

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/265579734>

La vision par ordinateur

Article

CITATIONS

0

READS

579

1 author:



R. Lepage

École de Technologie Supérieure

67 PUBLICATIONS 561 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Automatic road detection within radar satellite images [View project](#)



Remote Sensing for Major Disasters [View project](#)

La vision par ordinateur

Richard Lepage

Presses de l'École de technologie supérieure

© 2005

Table des matières

1. Les modèles computationnels en vision par ordinateur	1
1.1 Modélisation algorithmique du système de vision	2
1.1.1 Images d'éclairément	4
1.1.2 Croquis	4
1.1.3 Esquisse	6
1.1.4 Esquisse 2,5D	7
1.1.5 Description 3D	10
1.2 Modèle à base de connaissances	12
1.2.1 Détection des arêtes	14
1.2.2 Analyse de la profondeur	14
1.2.3 Analyse de l'éclairément	15
1.2.4 Fusion et interpolation	15
1.3 Modèle simplifié pour la reconnaissance	17
Conclusion	19
Références	20
2. Formation des images d'éclairément	25
2.1 Formation des images	25
2.1.1 Équation du rayon lumineux	26
2.1.2 Équation de formation des images	30
2.1.3 Surface lambertienne	33
2.2 Synthèse de l'image d'éclairément à partir d'une image 3D	35
2.2.1 Images de profondeur	35
2.2.2 Images d'éclairément	36

<i>Hypothèses sur la scène 3D</i>	37
<i>Géométrie</i>	37
<i>Équation de formation des images</i>	42
<i>Calcul des gradients de surface par dérivée partielle</i>	44
<i>Calcul des gradients de surface par la méthode du plan tangent.</i>	46
<i>Programmation de l'algorithme</i>	51
2.3 Génération d'images stéréos d'éclairément à partir d'une image 3D.	53
3. L'analyse multirésolution en vision informatique	57
3.1 Analyse d'une scène à multiple résolution spatiale	58
3.1.1 Primitives de l'image à multiple résolutions	61
3.1.2 Discrétisation de l'échelle d'espace.	65
3.1.3 Filtre passe-bas pour la génération des niveaux de résolution. .	65
3.2 Représentation pyramidale	68
Conclusion	73
Références.	75
4. Détection des arêtes d'une image.	81
4.1 Attributs de la scène observée.	82
4.2 Détection des variations brusques d'éclairément	84
4.2.1 Maximum de la dérivée première directionnelle	86
4.2.2 Maximum de la dérivée première isotrope	93
4.2.3 Passage par zéro de la dérivée seconde directionnelle.	94
4.2.4 Passage par zéro de la dérivée seconde isotrope	94
4.2.5 Détection des arêtes par apprentissage supervisé	97
4.2.6 Détection multirésolution des arêtes par réseau neuronique ..	100
4.3 Détection neuronique de l'amplitude du gradient.	100
4.4 Amincissement des arêtes.	101
4.5 Renforcement des arêtes	103
4.6 Résultats expérimentaux.	106
Conclusion	108
Références.	110

CHAPITRE 1 *Les modèles computationnels en vision par ordinateur*

Un système de vision peut être considéré comme un processus de traitement d'information : un stimulus est présenté à l'entrée du système et l'information ainsi reçue est traitée afin d'en extraire une représentation symbolique. Cette dernière pourra par la suite être utilisée comme entrée à un processus de traitement de plus haut niveau ou actionner des mécanismes de navigation ou d'adaptation à l'environnement.

Le système de vision peut donc être avantageusement analysé comme un système de traitement d'information (Marr, 1982; Jolion, 2001). Pour être complète, l'analyse du système de vision doit alors être entreprise à trois niveaux de compréhension, comme l'illustre le tableau 1.1. Cette analyse à trois niveaux assure que toutes les facettes du problème posé seront évaluées afin de proposer une solution qui tienne compte de ces différents niveaux d'abstraction. La section 1.1 présente une telle analyse et décrit un modèle algorithmique du système de vision. Dans ce modèle, le processus de vision est décomposé en représentations internes de l'information et en algorithmes qui permettent de passer d'une représentation interne à une autre. La section 1.1 précise les formes de représentations internes utilisées aux différents niveaux du système de vision. La section 1.2 propose une modélisation algorithmique modifiée du système de vision. C'est ce modèle modifié qui sera utilisé tout au long du cours. La section 1.2 décrit le modèle modifié de même que les différents algorithmes retenus pour sa réalisation.

Théorie du calcul	Représentations et algorithmes	Réalisation matérielle
<p>Quel est le but du calcul?</p> <p>Pourquoi est-il approprié?</p> <p>Quelle est la logique de la stratégie adoptée pour atteindre le but?</p>	<p>De quelle façon cette théorie du calcul à effectuer peut-elle être réalisée?</p> <p>Plus précisément, quelles sont les représentations nécessaires pour l'entrée et les différentes sorties intermédiaires et les algorithmes de transformation permettant de passer d'une représentation à la suivante?</p>	<p>Comment les représentations et les algorithmes sont-ils réalisés physiquement?</p>

Tableau 1.1 Niveaux de compréhension.

Les trois niveaux de compréhension nécessaires pour l'étude d'un mécanisme de traitement d'information.

1.1 MODÉLISATION ALGORITHMIQUE DU SYSTÈME DE VISION

Un système de vision extrait de l'information à partir d'une séquence d'images d'éclairage monoculaires ou stéréos. L'information est traitée afin d'en arriver à une compréhension raisonnable du monde environnant et permettre ainsi au mécanisme doté du système de vision d'effectuer des tâches comme par exemple de se diriger vers un but identifié, manipuler des objets de l'environnement ou éviter des collisions. Une modélisation algorithmique du système de vision consiste à décomposer tout le processus en plusieurs niveaux de représentations internes qui mettront en relief différents aspects de l'information visuelle analysée. Le passage d'une représentation interne à une autre est réalisé par un ou plusieurs algorithmes (Landy, & Movshon, 1991). La figure 1.1 illustre la modélisation algorithmique du système de vision telle que proposée par Marr (Brady, 1982; Cohen, & Feigenbaum, 1982; Marr, 1976; Rosenfeld, 1988; Rosenfeld, 1988; Yuille, & Ullman, 1990). Cette modélisation est caractérisée par un flux d'information qui procède du bas vers le haut. Le croquis, par exem-

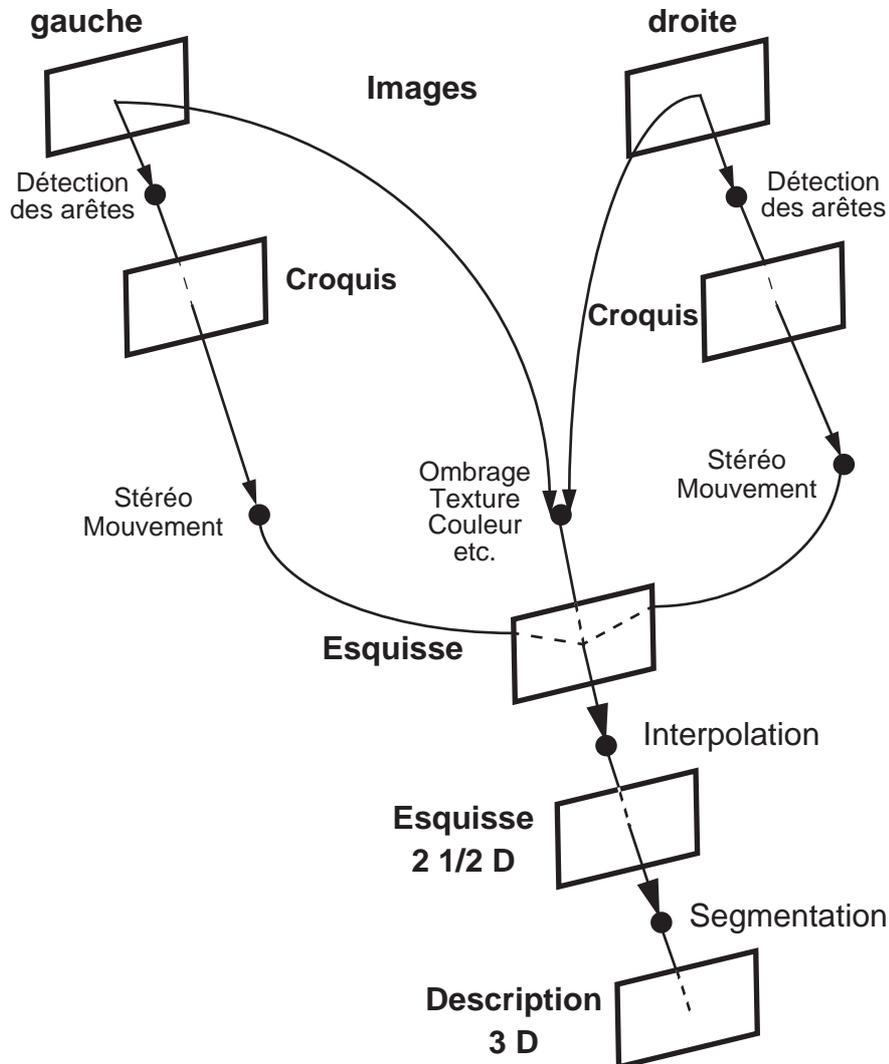


Figure 1.1 Modélisation algorithmique du système de vision. L'information présente dans l'image est codée à plusieurs niveaux de représentation interne. Des algorithmes permettent de passer d'une représentation à une autre.

ple, est généré à partir de l'image d'éclairage et n'est pas influencé par les niveaux supérieurs de représentation de données. Ce sens unique de bas vers le haut, utilisé pour construire les différents niveaux de représentation, est maintenant remis en question par de plus en plus de chercheurs (Weisstein, & Wong, 1987) qui proposent d'enrichir le modèle en y ajoutant des interactions haut

→ bas. Les sous-sections suivantes détaillent chacun des niveaux de représentation et les différents algorithmes qui peuvent être utilisés pour les générer.

1.1.1 Images d'éclairément

Le premier niveau de représentation interne est constitué des images d'éclairément gauche et droite qui sont produites par les capteurs correspondants du système de vision. Chaque image est la sortie directe du capteur et la représentation exacte dépend donc de la technologie utilisée. La plupart des capteurs produisent une image dont l'amplitude du signal de sortie à chaque pixel est proportionnelle à l'éclairément de l'image de la scène observée projetée sur le pixel correspondant du capteur.

La topologie des pixels du capteur est assumée comme étant une mosaïque rectangulaire, comme c'est le cas des capteurs commerciaux telles les caméras CCD. D'autres topologies sont possibles, comme par exemple la mosaïque hexagonale utilisée dans la construction d'un capteur VLSI spécialisé au laboratoire de Vision et Systèmes Numériques de l'Université Laval à Québec (Tremblay, & Poussart, 1992).

Les variations brusques d'éclairément dans l'image correspondent à des caractéristiques physiques définies de la scène observée. Le niveau suivant de représentation, le croquis, est une description des variations brusques d'éclairément détectées dans l'image d'entrée.

1.1.2 Croquis

Ce niveau de représentation rend explicite les informations importantes contenues dans l'image d'éclairément du niveau précédent. Ces informations importantes sont principalement les changements abruptes d'éclairément dans l'image, leur distribution géométrique et leur organisation. Les changements abruptes d'éclairément, ou arêtes, correspondent à des caractéristiques physiques de la scène observée tels un changement de l'orientation de la surface, un changement d'illumination (ombrage), une occlusion de surfaces ou un marquage de la surface (e.g. texture). Le croquis permet une nette diminution de la densité d'information à traiter afin de générer les niveaux ultérieurs de représentation.

(Lindeberg, & Eklundh, 1990; Lindeberg, & Eklundh, 1992) proposent d'utiliser des primitives plus simples pour construire le croquis. Ces primitives

sont les taches ou contours fermés qui sont obtenus à partir de l'image d'éclairément. Pour illustrer cette notion de tache, imaginons l'image d'éclairément comme un terrain vallonné dont la hauteur à chaque emplacement serait proportionnelle au niveau de gris de l'image au pixel correspondant. Supposons de plus que le terrain est immergé et que nous ayons le moyen d'écouler l'eau jusqu'à un certain niveau. Les taches du croquis correspondent tout simplement aux îlots du terrain immergé à un niveau donné. L'intérêt de cette méthode réside surtout dans l'analyse du comportement de ces taches à différents niveaux de résolution spatiale. Cette méthode résulte en une représentation plus simple mais est plus difficile à réaliser avec des réseaux neuroniques artificiels qui s'accommodent bien d'opérations locales et plus difficilement d'opérations plus régionales comme le demande le processus de formation des taches.

L'algorithme pour passer de la représentation par niveaux de gris (image d'éclairément) au croquis des changements abruptes d'éclairément appartient à la famille des détecteurs d'arêtes. De nombreux choix s'offrent à nous pour la réalisation du détecteur. L'un des algorithmes les plus utilisés est le Laplacien de Gaussienne, proposé par (Marr, 1982). Cet opérateur consiste à détecter les passages par zéro de l'image d'éclairément convoluée avec un filtre Laplacien de Gaussienne. Le paramètre σ du filtre gaussien permet de choisir l'échelle spatiale à laquelle on désire effectuer l'analyse tandis que l'opérateur du laplacien calcule la dérivée seconde isotrope de l'image filtrée. Le résultat est un opérateur simple qui détecte les arêtes pour toutes les directions dans l'image. Cet opérateur serait, croit-on, à la base de la détection d'arêtes dans le système visuel des mammifères. Marr propose de réaliser la détection des arêtes à plusieurs résolutions spatiales afin d'obtenir une description complète des informations importantes contenues dans les images d'éclairément. Cette approche multirésolution aurait également des fondements biologiques.

Plusieurs autres algorithmes sont possibles pour la détection des arêtes. Le choix d'un algorithme dépend de la méthode de réalisation choisie et des niveaux de performance que l'on désire atteindre en fonction de la complexité que la structure de réalisation peut supporter.

Le croquis fournit une description des caractéristiques importantes présentes dans l'image d'éclairément d'entrée. Cette information n'infère cependant aucune connaissance sur la troisième dimension de la scène observée, soit la profondeur. Pour reconstituer la troisième dimension, le système de vision doit disposer de plusieurs images (au minimum deux) provenant de caméras observant la scène de points de vue différents et/ou d'une séquence temporelle d'images représentant la scène d'un même point de vue à des intervalles de temps

fixes dans les cas de scènes dynamiques où des objets sont en mouvement. La représentation résultante, l'esquisse, décrit la profondeur des points caractéristiques de la scène.

1.1.3 Esquisse

L'esquisse est une représentation qui consiste en une estimation grossière de la profondeur des points caractéristiques de la scène situés sur les surfaces visibles des objets. Les points caractéristiques sont souvent situés sur les arêtes détectées du croquis. L'esquisse est donc une représentation clairsemée qui ne donne l'information de profondeur qu'aux points considérés importants, soient les arêtes qui indiquent les variations brusques d'éclairément.

Plusieurs algorithmes permettent de construire l'esquisse. Les algorithmes les plus utilisés sont la stéréoscopie (Ayache, 1989; Mayhew, & Frisby, 1991; Rosenfeld, 1986), l'ombrage (Horn, 1986; Horn, & Brooks, 1989) ou la couleur (Drew, 1992), le mouvement (Grossberg, & Rudd, 1989; Hutchinson, 1988; Lawton, Rieger, & Steenstrup, 1987; Marshall, 1990; Ullman, 1981; Vega-Riveros, & Jabbour, 1989), et la texture (Choe, & Kashyap, 1991; Connors, & Ng, 1989; Jau, & Chin, 1990).

La stéréoscopie utilise deux images (ou plutôt les croquis découlant de ces images ...) qui proviennent de caméras qui "observent" la scène d'un point de vue légèrement différent. La profondeur des arêtes identifiées est estimée à partir de la disparité des arêtes appariées sur les croquis gauche et droite. L'estimé de profondeur est obtenu par une simple triangulation découlant de la géométrie et de l'étalonnage des caméras. L'appariement peut être réalisé point par point sur une même ligne épipolaire sur les deux croquis ou segment par segment, un segment étant défini comme une arête connectée.

L'ombrage sur une image d'éclairément permet d'inférer la forme de l'objet observé. Le gradient d'éclairément de l'image, qui est une mesure de l'ombrage, nous permet d'avoir une idée sur la variation du gradient de surface de l'objet. Un estimé de la profondeur peut être obtenu à partir du gradient d'éclairément calculé et de la mesure de la profondeur aux discontinuités. Cette dernière mesure peut être obtenue par exemple par le module stéréo décrit précédemment. L'utilisation de la couleur permet d'avoir un estimé de la profondeur directement à partir d'une seule image en la décomposant selon les trois couleurs de base.

La profondeur peut également être estimée à partir du mouvement d'un ou plusieurs objets de la scène observée. Une séquence d'images doit alors être captée et une analyse du gradient spatio-temporel des arêtes contribue à la construction de l'esquisse. L'analyse du mouvement se ramène à une analyse stéréoscopique lorsque les images de la séquence sont analysées deux à la fois : l'analyse du gradient spatio-temporel se ramène alors à une mesure de la disparité des arêtes des deux croquis et la profondeur aux arêtes est obtenue par triangulation.

Une surface texturée comportant une structure régulière fournit des indices sur la profondeur de la surface. L'effet de perspective déforme les composantes de la structure et l'analyse de cette déformation permet d'en déduire un estimé de la profondeur. Une analyse fréquentielle de la structure de texture identifie les variations de fréquence spatiale à partir desquelles la profondeur peut être calculée.

Les différents algorithmes estiment la profondeur seulement à certains points caractéristiques de la scène. L'esquisse est donc une représentation partielle de la profondeur des surfaces visibles. Le niveau suivant de représentation, l'esquisse 2,5D, complète l'information manquante.

1.1.4 Esquisse 2,5D

L'esquisse 2,5D est une représentation caméra-centrique de la profondeur de toutes les surfaces visibles de la scène. Une profondeur est donc assignée à chaque coordonnée de la scène et cette information est codée dans un référentiel centré sur l'un des capteurs. À partir de cette information, plusieurs caractéristiques de la scène peuvent être déduites. Par exemple, la concavité ou la convexité d'un morceau de surface peut être différenciée, apportant ainsi une information primordiale sur la surface considérée (Marr, & Poggio, 1977). C'est en effet à ce niveau que peut se manifester une ambiguïté d'interprétation de certains objets comme par exemple le fameux cube de Necker (McClelland, & Rumelhart, 1989). L'orientation locale de surface de même que la forme de l'objet peuvent également être mesurées à partir des données de profondeur. Les paramètres 3D que l'on peut ainsi définir à partir des données de profondeur de même qu'à partir d'autres mesures sur les images d'éclairément constituent les *images intrinsèques* (Aloimonos, & Brown, 1988; Wechsler, 1990), une expression souvent utilisée pour désigner ce que Marr appelle l'esquisse 2,5D. Les images intrinsèques les plus largement utilisées et calculées sont la forme,

l'orientation locale de surface, la profondeur, la réflectance, le mouvement 3D et la direction de l'illumination.

L'existence même de l'esquisse 2,5D en tant que forme de représentation interne de données est contestée par l'école de Grossberg (Grimson, 1983; Grossberg, & Mingolla, 1987). Cette école préconise plutôt une forte interaction entre les différents algorithmes et des liens de contre-réaction entre les différents niveaux de représentation. Il en résulte que le modèle 3D est directement généré à partir des images d'éclairage et des croquis. Les interactions entre les algorithmes et entre les différentes représentations sont non-linéaires et complexes de sorte qu'il est très difficile de construire une modélisation complète du système de vision et surtout d'en analyser les différentes composantes. De nombreuses simplifications doivent être apportées au modèle pour parvenir à simuler expérimentalement ces structures non-linéaires. Deux recueils récents d'articles illustrent bien l'approche préconisée par Grossberg et son équipe (Carpenter, & Grossberg, 1992; Grossberg, 1988). Nous verrons dans les prochaines parties, telles la détection d'arêtes, la stéréoscopie et l'ombrage, les simplifications et les résultats obtenus par les différents chercheurs de l'école de Grossberg et ceux de l'école de Marr.

Un algorithme d'interpolation de surface permet de raffiner l'esquisse et ainsi générer l'esquisse 2,5D. Plusieurs approches permettent de passer d'une matrice clairsemée de valeurs de profondeur¹ à une représentation plus dense, l'esquisse 2,5D, où une valeur de profondeur sera calculée pour chaque pixel de l'image de profondeur résultante. (Besl, & Jain, 1985), et plus particulièrement (Bolle, & Vemuri, 1991) nous proposent des études fort complètes sur les différentes méthodes d'interpolation de surface. Il ressort de ces analyses que les méthodes les plus utilisées et les plus prometteuses sont, parmi les méthodes globales, celles de plaque mince de Grimson, de membrane ou de plaque invariants de Blake et Zisserman et à multi-grilles de Terzopoulos, et parmi les méthodes locales, celles de reconstruction paramétrique de surface et de triangulation de Delaunay.

(Grimson, 1981; Grimson, 1983) bâtit sa théorie de reconstruction de surface sur la contrainte de consistance de la surface : l'algorithme implicite de formation de l'ombrage sur l'image d'éclairage impose que l'opérateur de Laplacien de Gaussienne (que Grimson utilise pour détecter les arêtes de l'image) ne doive pas générer d'arêtes supplémentaires lorsqu'il est appliqué sur l'image reconstituée. Cette contrainte revient physiquement à utiliser une

1. Les valeurs de profondeur ne sont connues qu'aux positions des arêtes du croquis.

plaque flexible et mince pour interpoler entre les arêtes 3D et à minimiser l'énergie de déploiement. Pour ce faire, Grimson utilise une fonctionnelle qui représente l'énergie de la plaque mince. Cette fonctionnelle à minimiser mesure les variations quadratiques de la surface $f(x, y)$. Les variations quadratiques de surface sont calculées par une technique de différences finies.

La théorie de (Blake, & Zisserman, 1986) inclut explicitement les discontinuités de surface au moyen de contraintes faibles de continuité. Ils en arrivent ainsi à pouvoir utiliser pour l'interpolation des membranes ou des plaques qui sont invariantes du point d'observation¹. La surface reconstruite est ainsi indépendante du point de vue de l'observateur.

Terzopoulos s'inspire plutôt de la technique des éléments finis pour proposer une interpolation de surface qui part d'images d'éclairément à multiple résolutions spatiales afin d'augmenter de façon considérable l'efficacité de calcul (March, 1988; Terzopoulos, 1984; Terzopoulos, 1986). Les surfaces sont approximées par de petites pièces de métal mince sous tension, bordées par les discontinuités de profondeur et liées par des bandelettes de membrane le long des discontinuités d'orientation. Le problème de reconstruction, qui en est un mal posé au départ, est régularisé à l'aide de contraintes de continuité contrôlée afin de tenir compte des occlusions et des discontinuités de profondeur. Terzopoulos a développé une méthode de reconstruction à multiple grilles et multiple résolutions spatiales afin de pouvoir accélérer grandement l'efficacité de calcul (plusieurs ordres de grandeur). Cette méthode permet une intégration de la profondeur de surface et de la discontinuité d'orientation dans le problème de reconstruction de surface. (Szeliski, 1990) propose une amélioration de la technique de relaxation multi-grilles qui est beaucoup plus facile à réaliser. Cette méthode s'inspire de la technique de la descente selon le gradient conjugué avec l'aide d'un ensemble hiérarchique et multirésolution de fonctions de base. (Whitten, 1990) propose une approche multirésolution basée sur les modèles physiques lisses telles les feuilles déformables, les membranes élastiques minces et les plaques minces.

La reconstruction paramétrique de surface (Bolle, et al., 1991) permet une reconstruction qui est invariante au mouvement rigide 3D, ce qui n'est pas possible ni avec Grimson, ni avec Terzopoulos. Cette propriété d'invariance est importante surtout si l'objectif ultime du système de vision est la reconnais-

1. Il est à noter, cependant, que les membranes ou plaques utilisées ne sont pas indépendantes de l'ensemble des données, et donc des images de départ.

sance d'objet. De plus, cette invariance facilite grandement la construction du modèle 3D, qui est l'objet de la sous-section suivante.

La dernière méthode de reconstruction de surface que nous abordons brièvement ici est celle de triangulation de Delaunay (Faugeras, Le Bras-Mehlman, & Boissonnat, 1990; Le Bras-Mehlman, Schmitt, Faugeras, & Boissonnat, 1988). C'est une technique de reconstruction par surfaces polyédriques basée sur l'utilisation d'une triangulation contrainte de Delaunay. Cette méthode de premier ordre possède les avantages de préserver les discontinuités et d'être très simple d'application. De plus, cette méthode a son référentiel centré sur l'objet à reconstruire, ce qui assure une invariance au mouvement rigide, comme pour le cas de la reconstruction paramétrique décrit plus haut. La triangulation de Delaunay fonctionne avec peu ou beaucoup de données de profondeur au départ, la représentation étant tout simplement plus précise si on dispose de plus de données. L'algorithme s'ajuste bien avec l'apport de nouvelles données.

Nous avons brièvement étudié dans cette sous-section l'esquisse 2,5D et un échantillonnage d'algorithmes d'interpolation de surface qui nous permettent de bâtir cette représentation. Nous avons vu l'algorithme de minimisation des variations quadratiques de surface de Grimson, un algorithme particulièrement lent et qui ne tient pas tellement compte des discontinuités comme les occlusions. Terzopoulos utilise à l'entrée une représentation multirésolution et une fonctionnelle différente, ce qui lui permet d'obtenir un temps de calcul très bas et de tenir compte des discontinuités. Finalement, une méthode de reconstruction paramétrique et une autre basée sur la triangulation de Delaunay permettent toutes les deux une reconstruction qui tient compte des discontinuités mais qui est surtout invariante au mouvement, ce qui facilite la construction de la représentation suivante, la description 3D.

1.1.5 Description 3D

La description 3D constitue le dernier niveau du système de vision. Ce niveau décrit symboliquement la forme des objets visibles et leur organisation spatiale dans un système de coordonnées objet-centrique. La représentation des objets est hiérarchique et modulaire avec des primitives de volume et de surface. Cette forme de représentation est souvent réalisée en comparant le résultat de la segmentation de l'esquisse 2,5D à des modèles emmagasinés en mémoire. Cette forme de représentation se veut idéalement indépendante du point de vue et permet ainsi de reconnaître les objets composant la scène observée, ce qui constitue en somme le but ultime du système de vision.

La description 3D s'obtient en segmentant l'esquisse 2,5D. Les recherches sont nombreuses et très actives dans ce domaine. La segmentation consiste à identifier les différentes composantes des objets, à leur assigner une description indépendante du point de vue et de coder cette information afin d'effectuer la comparaison avec une banque de modèles ou de primitives comme les *geons* (Bergevin, & Levine, 1993). La description indépendante du point de vue d'une composante d'un objet peut commander un déplacement du système de vision afin de compléter la collecte d'information. La comparaison de la composante codée avec une banque de modèles contribue habituellement à l'enrichissement de la banque en cas d'insuccès - on dit alors que le système est en mode d'apprentissage.

La description 3D et l'algorithme pour y parvenir, la segmentation, constituent la vision de haut niveau. C'est à ce niveau qu'interviennent des fonctions cognitives telles la reconnaissance et la mémorisation. La vision de bas niveau et de niveau intermédiaire comprend toutes les formes de représentation avant le modèle 3D, soit des images d'éclairage à l'esquisse 2,5D. La vision de bas niveau, que l'on appelle également *vision pré-attentive*, est généralement considérée comme ne faisant pas intervenir des fonctions cognitives. C'est donc dire que la construction des différents niveaux de représentation ne fera pas intervenir des connaissances emmagasinées dans la mémoire du système de vision. Les contraintes imposées par notre monde tridimensionnel physique seront programmées directement dans la structure du système de vision de bas niveau. On peut penser à la programmation directe de ces contraintes dans la structure du système de vision comme le résultat de l'évolution : la structure du système de vision s'est modifiée au cours des millénaires afin de s'adapter au monde réel environnant. Cette adaptation se traduit par des liens pré-programmés et des structures de réseaux neuroniques adaptées à ces contraintes.

1.2 MODÈLE À BASE DE CONNAISSANCES

La figure 1.2 présente un schéma modifié de la modélisation algorithmique d'un système de vision. Ce schéma diffère du schéma classique proposé par Marr (figure 1.1) à trois niveaux :

1. Une représentation multirésolution spatiale est utilisée pour les images d'éclairément, le croquis et l'esquisse.
2. L'analyse de l'ombrage et de la couleur contribue directement à la génération de l'esquisse 2,5D.
3. Le contenu sémantique de la description 3D est analysé par un système à base de connaissances et les résultats permettent d'ajuster, au besoin, les paramètres des algorithmes des niveaux précédents (contre-réaction).

Les mêmes niveaux de représentation que sur le schéma de Marr sont utilisés, soient les images d'éclairément, le croquis, l'esquisse, l'esquisse 2,5D et la description 3D. Par contre, tous ces niveaux, sauf les niveaux finaux, l'esquisse 2,5D et l'interprétation 3D, sont représentés à plusieurs résolutions spatiales. Une image à une résolution donnée est obtenue en appliquant un filtre passe-bas sur l'image à pleine résolution. Cette opération a pour effet de brouiller l'image, comme si on observait la scène à travers une lentille hors-foyer. Cette mise hors-foyer élimine les variations rapides d'éclairément et permet donc d'observer la scène à plusieurs niveaux de détails. Certaines caractéristiques de la scène apparaissent mieux à certaines échelles et donc une représentation à multiple résolutions spatiales constitue une description plus complète de la scène (Saund, 1989). L'analyse multirésolution s'inspire des systèmes de vision biologiques qui analysent les caractéristiques d'une scène à l'aide de plusieurs canaux possédant des bandes passantes différentes (jusqu'à quatre canaux pour le système visuel des mammifères (Blakemore, & Campbell, 1969; Campbell, & Robson, 1968; Marcelja, 1980; Wilson, & Bergen, 1979)).

