

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.

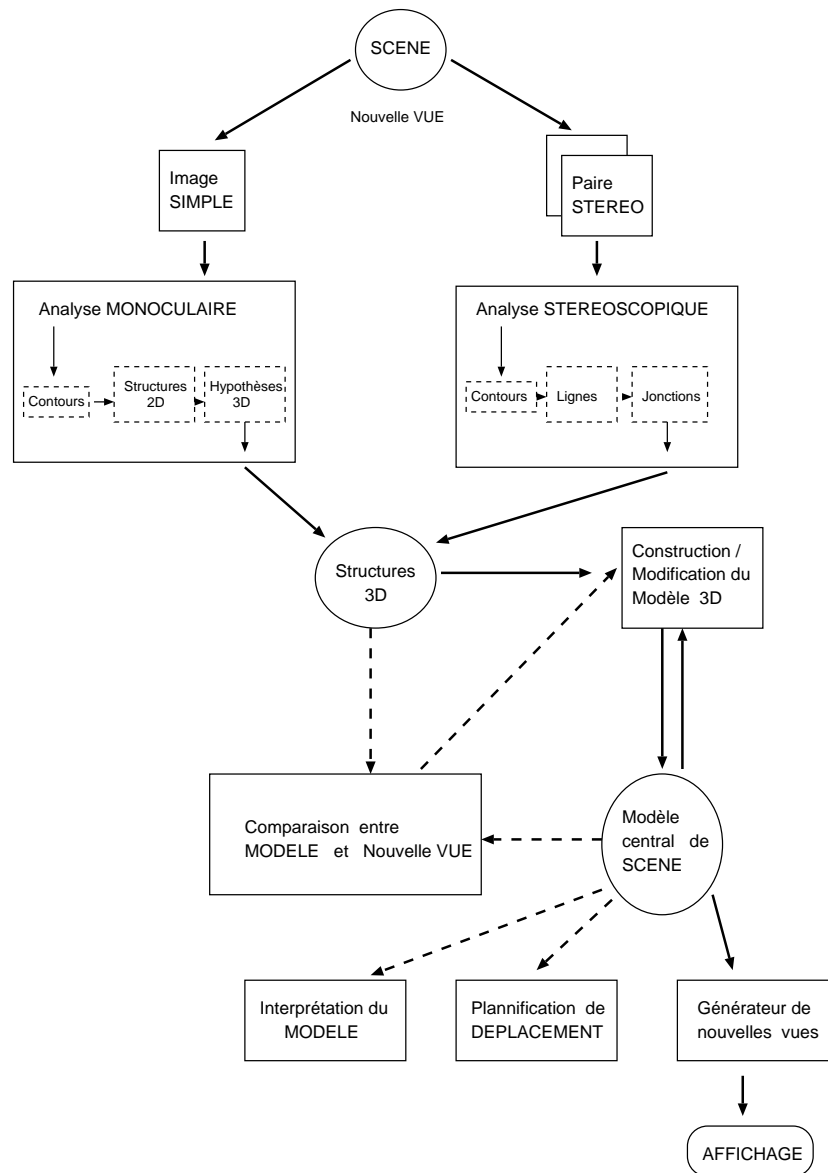


Figure 2.1 - Exemple de système visuel - tiré de "3D Mosaic Scene Understanding System" - Herman M. et Kanade, T. - 1986

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 contours¹. Les contours servent de représentation de base aux descriptions géométriques utilisant des segments (scènes polyédriques) 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] .

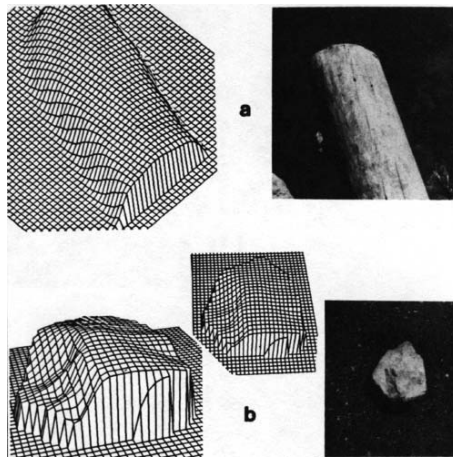


Figure 2.2 - Deux exemples de surfaces extraites à partir de l’illumination de la scène (*shape from shading*).

