Recognition Letters*, pages 95–102.

[Kovács, 1996] Kovács, I. (1996). Gestalten of today : Early processing of visual contours and surfaces. *Behav. Brain Res. Invited Review*, 82(1):1–11.

[Kropatsch, 1995] Kropatsch, W. G. (1995). Equivalent contraction kernels and the domain of dual irregular pyramids. Rapport de recherche 42, Technical University of Vienna, Institute for Automation, Treitlstr. 3/1832, A-1040 Vienna AUSTRIA.

[Lai et Chin, 1993] Lai, K. F. et Chin, R. T. (1993). On regularization, formulation and initialization of the active contour models. In *First Asian Conference on Computer Vision*, pages 542–545, Osaka.

[Laurentini, 1997] Laurentini, A. (1997). How many 2d silhouettes does it take to reconstruct a 3d object. *Computer Vision and Image Understanding*, 67(1):81–87.

[Leonardis et Bajcsy, 1992] Leonardis, A. et Bajcsy, R. (1992). Finding parametric curves in an image. pages 653–657, Santa Margherita Ligure, Italy. 2nd ECCV.

[Lindeberg et Li, 1997] Lindeberg, T. et Li, M. X. (1997). Segmentation and classification of edges using minimum description length approximation and complementary junction cues. *Computer Vision and Image Understanding*, 67(1):88–98.

[Lowe, 1985] Lowe, D. G. (1985). *Perceptual Organization and Visual Recognition*. Kluwer Academic publisher, Hingham MA 02043, USA.

[Lu et Jain, 1992] Lu, Y. et Jain, R. C. (1992). Reasoning about edges in scale space. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(4):450–467.

[Luong et Faugeras, 1993] Luong, Q. T. et Faugeras, O. (1993). Self-calibration of a stereo rig from unknown camera motion and point correspondances. Rapport de recherche 2014, INRIA, Sophia-Antipolis.

[Malik et Maydan, 1989] Malik, J. et Maydan, D. (1989). Recovering three-dimensional shape from a single image of curved objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):555–566.

[Malik et Perona, 1990] Malik, J. et Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of Optical Society of America*, 7(5):923–932.

[Mangin, 1994] Mangin, F. (1994). *Amélioration de la détection de contours en imagerie artificielle par un modèle coopératif multi-résolution.* PhD thesis, Université de Nice, Sophia Antipolis.

[Mangin *et al.*, 1992] Mangin, F., Berthod, M., et Zerubia, J. (1992). A cooperative network for contour grouping. The Hague, the Netherland. 11th ICPR.

[Marr, 1982] Marr, D. C. (1982). *Vision.* Freeman, Oxford.

[Matas et Kittler, 1993] Matas, J. et Kittler, J. (1993). Junction detection using probabilistic relaxation. *Image and Vision Computing*, 11(4):197–202.

[Merlet et Zerubia, 1996] Merlet, N. et Zerubia, J. (1996). New prospects in line detection by dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):426–431.

[Mohan et Nevatia, 1989] Mohan, R. et Nevatia, R. (1989). Using perceptual organization to extract 3d structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1121–1139.

[Mohan et Nevatia, 1992] Mohan, R. et Nevatia, R. (1992). Perceptual organization for scene segmentation and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):616–635.

[Mokhtarian et Mackworth, 1992] Mokhtarian, F. et Mackworth, A. K. (1992). A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):789–805.

[Monga *et al.*, 1995] Monga, O., Armande, N., et Montesinos, P. (1995). Crest lines and thin net extraction. volume 1, pages 287–295, Uppsala, Sweden. 9th SCIA.

[Monga et Horaud, 1993] Monga, O. et Horaud, R. (1993). *Vision par Ordinateur : outils fondamentaux.* Hermes.

[Montanari, 1971] Montanari, U. (1971). On the optimal detection of curves in noisy pictures. *Communications of the ACM*, 14(5):335–345.

[Montesinos et Alquier, 1996] Montesinos, P. et Alquier, L. (1996). Perceptual organization with active contour functions: application to aerial and medical images. In *13th ICPR*, volume 2, Vienna, Austria.