Les algorithmes pour passer d'une représentation à la suivante sont sensiblement les mêmes que dans le schéma de Marr. La détection des arêtes permet de générer le croquis. De nombreux algorithmes tels la stéréoscopie, l'analyse du mouvement et la texture, peuvent être utilisés pour passer du croquis à l'esquisse. L'analyse de l'ombrage et/ou de la couleur est ici utilisée pour con-

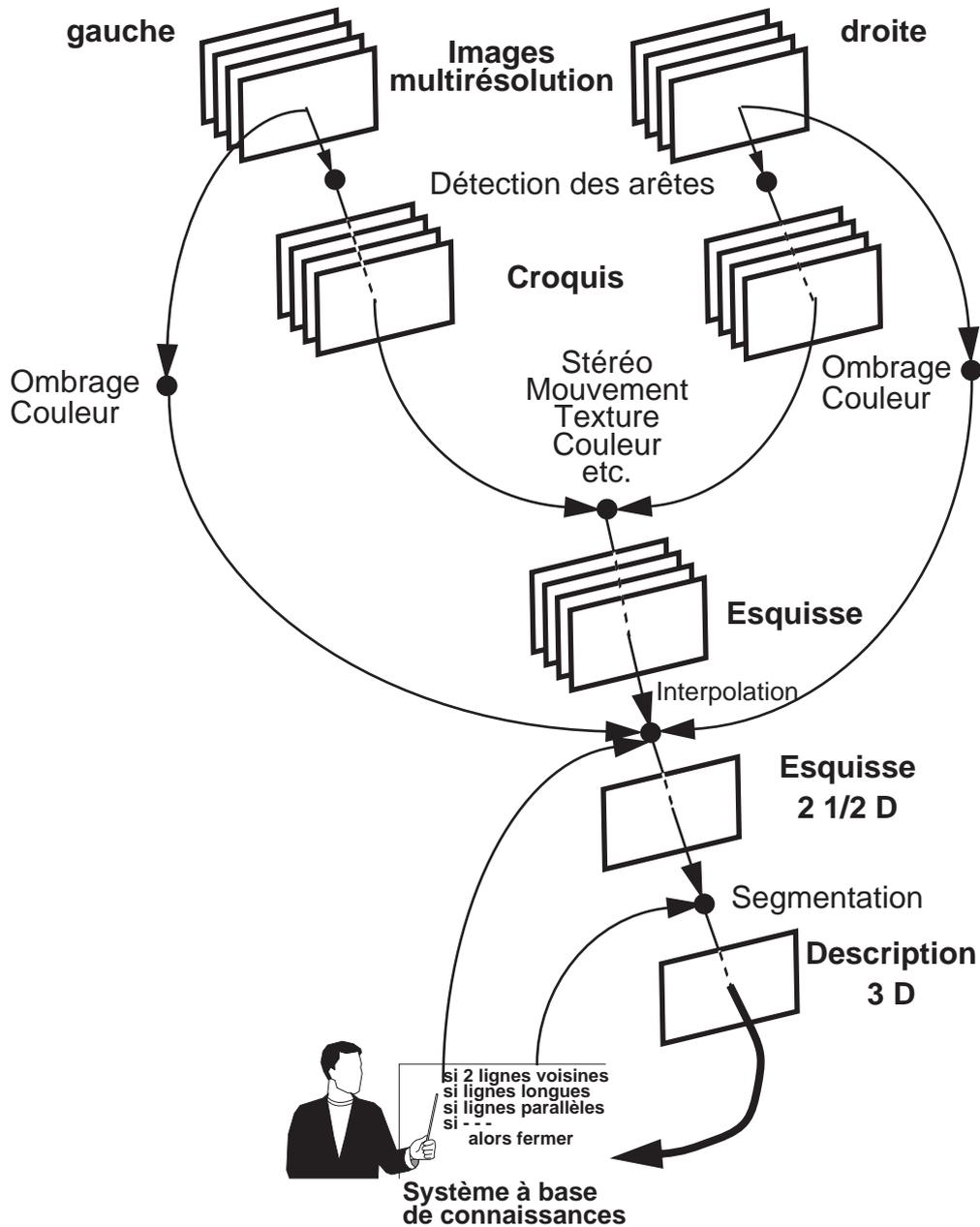


Figure 1.2 *Modélisation algorithmique du système de vision. L'information présente dans les images est codifiée à plusieurs niveaux de représentation interne. Des algorithmes permettent de passer d'une représentation à la suivante. Des règles d'interprétation du contenu symbolique de la description 3D permettent d'ajuster des algorithmes de plus bas niveau.*

tribuer directement à la formation de l'esquisse 2,5D par l'intermédiaire d'un algorithme de fusion ou d'interpolation.

1.2.1 Détection des arêtes

L'algorithme de détection d'arêtes permet de générer le croquis. Le croquis est une description des variations importantes d'éclairage dans l'image d'entrée. Ces variations se mesurent en détectant un maximum de la dérivée première directionnelle ou en détectant un changement de signe (passage par zéro) de la dérivée seconde. L'opération de dérivation est génératrice de bruit parce qu'elle amplifie les hautes fréquences et donc accentue les variations rapides d'éclairage. Nous proposons une détection à multiple résolutions spatiales afin d'éliminer les arêtes bruitées mises en évidence par le processus de différenciation. Une interaction des niveaux à faible résolution spatiale sur les niveaux à plus haute résolution permet d'identifier les arêtes correspondant à des caractéristiques physiques de la scène observée et contribue à l'élimination des arêtes causées par le bruit.

1.2.2 Analyse de la profondeur

La stéréoscopie consiste à calculer la profondeur de points bien identifiés de la scène en mesurant le déplacement des points projetés correspondants sur les deux caméras. En connaissant la géométrie exacte des caméras, la profondeur d'un point particulier se calcule facilement par triangulation inverse de la projection sur les deux plans images. La principale difficulté de la stéréoscopie consiste donc à s'assurer que les deux points que l'on essaie d'apparier correspondent au même point de la scène 3D. Pour un point d'intérêt identifié sur l'image gauche, il faut donc trouver le point correspondant sur l'image droite en se déplaçant sur la droite épipolaire (définie par la géométrie de la scène et des caméras), éviter les fausses cibles, mesurer le déplacement relatif du point d'intérêt sur l'image droite et calculer la profondeur de ce point par triangulation inverse. Nous proposons une approche hiérarchique pour diminuer l'appariement de fausses cibles. L'appariement est d'abord effectué avec les images d'arêtes à la plus faible résolution spatiale de sorte que la probabilité d'appariement de fausses cibles est faible étant donné que peu d'arêtes sont présentes sur ces images. L'estimation de la profondeur est approximative due aux échelles grossières d'échantillonnage utilisées à ces niveaux de représentation. Par contre, cette valeur approximative de profondeur limite la zone de recherche pour l'appariement des arêtes au niveau suivant de résolution spatiale : seuls les

pixels situés dans la plage de disparité définie par la valeur approximative trouvée précédemment sont l'objet de la recherche pour l'appariement du point caractéristique de l'image gauche. Cet appariement hiérarchique se poursuit de la sorte en allant vers les représentations à pleine résolution spatiale. La mesure de disparité est ainsi de plus en plus raffinée, partant d'une estimation grossière vers une estimation de plus en plus précise tout en évitant au maximum les faux appariements (qui génèrent des mesures erronées de disparité).

1.2.3 Analyse de l'éclairement

L'observation d'une photographie, même si elle n'est qu'une représentation en deux dimensions d'une scène tridimensionnelle, nous fournit généralement assez d'indices pour que nous puissions nous faire une idée assez précise de la forme des objets constituant la scène et de leur localisation spatiale relative. Une image contient donc suffisamment d'information pour pouvoir en extraire des données de profondeur (troisième dimension). Cette information de profondeur nous est fournie par les ombrages, ou variations d'éclairement, qui sont présents sur la photographie. L'équation de formation des images pour les corps lambertiens nous apprend que l'éclairement d'un pixel de l'image est proportionnelle au produit scalaire entre le vecteur pointant vers la source lumineuse et le vecteur normal à la surface, et donc est proportionnelle au cosinus de l'angle entre ces deux vecteurs. L'ombrage sur l'image, qui est mesuré par le gradient de l'éclairement, est donc proportionnel au taux de variation du vecteur normal (puisque le vecteur indiquant la direction de la source lumineuse est constant). Une mesure du gradient d'éclairement sur l'image indique donc une variation du vecteur normal à la surface, ce qui nous fournit une perception de la forme de l'objet observé sur l'image, puisqu'une variation du vecteur normal se traduit par une variation de la courbure de la surface, et donc par une variation de la forme tridimensionnelle de l'objet.

1.2.4 Fusion et interpolation

L'algorithme de stéréoscopie mesure la profondeur à certains points caractéristiques de la scène tridimensionnelle. Ces points caractéristiques correspondent aux arêtes de l'image, soient les variations brusques d'éclairement. L'algorithme de stéréo ne mesure la profondeur qu'aux positions des arêtes de l'image de profondeur résultante (esquisse 2,5D).

L'algorithme d'ombrage, pour sa part, mesure les variations de profondeur à partir des faibles variations d'éclairement sur une des images d'éclairement,

et donc les variations de profondeur entre les arêtes de l'image de profondeur résultante.

L'algorithme de fusion combine ces deux types d'informations complémentaires afin de générer l'esquisse 2,5D. L'algorithme peut utiliser un réseau de relaxation afin de fusionner les deux types d'information. Les contraintes sont imposées au réseau par les valeurs de profondeur calculées par l'algorithme de stéréoscopie. L'algorithme d'ombrage fournit, pour chaque position de pixel de l'esquisse 2,5D, une famille de morceaux de surface guidant des solutions de l'équation inverse de formation d'image. Le réseau de relaxation converge itérativement vers une solution qui tient compte des valeurs aux frontières trouvées par la stéréo et des orientations possibles des morceaux de surface imposées par l'algorithme d'ombrage.

Cette étude propose la fusion de la stéréoscopie et de l'ombrage afin de construire l'esquisse 2,5D. On démontre que cette fusion permet de construire l'esquisse 2,5D beaucoup plus rapidement qu'une méthode d'interpolation globale par minimisation d'une fonctionnelle du type utilisée par Grimson. Cette augmentation d'efficacité provient du fait que le processus d'interpolation de surface, ou de remplissage, terme préféré par l'école de Grossberg, est guidé par l'algorithme d'ombrage qui restreint de beaucoup l'espace des solutions possibles, et donc le temps de calcul pour parvenir à trouver une solution qui tienne compte des contraintes imposées par les algorithmes de stéréo et d'ombrage de même que les contraintes imposées par le monde physique. De plus en plus de chercheurs essaient de fusionner l'information provenant de différents algorithmes qui fournissent une solution partielle pour la génération de l'esquisse 2,5D. C'est ainsi que (Waxman, & Duncan, 1986) proposent de combiner les algorithmes de stéréoscopie et de détection du mouvement afin de faciliter la construction de l'esquisse 2,5D.

1.3 MODÈLE SIMPLIFIÉ POUR LA RECONNAISSANCE

Les deux objectifs fondamentaux d'un système de vision sont l'identification et la position des divers objets perçus par les capteurs du système tel qu'illustré à la figure 1.3. La modélisation complète du système de vision développée à la section précédente est alors appropriée.

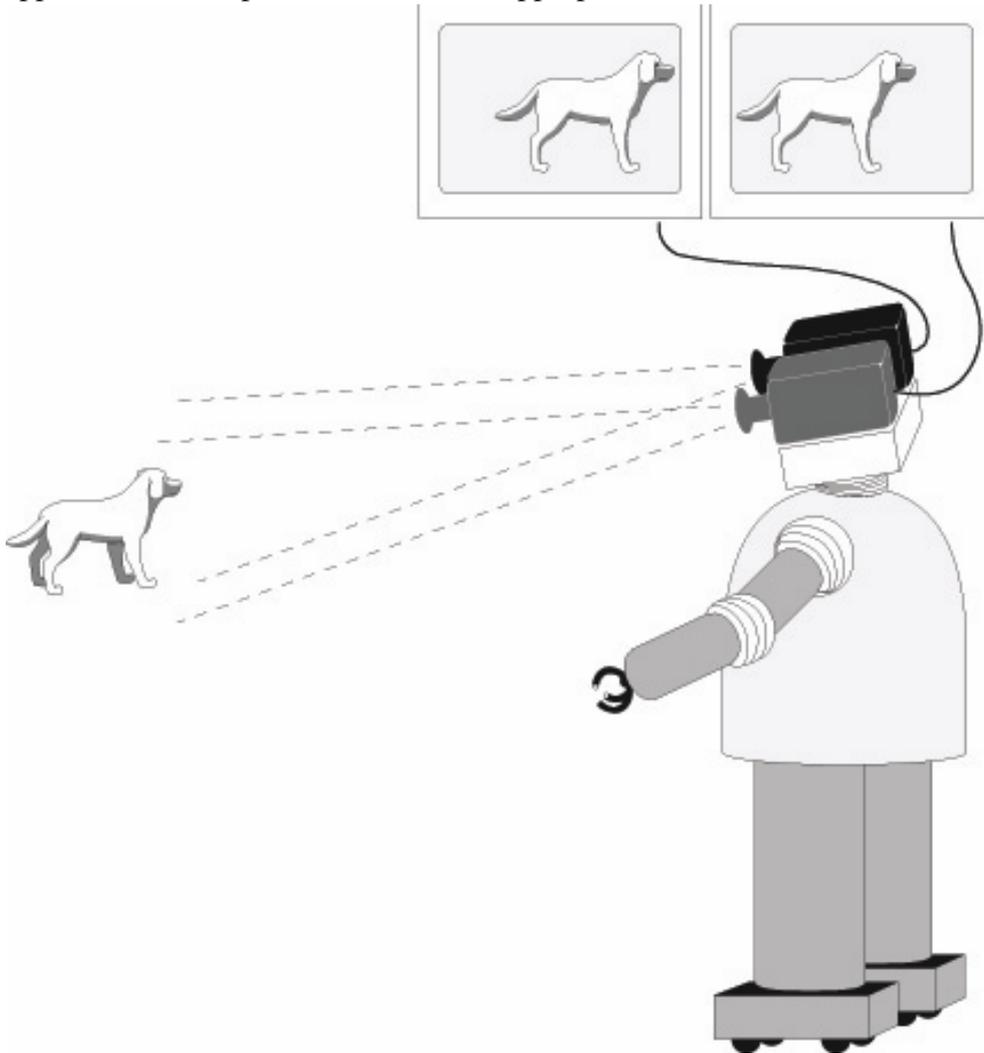


Figure 1.3 Système de vision par ordinateur.
Les deux objectifs généraux poursuivis par un système de vision par ordinateur sont d'identifier les objets présents dans le champ de vision et de les localiser.

Plusieurs applications, par contre, ne nécessitent que la reconnaissance des divers objets présents dans la scène observée, sans nécessiter leur localisation 3D. L'analyse de tels systèmes tirera avantage du modèle simplifié ramené à un seul canal d'information visuelle illustrée ci-dessous (figure 1.4).

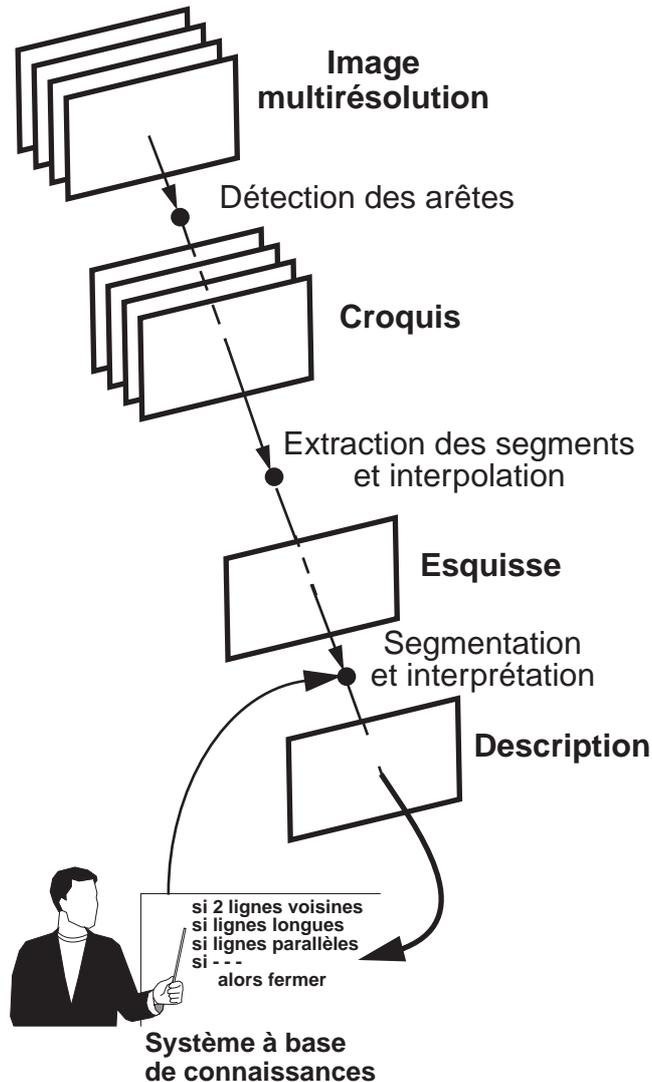


Figure 1.4 Modélisation algorithmique simplifiée du système de vision. L'information présente dans l'image est codée à plusieurs niveaux de représentation interne. Des algorithmes permettent de passer d'une représentation à la suivante. Des règles d'interprétation du contenu symbolique de la description des objets permettent d'ajuster la segmentation.

CONCLUSION

Le système de vision proposé est modélisé comme un système de traitement de données. Le système est divisé en plusieurs niveaux de représentation interne des données. Des algorithmes permettent de passer d'un niveau de représentation au suivant. La modélisation de Marr a été modifiée afin d'inclure les représentations à multiple résolutions spatiales et de fusionner les algorithmes de stéréoscopie et d'ombrage afin de générer plus efficacement et plus rapidement le niveau final (dans le cadre de notre étude) de représentation, soit l'esquisse 2,5D.

Nous avons étudié le modèle selon les trois niveaux d'abstraction proposé par Marr. Nous nous sommes d'abord intéressés à la théorie du système de vision proposé et les principaux buts recherchés. Nous avons ensuite identifié les représentations internes nécessaires pour atteindre nos objectifs de même que les algorithmes que nous comptons mettre au point afin de générer les différents niveaux de représentation.

RÉFÉRENCES

- Aloimonos, J. Y., & Brown, C. M. (1988). Robust computation of intrinsic images from multiple cues. In C. Brown (Ed.), *Advances in Computer Vision* (pp. 115-163). Hillsdale, N.J.: Lawrence Erlbaum Ass.
- Ayache, N. (1989). *Vision stéréoscopique et perception multisensorielle - applications à la robotique mobile*. Paris: InterEditions.
- Bergevin, R., & Levine, M. D. (1993). Generic Object Recognition: Building and Matching Coarse Descriptions from Line Drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(1), 19-36.
- Besl, P. J., & Jain, R. C. (1985). Three-Dimensional Object Recognition. *ACM Computing Surveys*, 17, 75-145.
- Blake, A., & Zisserman, A. (1986). Invariant surface reconstruction using weak continuity constraints. *Proceedings of the CVPR'86: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 22-26, Miami Beach, Florida, 62-67.
- Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, 203, 237-260.
- Bolle, R. M., & Vemuri, B. C. (1991). On Three-Dimensional Surface Reconstruction Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1), 1-13.
- Brady, M. (1982). Computational Approaches to Image Understanding. *ACM Computing Surveys*, 14(1), 3-71.
- Campbell, F. W. C., & Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551-566.
- Carpenter, G. A., & Grossberg, S. (Ed.). (1992). *Neural networks for vision and image processing*. Cambridge, Massachusetts: The MIT Press.
-

-
- Choe, Y., & Kashyap, R. L. (1991). 3-D Shape from a Shaded and Textural Surface Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9), 907-919.
- Cohen, P. R., & Feigenbaum, E. A. (Ed.). (1982). *The Handbook of Artificial Intelligence*. Reading, Mass.: Addison-Wesley.
- Connors, R. W., & Ng, C. T. (1989). Developing a Quantitative Model of Human Preattentive Vision. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6), 1384-1407.
- Drew, M. S. (1992). Shape from Color (No. CSS/LCCR TR 92-07). School of Computing Science, Simon Fraser University, Vancouver, B.C.
- Faugeras, O. D., Le Bras-Mehlman, E., & Boissonnat, J. D. (1990). Representing Stereo Data with the Delaunay Triangulation. *Artificial Intelligence*, 44, 41-87.
- Grimson, W. E. L. (1981). *From Images to Surfaces - A Computational Study of the Human Early Visual System*. Cambridge, Mass.: The MIT Press.
- Grimson, W. E. L. (1983). An Implementation of a Computational Theory of Visual Surface Interpolation. *Computer Vision, Graphics, and Image Processing*, 22, 39-69.
- Grossberg, S. (Ed.). (1988). *Neural Networks and Natural Intelligence*. Cambridge, Massachusetts: The MIT Press.
- Grossberg, S., & Mingolla, E. (1987). Neural Dynamics of Surface Perception: Boundary Webs, Illuminants, and Shape-from-Shading. *Computer Vision, Graphics, and Image Processing*, 37, 116-165.
- Grossberg, S., & Rudd, M. E. (1989). A Neural Architecture for Visual Motion Perception: Group and Element Apparent Motion. *Neural Networks*, 2, 421-450.
- Horn, B. K. P. (1986). *Robot Vision*. Cambridge, Mass.: The MIT Press.
- Horn, B. K. P., & Brooks, M. J. (Ed.). (1989). *Shape from Shading*. Cambridge, Mass.: The MIT Press.
-

-
- Hutchinson, J. (1988). Computing Motion Using Analog and Binary Resistive Networks. *Computer*, (March), 52-63.
- Jau, J. Y., & Chin, R. T. (1990). Shape from Texture Using the Wigner Distribution. *Computer Vision, Graphics, and Image Processing*, 52, 248-263.
- Jolion, J.-M. (2001). Sur la méthodologie de conception des systèmes de vision. Chapitre de livre, *Les systèmes de vision*. J.-M. Jolion. Paris, Hermes: 97-131.
- Landy, M. S., & Movshon, J. A. (Ed.). (1991). *Computational Models of Visual Processing*. Cambridge, Massachusetts: The MIT Press.
- Lawton, D. T., Rieger, J., & Steenstrup, M. (1987). Computational Techniques in Motion Processing. In M. A. Arbib & A. R. Hanson (Ed.), *Vision, Brain and Cooperative Computation* (pp. 419-488). Cambridge, Mass.: The MIT Press.
- Le Bras-Mehlman, E., Schmitt, M., Faugeras, O. D., & Boissonnat, J. D. (1988). How the Delaunay Triangulation can be used for Representing Stereo Data. *Proceedings of the Second International Conference on Computer Vision*, Tampa, Florida, December 5-8, 54-63.
- Lindeberg, T., & Eklundh, J.-O. (1990). Scale Detection and Region Extraction from a Scale-Space Primal Sketch. *Proceedings of the Third International Conference on Computer Vision*, Osaka, Japan, December 4-7, 416-426.
- Lindeberg, T., & Eklundh, J.-O. (1992). Scale-space primal sketch: construction and experiments. *Image and Vision Computing*, 10(1), 3-18.
- Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America A*, 70(11), 1297-1300.
- March, R. (1988). Computation of stereo disparity using regularization. *Pattern Recognition Letters*, 8(October), 181-187.
- Marr, D. (1976). Early Processing of Visual Information. *Philosophical Transactions of the Royal Society of London B*, 245, 483-519.
-

-
- Marr, D. (1982). *Vision - A Computational Investigation into the Human Representation and Processing of Visual Information* . San Francisco: Freeman.
- Marr, D. C., & Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Research Progress Bulletin*, 15(3), 470-488.
- Marshall, J. A. (1990). Self-Organizing Neural Networks for Perception of Visual Motion. *Neural Networks*, 3, 45-74.
- Mayhew, J. E. W., & Frisby, J. P. (Ed.). (1991). *3D Model Recognition from Stereoscopic Cues* . Cambridge, Massachusetts: The MIT Press.
- McClelland, J. L., & Rumelhart, D. E. (1989). *Explorations in Parallel Distributed Processing - A Handbook of Models, Programs, and Exercises* . Cambridge, Massachusetts: The MIT Press.
- Rosenfeld, A. (Ed.). (1986). *Techniques for 3-D Machine Perception* . Amsterdam: North-Holland.
- Rosenfeld, A. (1988). Computer Vision: Basic Principles. *Proceedings of the IEEE*, 76, 863-868.
- Rosenfield, I. (1988). *The Invention of Memory - A New View of the Brain* . New York: Basic Books.
- Saund, E. (1989). Adding Scale to the Primal Sketch. *Proceedings of the the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'89)*, San Diego, California, June 4-8, 70-78.
- Szeliski, R. (1990). Fast Surface Interpolation Using Hierarchical Basis Functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6), 513-528.
- Terzopoulos, D. (1984). Multilevel Reconstruction of Visual Surfaces: Variational Principles and Finite-Element Representations. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 237-310). New York: Springer-Verlag.
-

-
- Terzopoulos, D. (1986). Regularization of Inverse Visual Problems involving Discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4), 413-424.
- Tremblay, M., & Poussart, D. (1992). Focal plane VLSI processing for multiresolution edge extraction. *Proceedings of the SPIE: Intelligent Information Systems*, Orlando, Florida, 20-22 april, 1705, (12 p.).
- Ullman, S. (1981). Analysis of Visual Motion by Biological and Computer Systems. *Computer*, 14(8), 57-69.
- Vega-Riveros, J. F., & Jabbour, K. (1989). Review of motion analysis techniques. *IEE Proceedings I (Solid-State and Electron Devices)*, 136(6), 397-404.
- Waxman, A. M., & Duncan, J. H. (1986). Binocular Image Flows: Steps Toward Stereo-Motion Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), 715-729.
- Wechsler, H. (1990). *Computational Vision*. Boston: Academic Press.
- Weisstein, N., & Wong, E. (1987). Figure-Ground Organization Affects the Early Visual Processing of Information. In M. A. Arbib & A. R. Hanson (Ed.), *Vision, Brain and Cooperative Computation* (pp. 209-230). Cambridge, Mass.: The MIT Press.
- Whitten, G. (1990). A Framework for Adaptive Scale Space Tracking Solutions to Problems in Computational Vision. *Proceedings of the 3rd ICCV-90: Third International Conference on Computer Vision*, Osaka, Japan, Dec. 4-7, 210-220.
- Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for spatial vision. *Vision Research*, 19, 19-32.
- Yuille, A. L., & Ullman, S. (1990). Computational Theories of Low-Level Vision. In D. N. Osherson, S. M. Kosslyn & J. M. Hollerbach (Ed.), *Visual Cognition and Action* (pp. 5-39). Cambridge, Massachusetts: The MIT Press.
-

CHAPITRE 2 *Formation des images d'éclairément*

Ce chapitre décrit des algorithmes pour la synthèse d'images d'éclairément à partir d'images de profondeur. La section 2.1 explicite le processus de formation de l'image d'éclairément. La section 2.2 montre un algorithme de synthèse de l'image d'éclairément à partir d'une image de profondeur qui suppose que la surface de l'objet 3D est lambertienne tandis que la section 2.3 explicite les algorithmes utilisés pour générer des images stéréos d'éclairément à partir d'images de profondeur étalonnées.

2.1 FORMATION DES IMAGES

Nous nous intéressons dans cette section au processus de formation de l'image d'éclairément sur le plan image d'une caméra. Le processus de formation de l'image d'éclairément est illustré à la figure 2.1. Nous cherchons à calculer l'éclairément E qui frappe un point du plan image et correspond à la projection géométrique du rayon réfléchi à la surface de l'objet imagé en direction de l'orifice de la lentille sténopé. La formation des images est donc l'étude du lien entre la luminance L d'un point de la scène (rayon lumineux L réfléchi par la surface externe de la scène) et l'éclairément E de son point image sur le capteur. Nous débuterons par d'abord définir le concept de rayon lumineux.

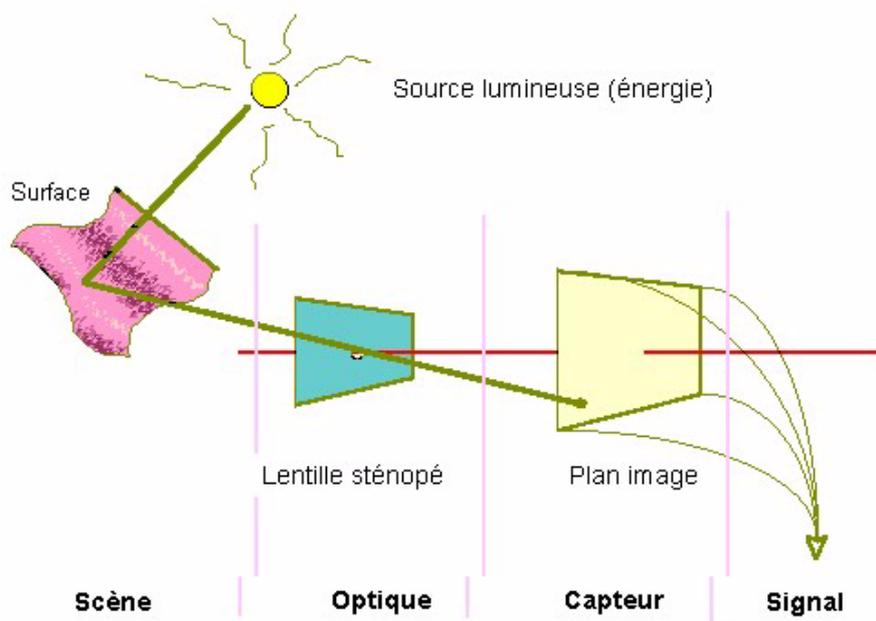
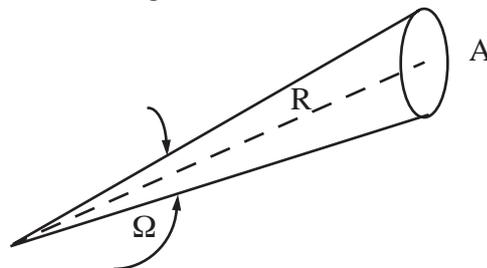


Figure 2.1 *Processus de formation de l'image.*
 La surface de l'objet réfléchit un rayon qui provient de la source lumineuse. Une fraction est réfléchiée en direction de la lentille et touche le plan image.

2.1.1 Équation du rayon lumineux

Quelques définitions pour bien poser le concept du rayon lumineux :

- **Angle solide Ω** : L'angle solide est une mesure de base en 3D pour les phénomènes rayonnant. C'est une extension du concept d'angle planaire étendu au 3D. C'est le rapport entre l'aire projetée sur une sphère et le carré du rayon de la sphère, comme l'illustre la figure ci-dessous.