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] .

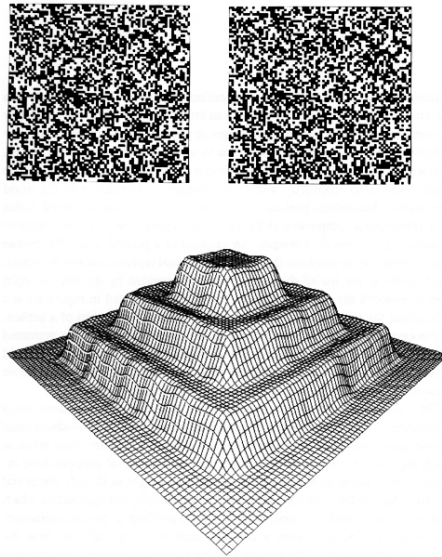


Figure 2.3 - *Déduction de la profondeur à partir de la disparité entre une paire stéréoscopique - ici, un stéréogramme.*

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

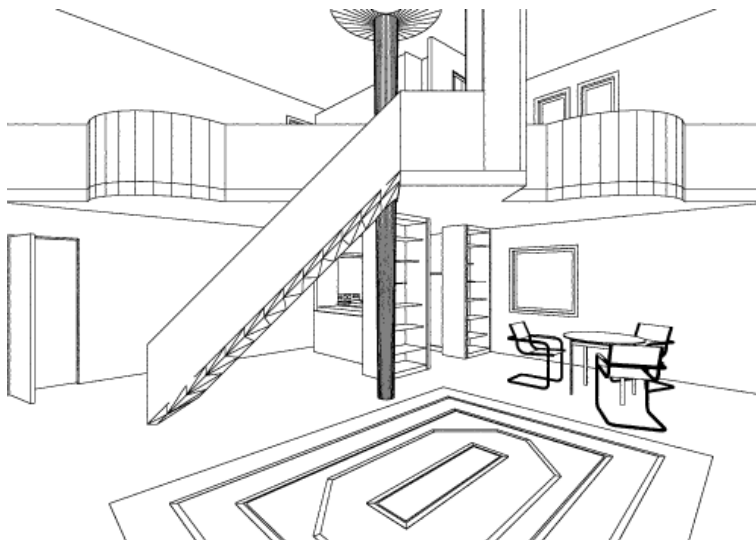


Figure 2.4 - *Les contours permettent souvent d'interpréter les objets d'une scène et sa structure 3D.*

rencontrées pour la représentation de scènes à partir de contours. Du type de scènes observées à l'acquisition d'images, les sources d'ambiguïtés 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 polyédrique simple, l'observation portant essentiellement sur des prismes, de surface homogène, posés sur un arrière plan uniforme. En l'absence d'ombres et de réflexions, 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 polyédrique 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].

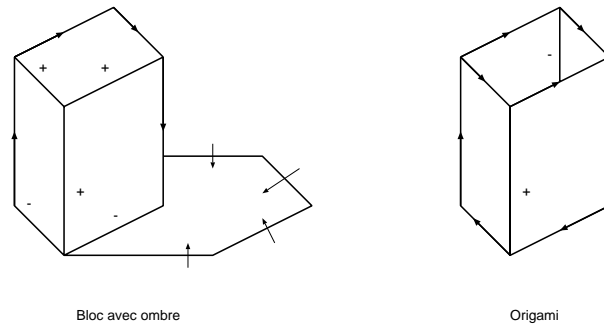


Figure 2.5 - Exemples simples d'étiquetage dans une scène de blocs et d'origami. Les (+) désignent des arêtes convexes les (-) des arêtes concaves. Les arêtes étiquetées par un \rightarrow signalent la présence de matière derrière la facette. Les flèches transversales désignent le cas particulier d'ombres.

Des contributions importantes à ces travaux ont été apportées par Waltz (1975), avec l'introduction d'ombres et l'étiquetage de polyèdres 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 polyédriques, il en est autrement pour des objets quelconques, présentant des parties courbes. L'ajout d'arêtes 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 modélisables à partir de modèles géométriques explicites [Pentland, 1986] .