[Montesinos et Blanc, 1994] Montesinos, P. et Blanc, H. V. (1994). Perceptual grouping of continuation. Rapport de recherche, EERIE/LERI, Nimes, France.

[Montesinos et Datteny, 1997] Montesinos, P. et Datteny, S. (1997). Sub-pixel accuracy using recursive filtering. Lappeenranta, Finland. 10th SCIA.

[Nalwa, 1988] Nalwa, V. S. (1988). Line-drawing interpretation: straight lines and conic sections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):514–529.

[Neuenschwander *et al.*, 1997] Neuenschwander, W. M., Fua, P., Iverson, L., Székely, G., et Kübler, O. (1997). Ziplock snakes. *International Journal of Computer Vision*, 25(3):191–201.

[Nevatia, 1982] Nevatia, R. (1982). *Machine Perception*. Prentice Hall.

[Nguyen et Levine, 1996] Nguyen, Q. L. et Levine, M. D. (1996). Representing 3-d objects in range images using geons. *Computer Vision and Image Understanding*, 63(1):158–168.

[Noble, 1988] Noble, J. A. (1988). Finding corners. *Image and Vision Computing*, 6:121–128.

[Palmer *et al.*, 1997] Palmer, P. L., Kittler, J., et Petrou, M. (1997). An optimizing line finder using a hough transform algorithm. *Computer Vision and Image Understanding*, 67(1):1–23.

[Palmer, 1983] Palmer, S. E. (1983). *Human and Machine Vision*, chapter The psychology of perceptual organization : a transformational approach, pages 269–339. New York.

[Parent et Zucker, 1989] Parent, P. et Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):823–839.

[Parodi, 1996] Parodi, P. (1996). The complexity of understanding line drawings of origami scenes. *International Journal of Computer Vision*, 18(2):139–170.

[Parodi et Piccioli, 1996] Parodi, P. et Piccioli, G. (1996). 3d shape reconstruction by using vanishing points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2):211–217.

[Pavlidis, 1981] Pavlidis, T. (1981). Algorithms for graphics and image processing. pages 275–298. Springer-Verlag, New York.

[Pentland, 1986] Pentland, A. P. (1986). Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28:293–331.

[Pentland, 1987] Pentland, A. P. (1987). A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:523–531.

[Pentland et Bichsel, 1994] Pentland, A. P. et Bichsel, M. (1994). *Handbook of Pattern Recognition and Image Processing : Computer Vision*, chapter Chapter 6 : Extracting shape from shading, pages 161–183. Academic Press, Inc.

[Perona et Malik, 1990] Perona, P. et Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. In *IEEE Transactions on Pattern Recognition and Machine Intelligence*, volume 12, pages 629–639.

[Perrett et Oram, 1993] Perrett, D. I. et Oram, M. W. (1993). Neurophysiology of shape processing. *Image and Vision Computing*, 11(6):317–333.

[Pollard *et al.*, 1991] Pollard, S. B., Porrill, J., et Mayhew, J. E. (1991). Recovering partial 3d wire frames descriptions from stereo data. *Image and Vision Computing*, 9(1):58–65.

[Pomerantz, 1981] Pomerantz, J. (1981). *Perceptual Organization*, chapter Perceptual Organization in Information Processing. Lawrence Erlbaum Associates, Hillsdale, N.J.

[Ponce, 1988] Ponce, J. (1988). Ribbons, symmetries and skewed symmetries. pages 1,074–1,079, Mass. Proc. Image Understanding Workshop.

[Pope, 1994] Pope, A. R. (1994). Model-based object recognition : A survey of recent research. Rapport de recherche 94-04, University of British Columbia.

[Pope et Lowe, 1993] Pope, A. R. et Lowe, D. G. (1993). Learning 3d object recognition models from 2d images. pages 35–39, Raleigh, North Carolina. AAAI, Fall Symposium : Machine Learning and Pattern Recognition.

[Pope et Lowe, 1996] Pope, A. R. et Lowe, D. G. (1996). Learning appearance models for object recognition. pages 201–219, Cambridge, England. International Workshop ob Object Representation for Computer Vision.

[Posch, 1992] Posch, S. (1992). Detecting skewed symmetries. volume 1, pages 602–606. IAPR.