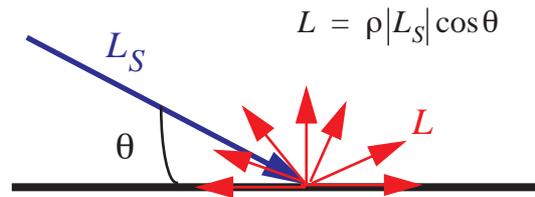


L'angle solide se mesure en stéradians (sr). L'angle solide est donné par :

$$\Omega = \frac{A}{R^2}$$

$$\Omega_{sphere} = 4\pi \text{ stéradians}$$

- **Source de lumière Φ** : Flux lumineux exprimé en lumens ou en candelas/sr (1 candela = 20,3 mW de lumière visible).
- **Surface lambertienne**: C'est une surface parfaitement diffusante et correspond à une surface recouverte de peinture mate. Un rayon lumineux est absorbé par la surface et une fraction du rayon est retransmise dans toutes les directions avec une amplitude égale.



Le facteur d'atténuation ρ est appelé l'albédo de la surface et est égal à $\frac{1}{\pi}$ pour une surface lambertienne.

On peut définir un rayon lumineux à partir d'une source de flux lumineux dont on commande l'amplitude et la direction à l'aide de deux écrans comme il est illustré à la figure 2.2. L'écran 1 permet de réduire le flux lumineux à un élément différentiel de flux $d\Phi$. L'écran 2 permet de réduire la surface orientée ΔA à un élément différentiel dA caractérisé par le vecteur normal à la surface \vec{n} . L'angle solide soutenu par la surface projetée dA est ainsi réduit à un élément différentiel d'angle solide $d\omega$. L'orientation de la surface par rapport à la direction du rayon lumineux \vec{L} est donnée par l'angle θ . L'unité de base de l'amplitude du rayon lumineux est la luminance L :

$$L = \frac{d^2 \Phi}{d\omega dA \cos \theta} \quad (2.1)$$

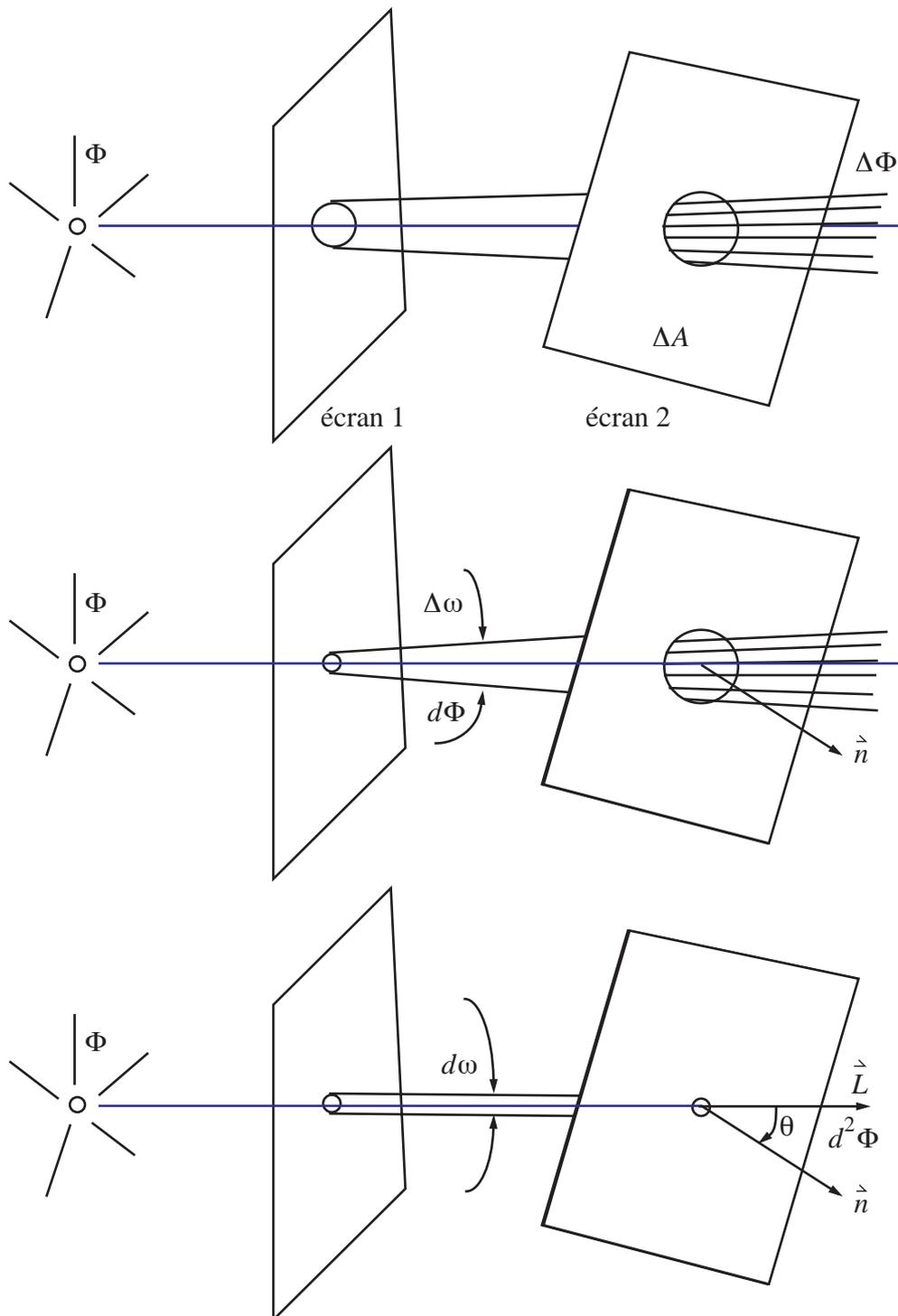


Figure 2.2 Définition d'un rayon lumineux L .

La luminance s'exprime en cd/m^2 ou en $lumens/m^2\text{str}$. La luminance nous donne la densité de flux par unité de surface projetée par unité d'angle solide.

L'éclairement lumineux E exprime la quantité de lumière que reçoit un élément de surface dA quelque soit l'orientation des rayons lumineux qui arrivent sur la surface. L'éclairement s'exprime en fonction de la luminance incidente :

$$E = \frac{d\Phi}{dA} \quad (2.2)$$

Les notions de luminance (rayon lumineux) et d'éclairement lumineux sont illustrés à la figure 2.3.

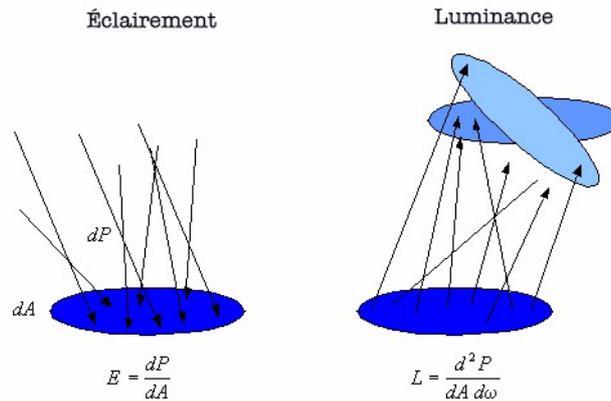


Figure 2.3 *Éclairement lumineux et luminance.*
L'unité fondamentale en radiométrie est le rayon lumineux L qui possède une direction et une amplitude. L'éclairement lumineux E mesure la quantité de lumière que reçoit un élément de surface.

2.1.2 Équation de formation des images

La géométrie de la formation de l'image est illustrée ci-dessous. L'élément

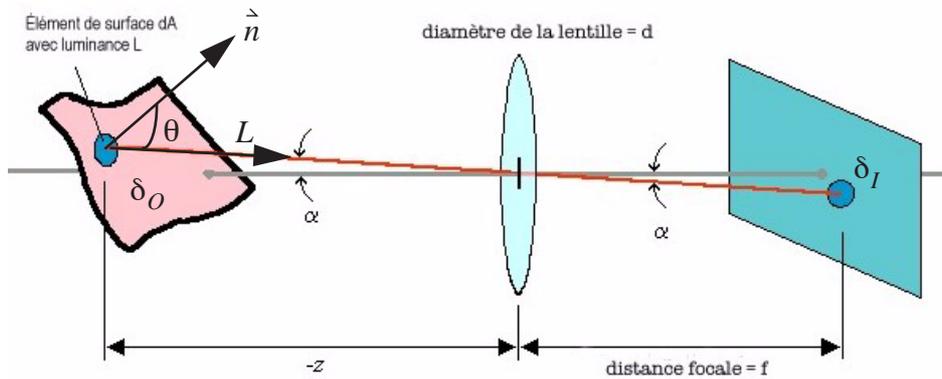


Figure 2.4 Géométrie de la formation des images.

de surface δ_O sur l'objet est projeté sur le plan image et apparaît en δ_I sur ce plan. Nous cherchons l'éclairement lumineux E qui sera absorbé et traduit en signal électrique par le pixel δ_I du plan image.

$$E = \frac{d\Phi}{\delta_I} \quad (2.3)$$

où $d\Phi$ représente le flux de lumière recueilli par la lentille de l'élément δ_O et concentré sur le plan image en l'élément δ_I .

La géométrie de la formation des images (Fig. 2.4) nous permet de trouver δ_I . Si l'angle solide soutenu par δ_O est appelé $d\omega_O$ et l'angle solide soutenu par δ_I est appelé $d\omega_I$, alors :

$$d\omega_O = \frac{\delta_O \cos \theta}{R_1^2} = \frac{\delta_O \cos \theta}{\left(\frac{z}{\cos \alpha}\right)^2}$$

$$d\omega_I = \frac{\delta_I \cos \alpha}{R_2^2} = \frac{\delta_I \cos \alpha}{\left(\frac{f}{\cos \alpha}\right)^2}$$

où R_1 est la distance source-lentille et R_2 est la distance plan-lentille.

Puisque les rayons passant par le centre de la lentille ne sont pas déviés, les deux angles solides sont égaux.

$$d\omega_O = d\omega_I \Rightarrow \frac{\delta_O}{\delta_I} = \left[\frac{z}{f} \right]^2 \frac{\cos \alpha}{\cos \theta} \quad (2.4)$$

Calculons la quantité de lumière $d\Phi$ provenant de l'élément de surface δ_O et concentré sur le plan image en δ_I par la lentille. La géométrie utilisée pour le calcul est illustrée ci-dessous.

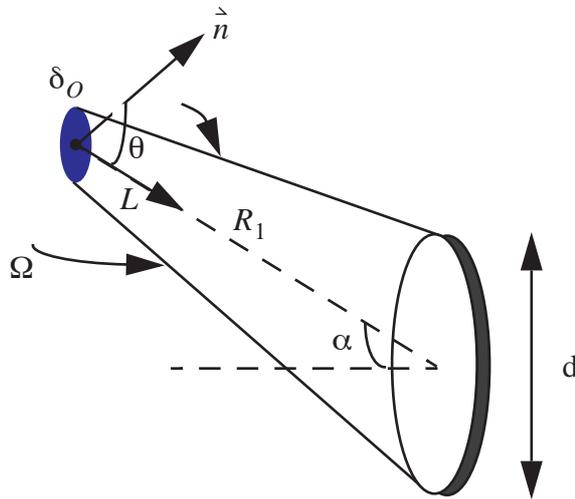


Figure 2.5 Géométrie de la lentille.
La lentille concentre tous les rayons provenant de l'élément de surface δ_O .

Soit d le diamètre de la lentille. La lentille sous-tend un angle solide Ω en δ_O . L'angle solide Ω est défini :

$$\Omega = \frac{A_{\text{lentille}} \cos \alpha}{R_1^2}$$

avec $A_{\text{lentille}} = \text{surface lentille} = \frac{\pi d^2}{4}$ et la distance source-lentille

$R_1 = \frac{z}{\cos \alpha}$. L'expression de l'angle solide devient :

$$\Omega = \frac{\frac{\pi d^2}{4} \cos \alpha}{\left(\frac{z}{\cos \alpha}\right)^2} = \frac{\pi}{4} \left(\frac{d}{z}\right)^2 \cos^3 \alpha \quad (2.5)$$

Le flux de lumière $d\Phi$ recueilli par la lentille est :

$$\begin{aligned} d\Phi &= L \Omega \delta_O \cos \theta \\ &= \frac{\pi}{4} \left(\frac{d}{z}\right)^2 \cos^3 \alpha \cos \theta L \delta_O \end{aligned}$$

selon la définition du rayon lumineux L (équation 2.1).

Le flux $d\Phi$ est concentré par la lentille sur l'élément de surface δ_I . Le flux de lumière absorbé par δ_I ne provient que de δ_O .

$$\begin{aligned} E &= \frac{d\Phi}{\delta_I} \\ &= \frac{\pi}{4} \left(\frac{d}{z}\right)^2 \cos^3 \alpha \cos \theta L \frac{\delta_O}{\delta_I} \end{aligned}$$

L'équation de formation de l'image s'obtient en substituant le résultat obtenu en (2.4) :

$$E = \frac{\pi}{4} \left(\frac{d}{f}\right)^2 \cos^4 \alpha L \quad (2.6)$$

Plusieurs observations peuvent être tirées de cette équation :

- $E \propto L$ L'éclairément d'un pixel est directement proportionnel à l'amplitude du rayon lumineux réfléchi par l'élément de surface imagé.

- E dépend fortement de l'angle α . Ce n'est donc que le centre de la lentille qui est utilisé afin d'assurer un éclairage sans atténuation du plan image.
- E ne dépend pas de z . L'éclairage d'un pixel sera le même et ne dépend que de l'orientation de la surface imagée par rapport à la source d'éclairage, que la surface imagée appartienne à un nuage, un édifice ou un objet à proximité de la caméra.
- Le paramètre $\frac{f}{d}$ est appelé l'ouverture du diaphragme et s'exprime par un chiffre qui indique le rapport entre la distance focale de l'objectif et l'ouverture de l'iris de la caméra. Ainsi, une ouverture $f:8$ pour une focale de 50 mm indiquera une ouverture effective de $\frac{50\text{mm}}{8}$, soit un diamètre de 6,25 mm. Les valeurs d'ouverture que l'on retrouve généralement sur une caméra sont : $f: 2, 2,8, 4, 5,6, 8, 11, 16, 22$, la plus grande ouverture étant $f:2$ et la plus petite, $f:22$. Dans l'équation de formation de l'image, plus le nombre d'ouverture est grand (e.g. $f:22$), plus le diamètre d'ouverture sera petit, et plus l'éclairage sur le plan image sera diminué pour un temps d'exposition donné.
Plus l'ouverture est grande (f: petit), plus le temps d'exposition doit être court (vitesse plus rapide) et plus la profondeur de champ est réduite.

2.1.3 Surface lambertienne

Lorsque la surface de l'objet à imager est lambertienne, l'équation (2.6) de formation des images revêt une forme intéressante. Lorsque la surface de l'objet est lambertienne, le rayon lumineux réfléchi en chaque point de la surface est d'amplitude constante dans toutes les directions et sa valeur est proportionnelle au cosinus de l'angle entre la direction de la source d'éclairage et la normale à la surface en ce point de l'objet. La géométrie utilisée pour les surfaces lamber-

tiennes est illustrée à la figure 2.6. Le rayon lumineux réfléchi par une surface

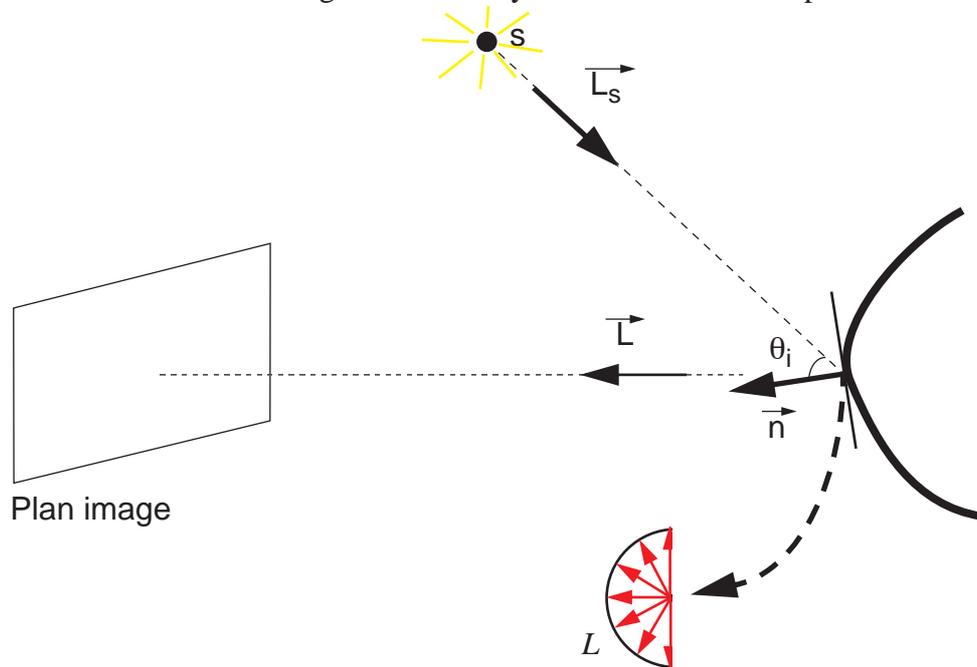


Figure 2.6 Formation des images pour des surfaces lambertiennes. Éclairage sur le plan image en fonction de la normale à la surface et la direction de la source lumineuse.

lambertienne éclairée par un rayon L_s a une amplitude :

$$L = \frac{1}{\pi} E_0 \cos \theta_i$$

L'éclairage maximum se produit lorsque la surface est perpendiculaire à la direction de la source d'éclairage. Notons cette valeur maximum E_0 .

L'équation de formation des images devient :

$$E = \left(\frac{d}{f}\right)^2 \cos^4 \alpha \frac{E_0}{4} \cos \theta_i \quad (2.7)$$

Le niveau de gris d'un pixel, proportionnel à l'éclairage absorbé, ne dépend que de l'angle entre la direction de la source et la normale à la surface.

2.2 SYNTHÈSE DE L'IMAGE D'ÉCLAIREMENT À PARTIR D'UNE IMAGE 3D

Nous allons dans cette section illustrer l'équation de formation des images au moyen d'une application pratique qui va nous permettre de synthétiser une image d'éclairément à partir d'une image de profondeur et de la position (qu'on peut varier) d'une source d'éclairage. Le procédé de synthèse nécessite au départ une image de profondeur qui décrit la distance entre le plan de la caméra et chaque point de la scène tridimensionnelle observée. Une image d'éclairément est ensuite générée à partir des spécifications de la surface visible de la scène 3D et des conditions d'éclairage.

2.2.1 Images de profondeur

Les images de profondeur proviennent d'une banque d'images construite au Conseil National de Recherche du Canada (Rioux, & Cournoyer, 1988). La caméra 3D mise au point au CNRC est un système laser à triangulation. La projection du faisceau laser et sa détection sont balayés à la même fréquence (détection synchronisée). Ce synchronisme résulte en une caméra plus compacte avec une grande résolution et une grande profondeur de champ. Ce type de caméra est équivalent à un système de triangulation laser avec une grande séparation entre la source laser et le récepteur. Le schéma de la caméra 3D du CNRC est illustré à la figure 2.7. L'image de profondeur résultante est constituée de valeurs de profondeur codées en format entier court (16 bits) mesurant la profondeur sur une grille régulière de 256x256 pixels. On obtient donc une formulation explicite des surfaces visibles de la scène observée $z = f(x, y)$. Les fichiers fournis par le CNRC ne contiennent que les valeurs de profondeur. Les fichiers sont d'abord convertis en fichiers-images de format KBV¹ qui contiennent dans leur en-tête les paramètres de numérisation Δx , Δy et Δz , paramètres que l'on retrouve dans le rapport du CNRC pour chacune des images de profondeur.

1. KBV est un logiciel de traitement et d'analyse d'images conçu et mis au point par Amerinex Advanced Imaging. Il a été remplacé par le logiciel Aphelion qui tourne sur PC (comparé à KBV qui tournait sous UNIX uniquement). Aphelion est une extension de KBV et de nombreux algorithmes de morphologie mathématique ont été notamment ajoutés.

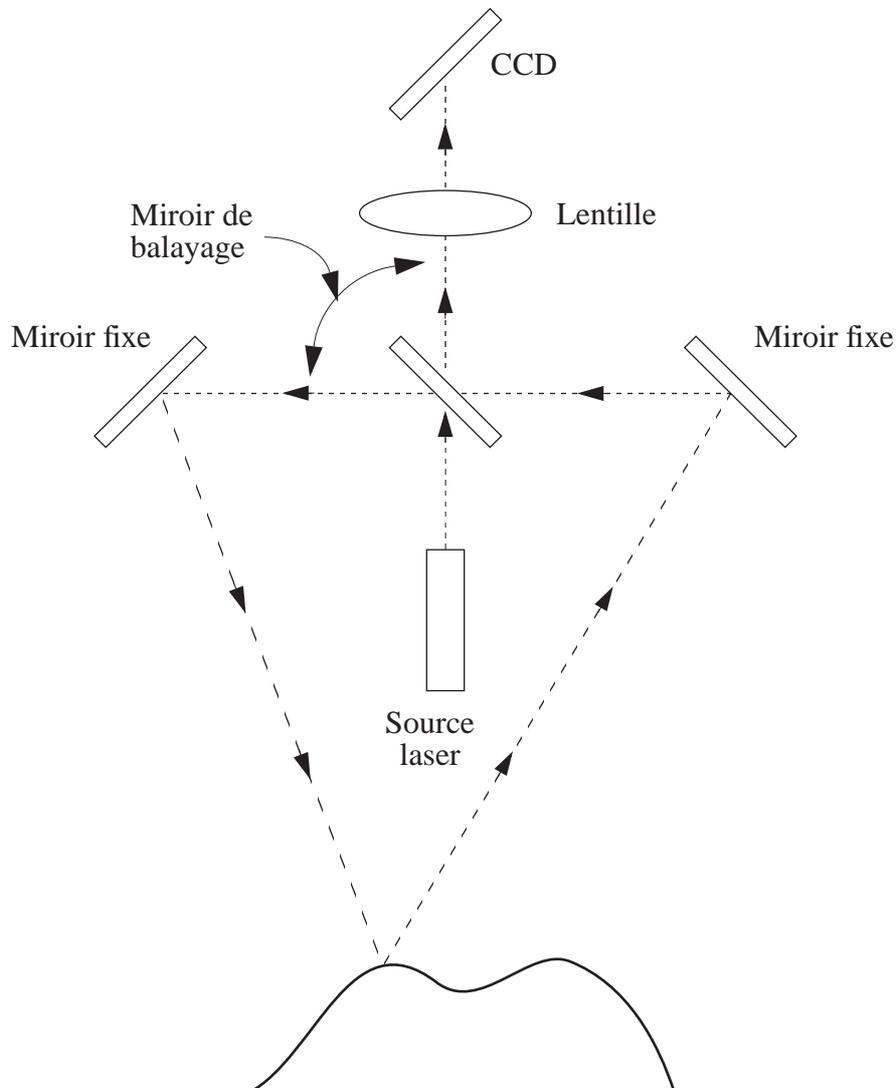


Figure 2.7 *Caméra 3D.*

*La caméra utilise une triangulation laser synchronisée.
La projection du faisceau laser et sa détection sont balayés à la même fréquence.*

2.2.2 Images d'éclairage

La première étape consiste à générer une image d'éclairage à partir de l'image de profondeur que l'on désire reconstituer et des conditions d'éclairage

que l'on désire fixer pour l'expérimentation. Les paramètres de l'expérimentation sont ainsi connus et nous avons un contrôle sur le bruit du système, le type de surface et l'illumination de la scène.

2.2.2.1 Hypothèses sur la scène 3D

Plusieurs hypothèses sont apportées sur la scène tridimensionnelle observée et son environnement. Ce sont:

1. L'éclairage de la scène est fourni par une seule source ponctuelle située loin des objets 3D observés.
2. Les surfaces de l'objet tridimensionnel sont lambertiennes. On assume donc que ces surfaces sont mates et réfléchissent le flux lumineux de telle sorte que la luminance diffusée soit constante pour toutes les directions d'observation. L'intensité lumineuse diffusée est proportionnelle au cosinus de l'angle entre la normale à la surface et la direction d'observation. La brillance, ou luminance, des points sur une image d'une surface lambertienne n'est pas fonction de la position de la caméra.
3. La projection sur le plan image est orthogonale (pas d'effets de perspective).

Ces hypothèses sont largement utilisées dans les travaux publiés portant sur le recouvrement de la forme à partir de l'ombrage. Ces hypothèses assurent un environnement simplifié pour lequel l'équation de formation des images est bien définie.

2.2.2.2 Géométrie

La géométrie utilisée pour décrire l'équation de formation des images est définie à la figure 2.8. La position de la source lumineuse est décrite par l'angle d'azimut (ou d'observation) θ et par l'angle d'élévation φ . Les composantes du

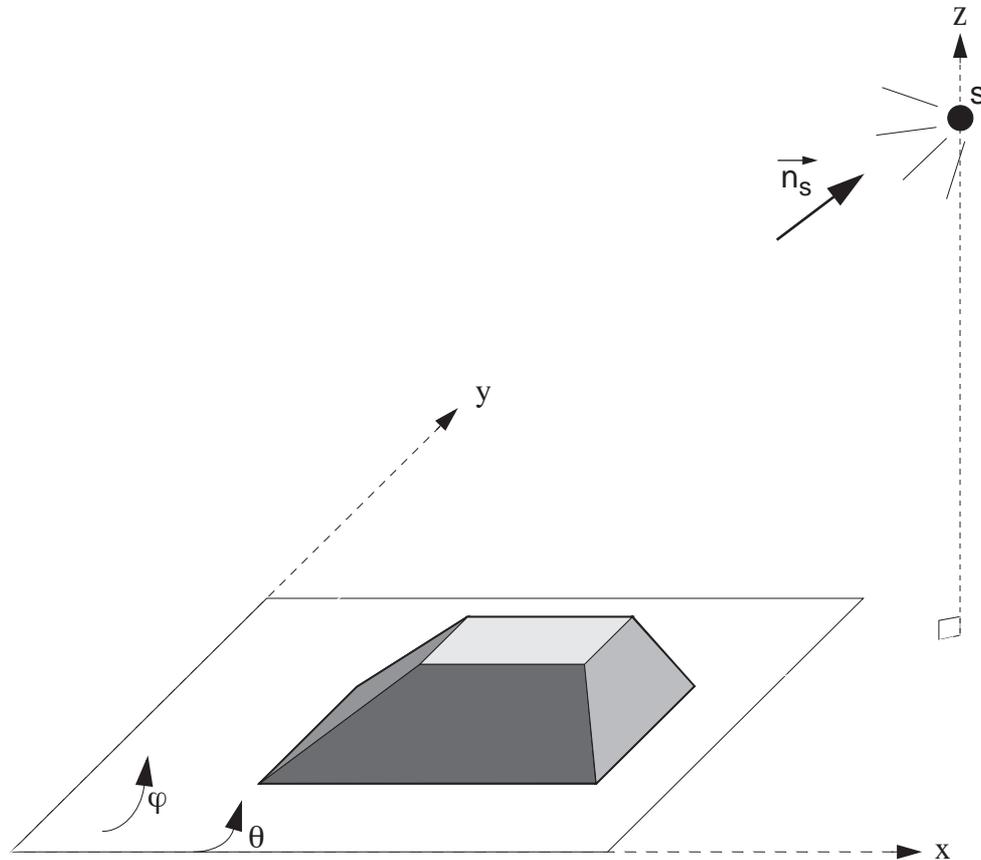


Figure 2.8 Géométrie pour la formation des images.
 La source lumineuse s est située à un angle d'azimut (ou d'observation) θ et à un angle d'élévation φ . On suppose que la projection sur le plan image x - y est orthogonale.

vecteur unitaire n_s qui pointe dans la direction de la source lumineuse sont (lorsque le vecteur est ramené à l'origine):

$$\vec{n}_s = (\cos\varphi \cos\theta, \cos\varphi \sin\theta, \sin\varphi) \quad (2.8)$$

Dans ce référentiel centré sur le plan image, l'objet y est défini comme une fonction explicite:

$$z = f(x, y) \quad (2.9)$$

La position de la source lumineuse et l'équation de description de la scène

tridimensionnelle sont établies dans un référentiel centré sur la scène observée. Afin d'étudier plus facilement les principes de la formation des images et ainsi relier l'apparence d'une image avec l'orientation des surfaces dans la scène, il faut utiliser un référentiel centré sur l'observateur plutôt que sur l'objet. Nous allons maintenant décrire la scène et la direction de la source lumineuse dans un référentiel centré sur l'une des caméras d'observation (par convention, nous utilisons la caméra gauche). De cette façon, nous pouvons décrire la scène et la direction de la source lumineuse dans le même référentiel que celui des images saisies. La géométrie utilisée est illustrée à la figure 2.9. Les surfaces visibles de la scène observée et la direction de la source lumineuse y sont exprimées sous forme paramétrique.

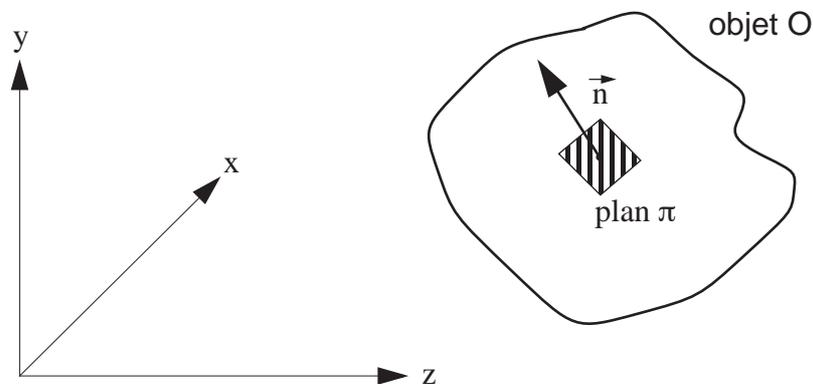


Figure 2.9 Référentiel centré sur l'observateur. Géométrie pour la formation des images dans un référentiel centré sur l'observateur. Pour un point à la surface d'un objet O sans variations brusques de surface, on peut faire passer un plan tangent π dont la normale est proche de celle à la surface de l'objet O .