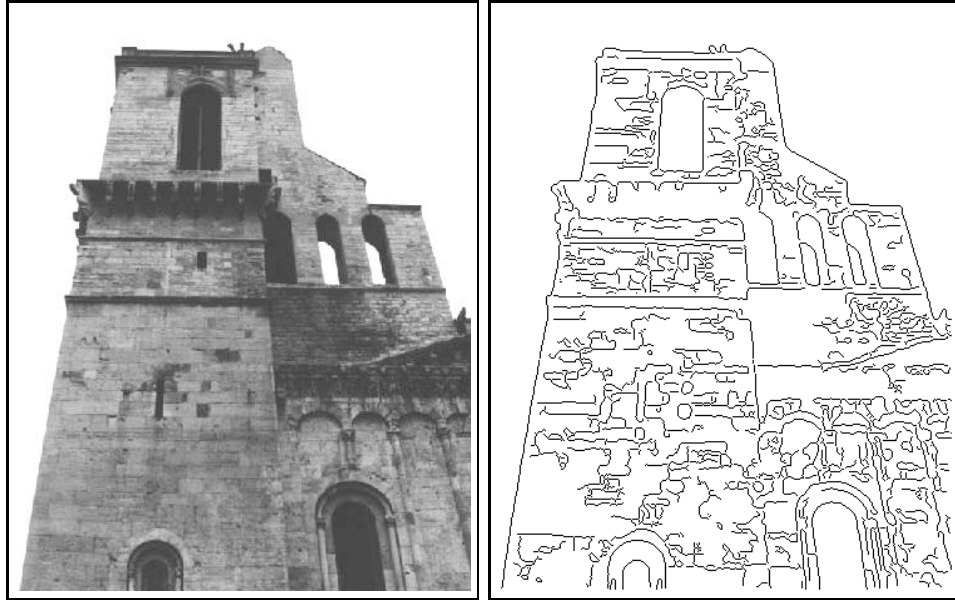


Figure 2.6 - Exemple de détection de contours

2.2.2 Sources d'ambiguïtés 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'ambiguïtés. 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.

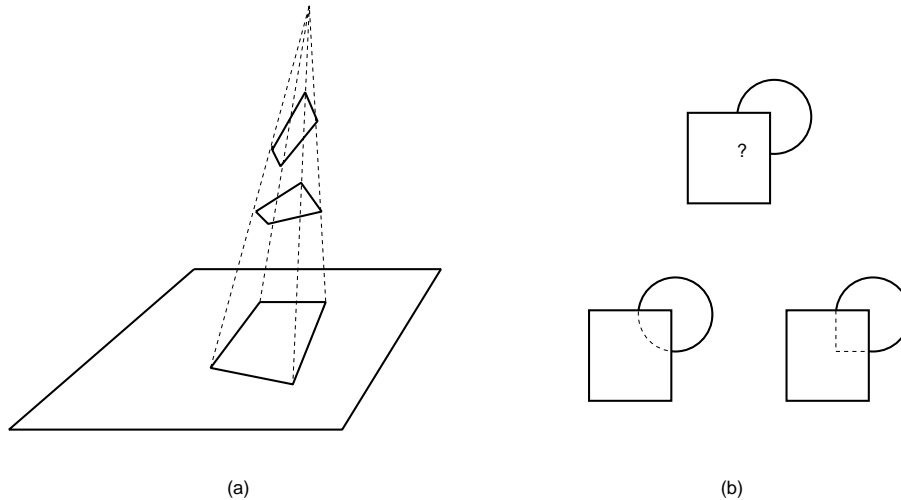


Figure 2.7 - *Ambiguïtés de projection (a) et d'occlusion (b).*

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 d'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

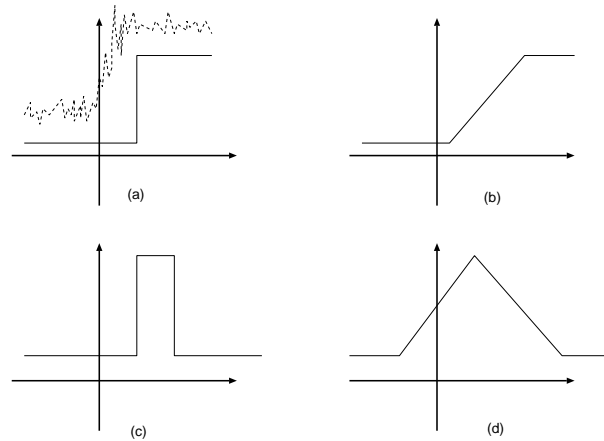


Figure 2.8 - Exemples classiques de contours d'intensité 1D - (a) marche, (b) rampe, (c) porte, (d) crête. La courbe en pointillés représente l'allure d'un contour réel, bruité.