[Princen *et al.*, 1994] Princen, J., Illingworth, J., et Kittler, J. (1994). Hypothesis testing : A framework for analysing and optimizing hough transform performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(4):329–341.

[Pun, 1992] Pun, T. (1992). *Advances in Machine Vision : Strategies and Applications*, chapter Electromagnetic Models for Perceptual Grouping. World Scientific Publishing Co.

[Ramesh, 1994] Ramesh, J. (1994). Expansive vision. *CVGIP : Image Understanding*, 60(1):86–88. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Rattarangsi et Chin, 1992] Rattarangsi, A. et Chin, R. T. (1992). Scale-based detection of corners of planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(4):430–449.

[Ray et Ray, 1992] Ray, B. et Ray, K. (1992). An algorithm for polygonal approximation of digitized curves. *Pattern Recognition Letters*, 13:489–496.

[Regier, 1991] Regier, T. (1991). Line labeling and junction labeling: A coupled system for image interpretation. *IJCAI*, 91:1305–1310.

[Richetin *et al.*, 1991] Richetin, M., Dhome, M., Lapresté, J. T., et Rives, G. (1991). Inverse perspective transform using zero-curvature contour points : application to the localization of some generalized cylinders from a single view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):185–192.

[Robert, 1997] Robert, A. (1997). From contour completion to image schemas : A modern perspective on gestalt psychology. Rapport de recherche CogSci. UCSD 97.02, Department of Cognitive Science, University of California, San Diego.

[Roberts, 1968] Roberts, L. G. (1968). *Optical and Electro-Optical Information Processing*, chapter Machine perception of three-dimensional solids, pages 159–197. MIT Press, Cambridge, Mass.

[Rohr, 1992] Rohr, K. (1992). Modeling and identification of characteristic intensity variations. *Image and Vision Computing*, 10(2):66–76.

[Rosin, 1997] Rosin, P. L. (1997). Techniques for assessing polygonal approximations of curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):659–666.

[Roth et Levine, 1992] Roth, G. et Levine, M. D. (1992). Geometric primitive extraction using a genetic algorithm. pages 640–643, Champaign, IL, USA. CVPR.

[Roth et Levine, 1993] Roth, G. et Levine, M. D. (1993). Extracting geometric primitives. *CVGIP Image Understanding*, 58(1):1–22.

[Sandini et Grosso, 1994] Sandini, G. et Grosso, E. (1994). Why purposive vision? *CVGIP : Image Understanding*, 60(1):109–112. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Sapiro, 1997] Sapiro, G. (1997). Color snakes. *Computer Vision and Image Understanding*, 68(2):247–253.

[Sarkar, 1994] Sarkar, S. (1994). Tracking 2d structures using perceptual organizational principles. Rapport de recherche, University of South Florida.

[Sarkar et Boyer, 1992] Sarkar, S. et Boyer, K. L. (1992). Computing perceptual organization using voting method and graphical enumeration. volume 1, pages 263–267. IAPR.

[Sarkar et Boyer, 1993a] Sarkar, S. et Boyer, K. L. (1993a). Integration, inference and management of spatial information using bayesian networks : Perceptual organization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):256–274.

[Sarkar et Boyer, 1993b] Sarkar, S. et Boyer, K. L. (1993b). Perceptual organization in computer vision : A review and a proposal for a classificatory structure. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2):382–399.

[Sarkar et Boyer, 1994] Sarkar, S. et Boyer, K. L. (1994). Using perceptual inference networks to manage vision processes. volume A, pages 808–810, Jerusalem, Israël. 12th ICPR.

[Saund, 1991] Saund, E. (1991). Identifying salient circular arcs on curves. *CVGIP Image Understanding*, 58(3):327–337.

[Schreiber et Ben-Bassat, 1996] Schreiber, I. et Ben-Bassat, M. (1996). Feg structures for representation and recognition of 3-d polyhedral objects. *International Journal of Computer Vision*, 18(3):211–232.

[Sekuler et Blake, 1985] Sekuler, R. et Blake, R. (1985). *Perception.* A.A.Knopf, New York.

[Shashua, 1988] Shashua, A. (1988). Structural saliency : The detection of globally salient structures using a locally connected network. Master's thesis, Weizmann Institute of Science, Rehovot, 76100, Israël.