Les surfaces visibles de la scène sont décrites par les gradients de surface p et q (Horn, 1977; Horn, 1986; Horn, & Sjoberg, 1989):

$$p = \frac{\partial z}{\partial x} \quad q = \frac{\partial z}{\partial y} \quad (2.10)$$

Les paramètres p et q décrivent le gradient selon l'horizontale et selon la verticale de la surface externe de l'objet au point (x, y) . L'espace des gradients p - q est illustré à la figure 2.10. Le vecteur normal à la surface, n , caractérise la surface en tout point du référentiel choisi. Définissons donc le vecteur n en fonction des gradients p et q grâce à la figure 2.11 ci-dessous. Les vecteurs r_x

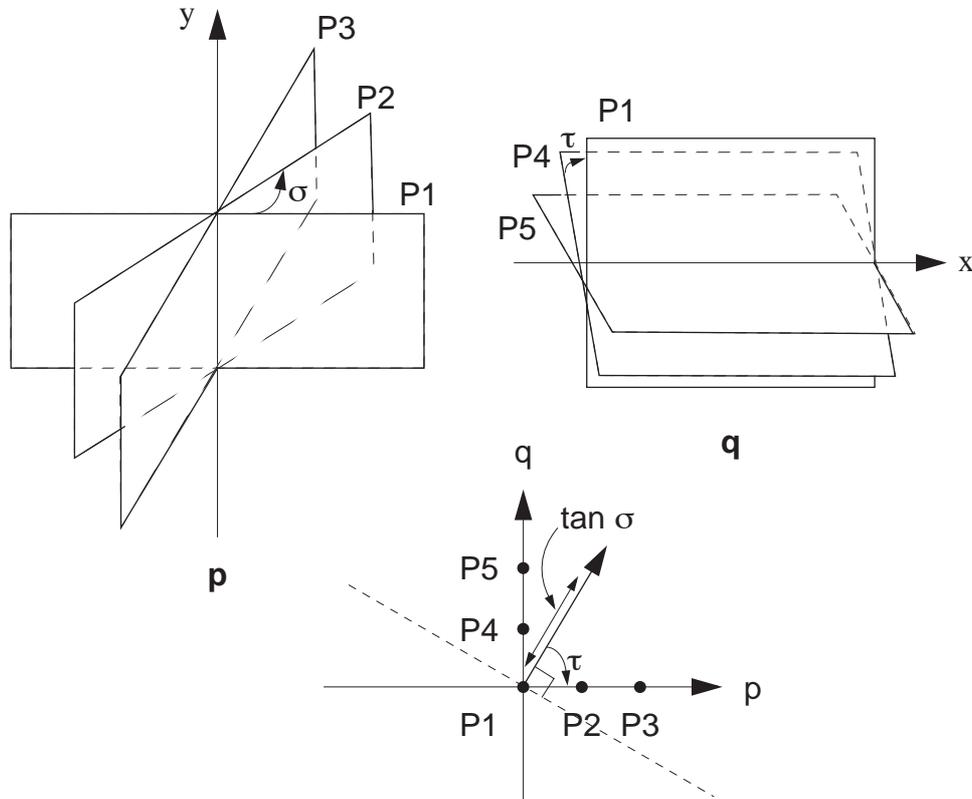


Figure 2.10 Espace des gradients.

La coordonnée p est associée à un rectangle pouvant tourner autour de l'axe vertical (y) tandis que la coordonnée q est associée à un rectangle en rotation autour de l'axe horizontal (x).

et \vec{r}_y bordent le plan π qui est tangent à la surface de l'objet au point (x, y) . Les deux vecteurs ont comme coordonnées:

$$\begin{aligned} \vec{r}_x &= (\delta x, 0, p\delta x) \\ \vec{r}_y &= (0, \delta y, q\delta y) \end{aligned} \quad (2.11)$$

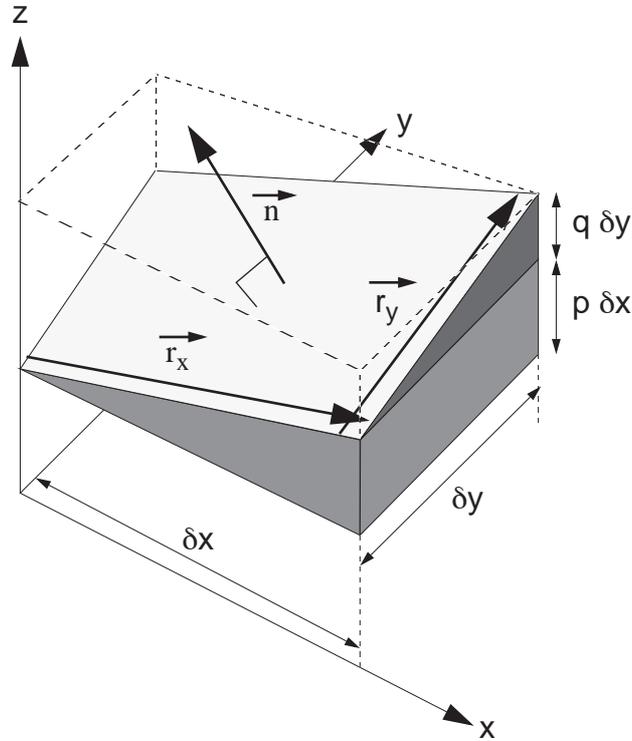


Figure 2.11 Gradients de la surface.
Définition du vecteur normal à la surface, \vec{n} , en fonction des paramètres p et q .

Nous obtenons en normalisant:

$$\begin{aligned}\vec{r}_x &= (1, 0, p) \\ \vec{r}_y &= (0, 1, q)\end{aligned}\tag{2.12}$$

Le produit vectoriel $r_x \times r_y$ résulte en un vecteur perpendiculaire au plan formé par r_x et r_y (plan π) et est donc colinéaire avec le vecteur normal n au point (x, y) . Après avoir calculé le produit vectoriel et après avoir normalisé, le vecteur unitaire normal à la surface est:

$$\vec{n} = \frac{(-p, -q, 1)}{\sqrt{1 + p^2 + q^2}}\tag{2.13}$$

Il est avantageux d'exprimer le vecteur unitaire n_s avec les mêmes paramètres. On cherche donc à exprimer la direction de la source lumineuse en terme des paramètres p et q . Il s'agit de trouver une surface dont la normale pointe dans la direction de la source lumineuse. L'équation 2.8 peut donc s'exprimer en fonction des gradients p_s et q_s qui définissent une surface dont la normale est parallèle au vecteur n_s (la surface ainsi trouvée est donc perpendiculaire au vecteur n_s):

$$\hat{n}_s = \frac{(-p_s, -q_s, 1)}{\sqrt{1 + p_s^2 + q_s^2}} \quad (2.14)$$

Nous avons exprimé les surfaces visibles de la scène et la direction de la source lumineuse en fonction de l'espace des gradients de surface. Il nous reste à aborder l'équation de formation des images qui relie l'éclairage produit en un point de l'image à l'orientation de la surface visible en ce point et à l'orientation de la source lumineuse.

2.2.2.3 Équation de formation des images

La géométrie utilisée pour décrire la formation des images est illustrée à la figure 2.12. L'éclairage fourni par une source lumineuse ponctuelle s sur une

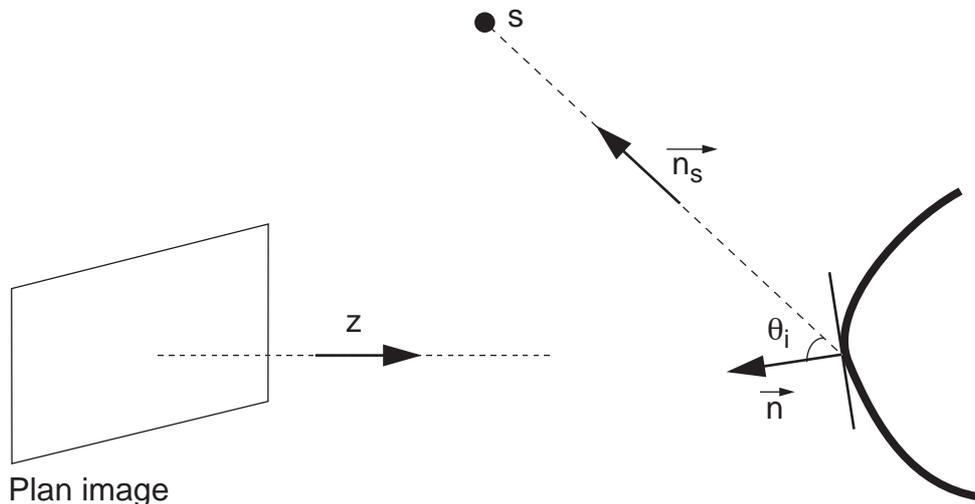


Figure 2.12 Éclairage sur le plan image.
Éclairage sur le plan image en fonction de la surface et de la direction de la source lumineuse.

surface orientée à un angle θ_i est simplement $E_0 \cos \theta_i$, où E_0 est l'éclairément maximal lorsque la source lumineuse éclaire à incidence normale. Comme nous assumons que la surface de l'objet est lambertienne, la luminance réfléchiée par la surface sera $L = \frac{1}{\pi}(E_0 \cos \theta_i)$, et ce, dans n'importe quelle direction d'observation, et donc dans la direction qui nous intéresse, $-z$. L'équation fondamentale de formation des images relie la luminance en un point de la scène à l'éclairément sur le plan image du même point projeté (Horn, 1986):

$$E_{image} = \frac{\pi}{4} \left(\frac{d}{f}\right)^2 \cos^4 \alpha \quad L = \left(\frac{d}{f}\right)^2 \cos^4 \alpha \quad \frac{E_0}{4} \cos \theta_i \quad (2.15)$$

où d est le diamètre de la lentille, f est la distance entre la lentille et le plan image et α est l'angle de visualisation de la caméra - tous des paramètres constants. L'éclairément en un point de l'image ne dépend donc que du cosinus de l'angle entre le vecteur normal à la surface et la direction de la source lumineuse. Ce cosinus se calcule par le produit scalaire entre ces deux vecteurs:

$$\cos \theta_i = \frac{\vec{n}_s \cdot \vec{n}}{\|\vec{n}_s\| \|\vec{n}\|} \quad (2.16)$$

En substituant les valeurs trouvées aux équations 2.13 et 2.14, nous pouvons exprimer le cosinus de θ_i en fonction seulement des gradients de surface:

$$R(p, q) = \cos \theta_i = \frac{1 + pp_s + qq_s}{\sqrt{1 + p_s^2 + q_s^2} \sqrt{1 + p^2 + q^2}} \quad (2.17)$$

Nous pouvons substituer dans l'équation 2.15 l'expression obtenue pour $\cos \theta_i$ et nous obtenons alors, à un facteur constant près, l'expression de l'éclairément sur le plan image en fonction des gradients de surface et de la direction de la source lumineuse:

$$E_{image} = R(p, q) = \frac{1 + pp_s + qq_s}{\sqrt{1 + p_s^2 + q_s^2} \sqrt{1 + p^2 + q^2}} \quad (2.18)$$

La relation $R(p, q)$, appelée *carte de réflectance*, rend explicite la relation entre l'orientation de surface de l'objet, la luminance à la surface, son éclairément

dans l'image et l'orientation de la source lumineuse¹.

Nous allons maintenant exprimer les gradients p_s et q_s en fonction des angles d'azimut θ et d'élévation φ utilisés pour localiser la source lumineuse ponctuelle. Pour ce faire, nous n'avons qu'à égaler terme à terme les équations 2.8 et 2.14 définissant le vecteur \vec{n}_s :

$$p_s = \frac{-\cos\theta}{\tan\varphi} \quad q_s = \frac{-\sin\theta}{\tan\varphi} \quad (2.19)$$

Les équations 2.18 et 2.19 définissent complètement le processus de formation des images. Les angles d'azimut et d'élévation sont fournis en paramètres. Il reste à déterminer les gradients de surface p et q à partir de l'image tridimensionnelle qui décrit la scène. Ces gradients peuvent être calculés soit en appliquant des masques de dérivée partielle sur les données 3D ou soit en calculant les paramètres du meilleur plan tangent qui passe par les données 3D au point considéré.

2.2.2.4 Calcul des gradients de surface par dérivée partielle

On cherche à déterminer le gradient de surface (p,q) à chaque coordonnée (x,y) de l'image. Afin de calculer le gradient local (pente de la surface selon x et selon y), il faut considérer les données de profondeur au voisinage du point (x,y) considéré. Les gradients p et q sont calculés à l'aide d'équations aux différences finies qui approximent l'opération de dérivation:

$$\begin{aligned} G_p(x, y) &= \frac{\partial}{\partial x} f(x, y) \cong \frac{f(x + \delta x, y) - f(x, y)}{\delta x} \delta z \\ G_q(x, y) &= \frac{\partial}{\partial y} f(x, y) \cong \frac{f(x, y + \delta y) - f(x, y)}{\delta y} \delta z \end{aligned} \quad (2.20)$$

Les masques qui sont appliqués sur l'image pour obtenir chacun des gradients

1. Dans le cas de n sources, le principe de superposition s'applique et
 $E = \cos\theta_1 + \cos\theta_2 + \dots + \cos\theta_n$

sont:

$$\begin{aligned} D_x &= \left(\frac{\delta z}{\delta x} \right) \begin{bmatrix} -1 & 1 \end{bmatrix} \\ D_y &= \left(\frac{\delta z}{\delta y} \right) \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{aligned} \quad (2.21)$$

Cette mesure de dérivée avec seulement deux valeurs est sujette au bruit local de saisie des données 3D et peut entraîner de fortes variations dans la mesure de la dérivée. Un moyen de réduire le bruit local est d'utiliser un voisinage plus large autour du point d'intérêt, ce qui revient à calculer la moyenne des valeurs utilisées pour le calcul de la dérivée. Par exemple, pour le calcul du gradient en x, on peut prendre pour le calcul la valeur moyenne de la colonne supérieure (d'une longueur de 3 pixels, pondérée ou non) moins la valeur moyenne de la colonne de gauche. Différents masques ont été créés, la pondération allouée aux différentes positions du masque servant à les différencier. Les masques les plus utilisés sont ceux de Sobel et de Prewitt (Besançon, 1988; Gonzalez, & Wintz, 1977). Ces masques sont de dimension 3x3, ce qui occasionne cependant un certain adoucissement des arêtes détectées par rapport au masque décrit plus haut (équation 2.21). Le masque de Sobel est défini:

$$\begin{aligned} D_x &= \frac{1}{8} \frac{\delta z}{\delta x} \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \\ D_y &= \frac{1}{8} \frac{\delta z}{\delta y} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \end{aligned} \quad (2.22)$$

Pour sa part, le masque de Prewitt est défini:

$$D_x = \frac{1}{6} \frac{\delta z}{\delta x} \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad (2.23)$$

$$D_y = \frac{1}{6} \frac{\delta z}{\delta y} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Le choix du masque approprié de différentiation est gouverné par le compromis suivant:

- un petit masque est utilisé pour détecter des arêtes minces, mais est plus sensible au bruit local de numérisation.
- un grand masque adoucit les arêtes détectées, mais assure un moyennage du bruit de numérisation.

Le masque est donc surtout choisi en fonction de la qualité de l'image de profondeur et du type d'arêtes rencontrées.

2.2.2.5 Calcul des gradients de surface par la méthode du plan tangent

Le calcul du gradient local de surface (p,q) à chaque position (x,y) de l'image peut aussi s'effectuer par la méthode du plan tangent. Cette dernière consiste à calculer les paramètres du plan qui est tangent à la surface au point d'intérêt. Ce plan est celui qui passe au mieux, selon un critère de moindres carrés, auprès des données de profondeur expérimentales dans un voisinage autour du point d'intérêt. Une approche de solution possible utilise la technique des pièces rectangulaires décrite en détail dans la thèse de doctorat de Denis Laurendeau. Le contenu de cette sous-section s'inspire largement de ces travaux (Laurendeau, 1985).

La technique des pièces rectangulaires consiste à projeter sur la surface de l'objet une grille définie autour du point d'intérêt sur le plan image, tel qu'illustré à la figure 2.13. Les points projetés sur la surface forment un nuage de points sur la surface externe. Les points projetés font partie d'un plan qui est tangent à

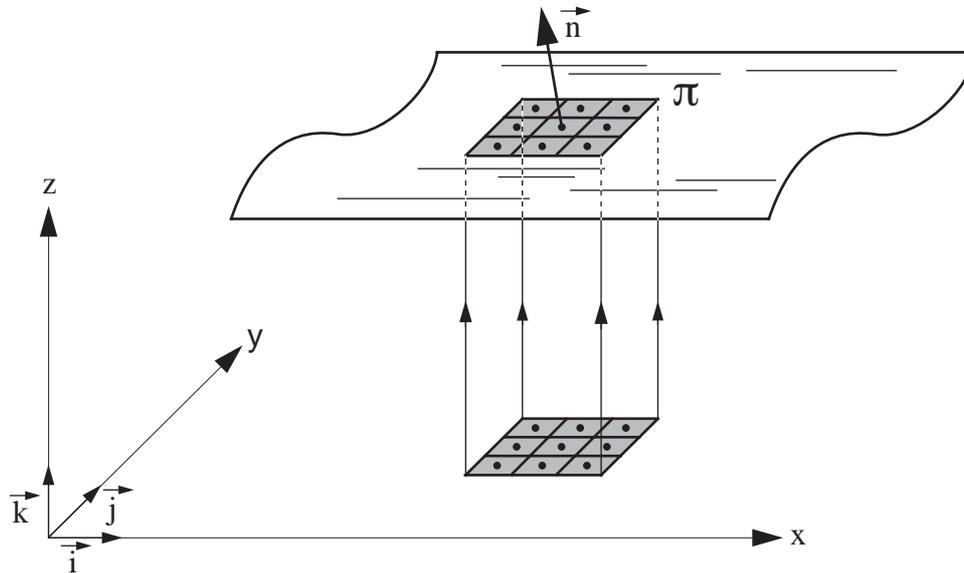


Figure 2.13 Méthode du plan tangent.

Calcul des gradients de surface par la méthode du meilleur plan tangent au sens des moindres carrés.

la surface au point (x, y) si la grille est suffisamment fine par rapport aux dimensions de l'objet tridimensionnel. L'équation du plan π est:

$$ax + by + cz + d = 0 \quad (2.24)$$

Le vecteur normal à ce plan a comme composantes:

$$\vec{n} = a\vec{i} + b\vec{j} + c\vec{k} \quad (2.25)$$

L'on cherche donc les paramètres de l'équation du plan π . Ces paramètres sont obtenus par la méthode des moindres carrés en calculant le plan qui passe le plus près possible des valeurs de profondeur échantillonnées au voisinage du point d'intérêt. L'équation du plan peut s'écrire:

$$z = \alpha x + \beta y + \gamma \quad (2.26)$$

$$\text{avec } p = \alpha = -\frac{a}{c} \quad q = \beta = -\frac{b}{c} \quad \gamma = -\frac{d}{c}$$

À chaque point (x_i, y_i) de la grille correspond une valeur de profondeur z_i . L'équation 2.26 permet donc d'obtenir 9 équations à 3 inconnues, ce qui

constitue un système surdéterminé. Le système d'équations s'écrit, sous forme matricielle:

$$Av = z \quad (2.27)$$

avec

$$A = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_9 \end{bmatrix} \quad v = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \quad x_i = [x_i \ y_i \ 1] \quad z = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_9 \end{bmatrix}$$

On résout le système d'équations en utilisant la méthode de Cholesky:

$$A^t A v = A^t z \quad (2.28)$$

La solution de cette équation nous fournit les paramètres α , β et γ du plan qui passe au mieux, selon un critère de moindre carrés, par les valeurs expérimentales de profondeur dans le voisinage du point d'intérêt. Afin de déterminer les paramètres a , b , c et d de ce plan, la contrainte suivante est utilisée:

$$a^2 + b^2 + c^2 + d^2 = 1 \quad (2.29)$$

Les paramètres du plan sont alors donnés:

$$\begin{aligned} a &= -\frac{\alpha}{\sqrt{1 + \alpha^2 + \beta^2 + \gamma^2}} & b &= -\frac{\beta}{\sqrt{1 + \alpha^2 + \beta^2 + \gamma^2}} \\ c &= \frac{1}{\sqrt{1 + \alpha^2 + \beta^2 + \gamma^2}} & d &= -\frac{\gamma}{\sqrt{1 + \alpha^2 + \beta^2 + \gamma^2}} \end{aligned} \quad (2.30)$$

La résolution de l'équation 2.28 fournit les paramètres α , β et γ pour chaque

pixel de l'image de profondeur. L'équation 2.28 s'écrit:

$$\begin{bmatrix} x_1 & x_2 & \dots & x_9 \\ y_1 & y_2 & \dots & y_9 \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \dots & \dots & \dots \\ x_9 & y_9 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_9 \\ y_1 & y_2 & \dots & y_9 \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_9 \end{bmatrix} \quad (2.31)$$

Pour simplifier la recherche de la solution, nous constatons que toutes les pièces rectangulaires sont semblables (même longueur et même largeur), sauf en ce qui concerne leur position dans l'image et les valeurs de profondeur (z_i) des différents éléments. Les calculs sont simplifiés si le centre de chaque pièce est déplacé à l'origine du système de coordonnées. De cette façon, la matrice des coefficients $A^t A$ est commune à toutes les pièces et seulement la matrice $A^t z$ doit être calculée séparément pour chaque pièce. La figure 2.14 illustre le déplacement d'une pièce rectangulaire vers l'origine du système de coordonnées.

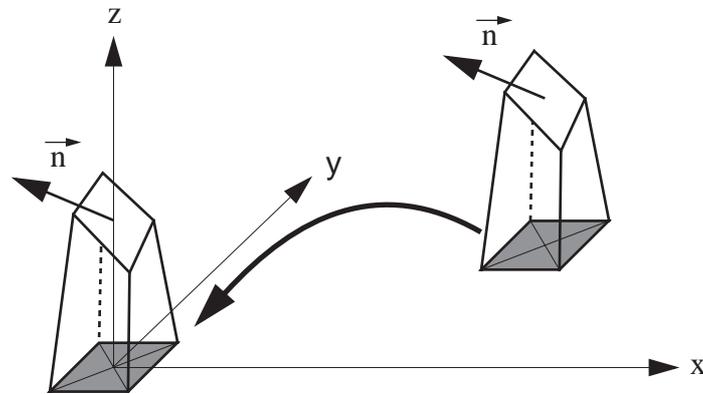


Figure 2.14 Déplacement à l'origine.
Déplacement d'une pièce rectangulaire et son plan tangent à l'origine.

La solution de l'équation 2.31 s'écrit:

$$\begin{aligned}\gamma &= \frac{\sigma_{x^2}\sigma_{y^2}\sigma_z - \sigma_z\sigma_{xy}^2 - \sigma_{y^2}\sigma_{xz}\sigma_x - \sigma_{x^2}\sigma_y\sigma_{yz} + \sigma_y\sigma_{xz}\sigma_{xy} + \sigma_{yz}\sigma_{xy}\sigma_x}{\sigma_{x^2}\sigma_{y^2}\sigma - \sigma\sigma_{xy}^2 - \sigma_{y^2}\sigma_x^2 - \sigma_{x^2}\sigma_y^2 + 2\sigma_y\sigma_{xy}\sigma_x} \\ q = \beta &= \frac{\sigma_{yz}\sigma_{x^2} - \sigma_{xz}\sigma_{xy} - \gamma(\sigma_{x^2}\sigma_y - \sigma_{xy}\sigma_x)}{\sigma_{x^2}\sigma_{y^2} - \sigma_{xy}^2} \\ p = \alpha &= \frac{\sigma_{xz} - \gamma\sigma_x - \beta\sigma_{xy}}{\sigma_{x^2}}\end{aligned}\quad (2.32)$$

Le paramètre σ indique le nombre de cellules d'évaluation de la pièce rectangulaire ($\sigma = 9$ pour une grille de 3×3 , tel qu'illustré à la figure 2.13). Les paramètres suivants ne sont calculés qu'une seule fois:

$$\begin{aligned}\sigma_{x^2} &= \sum_{i=1}^{\sigma} x_i^2 & \sigma_{y^2} &= \sum_{i=1}^{\sigma} y_i^2 \\ \sigma_{xy} &= \sum_{i=1}^{\sigma} x_i y_i \\ \sigma_x &= \sum_{i=1}^{\sigma} x_i & \sigma_y &= \sum_{i=1}^{\sigma} y_i\end{aligned}\quad (2.33)$$

Pour chaque pièce rectangulaire dont on calcule les gradients p et q , il faut calculer les paramètres suivants:

$$\begin{aligned}\sigma_z &= \sum_{i=1}^{\sigma} z_i \\ \sigma_{xz} &= \sum_{i=1}^{\sigma} x_i z_i & \sigma_{yz} &= \sum_{i=1}^{\sigma} y_i z_i\end{aligned}\quad (2.34)$$

À partir de ces deux systèmes d'équations, les gradients p et q sont calculés à l'aide du système d'équations 2.32.

2.2.2.6 Programmation de l'algorithme

La figure 2.15 illustre l'image d'éclairage résultante d'un pion de jeu

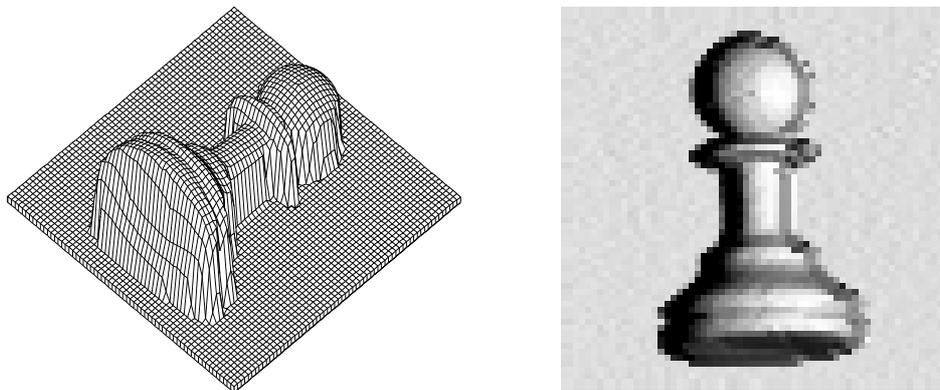


Figure 2.15 Image de profondeur et d'éclairage d'un pion.
La source lumineuse est située à 0° d'azimut et à 60° d'élévation. Les gradients de surface sont calculés avec un masque de Sobel.

d'échecs éclairé par une source lumineuse localisée à droite et à 60° d'élévation. La figure 2.16 illustre la synthèse de quatre images d'éclairage à partir de la même image de profondeur mais sous quatre conditions différentes d'illumination.

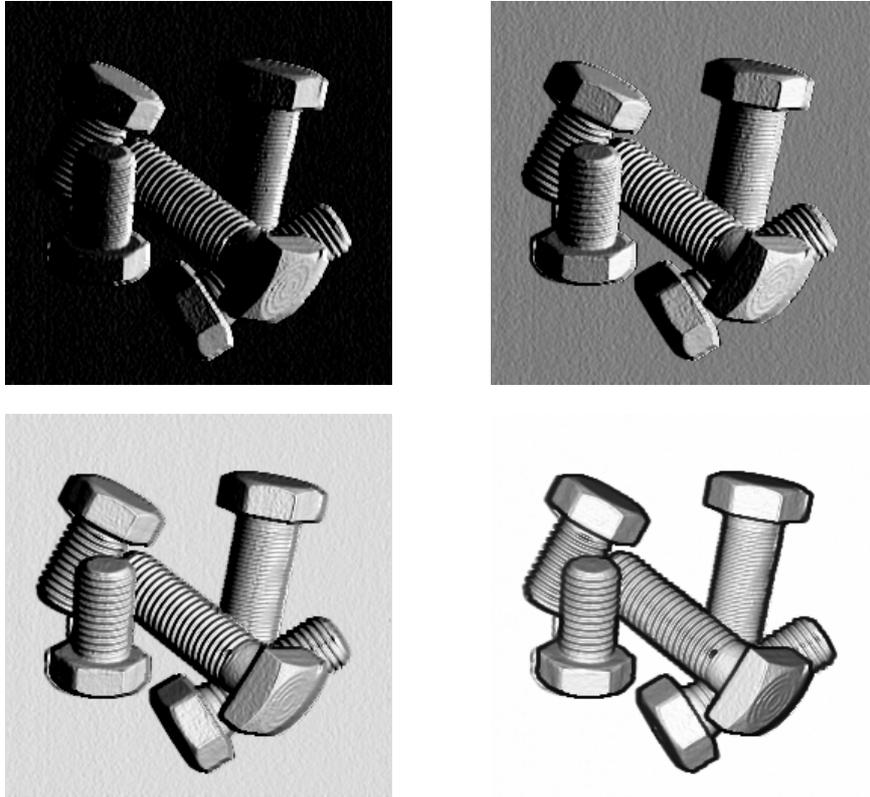
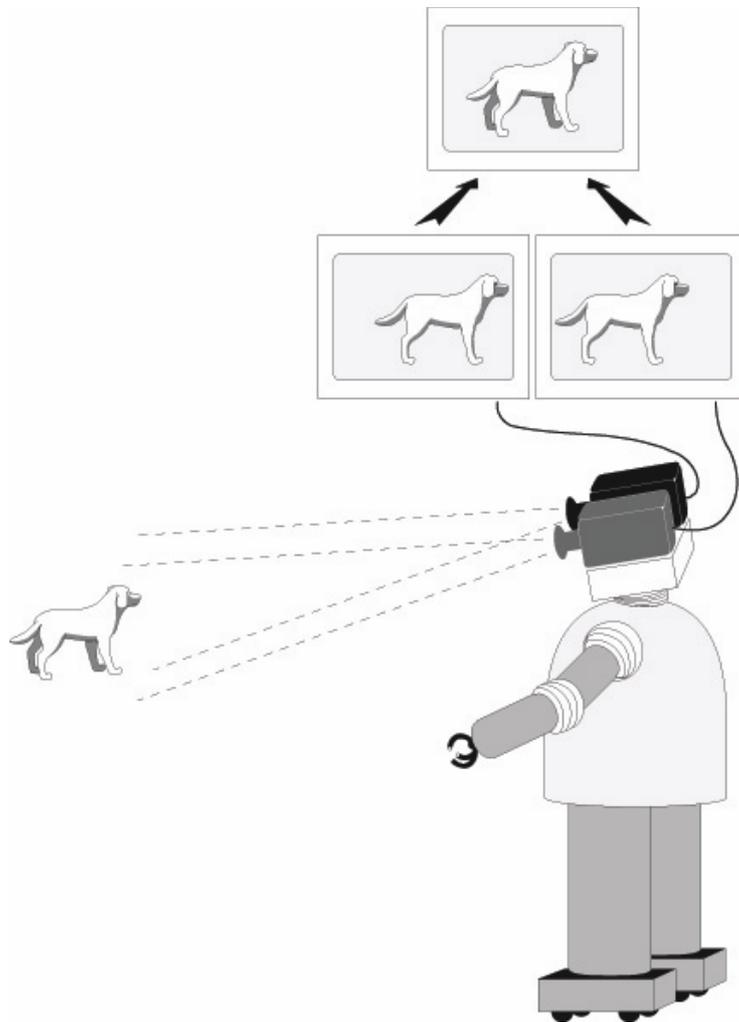


Figure 2.16 Jeu de boulons éclairé sous différents angles.
Images d'éclairage d'un jeu de boulons éclairé à différents angles d'élévation. La source lumineuse est située à droite (angle d'azimut nul), à un angle d'élévation passant de 0° , 30° , 60° et 90° sur les images balayées de la gauche vers la droite et du haut vers le bas.