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 ou 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.

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

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

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

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 appliquent 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'origine².

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 $w \rightarrow 0$, on obtient le filtre de dérivation suivant :

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

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

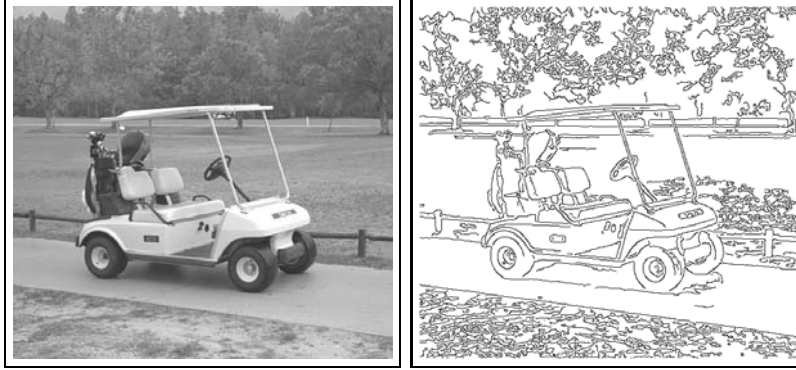


Figure 2.9 - *Détection de contours par application du filtre de Canny*

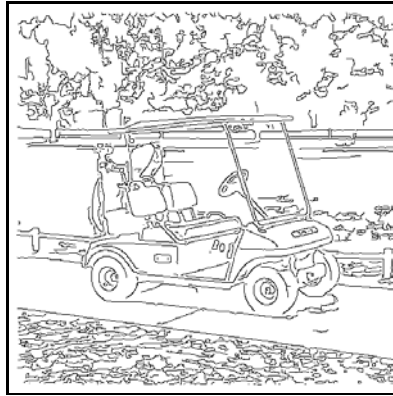


Figure 2.10 - *Détection de contours par application du filtre de Deriche.*

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 :

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

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 près 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.

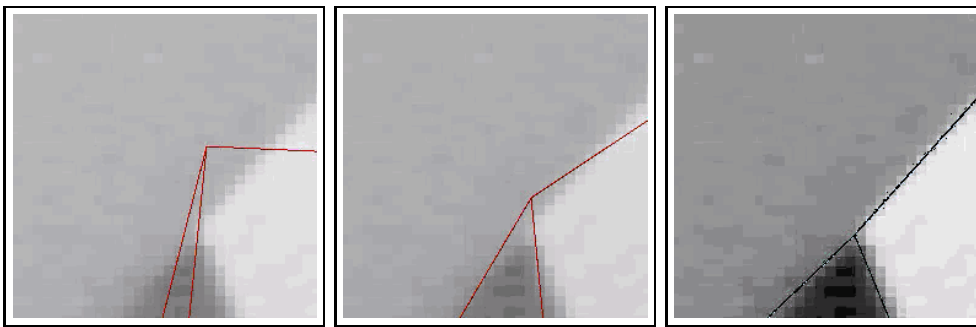


Figure 2.11 - *Détection précise de coins - convergence d'un modèle de coin vers une position optimale (à droite) - Méthode de Blaszk et Deriche.*

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.