[Shashua et Ullman, 1988] Shashua, A. et Ullman, S. (1988). Structural saliency: The detection of globally salient structures using a locally connected network. pages 321–327, Tampa, FL, USA. 2nd ICCV.

[Shashua et Ullman, 1991] Shashua, A. et Ullman, S. (1991). Grouping contours by iterated pairing network. In Lippmann, R., Moody, J., et Touretzky, D. e., editors, *Advances in Neural Information Processing Systems*, volume 3, pages 335–341. Morgan Kaufmann publishers.

[Shiu, 1990] Shiu, Y. C. (1990). Experiments with perceptual grouping. In *SPIE, Proc. Intelligent Robots and Computer Vision IX : Algorithms and Techniques*, volume 1381, pages 130–141, Boston, Massachusetts.

[Shpitalni et Lipson, 1996] Shpitalni, M. et Lipson, H. (1996). Identification of faces in a 2d line drawing projection of a wireframe object. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18(10):1000–1011.

[Shrikhande et Stockman, 1989] Shrikhande, N. et Stockman, G. (1989). Surface orientation from a projected grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):650–655.

[Sklansky et Gonzalez, 1980] Sklansky, J. et Gonzalez, V. (1980). Fast polygonal approximation of digitized curves. *Pattern Recognition*, 12:327–331.

[Smith et Brady, 1995] Smith, S. M. et Brady, J. M. (1995). Susan - a new approach to low level image processing. Rapport de recherche TR95SMS1c, Defense Research Agency, Farnborough, Hampshire, UK.

[Startchik *et al.*, 1994] Startchik, S., Bost, J. M., Rauber, C., Milanese, R., et Pun, T. (1994). Automatic construction of invariant geometric representations for object recognition. Rapport de recherche, University of Geneva.

[Straforini *et al.*, 1993] Straforini, M., Coelho, C., et Campani, M. (1993). Extraction of vanishing points from images of indoor and outdoor scenes. *Image and Vision Computing*, 11(2):91–99.

[Straforini *et al.*, 1992] Straforini, M., Coelho, C., Campani, M., et Torre, V. (1992). The recovery and understanding of a line drawing from indoor scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):298–303.

[Strauss, 1992] Strauss, O. (1992). *Perception de l'environnement par vision en lumière structurée : segmentation des images par poursuite d'indices.* PhD thesis, Université Montpellier II.

[Subirana-Vilanova et Sung, 1992] Subirana-Vilanova, J. B. et Sung, K. K. (1992). Multi-scale vector-ridge-detection for perceptual organization without edges. A.I. Memo 1318, Massachusetts Institute of Technology.

[Tai *et al.*, 1993] Tai, A., Kittler, J., Petrou, M., et Windeatt, T. (1993). Vanishing point detection. *Image and Vision Computing*, 11(4):240–245.

[Tarel, 1996] Tarel, J. P. (1996). Reconstruction globale et robuste de facettes 3d. Rapport de recherche, INRIA Rocquencourt.

[Tarr et Black, 1994a] Tarr, M. J. et Black, M. J. (1994a). A computational and evolutionary perspective on the role of representation in vision. *CVGIP : Image Understanding*, 60(1):65–73.

[Tarr et Black, 1994b] Tarr, M. J. et Black, M. J. (1994b). Reconstruction and purpose. *CVGIP : Image Understanding*, 60(1):113–118. Reply to responses about 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Terzopoulos, 1988] Terzopoulos, D. (1988). The computation of visible surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):417–439.

[Thornber et Williams, 1997] Thornber, K. K. et Williams, L. R. (1997). Orientation, scale, and discontinuity as emergent properties of illusory contour shape. Fort Lauderdale, FL. Association of Researchers in Vision and Opthalmology Annual Meeting.

[Tomita et Koizumi, 1992] Tomita, F. et Koizumi, M. (1992). A step toward generic object recognition. volume 1, pages 632–636. IAPR.

[Trivedi et Rosenfeld, 1989] Trivedi, M. et Rosenfeld, A. (1989). On making computers "see". *IEEE Trans. Systems Man Cybern.*, 19(6):1333–1335.