2.3 GÉNÉRATION D'IMAGES STÉRÉOS D'ÉCLAIREMENT À PARTIR D'UNE IMAGE 3D

Un système de vision stéréographique requiert deux images d'éclairément à son entrée: une image provenant de la caméra gauche et une autre provenant de la droite. Ces deux images correspondent à deux points de vue légèrement différents de la scène tridimensionnelle que l'on essaie d'interpréter, comme illustré à la figure 2.17.



*Figure 2.17 Illustration d'un système de vision.
Le système de vision utilise à son entrée deux images légèrement différentes de la scène tridimensionnelle observée.*

Les deux images d'entrée peuvent provenir d'images réelles obtenues de caméras aux paramètres géométriques connus ou peuvent être synthétisées à partir d'une image de profondeur étalonnée et qui représente finalement le résultat auquel on espère parvenir. Nous avons opté dans ce document pour la seconde solution pour les motifs suivants :

1. L'équation de formation des images bénéficie d'un cadre mathématique bien défini, de telle sorte que l'éclairément en un point est déterminée de façon univoque par les propriétés de la surface projetée, l'éclairage de la scène et par la géométrie des caméras.
2. Les résultats que l'on obtient à la sortie des algorithmes peuvent facilement être comparés avec l'image de profondeur de départ.
3. La synthèse des images d'éclairément nous permet un contrôle absolu des paramètres de l'expérimentation. L'expérimentateur peut ainsi fixer les paramètres reliés à la géométrie des caméras tel la distance focale, la séparation entre les deux caméras et l'orientation des plans images. De même, il peut fixer les paramètres reliés à la géométrie de la scène, notamment la distance entre la scène observée et les caméras.

Les deux images stéréo de gauche et de droite sont synthétisées à partir de l'image de profondeur et de l'image d'éclairément qui a été générée selon la procédure développée à la section précédente.

Références

- Besançon, J. E. (1988). *Vision par ordinateur en deux et trois dimensions* . Paris: Eyrolles.
- Gonzalez, R. C., & Wintz, P. (1977). *Digital Image Processing* . London: Addison-Wesley.
- Horn, B. K. P. (1977). Understanding Image Intensities. *Artificial Intelligence*, 8(2), 201-231.
- Horn, B. K. P. (1986). *Robot Vision* . Cambridge, Mass.: The MIT Press.
- Horn, B. K. P., & Sjoberg, R. W. (1989). Calculating the Reflectance Map. In B. K. P. Horn & M. J. Brooks (Ed.), *Shape from Shading* (pp. 215-244). Cambridge, Mass.: The MIT Press.
- Laurendeau, D. (1985). *Construction du modèle d'objets tri-dimensionnels à partir d'images 3D*. Thèse de doctorat (PhD), Université Laval.
- Rioux, M., & Cournoyer, L. (1988). Les fichiers d'images tridimensionnelles du CNRC (No. 29077). CNRC.
-

CHAPITRE 3 *L'analyse multirésolution en vision informatique*

Une image est constituée d'une multitude de primitives visuelles qui prennent la forme d'arêtes, de coins, de lignes ouvertes ou de contours fermés. La détection de ces primitives constitue la première étape de calcul dans le modèle computationnel du système de vision. Cette étape produit le *croquis*, qui est le premier niveau de représentation interne de données. Les primitives présentes dans l'image sont de dimensions très variées et leur détection nécessite des opérateurs de différentes grandeurs.

Une façon équivalente d'aborder ce problème de détection des primitives de l'image consiste à analyser l'image d'entrée à plusieurs niveaux de résolution spatiale (Rosenfeld, 1984a; Uhr, 1987). L'analyse multirésolution fait partie d'une théorie plus générale et plus moderne, l'analyse par ondelettes (Hlawatsch, & Bouderaux-Bartels, 1992; Mallat, & Zhong, 1992; Meyer, Jaffard, & Rioul, 1987; Rioul, & Vetterli, 1991). L'image d'entrée, à pleine résolution, est filtrée par un filtre passe-bas afin d'atténuer les primitives aux hautes fréquences (i.e. les variations brusques) présentes dans l'image. Nous obtenons ainsi une image à résolution plus grossière sur laquelle nous pouvons appliquer des opérateurs de détection de primitives moins sensibles aux petites variations d'éclairement qui sont souvent le résultat du bruit d'échantillonnage et/ou du bruit de saisie de données. Par contre, les primitives détectées à plus faible résolution sont moins bien localisées que sur l'image à pleine résolution. L'analyse multirésolution consiste à obtenir plusieurs niveaux de résolution spatiale en filtrant l'image d'entrée à pleine résolution à l'aide de filtres passe-bas de diffé-

rentes largeur de bande. Les primitives des images aux différentes résolutions spatiales sont par la suite détectées à chaque niveau. Les primitives trouvées à basse résolution sont ensuite projetées aux niveaux de plus haute résolution afin de guider l'identification et la localisation des primitives aux échelles de résolution de plus en plus fine. Cette recherche hiérarchique permet d'atténuer les primitives bruitées qui n'apparaissent qu'aux niveaux à haute résolution et renforce les primitives qui apparaissent à plusieurs résolutions spatiales.

La section 3.1 détaille l'analyse multirésolution d'une scène. Le comportement des primitives d'une image à plusieurs échelles spatiales de même que les conditions imposées sur le choix du filtre passe-bas et sur le nombre de niveaux de résolution spatiale y sont spécifiés. La section 3.2 décrit une représentation multirésolution compacte de l'image.

3.1 ANALYSE D'UNE SCÈNE À MULTIPLE RÉOLUTION SPATIALE

Une image contient une multitude de primitives de différentes grosseurs. La première étape du système de vision consiste à extraire ces primitives qui comprennent les arêtes, coins, contours fermés et ouverts présents dans l'image. La description de ces primitives forme le *croquis*, le premier niveau de représentation interne du système de vision. À pleine résolution, l'image comporte beaucoup trop de détails. Beaucoup des primitives détectées à haute résolution correspondent à du bruit de numérisation et de saisie de données et ne correspondent donc pas à des primitives physiques de la scène. La figure 3.1 montre une scène de route et les primitives détectées à pleine résolution. L'opérateur de détection d'arêtes est un Laplacien de Gaussienne qui filtre d'abord l'image avec un filtre passe-bas gaussien de $\sigma=1,75$ pixels avant de détecter les passages par zéro de la dérivée seconde de l'image d'éclairement. Le filtrage atténue les variations spatiales rapides, ce qui permet de diminuer le bruit présent sur l'image d'éclairement. Malgré la filtration inhérente, le croquis généré par le détecteur d'arêtes résulte en une représentation beaucoup trop détaillée et dense où il n'est pas possible de différencier les arêtes correspondant à des primitives physiques de la scène observée (tels les bordures de la route, le contour du tronc de l'arbre, etc.) des arêtes bruitées. La figure 3.2 montre la même scène de route, cette fois-ci observée à 4 niveaux de résolution différents. Une image à un niveau de résolution donné est obtenue en appliquant un filtre passe-bas gaussien d'une certaine valeur d'écart type σ qui spécifie le niveau de résolution

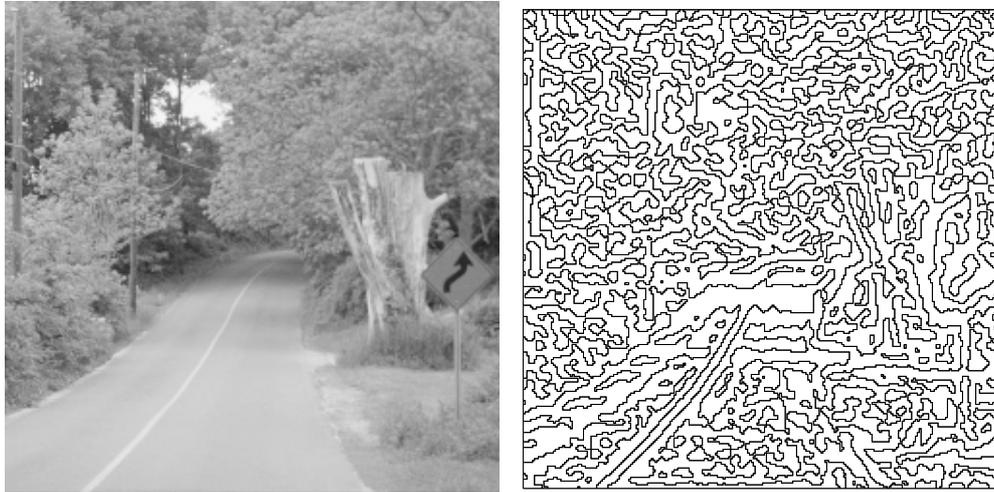


Figure 3.1 Scène de route.

L'image d'éclairage est à pleine résolution. L'image de droite représente le croquis obtenu par convolution de l'image d'éclairage avec un Laplacien de Gaussienne de $\sigma = 1,75$ pixels.

d'observation. Cette convolution de l'image par un filtre passe-bas correspond à une défocalisation de la caméra qui observe la scène. L'image à résolution grossière ne contient pas les nombreux petits détails (au sens spatial) que nous montre l'image à pleine résolution. Le niveau de résolution grossière ne contient que les primitives de grande échelle spatiale contenues dans la scène observée comme par exemple le contour des arbres, les bordures de la route, la forme sinueuse de la flèche du panneau de signalisation, etc. Nous n'y retrouvons pas les primitives de petite dimension comme par exemple les feuilles des arbres, la forme du panneau de signalisation, etc. Les primitives détectées sur l'image à faible résolution sont moins nombreuses, sont moins bien localisées à cause de l'effet de diffusion du filtre passe-bas, mais par contre, sont moins enclines à provenir de bruit de numérisation et donc correspondent pratiquement en totalité à des primitives physiques de la scène observée. Une description complète de la scène sera donc obtenue en procédant à une analyse multirésolution: les primitives détectées dans l'image et qui correspondent à des primitives physiques de la scène sont trouvées aux niveaux de résolution plus grossiers et leur position spatiale est précisée en projetant les primitives détectées au niveau de représentation à pleine résolution. Nous ne conserverons donc, au niveau de pleine résolution spatiale, que les primitives qui sont présentes au moins au niveau de représentation juste au dessus (niveau $\sigma = \sqrt{3}/2$ sur la figure 3.2). L'intensité, ou amplitude, de la primitive détectée, qui est une mesure de notre



Figure 3.2 Scène de route à multiple résolution spatiale.
Les représentations aux différentes résolutions sont obtenues en appliquant un filtre passe-bas gaussien de largeur de bande σ .

confiance que la primitive correspond à un phénomène physique de la scène, est renforcée par le nombre de niveaux supérieurs qui contiennent cette même primitive.

Une représentation multirésolution d'une image de scène observée consiste à filtrer l'image à pleine résolution à l'aide de filtres passe-bas de largeur de

bande de plus en plus étroite de façon à éliminer dans l'image résultante les primitives bruitées de petite dimension. Nous nous intéressons maintenant au comportement des primitives de l'image aux différentes résolutions, au nombre de niveaux de résolution différents qui sont nécessaires pour diminuer la quantité de primitives bruitées détectées et le type de filtre passe-bas optimal pour la génération des différents niveaux de résolution.

3.1.1 Primitives de l'image à multiple résolutions

Witkin fut un des premiers chercheurs à établir les fondements de l'analyse multirésolution (Witkin, 1983; Witkin, 1984). Ses travaux reposaient sur des signaux temporels (donc, à une dimension) et il cherchait à prédire le comportement des passages par zéro de la dérivée seconde du signal filtré par un filtre passe-bas. Le passage par zéro de la dérivée seconde est une mesure qui sert à caractériser un signal. Il indique les maxima et minima locaux du signal de même que leur dérivées. L'échelle d'espace du signal est une mesure de la largeur de bande du filtre passe-bas utilisé pour atténuer les variations rapides du signal. Witkin a trouvé qu'on pouvait caractériser complètement un signal en étudiant la signature des passages par zéro de la dérivée seconde en fonction de l'échelle d'espace. Le comportement des passages par zéro d'un signal en fonction de la largeur de bande du filtre passe-bas est illustré à la figure . L'examen de cette figure nous montre qu'en choisissant un filtre passe-bas approprié (par exemple, dans Witkin, un filtre passe-bas gaussien), aucune primitive du signal n'est générée à mesure que le facteur d'échelle augmente. C'est donc dire que les primitives du signal sont de moins en moins nombreuses à mesure que l'échelle augmente, c'est-à-dire à mesure que le signal devient de plus en plus filtré, brouillé. Le principe d'analyse suggéré par cette figure est que les primitives devraient être identifiées à une échelle spatiale de faible résolution, puis chacune des primitives devrait être poursuivie à mesure que l'échelle spatiale augmente de précision, permettant ainsi d'être parfaitement localisée lorsque nous sommes rendus à pleine résolution. Les travaux de Witkin ont permis de jeter les bases de l'analyse multirésolution :

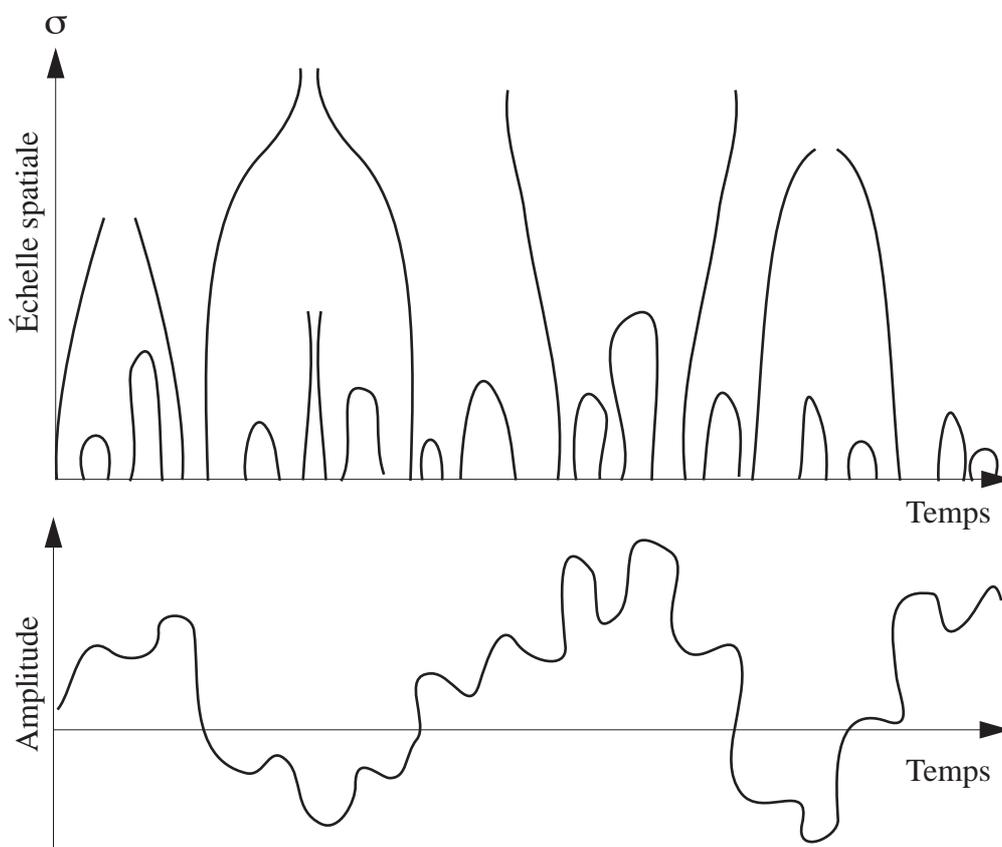


Figure 3.3 Analyse à échelle spatiale continue.

Le signal du bas représente un signal temporel dont on veut étudier le comportement des passages par zéro de la dérivée seconde en fonction de l'échelle d'analyse, soit la fréquence de coupure σ du filtre passe-bas. Le graphique du haut illustre le comportement des passages par zéro de la dérivée seconde (maximums/minimums locaux de la courbe du bas) en fonction de l'échelle spatiale. Le signal est de plus en plus filtré à mesure qu'on progresse le long de l'échelle spatiale. Le graphique montre qu'il y a de moins en moins de passages par zéro à mesure que la résolution spatiale est diminuée. Il n'y a pas de création de nouveaux passages par zéro lorsqu'il y a diminution de la résolution.

- **Causalité** : il n'y a pas d'arêtes générées lorsqu'on passe d'une résolution donnée à une résolution plus grossière.
- **Génération des niveaux** : un filtre passe-bas gaussien assure le principe de causalité.
- **Diffusion anisotrope** : les différents niveaux à multiple résolutions obéissent à la loi de diffusion anisotrope lorsqu'un filtre gaussien est utilisé. La loi de diffusion anisotrope dicte le comportement des arêtes d'une image à mesure que la résolution est diminuée et est similaire à la loi de transfert de la chaleur en thermodynamique. D'autres auteurs ont souligné l'importance du principe de diffusion anisotrope dans les représentations multirésolution (Koenderink, 1984; Perona, & Malik, 1990).
- **Analyse grossier -> fin** : les passages par zéro de la dérivée seconde sont moins nombreux, donc mieux identifiables, à échelle grossière. Par contre, ils sont moins bien localisés. L'analyse multirésolution consiste donc à identifier des primitives à échelle grossière puis à les poursuivre dans l'espace d'échelles dans le but de les localiser précisément.

Les résultats des travaux de Witkin ont été étendus aux images bidimensionnelles (Yuille, & Poggio, 1986). La poursuite des primitives le long de l'échelle d'espace a d'abord été exploitée avec succès par (Bergholm, 1987) qui propose une poursuite continue en partant d'une représentation grossière vers les représentations plus fines. Les arêtes qui correspondent à des primitives physiques de la scène sont localisées avec précision. Le changement d'échelle est continu et la variation du facteur d'échelle est tel que la position de l'arête poursuivie ne change pas plus d'un pixel. (Liu, & Huang, 1991) prolongent les travaux de Bergholm et proposent un algorithme pour apparier les lignes ainsi détectées sur des images stéréo. Plusieurs chercheurs ont par la suite continué les recherches sur la poursuite continue d'arêtes (Chen, 1992; Lu, & Jain, 1989; Lu, & Jain, 1992; Narbel, & Pun, 1989; Williams, & Shah, 1990; Ziou, & Tabbone, 1991; Zuerndorfer, & Wakefield, 1990). (Saint-Marc, Chen, & Medioni, 1991), pour leur part, utilisent un filtre passe-bas qui permet d'adoucir l'image tout en conservant les discontinuités (les discontinuités ne sont pas brouillées lors de la génération des images à moindre résolution). L'avantage de cette méthode con-

siste à obtenir, après quelques itérations, les arêtes de l'image qui sont correctement localisées.

La figure 3.4 illustre les principaux concepts liés à l'analyse à multiple réso-

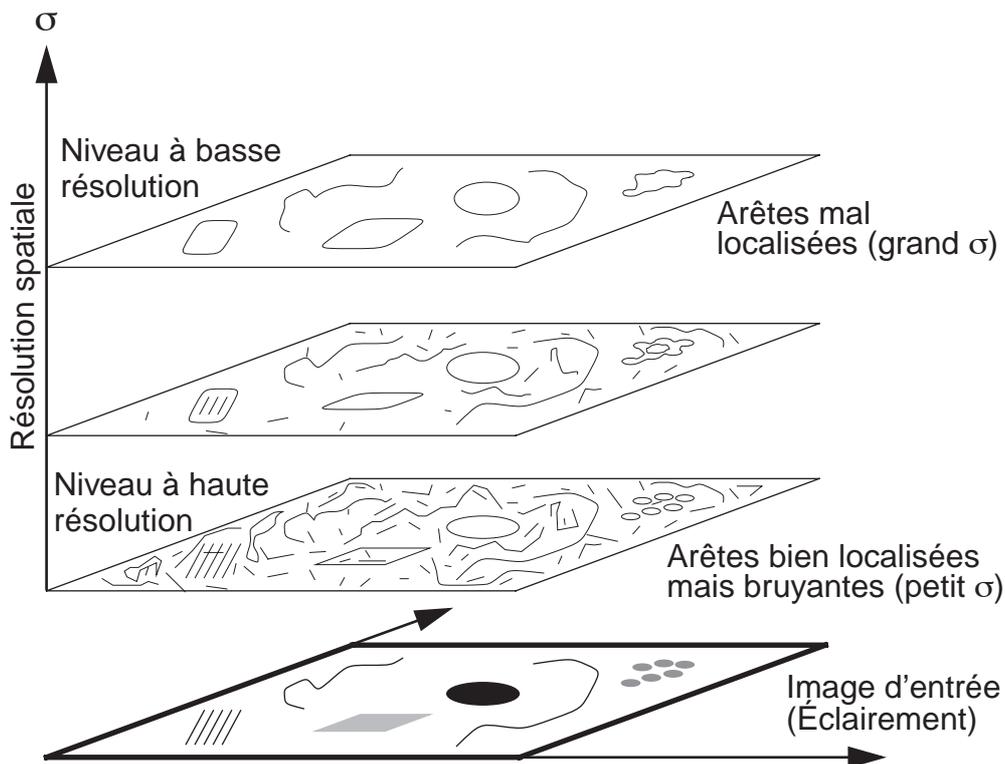


Figure 3.4 Représentation multirésolution des primitives.

Les primitives (arêtes) d'une image d'éclaircement peuvent être visualisées à plusieurs résolutions spatiales. La représentation à un niveau donné de résolution spatiale s'obtient en filtrant l'image d'éclaircement avec un filtre passe-bas puis en détectant les primitives de l'image filtrée.

©Marc Tremblay, 1992. Reproduit avec permission.

lutions. L'image de départ est l'image d'éclaircement à pleine résolution. L'image d'éclaircement est filtrée et les primitives (arêtes) des images résultantes sont détectées. Les arêtes sont moins nombreuses à basse résolution, mais par contre sont mal localisées à cause de l'effet de diffusion du filtre. L'analyse multirésolution consiste donc à générer une représentation de l'image de départ à plusieurs résolutions spatiales, puis à détecter les primitives de ces images, puis enfin à poursuivre à travers l'échelle d'espace les primitives détectées à basse

résolution afin de les localiser avec précision. Nous nous intéressons maintenant aux nombres de niveaux nécessaires à une bonne représentation multirésolution et aux catégories de filtre qui peuvent être utilisés pour la génération des différents niveaux.

3.1.2 Discrétisation de l'échelle d'espace

Un pixel d'une représentation multirésolution est identifié par 3 paramètres : par les coordonnées x et y dans l'image de même que par le facteur d'échelle d'espace σ qui indique la largeur de bande du filtre passe-bas utilisé. Le facteur d'échelle spatiale peut être continu ou discret. La plupart des travaux de recherche en analyse multirésolution s'inspirent, comme nous l'avons vu, des travaux de Bergholm, qui utilisait une échelle spatiale continue. (Jeong, & Kim, 1992) proposent de trouver d'abord les échelles spatiales optimales, puis de procéder à la détection des arêtes. L'étude du système visuel des mammifères suggère qu'il est suffisant de n'utiliser que trois ou quatre niveaux, les uns séparés des autres d'environ une octave (Barrilleaux, 1990; Marr, 1982). Cette discrétisation de l'échelle spatiale permet une réalisation efficace. De plus, la séparation d'une octave entre les niveaux permet de construire la représentation multirésolution sous forme de pyramide, comme nous le verrons à la section suivante (section 3.2). Une pyramide est une représentation multirésolution très compacte, qui ne nécessite que 33% de neurones additionnels par rapport à l'image à pleine résolution.

3.1.3 Filtre passe-bas pour la génération des niveaux de résolution

La génération des images à multiples résolutions spatiales requiert qu'un filtre passe-bas soit utilisé. Toute la théorie de l'analyse multirésolution repose sur la prémisse qu'aucune primitive n'est générée lorsqu'on se dirige vers les plus basses résolutions le long de l'échelle d'espace. Il faut donc que le filtre passe-bas ne génère pas de nouvelles primitives lorsque la résolution est baissée. Witkin avait trouvé que le filtre gaussien répond à cette exigence sévère pour les signaux à une dimension. (Babaud, Witkin, Baudin, & Duda, 1986) ont démontré que le filtre gaussien est le *seul* filtre parmi une très large classe à assurer le principe de causalité. (Yuille, & Poggio, 1988) ont étendu cette unicité du filtre gaussien aux images. De plus, le filtre gaussien est optimal dans la localisation des arêtes de l'image filtrée car il minimise le principe d'incertitude $x \cdot \omega = cte$.¹ Le filtre de Gabor (Daugman, 1985a; Daugman, 1985b; Daug-

man, & Kammen, 1987; Koenderink, & van Doorn, 1992; Rauber, 1989-1990) est une combinaison d'un filtre gaussien et d'une fonction d'éclairage $f(x,y)$ qui est une sinusoïde, orientée par rapport à l'axe horizontal, à une fréquence ω . C'est un filtre passe-bande qui exploite au maximum le principe d'incertitude et est aussi beaucoup utilisé en analyse multirésolution, surtout pour les travaux de recherche s'inspirant de la vision biologique chez les mammifères. Dans un autre ordre d'idées, (Wu, & Xie, 1990) réussissent à inclure les filtres exponentiels et les filtres moyenneurs en plus du filtre gaussien en représentant les pixels de l'image d'entrée comme des fonctions polynomiales au lieu des traditionnelles fonctions de Dirac pondérées. Récemment, des chercheurs ont proposé d'utiliser des filtres morphologiques pour construire la représentation multirésolution (Chen, & Yan, 1989; Toet, 1989b).

Un filtre gaussien discret (Ben-Arie, & Rao, 1991; Lindeberg, 1990; Meer, 1988a; Wilson, & Bhalerao, 1992) permet d'améliorer considérablement le temps de calcul. La meilleure approximation d'une réponse impulsionnelle gaussienne, lorsqu'on ne peut utiliser qu'un nombre limité de termes, nous est fournie par le théorème de De Moivre - Laplace, appelé aussi théorème de la centrale limite en probabilité :

$$\binom{n}{k} p^k q^{n-k} \cong \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} \quad (3.1)$$

Le premier membre de l'équation est l'expression de la densité binomiale et exprime la probabilité que k événements de probabilité p se produisent dans n essais. La première expression de l'équation représente les coefficients binomiaux et indique le nombre de combinaisons de k éléments pris parmi n éléments. Les coefficients binomiaux sont calculés par l'expression :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$$

Le tableau 3.1 indique la valeur des coefficients binomiaux pour les premières valeurs de n et k . La réponse impulsionnelle du filtre gaussien que l'on veut

1. Le principe d'incertitude stipule que la fréquence spatiale d'un signal (ou d'une image) ω ne peut être calculée qu'avec une certaine précision qui dépend de la largeur de la fenêtre d'analyse, x et du type de fenêtre appliquée. Le filtre gaussien correspond à une fenêtre d'analyse qui est aussi gaussienne.

$\downarrow n/k \rightarrow$	0	1	2	3	4	5
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

Tableau 3.1 Premiers coefficients binomiaux.

approcher s'écrit :

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (3.1)$$

En combinant les équations 3.1 et 3.1 et en posant $x=k$, nous obtenons la largeur de bande σ en fonction de la grandeur n du vecteur de coefficients binomiaux :

$$\sigma = \frac{\sqrt{n}}{2} \quad (3.2)$$

Lorsque la grandeur du vecteur de coefficients est choisie (en fonction de la largeur de bande désirée), nous obtenons un vecteur horizontal pour approcher la gaussienne selon x , $G(x)$, et un vecteur vertical pour approcher la gaussienne selon y , $G(y)$. Le filtre gaussien bidimensionnel recherché, $G(x, y)$, est séparable et s'obtient donc en calculant le produit vectoriel externe de ces deux vecteurs.

Ainsi, pour une valeur de $\sigma = \frac{\sqrt{3}}{2}$, le noyau générateur du filtre

bidimensionnel gaussien est :

$$G(x, y) = \begin{bmatrix} 1 \\ 3 \\ 3 \\ 1 \end{bmatrix} \times [1 \ 3 \ 3 \ 1] = \begin{bmatrix} 1 & 3 & 3 & 1 \\ 3 & 9 & 9 & 3 \\ 3 & 9 & 9 & 3 \\ 1 & 3 & 3 & 1 \end{bmatrix} \quad (3.3)$$

3.2 REPRÉSENTATION PYRAMIDALE

Une représentation multirésolution est donc une série d'images représentant une même scène observée à différentes résolutions spatiales. Une image à une résolution donnée s'obtient en appliquant un filtre passe-bas sur l'image pleine-résolution de base. L'opération de filtrage permet, grâce au théorème d'échantillonnage de Nyquist, de sous-échantillonner l'image résultante sans perte d'information. Le sous-échantillonnage de l'image filtrée a donné naissance à une catégorie de structures de données très compacte, les pyramides (Ballard, & Brown, 1982; Burt, 1984; Burt, & Adelson, 1983; Dyer, 1987; Hummel, 1987; Klinger, 1984; Peleg, Federbusch, & Hummel, 1987; Rosenfeld, 1984b; Samet, 1984). La figure 3.5 est une illustration d'une structure pyramidale.

Une structure pyramidale est facilement construite. La base de la pyramide contient l'image d'éclairage d'entrée, qui provient par exemple d'une caméra. Les niveaux successifs de la pyramide contiennent les images à des résolutions de plus en plus basses. L'image à chaque niveau est obtenue en calculant, pour chaque position de pixel, la somme pondérée des pixels situés juste en dessous. Les coefficients correspondent aux paramètres du filtre passe-bas utilisé. Pour plusieurs types de pyramide, chaque valeur de pixel est calculée en prenant la moyenne des quatre pixels juste en dessous : le filtre passe-bas est alors un filtre moyenneur et les coefficients du filtre sont simplement égaux à 0,25. Nous utilisons, pour notre part, un filtre gaussien, tel que discuté à la section 3.1.3. Le filtre moyenneur ou le filtre gaussien permettent de sous-échantillonner l'image résultante par un facteur 2 dans chacune des directions¹. La

1. Pour pouvoir sous-échantillonner l'image résultante par un facteur 2, l'atténuation du filtre passe-bas à la fréquence normalisée 0,5 doit être suffisante pour éviter les effets de recouvrement. On peut démontrer que le filtre moyenneur et le filtre gaussien respectent cette condition.