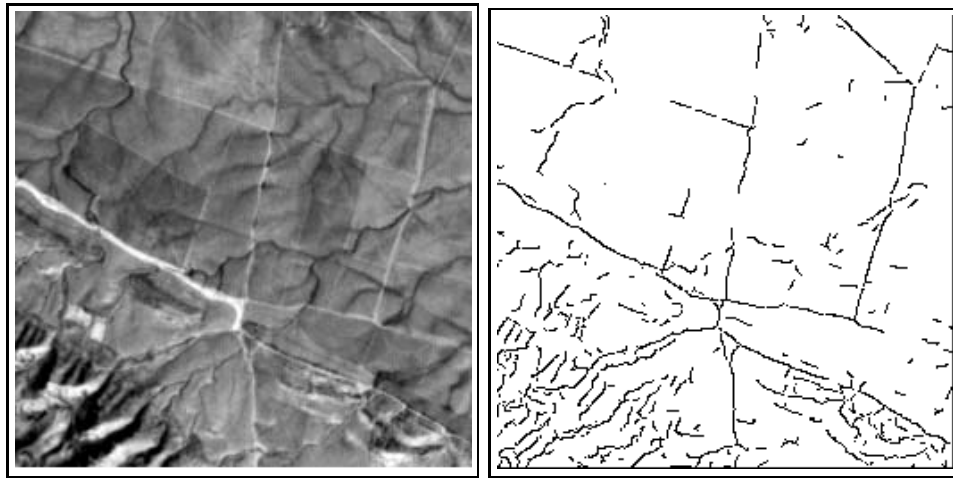


Figure 2.12 - *Détection de réseaux fins sur une image satellitaire. Un paramètre d'échelle σ permet de définir la largeur maximale des structures détectées. Ici, $\sigma = 1$. - Méthode de Armande, Monga et Montesinos.*

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 régions³.



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 maximales⁴.

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 naturelle⁵. 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.

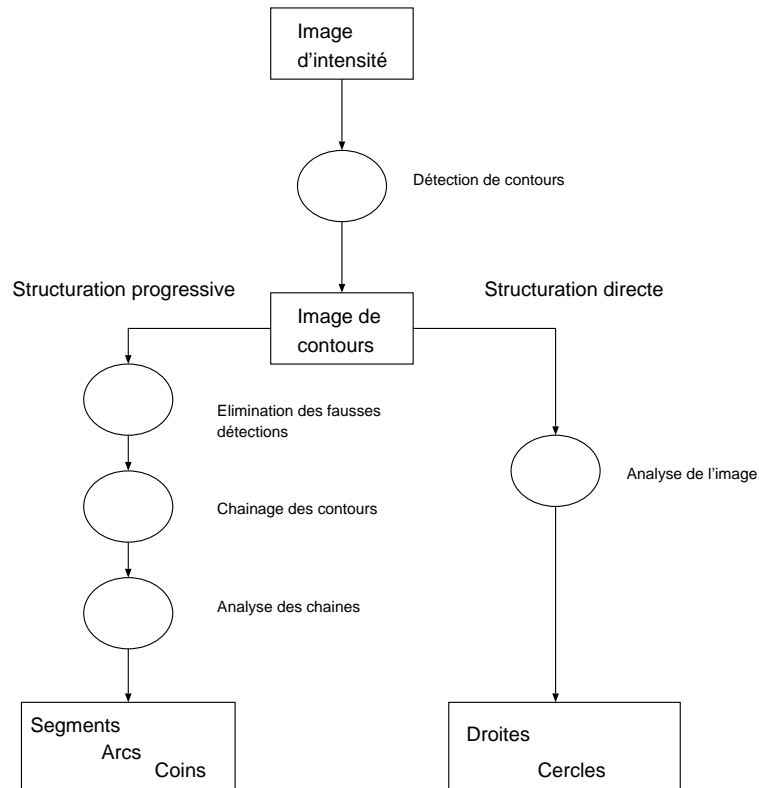


Figure 2.15 - *Structuration de contours après détection - deux niveaux d'application.*

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. **Élimination 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. **Châinage 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] .

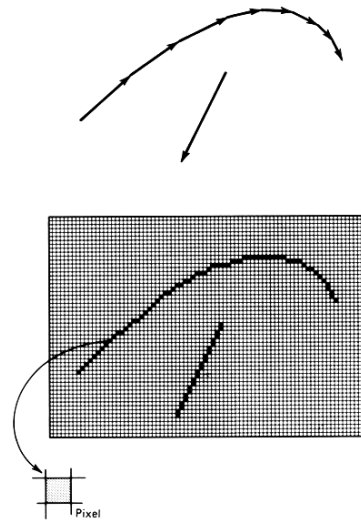


Figure 2.16 - *Ambiguïtés entre segments et arcs en géométrie discrète*

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 Boulton, 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 reste 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.