[Tsang *et al.*, 1994] Tsang, W., Yuen, P., et Lam, F. (1994). Detection of dominant points on object boundary : a discontinuity approach. *Image and Vision Computing*, 12(9):547–57.

[Tsotsos, 1994] Tsotsos, J. K. (1994). There is no one way to look at vision. *CVGIP : Image Understanding*, 60(1):95–97. Reply to 'A Computational and Evolutionary Perspective on the Role of Representation in Vision', by Tarr, M. J. and Black, M. J.

[Tupin *et al.*, 1996] Tupin, F., Gouinaud, C., Maître, H., Crettez, J.-P., et Nicolas, J.-M. (1996). Détection de structures linéaires sur des images ros. *Reconnaissance des Formes et Intelligence Artificielle - Traitement du Signal*, 13(6):635–650.

[Ulupinar et Nevatia, 1993] Ulupinar, F. et Nevatia, R. (1993). Perception of 3-d surfaces from 2-d contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(1):3–18.

[Urago *et al.*, 1992] Urago, S., Berthod, M., et Zerubia, J. (1992). Restauration d'images de contours incomplets par modélisation de champs de markov. Rapport de recherche 1688, INRIA.

[Urago *et al.*, 1995] Urago, S., Zerubia, J., et Berthod, M. (1995). A markovian model for contour grouping. *Pattern Recognition*, 28(5):683–693.

[Venkateswar et Chellappa, 1995] Venkateswar, V. et Chellappa, R. (1995). Hierarchical stereo and motion correspondance using feature groupings. *International Journal of Computer Vision*, 15:245–269.

[Wall et Danielson, 1984] Wall, K. et Danielson, P. E. (1984). A fast sequential method for polygonal approximation of digitized curves. *CVGIP*, 28:220–227.

[Wechsler, 1990] Wechsler, H. (1990). *Computational Vision*. Academic Press.

[Weinshall et Werman, 1997] Weinshall, D. et Werman, M. (1997). On view likehood and stability. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):97–108.

[Wertheimer, 1923] Wertheimer, M. (1923). *Untersuchungen zur Lehre von der Gestalt II*, pages 301–350. Number 4. (Republished in Ellis, W.D. (ed) (1938) A Source Book of Gestalt Psychology, London:Routledge and Kegan Paul).

[Williams et Shah, 1992] Williams, D. J. et Shah, M. (1992). A fast algorithm for active contours and curvature estimation. *CVGIP: Computer Vision and Image Understanding*, 55(1):14–26.

[Williams et Thornber, 1997] Williams, L. et Thornber, K. (1997). A comparison of measures for detecting natural shapes in cluttered backgrounds. Ft. Lauderdale, FL. Assoc. of Researchers in Vision and Opthalmology (ARVO) Annual Meeting.

[Williams et Jacobs, 1994] Williams, L. R. et Jacobs, D. W. (1994). Stochastic completion fields : A neural model for illusory contour shape and salience. (référence incomplète).

[Witkin et Tenenbaum, 1983] Witkin, A. P. et Tenenbaum, J. M. (1983). *Human and Machine Vision*, chapter On the role of structure in vision, pages 481–543. Beck, Hope and Rosenfeld.

[Witkin et Tenenbaum, 1986] Witkin, A. P. et Tenenbaum, J. M. (1986). *From Pixels to Predicates*, chapter 7 - On Perceptual Organization, pages 149–169. Alex P. Pentland.

[Wolfson, 1990] Wolfson, H. J. (1990). Model based object recognition by geometric hashing. Antibes, France. 1st ECCV.

[Wong *et al.*, 1991] Wong, K. C., Kittler, J., et Illingworth, J. (1991). Heuristically guided polygon finding. pages 400–407. BMVC.

[Worring et Smeulders, 1993] Worring, M. et Smeulders, A. W. M. (1993). Digital curvature estimation. *CVGIP Image Understanding*, 58(3):366–382.

[Wu et Wang, 1993] Wu, W.-Y. et Wang, M.-J. J. (1993). Detecting the dominant points by the curvature-based polygonal approximation. *CVGIP: Graphical Models and Image Processing*, 55(2):79–88.