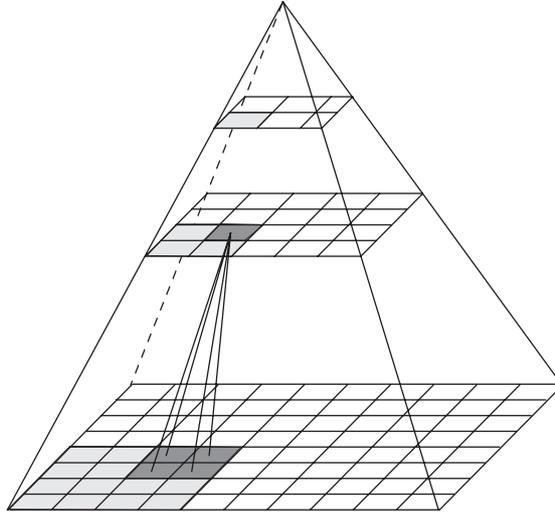


Figure 3.5 Représentation pyramidale.

Chaque pixel d'une couche de la pyramide est généré à partir des pixels en dessous au moyen d'un filtre passe-bas. L'image ainsi filtrée est sous-échantillonnée par un facteur 2. La base de la pyramide représente l'image à pleine résolution.

grandeur des images diminue donc d'un facteur 4 à mesure que l'on monte le long de la pyramide.

En résumé, une structure pyramidale offre les avantages suivants :

1. **Représentation compacte.** La structure ne nécessite que 33% plus d'espace en mémoire que l'image à pleine résolution. Cette grande compacité provient du sous-échantillonnage à chaque niveau de la pyramide qui divise par 4 les dimensions de l'image résultante par rapport aux dimensions de l'image au niveau précédent. La dimension totale relative de la structure est donc (pour une dimension totale unitaire de l'image pleine-résolution) :

$$1 + \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \dots \Rightarrow 1\frac{1}{3}$$

2. **Filtrage simple et rapide.** Chaque niveau de la pyramide est généré à partir du niveau précédent en appliquant toujours le même filtre passe-bas. Le même filtre peut être utilisé car il est appliqué sur une image qui est déjà sous-échantillonnée, ce qui déplace la fréquence de coupure effective du filtre. Le même filtre, de petite taille, est donc appliqué de couche en couche, résultant en une génération très rapide des images aux diverses résolutions spatiales puisque la somme pondérée utilisée pour calculer la valeur de chaque pixel est limitée en nombre de termes. Cette propriété a d'ailleurs été utilisée avec succès pour la construction de filtres passe-bas avec de très basses fréquences de coupure (Burt, 1981; Ranganath, 1991).
3. **Information globale disponible.** L'information contenue dans chaque pixel des niveaux supérieurs de la pyramide est une combinaison linéaire des valeurs du groupe de pixels localisés en dessous, aux niveaux plus bas de la pyramide. Donc, à mesure que l'on monte le long de la pyramide, les pixels ont un champ récepteur de plus en plus grand sur la couche de base. Cette information, de plus en plus globale, pourrait être utilisée, par exemple, pour fixer des niveaux de seuil qui tiendraient compte de propriétés communes à toute l'image.

La représentation multirésolution de la scène de route (figure 3.1) peut donc être avantageusement générée sous forme de pyramide. Cette représentation pyramidale est illustrée à la figure 3.6. L'effet multirésolution peut être mieux perçu si l'on observe les différents niveaux de la pyramide à des grandeurs identiques d'image, comme il est illustré à la figure 3.7 qui reprend la même représentation multirésolution, mais à des grandeurs fixes de l'image. Une comparaison de cette figure avec la figure 3.2 nous montre que la représentation pyramidale approxime bien la représentation multirésolution formelle, mais à un coût et une complexité de calcul beaucoup moindres. De plus, la représentation pyramidale se prête admirablement bien à une réalisation neuronique, comme nous le verrons à la section suivante.



Figure 3.6 Représentation pyramidale de la scène de route. Les représentations aux différentes résolutions sont obtenues en appliquant le même filtre passe-bas gaussien de largeur de bande $\sigma = \sqrt{3}/2$ entre chaque niveau consécutif de la pyramide. La valeur de σ indiquée à chaque niveau de la pyramide est la fréquence de coupure effective.

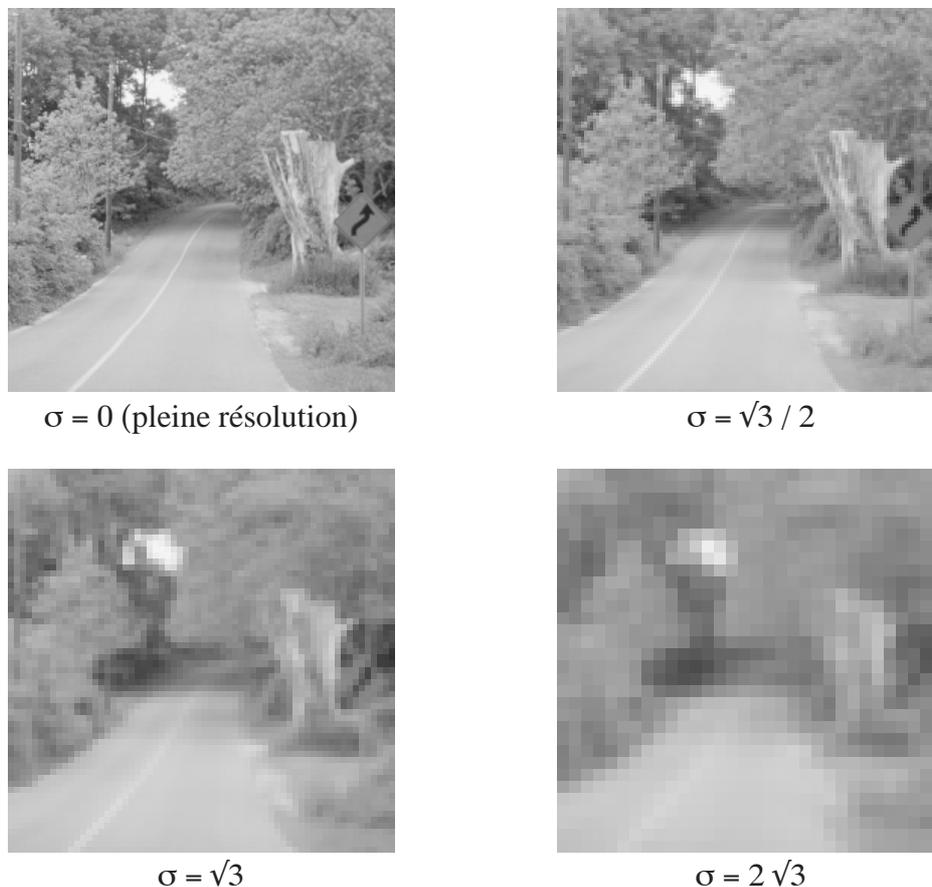


Figure 3.7 Niveaux de la pyramide avec grandeurs fixes d'images. Les représentations aux différentes résolutions sont obtenues en appliquant un filtre passe-bas gaussien de largeur de bande σ .

Quelques variations de la structure de la pyramide ont été proposées récemment. Lors de l'opération de sous-échantillonnage, les pixels éliminés le sont généralement de façon systématique, sans se préoccuper de leur contenu. (Jolion, & Montanvert, 1992) proposent de conserver les pixels qui renferment le plus d'information. Dans le cas d'images d'arêtes, par exemple, un pixel indiquant la présence d'une arête sera conservé au détriment d'un autre pixel qui n'en indiquerait pas. (Meer, 1988b; Meer, 1989) s'intéresse quant à lui à des structures pyramidales sans sous-échantillonnage : les images aux différents niveaux de la pyramide sont de mêmes dimensions que l'image de base à pleine résolution. L'efficacité de calcul offerte par les structures pyramidales est rete-

nue puisque les images aux différents niveaux de résolution sont générées par un même filtre de petite taille appliqué entre chaque niveau. Un algorithme efficace d'interpolation permet par la suite de générer les valeurs de pixel qui manquent. Plusieurs chercheurs utilisent des pyramides d'arêtes (Clark, & Lawrence, 1984; Shneier, 1984; Unser, 1992). Dans ce type de pyramide, l'image d'arêtes est directement générée à chaque niveau de résolution de la pyramide à l'aide d'un filtre passe-bande au lieu d'un filtre passe-bas. Enfin, quelques chercheurs utilisent le rapport ou encore la différence entre deux niveaux consécutifs de la pyramide pour générer les images d'arêtes multirésolution (Crowley, 1984; Toet, 1989a).

Nous avons montré, dans cette section, une structure de données très efficace pour représenter une image à plusieurs résolutions spatiales. La structure pyramidale se prête admirablement bien, en plus, à une implémentation sur réseaux neuroniques artificiels.

CONCLUSION

Nous avons vu, dans ce document, les éléments de base de l'analyse multirésolution. Nous avons d'abord présenté les grandes lignes de l'analyse multirésolution de même que les principaux avantages qu'elle procure. Nous avons par la suite examiné les différentes méthodes pour générer les représentations visuelles à multiples résolutions spatiales. Nous nous sommes attardé au choix du filtre de génération. Nous avons par la suite étudié une représentation multirésolution particulièrement compacte et bien adaptée à une réalisation par réseaux neuroniques : la structure pyramidale.

La génération de la représentation multirésolution est le tout premier algorithme appliqué par le système de vision aux images d'éclairément d'entrée. Cette représentation à plusieurs résolutions spatiales est maintenue pour toutes les représentations internes du système, excepté pour les représentations finales, l'esquisse 2.5D et la description 3D, qui sont générées à haute résolution seulement.

Disposant d'une représentation des images d'éclairément à plusieurs résolutions spatiales, il s'agit maintenant de générer la première représentation interne du système de vision, le croquis. Il s'agit donc de détecter les primitives des images d'entrée à chacun des niveaux de résolution puis de se servir de l'information présente à multiple résolutions pour pouvoir générer un croquis

qui reflète le mieux les primitives physiques de la scène tridimensionnelle observée.

RÉFÉRENCES

- Babaud, J., Witkin, A. P., Baudin, M., & Duda, R. O. (1986). Uniqueness of the Gaussian Kernel for Scale-Space Filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-8**(1), 26-33.
- Ballard, D. H., & Brown, C. M. (1982). *Computer Vision*. Englewood Cliffs, NJ: Prentice Hall.
- Barrilleaux, J. M. (1990). A biologically motivated algorithm for image interpretation based on multi-pass multi-resolution techniques. *Proceedings of the IJCNN - IEEE INNS International Joint Conference on Neural Networks*, San Diego, **2**, II-813 - II-818.
- Ben-Arie, J., & Rao, K. R. (1991). Gaussian groups and lattice architectures for image representation and transforms (No. IIT ECE TR91001). Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago.
- Bergholm, F. (1987). Edge Focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-9**(6), 726-741.
- Burt, P. J. (1981). Fast Filter Transforms for Image Processing. *Computer Graphics & Image Processing*, **16**, 20-51.
- Burt, P. J. (1984). The Pyramid as a Structure for Efficient Computation. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 6-35). New York: Springer-Verlag.
- Burt, P. J., & Adelson, E. H. (1983). The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, **COM-31**(4), 532-540.
- Chen, J.-S. (1992). Generalized Adaptive Smoothing for multiscale edge detection. *Proceedings of the SPIE Applications of Artificial Intelligence X: Machine Vision and Robotics*, Orlando, Florida, 22-24 april, **1708**, 341-352.
- Chen, M.-H., & Yan, P.-F. (1989). A Multiscaling Approach Based on Morphological Filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(7), 694-700.
-

-
- Clark, J. J., & Lawrence, P. D. (1984). A Hierarchical Image Analysis System Based Upon Oriented Zero Crossings of Bandpassed Images. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 148-168). New York: Springer-Verlag.
- Crowley, J. L. (1984). A Multiresolution Representation for Shape. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 169-189). New York: Springer-Verlag.
- Daugman, J. G. (1985a). Image analysis by local 2-D spectral signatures. *Journal of the Optical Society of America A*, **2**(13), P74.
- Daugman, J. G. (1985b). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, **2**(7), 1160-1169.
- Daugman, J. G., & Kammen, D. M. (1987). Image Statistics, Gases, and Visual Neural Primitives. *Proceedings of the IEEE First International Conference on Neural Networks*, San Diego, California, June 21-24, **4**, IV-163 - IV-175.
- Dyer, C. R. (1987). Multiscale Image Understanding. In L. Uhr (Ed.), *Parallel Computer Vision* (pp. 171-213). Boston: Academic Press.
- Hlawatsch, F., & Bouderaux-Bartels, G. F. (1992). Linear and Quadratic Time-Frequency Signal Representations. *IEEE Signal Processing Magazine*, **9**(2), 21-67.
- Hummel, R. (1987). The Scale-space Formulation of Pyramid Data Structures. In L. Uhr (Ed.), *Parallel Computer Vision* (pp. 107-123). Boston: Academic Press.
- Jeong, H., & Kim, C. I. (1992). Adaptive Determination of Filter Scales for Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(5), 579-585.
- Jolion, J. M., & Montanvert, A. (1992). The Adaptive Pyramid: A Framework for 2D Image Analysis. *Computer Vision, Graphics and Image Processing: Image Understanding*, **55**(3), 339-348.
-

-
- Klinger, A. (1984). Multiresolution Processing. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 77-85). New York: Springer-Verlag.
- Koenderink, J. J. (1984). The Structure of Images. *Biological Cybernetics*, **50**, 363-370.
- Koenderink, J. J., & van Doorn, A. J. (1992). Generic Neighborhood Operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(6), 597-605.
- Lindeberg, T. (1990). Scale-Space for Discrete Signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(3), 234-254.
- Liu, Y., & Huang, T. S. (1991). Determining straight line correspondences from intensity images. *Pattern Recognition*, **24**(6), 489-504.
- Lu, Y., & Jain, R. C. (1989). Behavior of Edges in Scale Space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(4), 337-356.
- Lu, Y., & Jain, R. C. (1992). Reasoning about Edges in Scale Space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(4), 450-468.
- Mallat, S., & Zhong, S. (1992). Characterization of Signals from Multiscale Edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(7), 710-732.
- Marr, D. (1982). *Vision - A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- Meer, P. (1988a). Efficient computation of two-dimensional Gaussian windows. *Pattern Recognition Letters*, **7**(April), 227-229.
- Meer, P. (1988b). Simulation of constant size multiresolution representations on image pyramids. *Pattern Recognition Letters*, **8**(November), 229-236.
- Meer, P. (1989). Stochastic Image Pyramids. *Computer Vision, Graphics, and Image Processing*, **45**, 269-294.
-

-
- Meyer, Y., Jaffard, S., & Rioul, O. (1987). L'analyse par ondelettes. *Pour la science*, **119**(Septembre), 28-37.
- Narbel, P., & Pun, T. (1989). *Journey through scale-space* (No. 89.01). Computer Science Center, University of Geneva.
- Peleg, S., Federbusch, O., & Hummel, R. (1987). Custom-Made Pyramids. In L. Uhr (Ed.), *Parallel Computer Vision* (pp. 125-146). Boston: Academic Press.
- Perona, P., & Malik, J. (1990). Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(7), 629-639.
- Ranganath, S. (1991). Image Filtering Using Multiresolution Representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**(5), 426-440.
- Rauber, C. (1989-1990). *La transformée de Gabor et ses applications*. Travail de diplôme, Université de Genève.
- Rioul, O., & Vetterli, M. (1991). Wavelets and Signal Processing. *IEEE Signal Processing Magazine*, **8**(4), 14-38.
- Rosenfeld, A. (Ed.). (1984a). *Multiresolution Image Processing and Analysis*. New York, NY: Springer-Verlag.
- Rosenfeld, A. (1984b). Some Useful Properties of Pyramids. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 2-5). New York: Springer-Verlag.
- Saint-Marc, P., Chen, J.-S., & Medioni, G. (1991). Adaptive Smoothing: A General Tool for Early Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**(6), 514-529.
- Samet, H. (1984). A Tutorial on Quadtree Research. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 212-223). New York: Springer-Verlag.
-

-
- Shneier, M. (1984). Multiresolution Feature Encodings. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 190-199). New York: Springer-Verlag.
- Toet, A. (1989a). Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters*, **9**, 245-253.
- Toet, A. (1989b). A morphological pyramidal image decomposition. *Pattern Recognition Letters*, **9**, 255-261.
- Uhr, L. (Ed.). (1987). *Parallel Computer Vision*. Boston: Academic Press.
- Unser, M. (1992). An improved least squares Laplacian pyramid. *Signal Processing*, **27**, 187-203.
- Williams, D. J., & Shah, M. (1990). Edge Contours Using Multiple Scales. *Computer Vision, Graphics, and Image Processing*, **51**, 256-274.
- Wilson, R., & Bhalerao, A. H. (1992). Kernel Designs for Efficient Multiresolution Edge Detection and Orientation Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(3), 384-390.
- Witkin, A. P. (1983). Scale-space filtering. *Proceedings of the the Ninth International Joint Conf. on Artificial Intelligence*, Karlsruhe (West Germany), 1019-1022.
- Witkin, A. P. (1984). Scale space filtering: a new approach to multi-scale description. In S. Ullman & W. Richards (Ed.), *Image Understanding 1984*. Norwood, NJ: Ablex Publishing Corporation.
- Wu, L., & Xie, Z. (1990). Scaling Theorems for Zero-Crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(1), 46-54.
- Yuille, A. L., & Poggio, T. (1988). Scaling and fingerprint theorems for zero-crossings. In C. Brown (Ed.), *Advances in Computer Vision* (pp. 47-78). Hillsdale, N.J.: Lawrence Erlbaum Ass.
- Yuille, A. L., & Poggio, T. A. (1986). Scaling Theorems for Zero Crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-8**(1), 15-25.
-

- Ziou, D., & Tabbone, S. (1991). Fusion des contours dans un espace échelle. *Proceedings of the Treizième Colloque sur le Traitement du Signal et des Images*, Juan-les-Pins, **2**, 1013 - 1016.
- Zuerndorfer, B., & Wakefield, G. H. (1990). Extensions of Scale-Space Filtering to Machine-Sensing Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(9), 868-882.
-

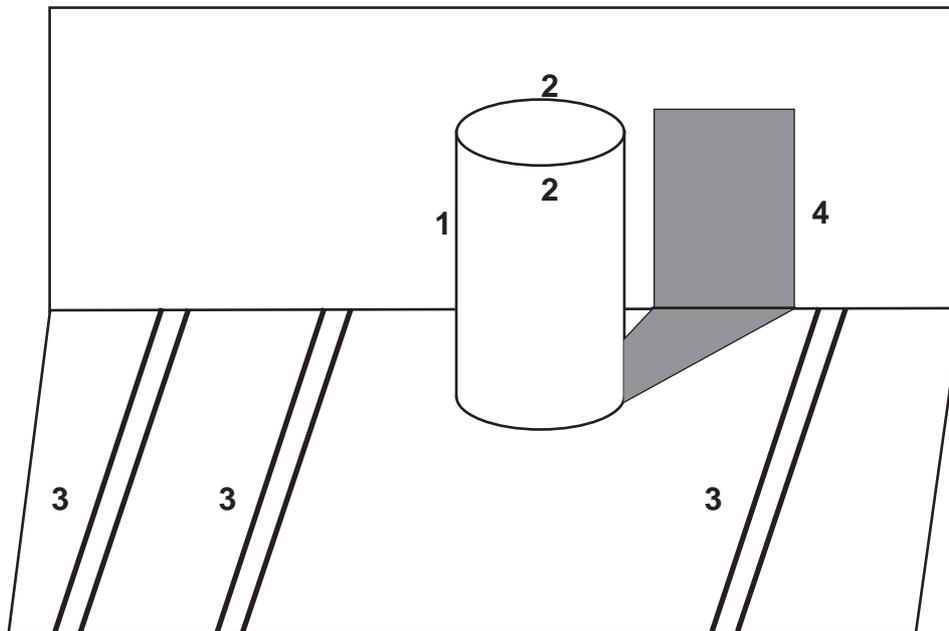
CHAPITRE 4 *Détection des arêtes d'une image*

Une image comporte beaucoup trop d'informations pour que l'on puisse tenter une analyse de cette dernière au niveau de ses valeurs d'éclairement en chaque pixel. Ainsi, une (petite) image de dimensions 256x256 comporte 65,536 pixels contenant chacun une valeur d'éclairement. Par contre, nous pouvons faire porter l'analyse sur les arêtes contenues dans l'image, ce qui limite de beaucoup la complexité de l'analyse. Les arêtes sont définies comme les variations brusques de l'éclairement et correspondent à des attributs physiques de la scène tridimensionnelle observée. La détection des arêtes nous permet de construire le **croquis**, le premier niveau de représentation interne du modèle computationnel du système de vision. Un algorithme de stéréoscopie pourrait être appliqué par la suite et consiste à identifier une même primitive sur les deux images d'éclairement gauche et droite et à mesurer la différence de position des deux primitives, ce qui sera inversement proportionnel à la profondeur de ce point caractéristique.

La section 4.1 établit la relation entre les différents attributs physiques de la scène et les primitives détectées sur les images d'éclairement. La détection des primitives sur l'image est détaillée à la section 4.2 tandis que la section 4.3 précise une méthode de détection multirésolution réalisée à l'aide de réseaux neuroniques. L'algorithme de stéréoscopie nécessite une localisation précise des primitives appariées. Une méthode pour amincir les arêtes détectées est présentée à la section 4.4. Finalement, une analyse multirésolution permet d'atténuer la présence des arêtes bruitées et fait l'objet de la section 4.5.

4.1 ATTRIBUTS DE LA SCÈNE OBSERVÉE

La scène tridimensionnelle sous observation comporte plusieurs attributs que nous désirons détecter afin de construire le **croquis**, le premier niveau de représentation interne du modèle computationnel d'un système de vision. Ces primitives qui apparaissent sur les images d'éclairage d'entrée sont engendrées par des phénomènes physiques qui se produisent dans la scène observée comme le montre la figure 4.1. Ces attributs physiques sont projetés sur les deux



- | | |
|---|---|
| 1. Occultations: masquage d'un objet par un autre | 3. Marques : texture, albédo, etc. |
| 2. Arêtes 3D: discontinuité du vecteur normal \vec{n} | 4. Ombre: discontinuité de l'éclairage incident |

Figure 4.1 Principaux attributs d'une scène tridimensionnelle. Phénomènes physiques pouvant intervenir dans une scène tridimensionnelle. Ces différents phénomènes sont projetés sur l'image d'éclairage sous forme de variations brusques d'éclairage.

plans-images et apparaissent comme des arêtes, ou variations brusques de

l'éclairage. Les attributs physiques de la scène qui sont traduits par des arêtes sur les images d'éclairage sont:

1. Le contour d'un objet qui masque partiellement un autre objet de la scène observée. Ce contour d'occultation empêche de détecter les attributs de l'objet masqué qui est derrière.
2. Les arêtes 3D, qui correspondent à une discontinuité du vecteur normal à la surface externe de l'objet.
3. Les différentes marques qui apparaissent sur des surfaces visibles de la scène. Ces marques peuvent provenir de la texture de la surface ou d'un changement d'albédo causé, par exemple, par un trait de crayon.
4. Le contour de l'ombre projetée sur des objets de la scène et provoquée par une discontinuité de l'éclairage incident.

Les arêtes de l'image sont donc associées à des attributs identifiables de la scène tridimensionnelle sous observation. La détection et la localisation précise de ces attributs est une étape cruciale puisqu'elle permet d'alimenter l'algorithme de stéréo, entre autre. Cet algorithme, rappelons le, consiste à identifier les mêmes primitives sur les deux images d'éclairage et à mesurer leur position relative. Cette différence de position, appelée disparité, restitue la profondeur de cet attribut dans la scène.

Une description plus complète des attributs d'une scène implique une analyse de ces attributs à plus d'une résolution spatiale. L'analyse à plus d'une résolution spatiale permet de mieux identifier les primitives composant la scène et d'éliminer les fausses arêtes qui ne correspondent pas à des attributs physiques de la scène mais sont plutôt causées par des artefacts tels les bruits de numérisation et de quantification. Une stratégie d'extraction des primitives d'une image pourrait donc consister à identifier les arêtes à chacun des niveaux de résolution spatiale de la représentation multirésolution des images d'éclairage, d'amincir les arêtes détectées à chaque niveau afin de mieux les localiser, puis d'utiliser l'information détectée aux autres niveaux de résolution afin de renforcer les arêtes correspondant à des attributs physiques de la scène et éliminer le plus possible les arêtes bruitées.

4.2 DÉTECTION DES VARIATIONS BRUSQUES D'ÉCLAIREMENT

Un attribut physique de la scène tridimensionnelle sous observation est projetée sur le plan image par le processus de formation des images comme une variation brusque d'éclairage, ou arête. La figure 4.2 illustre une primitive de

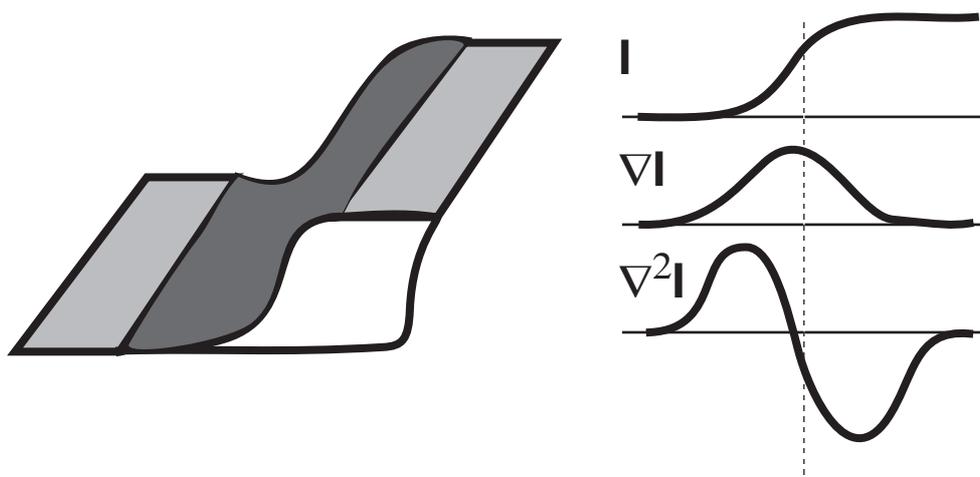


Figure 4.2 *Imagerie d'une variation de normale à la surface. Attribut de scène (variation de la normale à la surface) et sa traduction sur l'image d'éclairage I . La variation d'éclairage peut être détectée soit comme un maximum de la dérivée première de l'image ou soit comme un passage par zéro de la dérivée seconde.*

scène et sa traduction sur le plan image. La variation résultante d'éclairage peut essentiellement être détectée de deux façons :

1. La variation peut être détectée par un maximum de la dérivée première directionnelle de l'image d'éclairage. Cette méthode possède l'avantage d'être relativement peu bruitée (il est reconnu que l'opération de dérivation est bruitée parce qu'elle amplifie les variations brusques) et se prête bien à une réalisation neuronique puisque l'activité d'un neurone représentant une arête sera justement maximale lorsqu'une arête sera détectée à cette position. On n'a donc pas besoin d'une couche supplémentaire dont les neurones

sont activés en présence d'arête. Par contre, une arête détectée comme un maximum de la dérivée première aura tendance à être large, surtout si les variations d'éclairement sont relativement douces. Il faudra donc prévoir un mécanisme d'amincissement afin de pouvoir localiser précisément les arêtes.

2. La variation peut être détectée par un passage par zéro de la dérivée seconde de l'image d'éclairement. Cette méthode possède l'avantage de permettre une localisation précise de l'arête puisque c'est exactement le passage par zéro qui détermine la localisation de l'arête. Par interpolation, il est même possible de localiser l'arête entre les pixels. De plus, il existe un opérateur simple, le Laplacien, qui permet de rendre la détection isotrope, c'est-à-dire qu'une connaissance *a priori* de l'orientation de l'arête n'est pas requise. Par contre, l'opération de dérivée seconde génère beaucoup plus de bruit que l'opération de dérivée première. Il s'ensuit qu'un plus grand nombre d'arêtes bruitées sont détectées; ces arêtes ne correspondent pas à des attributs physiques de la scène observée et il faudra donc trouver un moyen de les éliminer, ou tout au moins d'en atténuer la quantité.

Le choix d'une méthode axée sur la détection d'un maximum de la dérivée première ou sur un passage par zéro de la dérivée seconde dépend principalement de la précision recherchée, du type d'information recherchée et du support choisi pour la réalisation matérielle. La recherche scientifique dans le domaine de la détection d'arêtes a été et est encore passablement active tant du côté de la détection par la dérivée première que dans celle par la dérivée seconde. La plupart de ces méthodes décrites dans la littérature filtrent d'abord l'image à analyser afin d'atténuer le bruit inhérent au processus de différentiation. Cette filtration correspond à analyser l'image à un niveau de résolution spatiale donné.

La littérature sur la détection d'arêtes est vaste. Nous nous limitons aux méthodes différentielles et mettons volontairement de côté d'autres paradigmes

de détection d'arêtes telles les méthodes morphologiques (Chen, & Yan, 1989; Jang, & Chin, 1990; Moran, 1990), les méthodes utilisant la transformée de Hough (Ballard, 1981; Princen, Illingworth, & Kittler, 1990), les méthodes basées sur la minimisation d'un facteur *coût* (Tan, Gelfand, & Delp, 1992) ou les méthodes utilisant une approximation par splines (Zucker, David, Dobbins, & Iverson, 1988).

Plusieurs bons articles ont été rédigés pour synthétiser les travaux sur la détection des arêtes. Le lecteur intéressé peut consulter (Basseville, 1979), l'article récent de (Boyer, & Sarkar, 1992) et l'article synthèse de (Torre, & Poggio, 1986). La plupart des livres qui traitent de vision informatique abordent la détection d'arêtes comme étant une étape cruciale du processus de vision. Les livres suivants nous ont inspiré dans nos travaux sur la détection d'arêtes : les deux manuels de référence classiques en intelligence artificielle (Barr, Cohen, & Feigenbaum, 1989; Cohen, & Feigenbaum, 1982), le livre en français de (Besançon, 1988), celui, classique, de (Gonzalez, & Wintz, 1977), un autre classique sur la vision informatique et biologique, le texte de (Levine, 1985) et enfin, un texte récent de (Jain, 1989).

Comme nous l'avons mentionné plus haut, la détection d'arêtes peut être réalisée soit en détectant un maximum de la dérivée première ou un passage par zéro de la dérivée seconde. De plus, la détection peut être directionnelle ou isotrope. Une détection directionnelle indique la présence d'une arête dans une direction spécifiée tandis qu'une détection isotrope indique la présence d'une arête indépendamment de sa direction. Les sous-sections suivantes précisent l'état de l'art dans chacun de ces domaines de détection d'arêtes.