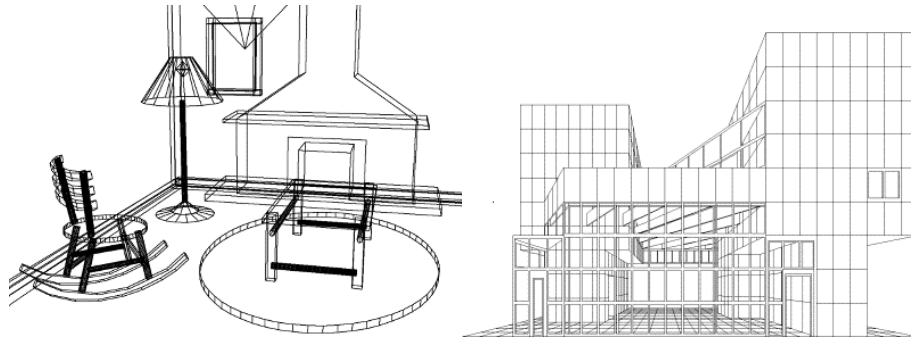


Figure 2.17 - *Modélisation de scène par représentation en fil de fer et par frontières. Le modèle “fil de fer” présente trop d’ambiguïtés pour représenter correctement la profondeur.*

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 ambiguë 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 polyédriques.

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 polyédriques 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 empêche 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 "événements 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 polyédriques. 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, boîtes), 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 polyédriques, 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 boîtes 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] .

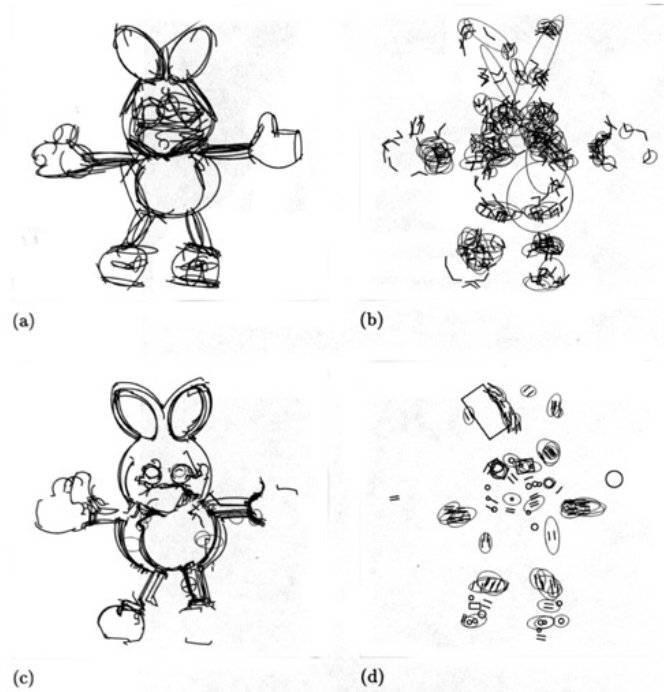


Figure 2.18 - Graphes de caractéristiques visuelles construits lors de l'apprentissage du modèle - (a) segments et arcs, (b) jonctions, (c) groupes de segments, (d) segments parallèles.

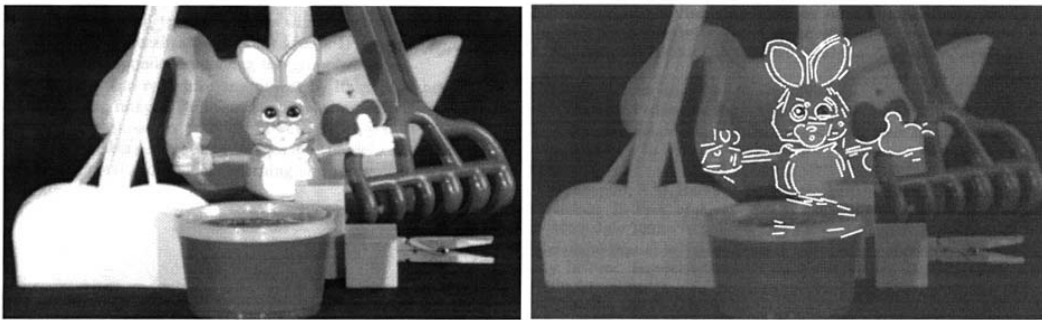


Figure 2.19 - Reconnaissance de l'objet malgré d'importantes occlusions - Méthode de Pope et Lowe, 1993

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.