[Wuescher et Boyer, 1991] Wuescher, D. W. et Boyer, K. L. (1991). Robust contour decomposition using a constant curvature criterion. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 13(1):41–51.

[Ylä-Jääski et Ade, 1992] Ylä-Jääski, A. et Ade, F. (1992). Line segment ribbons and their grouping. volume 1, pages 750–754. IAPR.

[Zakia, 1997] Zakia, R. D. (1997). *Perception and Imaging*. Focal Press.

[Zeki, 1977] Zeki, S. (1977). Colour coding in the superior temporal sulcus of the rhesus monkey visual cortex. Number 197, pages 195–223. Proc. Roy. Soc. London. Sec. B.

[Zerroug et Nevatia, 1996a] Zerroug, M. et Nevatia, R. (1996a). Three-dimensional descriptions based on the analysis of the invariant and quasi-invariant properties of some curved-axis generalized cylinders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):237–253.

[Zerroug et Nevatia, 1996b] Zerroug, M. et Nevatia, R. (1996b). Volumetric descriptions from a single intensity image. *International Journal of Computer Vision*, 20(1):11–42.

[Zhang *et al.*, 1994] Zhang, R., Tsai, P. S., Cryer, J. E., et Shah, M. (1994). Analysis of shape from shading techniques. pages 377–384, Seattle, WA. CVPR.

[Zhang, 1993] Zhang, Z. (1993). Le problème de la mise en correspondance: L'état de l'art. Rapport de recherche 2146, INRIA, Sophia-Antipolis.

[Ziou, 1991] Ziou, D. (1991). Line detection using an optimal iir filter. *Pattern Recognition*, 24(6):465–478.

[Zucker, 1983] Zucker, S. W. (1983). *Human and Machine Vision*, chapter Computational and psychological experiments in grouping: early orientation selection, pages 545–567. Beck, Hope and Rosenfeld.

[Zucker *et al.*, 1989] Zucker, S. W., Dobbins, A., et Iverson, L. (1989). Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, 1:68–81.

Index

Résumé

Une définition possible pour la vision par ordinateur est l'élaboration automatique de raisonnements à partir d'images à l'aide d'ordinateurs. Dans ce contexte, établir une ou plusieurs représentations à partir de l'image d'une scène joue un rôle fondamental.

Le système visuel humain présente de nombreux mécanismes dont le rôle est de guider en permanence la perception dans un flot continu d'informations visuelles. A un niveau psychologique le groupement perceptuel désigne la faculté, démontrée par la vision humaine, à organiser certains éléments visuels en groupes perceptuellement significatifs.

Notre travail se place dans le contexte d'une utilisation de groupement perceptuel en vision artificielle afin de rendre compte de la structure d'une image à partir de ses contours. Nous proposons une approche récursive pour extraire les éléments visuels les plus importants et fournir une base de représentation pour des processus de haut niveau d'interprétation.

Nous montrons finalement comment ces éléments perceptuels de représentation permettent de définir des candidats à la mise en correspondance de deux images. Des résultats sur des images synthétiques et sur des scènes réelles illustrent l'intérêt de notre approche.

Mots-Clés : Vision par ordinateur, groupement perceptuel, analyse de scènes, détection de courbes, représentation hiérarchique, mise en correspondance, optimisation combinatoire, programmation dynamique, saillance structurelle.

Abstract

A possible definition for Computer Vision is the automatic inference of decisions from pictures with the help of computers. In this context, extracting one or many representations from the image of a scene plays a fundamental role.

Human Vision shows the existence of numerous mechanisms, the purpose of which is to continuously guide perception through the constant flow of visual information. At a psychological level, more difficult to express in computational terms, one possible contribution to Computer Vision comes from Perceptual Grouping.

The work described in this manuscript uses the perspective of perceptual grouping to structure images into significant visual elements. We propose a progressive approach to extract visually important features from the contours of an image and give useful elements of representation of the scene for a higher recognition process.

We finally show how these visual elements can be used for the detection and matching of junctions between two images. Various results on synthetic and real images confirm the relevance of our approach.

Keywords : Computer vision, perceptual grouping, scene representation, curve detection, hierarchical representation, feature matching, combinatorial optimization, dynamic programming, hierarchical representation, structural saliency.