4.2.1 Maximum de la dérivée première directionnelle

La détection d'arêtes par une mesure du maximum de la dérivée première de l'image calculée dans une direction spécifiée est une méthode de détection qui a fait l'objet de nombreuses recherches. Cette méthode offre l'avantage d'être peu bruyante à cause de l'utilisation de la dérivée première et offre la possibilité d'obtenir des arêtes minces en choisissant soigneusement la direction selon laquelle la dérivée est calculée. Il existe deux écoles dominantes : ceux qui calculent la dérivée en chaque point dans une direction précise et ceux qui calculent la dérivée en chaque point selon plusieurs directions déterminées à l'avance et qui choisissent la direction dominante comme étant la direction de l'arête.

Les méthodes basées sur le calcul de la dérivée en chaque point selon une direction spécifiée sont issues principalement des travaux de (Canny, 1986). Ce dernier a trouvé que l'opérateur optimal pour trouver une arête bruitée est la dérivée première directionnelle (selon la direction du gradient) d'une fonction gaussienne G_σ de résolution σ . Une arête est alors définie comme le maximum local (dans la direction du gradient) de l'opérateur G_n convolué avec l'image I . L'opérateur G_n est défini :

$$G_n = \frac{\partial}{\partial n} G_\sigma \quad (4.1)$$

où n est la direction du gradient. L'image est donc convoluée avec l'opérateur G_n , ou de façon équivalente, puisque les opérations de dérivation et de convolution sont linéaires :

$$I * G_n = I * \frac{\partial}{\partial n} G_\sigma = \frac{\partial}{\partial n} (I * G_\sigma) \quad (4.2)$$

L'image peut d'abord être filtrée par le filtre passe-bas gaussien, puis la dérivée directionnelle est calculée en chaque pixel de l'image. C'est cette dernière approche qui est habituellement préconisée lors de la réalisation d'un détecteur de Canny, tel par exemple dans le système de traitement et d'analyse d'images AphelionTM. Les étapes pour la réalisation d'un détecteur de Canny sont alors :

1. $I * G_\sigma$ Convolution de l'image avec un filtre passe-bas gaussien.
2. $\nabla(I * G_\sigma) = D_x, D_y$ Calcul du gradient selon l'axe horizontal et vertical.
3. $|\nabla(I * G_\sigma)| = \sqrt{D_x^2 + D_y^2}$ Calcul de l'amplitude du gradient.
4. Un pixel est activé (indique la présence d'une arête) si son amplitude de gradient est maximum dans la direction du gradient. La valeur de l'amplitude du pixel doit donc être la plus grande par rapport aux deux valeurs d'amplitude (obtenues par interpolation linéaire) de part et d'autre du pixel dans la direction du gradient.

Cette opération de sélection a pour effet d'amincir considérablement la largeur détectée des arêtes.

5. Sélection des arêtes significatives. Une opération de seuillage avec hystérésis est appliquée sur l'image d'arêtes obtenue à l'étape précédente. Les pixels de l'image d'arêtes qui ont été activés à l'étape précédente :

- sont conservés si leur valeur d'amplitude de gradient est supérieure à la valeur supérieure de seuil
- ne sont pas conservés (mis à zéro) si l'amplitude est inférieure à la valeur inférieure de seuil
- sont conservés si la valeur d'amplitude est comprise entre les deux valeurs de seuil et qu'un des pixels d'arêtes voisins a été conservé parce que sa valeur d'amplitude dépassait la valeur supérieure du seuil. C'est un processus itératif qui évolue de telle sorte que les pixels dont l'amplitude de gradient est située entre les deux valeurs de seuil sont convertis en pixels d'arête s'il existe un parcours pouvant les relier à un pixel d'arête dont l'amplitude est supérieure au seuil supérieur.

(Shu, 1989) utilise une technique semblable à celle de Canny, mais rétrécit les arêtes tant selon l'axe horizontal que selon l'axe vertical, au lieu de rétrécir l'arête selon l'axe dans la direction du gradient. Cette modification permet une simplification de la réalisation de l'algorithme par réseaux neuroniques puisque l'amincissement peut alors être réalisé par deux circuits de compétition pré-connectés, un selon l'axe horizontal et l'autre selon l'axe vertical.

Le détecteur de Canny est potentiellement fort intéressant parce qu'il effectue automatiquement, avant même d'appliquer le seuil avec hystérésis, un amincissement très efficace des arêtes. Par contre, le choix du seuil supérieur pour enlever les arêtes de bruit est non-trivial et dépend beaucoup de l'image à ana-

lyser. Dans la méthode proposée par Canny, une arête de bruit sera définie en relation avec le choix d'une valeur maximum et minimum pour le seuil. Nous proposons au contraire, dans notre démarche, qu'une arête de bruit soit plutôt définie par rapport à une représentation multirésolution de l'image à analyser. De cette façon, une arête de bruit est identifiée par rapport au contenu de l'image, et non en fonction de la valeur arbitraire d'un seuil.

La seconde école dominante, inspirée par l'équipe de Grossberg (Carpenter, Grossberg, & Mehanian, 1989; Gaussier, & Cocquerez, 1992; Grossberg, & Mingolla, 1985; Grossberg, & Todorovic, 1988; Grossberg, & Wyse, 1991), prône de débiter l'analyse par une détection des arêtes selon plusieurs directions pré-définies puis de choisir l'arête en chaque pixel selon l'orientation dominante détectée et selon l'orientation des arêtes dans le voisinage. Ces travaux s'inspirent largement de ce qui est connu du fonctionnement du cortex visuel chez les mammifères (Hubel, & Wiesel, 1959) et se prêtent donc particulièrement bien à une réalisation par réseaux neuroniques.

La figure 4.3 illustre le fonctionnement de ce système de détection d'arêtes. Les principales étapes de la détection, qui correspondent aux différentes couches de cellules, sont :

1. **Cellules simples :** Cette couche constitue le premier étage du système de détection de contour. Les cellules simples détectent une arête dans une direction et un sens donnés et à une résolution spatiale spécifique qui dépend de l'étendue du champ récepteur. Les arêtes sont détectées selon 4 orientations : horizontale, 45° , verticale (tel qu'illustré sur la figure), et 135° . (Carpenter, et al., 1989) utilisent deux grosseur de champ récepteur, et donc deux échelles spatiales d'analyse. La spécification du champ récepteur détermine l'opération effectuée par la cellule. Une opération simple de différentiation entre le côté droit et gauche de la cellule est calculée dans le type de détecteur proposé par Carpenter *et al*, mais d'autres champs récepteurs sont possibles, telle l'utilisation d'une cellule de Gabor (Daugman, 1987) qui modélise assez bien les cellules simples biologiques ou l'utilisation

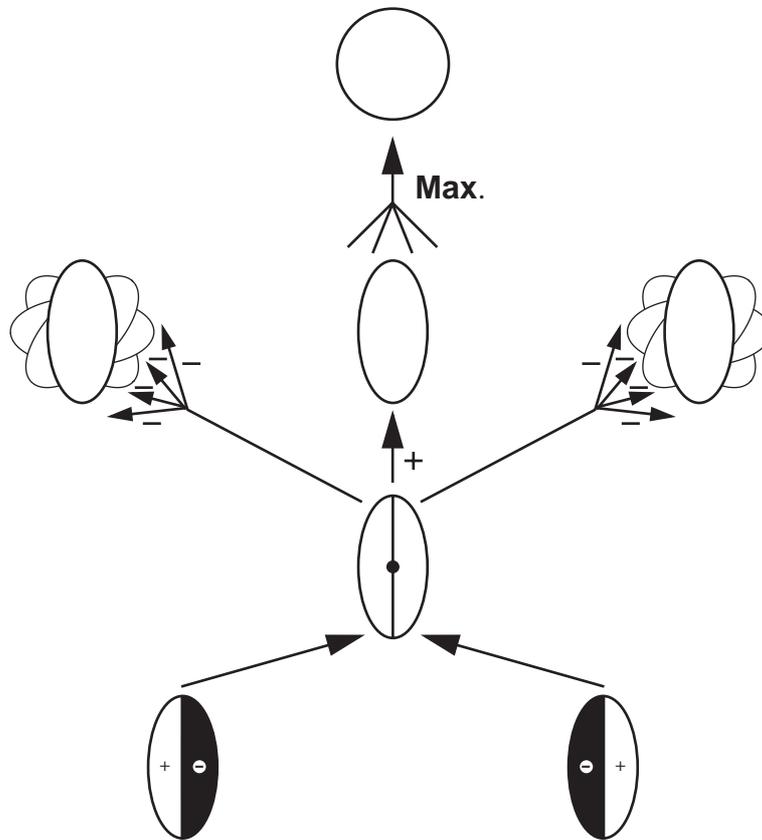


Figure 4.3 Système BCS de Grossberg.
 Version simplifiée du système BCS (“Boundary Contour System”) développé par Grossberg et son équipe. La première rangée est constituée de cellules simples qui détectent les arêtes dans une direction et une orientation données. La seconde couche est constituée de cellules complexes qui ne sont sensibles qu’à l’orientation de l’arête. Les troisième et quatrième couches constituent les deux couches de cellules hypercomplexes. Les deux couches sont des étages de compétition/coopération qui diminuent le bruit près des arêtes et indiquent la présence d’une arête indépendamment de sa direction ou de son orientation.

d’une cellule de Grossberg-Todorovic (Grossberg, et al., 1988) qui est construite à partir d’une combinaison pondérée de filtres orientés simulant des Différences de Gaussiennes (plus de

détails à la sous-section 4.2.3). La fonction d'activation de la cellule simple en est une de rectification qui ne peut produire qu'une valeur de sortie positive ou nulle.

2. **Cellules complexes :** Le second étage du détecteur est composé de cellules complexes qui produisent une détection insensible à la direction du contraste. Pour ce faire, chaque cellule complexe est connectée aux deux cellules simples de même orientation et de même résolution spatiale mais de direction de contraste opposée. Les sorties de chacune des deux cellules simples sont alors additionnées et la cellule complexe est activée lorsqu'une des deux cellules simples l'est. À chaque pixel de l'image, on a une cellule complexe pour chacune des 4 orientations et pour chacune des 2 échelles spatiales, soit la moitié du nombre de cellules simples.
3. **Cellules hyper-complexes :** Les cellules hyper-complexes forment les deux derniers étages du détecteur. Ce sont deux étages de compétition qui servent à supprimer le bruit près des arêtes alors que le bruit loin des arêtes est atténué grâce à l'effet moyenneur des cellules simples et complexes.

Le premier étage de compétition contribue spécifiquement à la diminution du bruit près des arêtes. Cet étage est un réseau de type "centre excitateur - voisinage inhibiteur". Les cellules hyper-complexes au voisinage d'une cellule complexe sont connectées par des liens de poids positif (liens d'excitation) si les cellules sont dans l'axe d'orientation de la cellule complexe et en plus, sont elles-mêmes orientées dans la même direction que la cellule complexe. Toutes les autres cellules hyper-complexes dans le voisinage de la cellule complexe sont connectées à cette dernière par des liens de poids négatifs

(liens d'inhibition). Les dimensions du voisinage sont choisies en fonction des échelles spatiales utilisées. Cet étage de compétition renforce les arêtes connectées, et donc les contours, et inhibe les arêtes isolées qui sont habituellement des arêtes bruitées.

Le second étage de compétition est simplement un réseau de type "le gagnant emporte tout" qui sélectionne l'activité maximale à chaque position de pixel. Les cellules hyper-complexes du dernier étage du détecteur indiquent donc pour chaque position de pixel la présence d'une arête, quelque soit son orientation ou sa direction.

4. **Interactions entre les deux échelles spatiales :**

Les cellules hyper-complexes possèdent des champs récepteurs dont les dimensions dépendent de l'échelle spatiale. Les cellules hyper-complexes à échelles fines suppriment le bruit près des arêtes et indiquent précisément la localisation des arêtes. Les cellules hyper-complexes à échelle grossière suppriment bien le bruit loin des arêtes mais n'indiquent pas précisément la localisation des arêtes. Une combinaison de ces deux cellules permet d'éliminer encore plus le bruit près et loin des arêtes.

5. **Complétion du contour :** Une dernière couche de cellules, appelées cellules bipolaires et non illustrée à la figure 4.3, permet de compléter le contour. Ces cellules sont activées s'il y a assez d'activation orientée de chaque côté (ou pôle) du centre du champ récepteur.

Le modèle de Grossberg est intéressant parce qu'assez près du modèle biologique tel que nous le connaissons. Par contre, l'opération du réseau complet dépend énormément de facteurs qu'il faut fixer assez intuitivement et le réseau lui-même nécessite une quantité très considérable de neurones et de connexions, ce qui provoque de nombreux problèmes lors des simulations sur ordinateurs classiques.

De nombreuses applications ne nécessitent pas une détection d'arêtes orientées aussi précise et compliquée que celle préconisée par l'équipe de Grossberg. Dans ces cas, un simple masque orienté (par exemple, Sobel ou une dérivée de gaussienne (Wilson, & Bhalerao, 1992)) est convolué avec l'image afin de découvrir les arêtes sur l'image à une orientation donnée. La description des arêtes orientées contenues dans l'image à analyser peut alors être utilisée, par exemple, pour caractériser et classifier une image texturée (Lepage, Laurendeau, & Gagnon, 1992) ou pour servir d'entrée à un processus d'analyse de l'image de plus grande complexité tel un analyseur de Fourier (Freeman, & Adelson, 1990). Une autre application (Brzakovic, Patton, & Wang, 1991) utilise des masques orientés pour trouver des arêtes candidates tandis que les arêtes résultantes sont choisies selon des règles heuristiques.

4.2.2 Maximum de la dérivée première isotrope

La détection d'arêtes par une mesure du maximum de la dérivée première isotrope consiste à détecter la présence d'une arête par une seule mesure indépendante de la direction de l'arête. L'amplitude du gradient local est généralement utilisée comme paramètre mesuré. Cette méthode permet une réalisation simple du circuit de détection mais n'offre pas la robustesse de la méthode du maximum de la dérivée première directionnelle lorsqu'utilisée seule.

La méthode de mesure du gradient utilisée seule n'offre pas une localisation précise de l'arête et est plus susceptible de détecter de fausses arêtes parce que c'est une méthode de détection ponctuelle qui ne tient pas compte des arêtes détectées dans le voisinage. Le détecteur d'arêtes proposé dans ce document (Lepage, & Poussart, 1991; Lepage, & Poussart, 1992) utilise cette méthode de détection surtout à cause de la simplicité du réseau neuronique résultant. L'amincissement des arêtes de même que la diminution du nombre de fausses arêtes sont obtenus grâce à une représentation multirésolution des images d'éclairément gauche et droite et à une interaction entre les arêtes détectées aux différentes échelles spatiales, comme il sera précisé à la section suivante. D'autres auteurs proposent des méthodes qui vont dans le même sens : (Chen, 1992) propose un opérateur morphologique qui préserve les arêtes lors d'un changement de l'échelle spatiale d'analyse; (Fleck, 1992) mesure la dérivée par une différence finie et améliore la détection en combinant des mesures de différence finie de différentes largeurs; et (McMichael, 1992) pour sa part, mesure le gradient de plusieurs images d'une scène éclairée de différentes façons et combine ces mesures, ce qui atténue l'effet de l'ombrage et renforce la détection des arêtes de l'image causées par des primitives de la scène.

4.2.3 Passage par zéro de la dérivée seconde directionnelle

La détection d'arêtes par une mesure du passage par zéro de la dérivée seconde de l'image calculée dans une direction spécifiée est une méthode de détection qui a fait l'objet de relativement peu de recherche. Ce manque d'intérêt provient en grande partie de la popularité et de l'efficacité des méthodes de détection du passage par zéro de la dérivée seconde non-directionnelle que nous étudions à la sous-section suivante. Les méthodes non-directionnelles résultent en un réseau neuronique de détection plus simple et linéaire, ce qui n'est pas le cas pour une mesure de la dérivée seconde directionnelle. De plus, l'utilisation de la dérivée seconde génère intrinsèquement plus de bruit que pour l'utilisation de la dérivée première. Quelques auteurs ont néanmoins exploré cette avenue, parmi lesquels nous pouvons mentionner (Clark, & Lawrence, 1984)

4.2.4 Passage par zéro de la dérivée seconde isotrope

La détection d'arêtes par une mesure du passage par zéro de la dérivée seconde de l'image calculée indépendamment de la direction est une méthode de détection qui fait l'objet d'une recherche diversifiée et qui suscite encore aujourd'hui énormément d'intérêt (Narbel, 1990; Ulupinar, & Medioni, 1990). Cet engouement pour cette méthode provient du fait qu'il existe un opérateur linéaire simple du second ordre qui détecte une variation brusque de l'éclairement indépendamment de la direction de la variation d'éclairement. Cet opérateur se nomme le Laplacien et opère sur un voisinage restreint du pixel d'intérêt. L'intérêt pour cette méthode provient aussi du fait que l'opérateur de Laplacien peut facilement être combiné avec un filtre passe-bas gaussien et ainsi rendre la détection d'arêtes sensible à l'échelle spatiale. Un dernier avantage apprécié de cette méthode est qu'elle permet, en principe, une localisation intra-pixel précise de l'arête puisque c'est un passage par zéro de l'image convoluée qui indique la présence d'une arête et que ce passage peut être interpolé à partir des valeurs positive et négative mesurées.

La méthode de détection du passage par zéro de la dérivée seconde isotrope est issue principalement des travaux de Marr et Hildreth (Hildreth, 1983; Marr, & Hildreth, 1980). Le détecteur de Marr-Hildreth, comme il est habituellement appelé, s'inspire des connaissances physiologiques du système visuel des mammifères : on croit que les contrastes de l'image rétinienne à certaines échelles spatiales sont d'abord accentués par un opérateur neuronique semblable à celui proposé par Marr et Hildreth et que les passages par zéro de l'image résul-

tante de même que leur orientation sont ensuite détectés pour former le croquis. Voyons plus en détail l'opérateur de Marr-Hildreth.

Les changements d'éclairément dans une image d'éclairément interviennent à différentes échelles spatiales. Leur détection optimale nécessite donc l'utilisation d'opérateurs de différentes largeurs. Le filtre passe-bas gaussien est un opérateur qui permet d'analyser, de « voir » l'image à différentes résolutions spatiales. C'est ainsi qu'un filtre gaussien à bande passante étroite brouille l'image de telle sorte que seuls les changements grossiers (à grande échelle) ressortent de l'image filtrée. Le filtre gaussien s'écrit :

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \quad (4.3)$$

Le paramètre σ indique le niveau de résolution auquel l'image est observée. L'image à la résolution σ est donc :

$$I_\sigma(x, y) = G_\sigma(x, y) * I(x, y) \quad (4.4)$$

Les arêtes sont ensuite trouvées sur l'image filtrée à l'aide du Laplacien, qui est un opérateur différentiel isotrope de second ordre, le plus petit ordre qui assure que l'opérateur soit linéaire. Le Laplacien s'écrit :

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \equiv \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (4.5)$$

L'isotropie assure qu'une arête, c'est-à-dire une variation brusque d'éclairément, soit détectée quelle que soit son orientation dans l'image. L'image résultante est alors la convolution du Laplacien et de l'image filtrée :

$$I_{sortie}(x, y) = \nabla^2(G_\sigma(x, y) * I(x, y)) \quad (4.6)$$

Puisque les deux filtres, le Laplacien ∇^2 et le filtre gaussien G_σ , sont linéaires et invariants, ils peuvent être combinés en un seul opérateur, le Laplacien de

gaussienne appelé aussi *chapeau mexicain* à cause de la forme de l'opérateur :

$$\nabla^2 G_\sigma(r) = \frac{-1}{2\pi\sigma^4} \left(2 - \frac{r^2}{\sigma^2} \right) e^{-\frac{r^2}{2\sigma^2}} \quad (4.7)$$

avec $r = \sqrt{x^2 + y^2}$ la distance radiale. La figure 4.4 illustre la forme (inversée) de l'opérateur de Laplacien de gaussienne. La largeur w de la portion négative de l'opérateur est donnée par :

$$w = 2\sqrt{2}\sigma \quad (4.8)$$

La largeur du lobe central de l'opérateur est donc proportionnelle à l'échelle spatiale d'observation σ . (Grimson, 1981) propose 4 canaux différents séparés chacun du précédent d'une octave pour l'analyse des arêtes de deux images stéréo d'entrée. Les canaux utilisés ont des largeurs du lobe central w de 4, 9, 17 et 35 pixels, ce qui correspondrait aux canaux d'analyse du système visuel humain (Wilson, & Bergen, 1979).

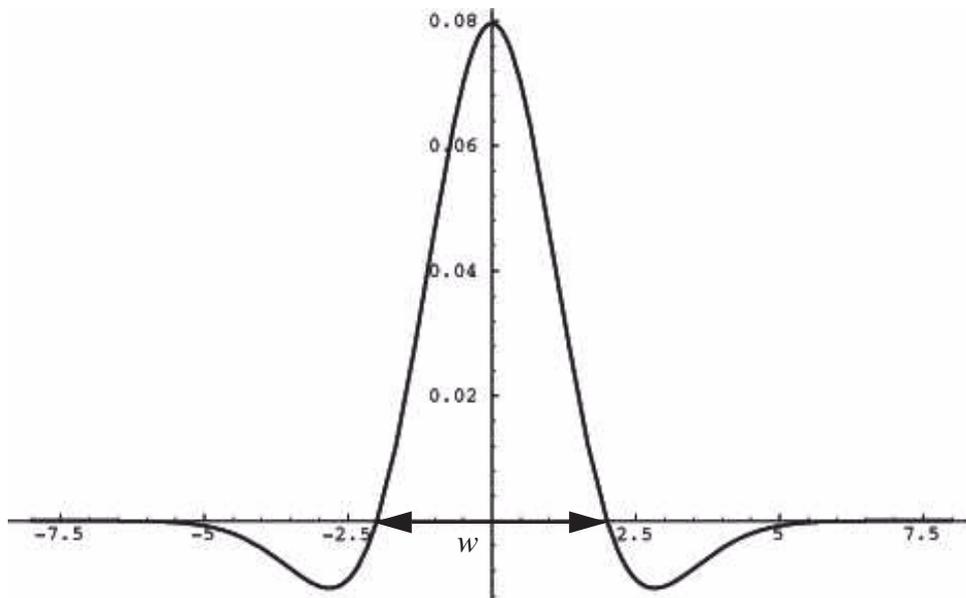


Figure 4.4 Laplacien de gaussienne. Opérateur « chapeau mexicain » qui détecte les variations d'éclairement en toute direction. L'opérateur de Marr-Hildreth est illustré ici pour un facteur d'échelle spatiale $\sigma = \sqrt{2}$.

Selon Marr et Hildreth, le Laplacien de gaussienne est la limite d'une différence de gaussiennes DdG :

$$\nabla^2 G_{\sigma}(r) \cong DdG(\sigma_e, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_e} e^{-\frac{r^2}{2\sigma_e^2}} - \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{r^2}{2\sigma_i^2}} \quad (4.9)$$

où $\sigma_e = \sigma$ représente la variance d'excitation (filtre gaussien positif) tandis que σ_i représente la variance d'inhibition (filtre gaussien négatif). L'équation 4.9 approxime un Laplacien de gaussienne par une différence de deux distributions gaussiennes lorsque :

$$\frac{\sigma_i}{\sigma_e} = 1,6 \quad (4.10)$$

Les cellules ganglionnaires, qui constituent le premier étage de traitement du système visuel chez les mammifères, ont un champ récepteur qui ressemble beaucoup à une différence de gaussiennes : le champ récepteur excitateur est une gaussienne de variance σ_e alors que le champ récepteur inhibiteur est une gaussienne plus large dans une proportion qui laisse croire que la fonction de ces cellules est effectivement d'appliquer l'équivalent d'un Laplacien de gaussienne sur l'image rétinienne afin d'accentuer les contrastes contenus dans l'image.

La détection des arêtes d'une image est illustrée à la figure 4.5 où une cellule de type *on-center* (cellule P) est utilisée en conjonction avec une cellule de type *off-center* (cellule Q). Ce principe de détection peut facilement être étendu à la détection d'arêtes orientées de longueurs variables, comme il est illustré à la figure 4.6.

4.2.5 Détection des arêtes par apprentissage supervisé

Les arêtes correspondent à des variations brusques d'éclairément. De telles variations sont détectées avec un opérateur de différentiation du premier ou du second ordre, opérateur qui sera le plus souvent combiné avec un filtre passe-bas afin de diminuer les artefacts générés par l'opération de différentiation. La traduction neuronique de ces opérateurs résulte en des connexions qui relient la couche image à la couche du croquis avec des poids fixes. Des recherches récentes explorent différentes formes d'apprentissage afin de trouver d'autres valeurs

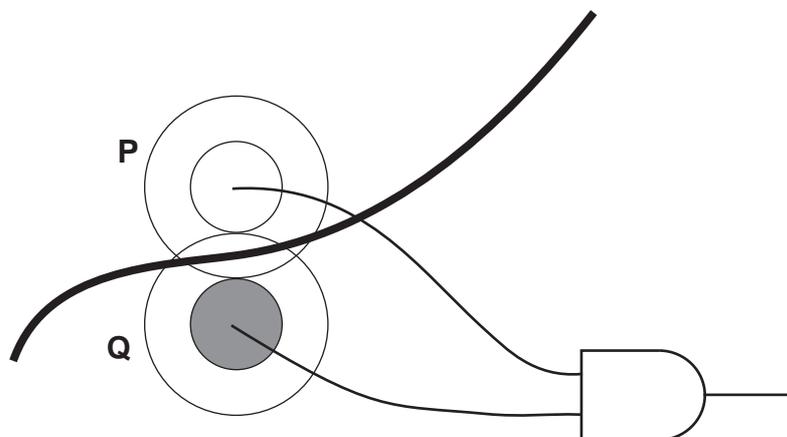


Figure 4.5 *Différence de gaussiennes. Détection d'arêtes par deux opérateurs de différence de gaussiennes. Chaque cellule de détection comprend un champ récepteur d'excitation (cercle central) et un champ récepteur d'inhibition (cercle extérieur). P et Q sont deux cellules complémentaires qui produisent une sortie inversée l'une par rapport à l'autre. La sortie de la cellule ET est activée lorsque les deux cellules P et Q sont activées et donc qu'une arête est située entre les deux cellules.*

de poids des connexions qui parviennent à la couche du croquis. Moura et Martins (Moura, & Martins, 1991) utilisent un réseau à 3 étages. Les deux premiers étages sont interreliés par des liens à poids fixes qui réalisent un filtre passe-bas gaussien pour le premier étage et un masque de Sobel pour le second afin de détecter l'amplitude et la direction du gradient. Le dernier étage en est un de satisfaction de contraintes pour éliminer le bruit et construit des liens excitateurs entre des neurones voisins orientés perpendiculairement au gradient et des liens inhibiteurs entre des neurones voisins orientés parallèlement au gradient. Bhatia $\epsilon\tau\alpha\lambda$ (Bhatia, Srinivasan, & Ong, 1991) proposent, à la même époque, un réseau à un seul étage, l'étage du croquis, relié à la couche de l'image d'éclaircissement. Chaque emplacement du croquis est occupé par quatre neurones, qui indiquent l'éclaircissement moyen de même que la force et la direction de l'éventuelle arête détectée. Chaque groupe de 4 neurones du croquis sont interreliés par des liens inhibiteurs pour augmenter la compétition et chaque groupe est relié à un champ récepteur de 16x16 pixels en dessous, sur l'image d'éclaircissement. Deux séries de connexions composent le champ récepteur, une première à poids fixes réalise un filtre gaussien et l'autre série doit être entraînée de façon non-supervisée à l'aide d'une règle hebbienne modifiée. L'image uti-

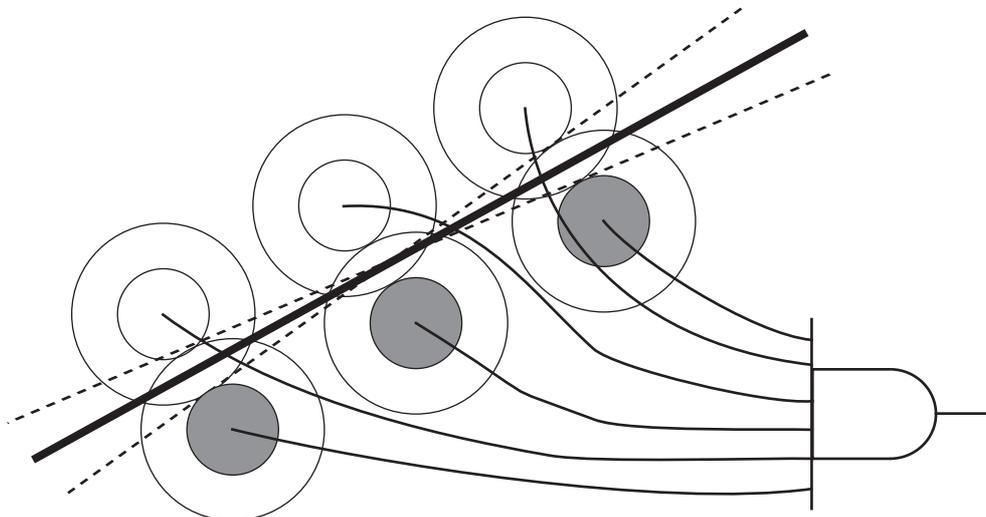


Figure 4.6 Différence de gaussiennes orientée. Détection d'arêtes orientées par des opérateurs alignés et complémentaires de différence de gaussienne. La sortie de la cellule ET est activée lorsque toutes les cellules connectées à la porte sont activées. La configuration des opérateurs détermine la plage d'orientations qui est détectée (lignes pointillées sur la figure).

lisée pour la phase d'entraînement est constituée de cercles concentriques pour avoir un bon échantillonnage d'arêtes de force et d'orientations variées.

Par la suite, Etemad et Chellappa (Etemad, & Chellappa, 1993) proposent un réseau neuronique approximateur de fonctions et entraîné par rétro-propagation du gradient d'erreur pour résoudre le problème de détection d'arêtes. Le réseau proposé comporte une couche d'entrée de 3x3 qui balaie l'image d'éclairément, d'une couche cachée et d'une couche de sortie également de 3x3 qui représente l'arête détectée débarrassée du bruit. Un ingénieux système de vote entre les couches de sortie voisines permet de diminuer le bruit et les artefacts. Mkaouar et Lepage (Mkaouar, & Lepage, 1996) proposent le même type de réseau à rétro-propagation du gradient d'erreur, mais cette fois en utilisant une couche d'entrée de 5x5, 2 couches cachées, et un seul neurone à la sortie pour indiquer la présence ou l'absence d'une arête au centre de la couche d'entrée. Les résultats sont encourageants et des travaux se poursuivent pour interpréter les configurations de poids obtenues et améliorer les performances en diminuant le temps nécessaire pour l'apprentissage.

4.2.6 Détection multirésolution des arêtes par réseau neuronique

Les réalisations neuroniques de détecteurs d'arêtes multirésolution ne font pas l'objet de recherche répandue probablement à cause de la complexité des modes de représentation de l'information visuelle multirésolution. (Lepage, & Crevier, 1994) proposent une extraction multirésolution des arêtes basée sur une représentation pyramidale et utilisant un neurone de vigilance à chaque niveau de la pyramide pour contrôler le renforcement des arêtes dans une direction basse \rightarrow haute résolution. (Mkaouar, & Lepage, 1995) proposent une extraction multirésolution des arêtes en utilisant une pyramide pour accélérer les simulations et un filtre gaussien récursif pour la génération des niveaux de représentation multirésolution. Ce modèle leur sert de base pour générer la base d'apprentissage dans (Mkaouar, et al., 1996).

4.3 DÉTECTION NEURONIQUE DE L'AMPLITUDE DU GRADIENT

Nous illustrons en détail dans ce document une détection d'arêtes basée sur la mesure du maximum de la dérivée première de l'image d'éclairément. Afin de diminuer le nombre d'arêtes bruitées détectées et afin d'amincir les arêtes détectées, la détection d'arêtes est effectuée à plusieurs résolutions spatiales et une interaction entre les arêtes détectées aux différentes échelles aide à obtenir un croquis précis et peu bruité.

Les images d'éclairément gauche et droite sont décomposées en représentations multirésolutions selon une structure pyramidale. Chaque niveau de la pyramide représente l'image d'éclairément correspondante à une résolution spatiale σ . Les arêtes de l'image sont d'abord détectées à chacun des niveaux des deux pyramides. Afin de détecter le gradient de l'arête en chaque position de pixel de chacune des images, la dérivée partielle selon l'horizontale et la dérivée partielle selon la verticale sont calculées à l'aide de connexions latérales dans le voisinage 3×3 du pixel d'intérêt. Des masques de Sobel sont utilisés pour la mesure des dérivées partielles. Le champ récepteur pour la mesure des dérivées partielles est illustré à la figure 4.7. La sortie du neurone au pixel d'intérêt est égal à la somme de la valeur absolue des deux dérivées partielles, ce qui constitue une approximation de l'amplitude du gradient à cette position. Cette approximation s'avère suffisante pour les besoins de notre détecteur puisqu'elle sera toujours positive, comme l'amplitude, et qu'elle sera toujours plus grande

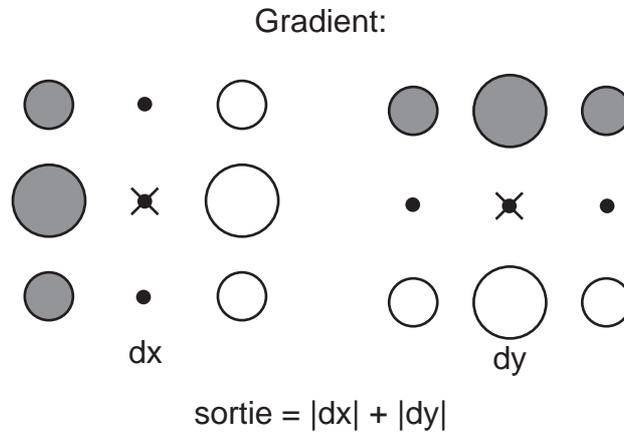


Figure 4.7 Champ récepteur de gradient.

Champ récepteur pour la mesure d'une approximation du gradient. Les masques sont des masques de Sobel. Un cercle noir indique une connexion avec un poids négatif dont l'amplitude est proportionnelle à la grandeur du cercle.

que la valeur de l'amplitude, et donc qu'elle accentuera la présence d'une arête. Les masques de Sobel utilisés ont comme valeurs de poids :

$$dx = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad dy = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (4.11)$$

Les arêtes détectées par une méthode de gradient sont la plupart du temps trop larges pour permettre une localisation précise de celle-ci. Le module suivant de traitement, le module stéréo, nécessite une localisation précise des arêtes appariées afin d'obtenir une mesure fiable et précise de la disparité. Nous devons donc amincir les arêtes détectées de telle sorte qu'un neurone activé représente une arête, et donc un maximum de la dérivée première.

4.4 AMINCISSEMENT DES ARÊTES

L'amincissement des arêtes peut être réalisé de différentes façons (Jang, et al., 1990; Lam, Lee, & Suen, 1992). Nous proposons un simple réseau de compétition qui favorise les neurones les plus activés dans le groupe de compétition.

Un réseau de compétition est construit pour tous les groupes de 2x2 neurones de toutes les couches des deux pyramides gauche et droite. Le réseau de compétition est illustré à la figure 4.8 pour un groupe de 4 neurones. Le poids

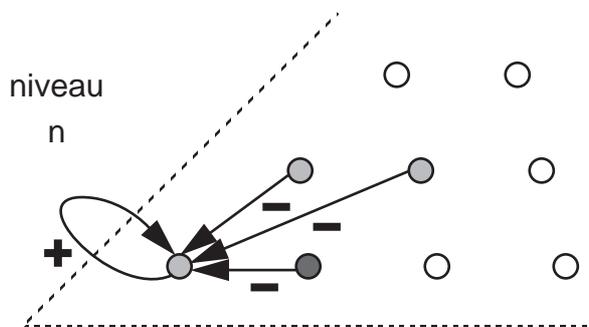


Figure 4.8 Réseau de compétition.

Chaque regroupement de 2x2 neurones est interconnecté en réseau de compétition. La figure ne montre les connexions que pour un des neurones. Le même schéma se répète pour les autres neurones. Le lien d'excitation a un poids de 0,3 tandis que les liens d'inhibition provenant des autres neurones ont un poids de -0,1. La fonction d'activation des neurones est la fonction iac de Rumelhart-McClelland. Le circuit fonctionne de telle sorte que les un ou deux neurones les plus activés tendent vers une sortie pleinement activée tandis que la sortie des autres neurones décroît vers zéro.

des connexions d'auto-excitation est de 0,3 tandis que le poids des connexions d'inhibition est de -0,1¹. La fonction d'activation des neurones à cette étape de la simulation est la fonction d'activation interactive et compétitive de Rumelhart & McClelland. L'équation d'activation iac est décrite par l'équation différentielle :

$$\begin{aligned} dx &= (B - x)net - Ax & net > 0 \\ dx &= (x - C)net - Ax & net \leq 0 \end{aligned} \quad (4.12)$$

avec : $dx : x(n+1) - x(n)$. Différentielle de la valeur d'activation

1. Plusieurs valeurs de poids pour les liens d'inhibition et d'excitation ont été expérimentées. La combinaison mentionnée dans le texte est celle qui a donné les meilleurs résultats tant du point de vue vitesse de convergence vers un état stable que du point de vue capacité à conserver les arêtes les plus activées. Cette combinaison a été trouvée de façon plutôt empirique, quoiqu'un bon fonctionnement du réseau demande que le poids du lien d'excitation soit égal ou supérieur à la somme de la valeur absolue des poids d'inhibition.

entre l'instant présent $n+1$ et l'instant précédent n .

x : la valeur d'activation au temps n .

A : facteur de décroissance. Paramètre de contre-réaction qui fait que le neurone se comporte comme un système discret du 1er ordre. Le facteur $(1-A)$ détermine le degré de contre-réaction. Une valeur de A unitaire indique l'absence de contre-réaction.

B : valeur maximale atteinte par la valeur d'activation.

C : valeur minimale atteinte par la valeur d'activation.

net : $\sum_j W_{ij} O_j$: valeur nette à l'entrée du neurone.

Le comportement du réseau qui résulte de cette configuration est tel que la sortie des un ou deux neurones les plus activés tend vers une valeur de pleine activation tandis que la sortie des autres neurones décroît vers la valeur minimum d'activation du neurone, soit zéro.

L'amincissement des arêtes assure une localisation précise des arêtes mais ne garantit pas que l'arête détectée soit générée par une primitive physique de la scène observée, et non une arête de bruit générée par la caméra ou le système d'échantillonnage. L'information détectée aux autres niveaux de la pyramide aide dans la décision de garder ou de rejeter une arête détectée.

4.5 RENFORCEMENT DES ARÊTES

Selon (Marr, 1982), un attribut physique de la scène observée apparaît à au moins deux échelles spatiales consécutives sur l'image. Une arête qui n'apparaît que sur la couche à haute résolution est probablement une arête de bruit et doit donc être éliminée du croquis qui sera soumis à l'appariement stéréoscopique.

La représentation des arêtes aux différents niveaux de résolution spatiale de la pyramide est illustrée à la figure 4.9.

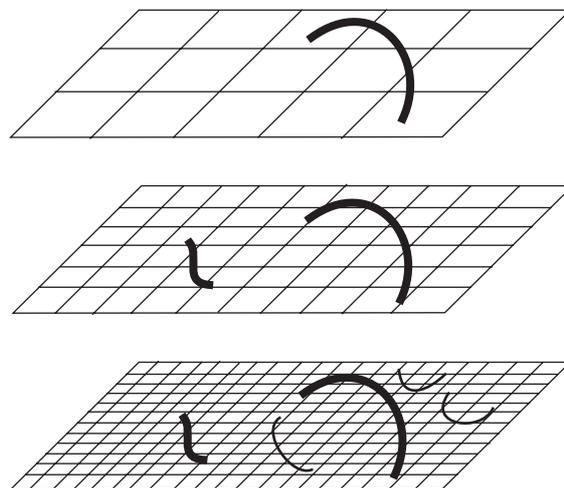


Figure 4.9 Comportement des arêtes à multiple résolutions. Comportement des arêtes dans une représentation pyramidale multirésolution. Les arêtes qui correspondent à des caractéristiques physiques de la scène observée sont présentes à au moins deux niveaux consécutifs. Les arêtes de bruit produites par le système de caméra ou par la nature discrète de l'échantillonnage n'apparaissent qu'aux niveaux à haute résolution.

La recherche sur le comportement des arêtes le long de l'espace d'échelle est abondante et diversifiée. L'ensemble de ces travaux a comme objectif d'identifier une arête sur une image à faible résolution et de projeter cette arête sur l'image à haute résolution afin de la localiser avec précision. (Bergholm, 1987), dans un papier majeur pour la poursuite d'arêtes à travers l'espace d'échelles, propose une poursuite continue d'une arête à partir d'une échelle à résolution grossière. Le changement d'échelle est continu et la variation du facteur d'échelle est telle que la position de l'arête poursuivie ne change pas plus d'un seul pixel. Lu et Jain (Lu, & Jain, 1989; Lu, & Jain, 1992) décrivent le comportement des arêtes le long de l'échelle spatiale. Ils détaillent le comportement de deux arêtes voisines qui peuvent se disloquer, se fusionner ou disparaître. (Mallat, & Zhong, 1992), pour leur part, étudient le comportement d'un détecteur de Canny à travers l'espace d'échelles spatiales. Un détecteur de Canny est équivalent à trouver les maxima locaux d'une transformée d'ondelettes. L'article est une étude des propriétés des arêtes multirésolutions à travers la

théorie des ondelettes. Finalement, (Ziou, & Tabbone, 1991) d'une part, puis (Perl, & Marsland, 1992) utilisent seulement deux échelles spatiales pour la poursuite des arêtes à travers l'espace d'échelles.

La recherche des arêtes à multiples résolutions spatiales présuppose qu'aucune nouvelle arête n'est générée à mesure que la résolution est baissée. Le choix du filtre gaussien pour la génération de la représentation multirésolution assure qu'il n'y aura pas de nouvelles arêtes. De plus, l'opérateur de détection des arêtes doit être un opérateur de différentiation *linéaire*, par exemple le passage par zéro du laplacien, qui est un opérateur de dérivée seconde non-directionnelle, et le maximum de la dérivée première (Tihanyi, & Barron, 1990). Cette contrainte exclut explicitement le détecteur de Canny, qui est un opérateur directionnel et non-linéaire. L'opérateur utilisé pour les travaux de cette thèse est la somme de l'amplitude absolue des dérivées partielles selon l'horizontale et la verticale. L'opérateur de valeur absolue, qui n'est pas linéaire, ne sert qu'à estimer l'amplitude de la dérivée première de l'image. L'opération de détection des arêtes par approximation de l'amplitude du gradient remplit donc la condition de non-génération d'arêtes à mesure que la résolution diminue puisque ce sont les dérivées partielles de l'image filtrée qui sont calculées, des opérations de différentiation qui sont linéaires.

Nous utilisons, pour le renforcement des arêtes, les quatre échelles spatiales utilisées pour la représentation multirésolution de chaque image d'éclairage gauche et droite. À un niveau de résolution spatiale donné, une arête est conservée (renforcement) si cette dernière est présente au niveau de la pyramide juste au dessus (et donc de résolution moindre). L'arête doit être écartée, particulièrement si elle se trouve au niveau de plus haute résolution, si elle n'est pas présente au niveau au dessus : dans ce cas, on n'est vraisemblablement pas en présence d'une arête causée par une primitive physique de la scène observée. Le renforcement d'arêtes selon leur présence aux niveaux de plus basse résolution est réalisé selon la **règle d'apprentissage de renforcement**, soit en modifiant les poids du circuit d'amincissement décrit à la section précédente selon que le neurone au niveau juste au dessus est activé ou non. Le processus de renforcement d'arêtes est illustré à la figure 4.10. Le renforcement d'arêtes est réalisé en modulant les poids du circuit d'amincissement par la valeur de sortie du neurone situé juste au dessus, au niveau suivant de moindre résolution de la pyramide. Si ce neurone est activé, indiquant que l'arête se retrouve à ce niveau de résolution, le réseau d'amincissement opère normalement et son comportement dynamique est tel que décrit à la section précédente. Les liens partant du neurone au niveau $n+1$ et qui modulent les poids du circuit de compétition ont un poids de 5. Si le neurone au niveau au dessus n'est pas ou est peu activé, indiquant l'absence

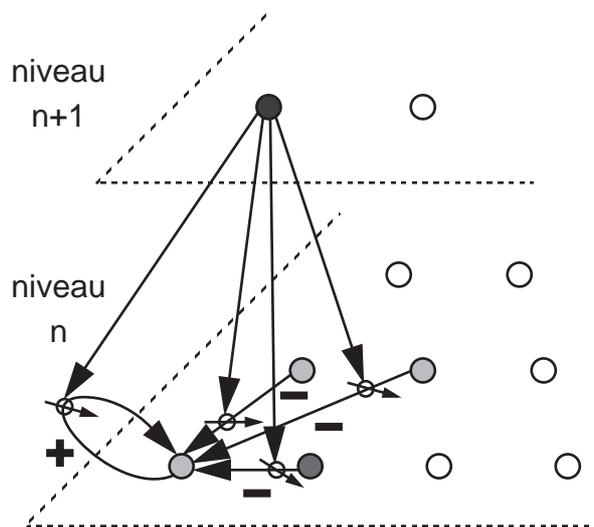


Figure 4.10 Renforcement des arêtes.

Le circuit d'amincissement d'arêtes opère normalement si le neurone au niveau de résolution juste au dessus est activé, indiquant la présence de l'arête à ce niveau de résolution moindre. Si l'arête détectée n'apparaît pas au niveau au dessus, elle est considérée comme une arête de bruit et le poids de toutes les connexions du circuit de compétition tend vers zéro, entraînant que l'activité des 4 neurones du groupe tendra vers zéro, effaçant l'arête de bruit détectée.

d'arêtes à cet endroit, les poids du circuit de compétition en dessous ont une valeur tendant vers zéro, ce qui entraîne que l'activité des 4 neurones du groupe décroît vers zéro, effaçant ainsi la présence d'une arête de bruit.

4.6 RÉSULTATS EXPÉRIMENTAUX

Le réseau de détection d'arêtes fut d'abord testé sur une image passablement complexe qui est largement utilisée dans la littérature pour vérifier et comparer les résultats de divers algorithmes de traitement d'image. L'image *Lenna* est illustrée à la figure 4.11. La première étape de détection latérale des arêtes à chaque niveau de la pyramide est illustrée à la figure 4.12. Le croquis ainsi généré n'est le résultat que de la détection latérale des arêtes, c'est-à-dire que la sortie de chaque neurone est une approximation de l'amplitude du gradient à cet emplacement. Les dérivées partielles selon x et y nécessaires au calcul du gradient sont approximées par des champs récepteurs ayant la forme de masque de



Figure 4.11 Image de Lenna.

Image pour illustrer l'algorithme neuronique de détection d'arêtes.

Sobel. Durant les étapes subséquentes de l'algorithme, les arêtes sont amincies grâce aux réseaux de compétition 2x2 et sont renforcées lorsque l'arête est présente au niveau juste au dessus, et encore plus si l'arête est présente à plus d'un niveau au dessus dans la pyramide. Le résultat de l'amincissement et du renforcement apparaît à la figure 4.13, où le croquis à pleine résolution aminci et ren-



Figure 4.13 Amincissement et renforcement des arêtes.

Le résultat de la détection latérale d'arêtes, à gauche, est comparé au résultat du renforcement et de l'amincissement après deux tops d'horloge.

forcé après deux tops d'horloge est comparé au croquis pleine résolution résultant de la détection latérale seulement. Les arêtes correspondant aux attributs physiques de la scène originale ont été conservées tandis que celles correspondant à des détails non-significatifs ou produites par l'erreur de quantification dans la production de l'image ont été éliminées.

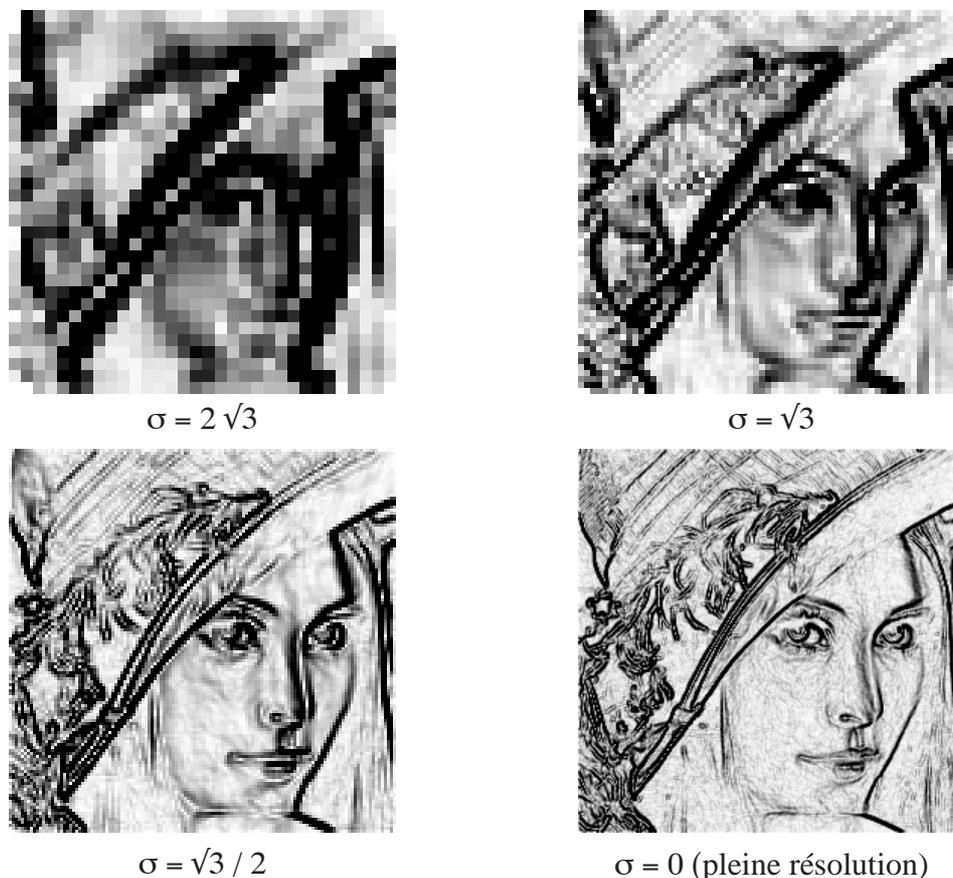


Figure 4.12 Croquis multirésolution de l'image Lenna.

Croquis de l'image Lenna aux 4 niveaux de résolution de la pyramide. Aucun amincissement ni renforcement ne sont appliqués.

CONCLUSION

La détection des arêtes constitue la première étape du système de vision et produit le **croquis**, la première forme de représentation interne de données du système. Nous nous sommes d'abord interrogé sur la nature des arêtes dans une image et leur relation avec des attributs physiques de la scène observée dont nous étudions une représentation par image d'éclairage produite par une caméra. Nous avons ensuite procédé à une analyse des méthodes classiques de détection d'arêtes. Puis, nous avons décrit la méthode de détection proposée dans ce document, soit une technique neuronique multirésolution qui utilise

l'information présente à plusieurs niveaux de résolution spatiale afin d'en arriver à la construction d'un croquis précis qui représente les primitives physiques de la scène sous observation.

Le croquis est la première forme de représentation interne du système de vision. Cette représentation interne sert d'entrée à des algorithmes tel la stéréoscopie et l'analyse du mouvement. Ces algorithmes nécessitent une localisation précise des arêtes de chacune des images gauche et droite et produisent l'**esquisse**, une forme de représentation interne qui mesure la profondeur de la scène aux différents points caractéristiques de cette dernière.

RÉFÉRENCES

- Ballard, D. H. (1981). Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, **13**(2), 111-122.
- Barr, A., Cohen, P. R., & Feigenbaum, E. A. (Ed.). (1989). *The Handbook of Artificial Intelligence*. Reading, Mass.: Addison-Wesley.
- Basseville, M. (1979). Détection de contours: méthodes et études comparatives. *Ann. Télécommunic.*, **34**(11-12), 559-579.
- Bergholm, F. (1987). Edge Focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-9**(6), 726-741.
- Besançon, J. E. (1988). *Vision par ordinateur en deux et trois dimensions*. Paris: Eyrolles.
- Bhatia, P., Srinivasan, V., & Ong, S. H. (1991). Single-layer edge detector with competitive unsupervised learning. *Proceedings of the International Joint Conference on Neural Networks*, Singapore, **1**, 634-639.
- Boyer, K. L., & Sarkar, S. (1992). Assessing the State of the Art in Edge Detection: 1992. *Proceedings of the SPIE Applications of Artificial Intelligence X: Machine Vision and Robotics*, Orlando, Florida, 22-24 april, **1708**, 353-362.
- Brzakovic, D., Patton, R., & Wang, R. L. (1991). Rule-Based Multitemplate Edge Detector. *Computer Vision, Graphics and Image Processing: Graphical Models and Image Processing*, **53**(3), 258-268.
- Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-8**(6), 679-698.
- Carpenter, G. A., Grossberg, S., & Meharian, C. (1989). Invariant Recognition of Cluttered Scenes by a Self-Organizing ART Architecture: CORT-X Boundary Segmentation. *Neural Networks*, **2**, 169-181.
-

-
- Chen, J.-S. (1992). Generalized Adaptive Smoothing for multiscale edge detection. *Proceedings of the SPIE Applications of Artificial Intelligence X: Machine Vision and Robotics*, Orlando, Florida, 22-24 april, **1708**, 341-352.
- Chen, M.-H., & Yan, P.-F. (1989). A Multiscaling Approach Based on Morphological Filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(7), 694-700.
- Clark, J. J., & Lawrence, P. D. (1984). A Hierarchical Image Analysis System Based Upon Oriented Zero Crossings of Bandpassed Images. In A. Rosenfeld (Ed.), *Multiresolution Image Processing and Analysis* (pp. 148-168). New York: Springer-Verlag.
- Cohen, P. R., & Feigenbaum, E. A. (Ed.). (1982). *The Handbook of Artificial Intelligence*. Reading, Mass.: Addison-Wesley.
- Daugman, J. G. (1987). Image Analysis and Compact Coding by Oriented 2D Gabor Primitives. *Proceedings of the SPIE: Image Understanding and the Man-Machine Interface*, **758**, 19-30.
- Etemad, K., & Chellappa, R. (1993). A Neural Network Based Edge Detector. *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, CA, March 28 - April 1, **1**, 132-137.
- Fleck, M. M. (1992). Multiple Widths Yield Reliable Finite Differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(4), 412-429.
- Freeman, W. T., & Adelson, E. H. (1990). Steerable Filters for Early Vision, Image Analysis, and Wavelet Decomposition. *Proceedings of the 3rd ICCV-90: Third International Conference on Computer Vision*, Osaka, Japan, Dec. 4-7, 406-415.
- Gaussier, P., & Cocquerez, J.-P. (1992). Neural networks for complex scene recognition: simulation of a visual system with several cortical areas. *Proceedings of the IJCNN '92 Baltimore - International Joint Conference on Neural Networks*, Baltimore, Maryland, June 7-11, **III**, III-233 - III-259.
- Gonzalez, R. C., & Wintz, P. (1977). *Digital Image Processing*. London: Addison-Wesley.
-

-
- Grimson, W. E. L. (1981). *From Images to Surfaces - A Computational Study of the Human Early Visual System*. Cambridge, Mass.: The MIT Press.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, **38**(2), 141-171.
- Grossberg, S., & Todorovic, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics*, **43**, 241-277.
- Grossberg, S., & Wyse, L. (1991). Invariant Recognition of Cluttered Scenes by a Self-Organizing ART Architecture: Figure-Ground Separation. *Proceedings of the IEEE INNS International Joint Conference on Neural Networks - Seattle*, Seattle, Washington, July 8-12, **1**, I-633 - I-638.
- Hildreth, E. C. (1983). The Detection of Intensity Changes by Computer and Biological Vision Systems. *Computer Vision, Graphics, and Image Processing*, **22**, 1-27.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, **148**, 574-591.
- Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice Hall.
- Jang, B.-K., & Chin, R. T. (1990). Analysis of Thinning Algorithms Using Mathematical Morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(6), 541-551.
- Lam, L., Lee, S.-W., & Suen, C. Y. (1992). Thinning Methodologies - A Comprehensive Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(9), 869-885.
- Lepage, R., & Crevier, D. (1994). Multiresolution neural network for the extraction of the primal sketch. *Proceedings of the SPIE Intelligent Robots and Computer Vision XIII: Algorithms and Computer Vision*, Boston, Massachusetts, 31 October - 2 November, **2353**, 375-384.
-

-
- Lepage, R., Laurendeau, D., & Gagnon, R. (1992). Extraction of texture features with a multiresolution neural network. *Proceedings of the SPIE: Applications of Artificial Neural Networks III*, Orlando, Florida, 21-24 april, **1709**, 64-75.
- Lepage, R., & Poussart, D. (1991). A Neural Network for Multiresolution Edge Detection. *Proceedings of the Canadian Conference on Electrical and Computer Engineering*, Québec (Québec) Canada, September 25-27, **1**, 25.1.1 - 25.1.4.
- Lepage, R., & Poussart, D. (1992). Multiresolution Edge Detection. *Proceedings of the IJCNN '92 Baltimore - International Joint Conference on Neural Networks*, Baltimore, Maryland, June 7-11, **IV**, IV-438 - IV-443.
- Levine, M. D. (1985). *Vision in man and machine*. New York: McGraw-Hill.
- Lu, Y., & Jain, R. C. (1989). Behavior of Edges in Scale Space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(4), 337-356.
- Lu, Y., & Jain, R. C. (1992). Reasoning about Edges in Scale Space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(4), 450-468.
- Mallat, S., & Zhong, S. (1992). Characterization of Signals from Multiscale Edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(7), 710-732.
- Marr, D. (1982). *Vision - A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, **207**, 187-217.
- McMichael, D. (1992). Fusing Multiple Images and Extracting Features for Visual Inspection (No. 762). Control Systems Centre, UMIST, Manchester, UK.
- Mkaouar, M., & Lepage, R. (1995). Extraction des caractéristiques d'une image par analyse à multiple résolutions spatiales. *Proceedings of the Congrès canadien en génie électrique et informatique*, Montréal, Québec, du 5 au 8 septembre, 1176-1179.
-

- Mkaouar, M., & Lepage, R. (1996). Le Perceptron multi-couches pour la détection des arêtes d'une image. *Proceedings of the Vision Interface '96*, Toronto (Ontario) Canada, 21-24 mai, 41-48.
- Moran, C. J. (1990). A Morphological Transformation for Sharpening Edges of Features before Segmentation. *Computer Vision, Graphics, and Image Processing*, **49**, 85-94.
- Moura, L., & Martins, F. C. M. (1991). Edge Detection Through Cooperation and Competition. *Proceedings of the International Joint Conference on Neural Networks*, Singapore, **3**, 2588-2593.
- Narbel, P. (1990). The zero-crossings of the Laplacian of Gaussian and their limitations (No. 90.05). Geneva University.
- Perl, D., & Marsland, T. A. (1992). An adaptive neural network for image enhancement (ANNIE). *Proceedings of the SPIE: Applications of Artificial Neural Networks III*, Orlando, Florida, 21-24 april, **1709**, 76-86.
- Princen, J., Illingworth, J., & Kittler, J. (1990). A Hierarchical Approach to Line Extraction Based on the Hough Transform. *Computer Vision, Graphics, and Image Processing*, **52**, 57-77.
- Shu, J. S.-P. (1989). One-pixel edge detection. *Pattern Recognition*, **22**(6), 665-673.
- Tan, H. L., Gelfand, S. B., & Delp, E. J. (1992). A Cost Minimization Approach to Edge Detection Using Simulated Annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(1), 3-18.
- Tihanyi, E. Z., & Barron, J. L. (1990). Spatio-Temporal Edge Focusing (No. 267). Department of Computer Science, the University of Western Ontario, London, Ontario.
- Torre, V., & Poggio, T. A. (1986). On Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-8**(2), 147-163.
- Ulupinar, F., & Medioni, G. (1990). Refining Edges Detected by a LoG Operator. *Computer Vision, Graphics, and Image Processing*, **51**, 275-298.
-

-
- Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for spatial vision. *Vision Research*, **19**, 19-32.
- Wilson, R., & Bhalerao, A. H. (1992). Kernel Designs for Efficient Multiresolution Edge Detection and Orientation Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(3), 384-390.
- Ziou, D., & Tabbone, S. (1991). Fusion des contours dans un espace échelle. *Proceedings of the Treizième Colloque sur le Traitement du Signal et des Images*, Juan-les-Pins, **2**, 1013 - 1016.
- Zucker, S. W., David, C., Dobbins, A., & Iverson, L. (1988). The Organization of Curve Detection: Coarse tangent fields and fine spline coverings. *Proceedings of the Second International Conference on Computer Vision*, Tampa, Florida, December 5-8, 568-577.
-

