Première partie

Perception de structures curvilignes en vision par ordinateur

Chapitre 1

Vision naturelle et artificielle

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

1.1 Problématique de la vision par ordinateur

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

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

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

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

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

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

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

Pour citer Y. Aloimonos :

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

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

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

1.2 Théories de la perception visuelle

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

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

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

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

1.2.1 Comment définir la vision ?

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

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

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

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

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

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

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

ou bien :

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

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

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

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

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

ou encore, d'une façon plus élaborée :

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

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

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

1.2.2 Psychologie de la perception visuelle

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

1.2.2.1 Approche psycho-physique et concept de seuil

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

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

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

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

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

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

1.2.2.2 Théorie du Gestalt

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

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

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

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

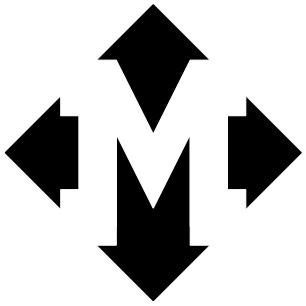

Figure 1.1 - *Illustration de la* Gestaltqualität - *cet arrangement de flèches noires représente quelque chose de plus que la somme de quatre flèches.*

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

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

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

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

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

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

1.2.2.3 Fonctionnalisme probabiliste de Brunswik

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

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

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

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

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

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

1.2.2.4 Empirisme ou paradigme constructioniste

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

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

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

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

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

Un certain nombre d'arguments semblent confirmer cette théorie:

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

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

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

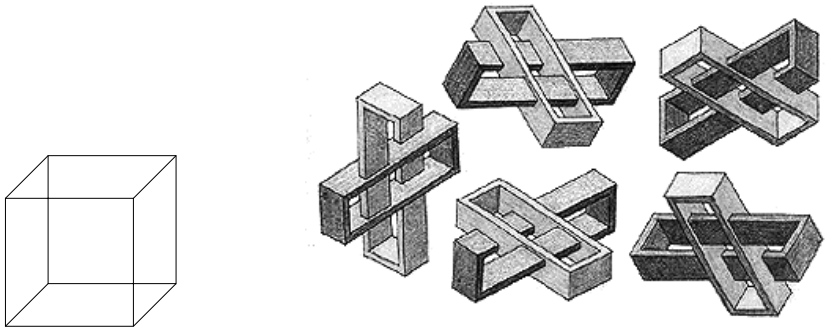

Figure 1.2 - *Cube de Necker et figures impossibles de Oscar Reutersvärd*

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

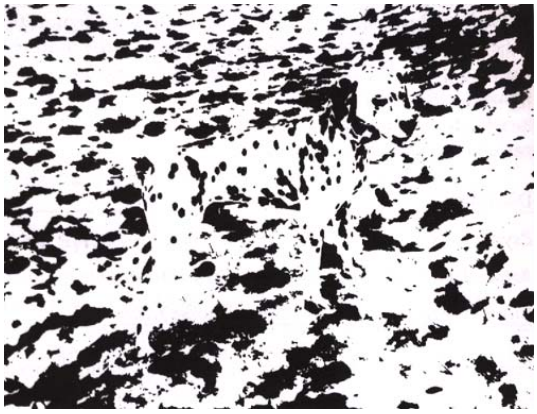

Figure 1.3 - *Dalmatien - Exemple de séparation d'un objet familier avec un arrière plan complexe.*

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

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

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

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

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

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

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

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

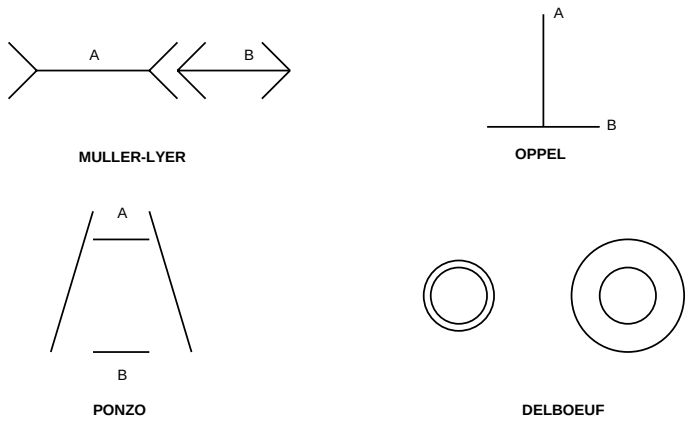

Figure 1.5 - *Illusions géométriques - Dans chaque cas, les segments A et B ont la même longueur. De même, les cercles intérieurs ont le même rayon.*

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

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

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

1.2.2.5 Perception immédiate

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

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

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

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

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

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

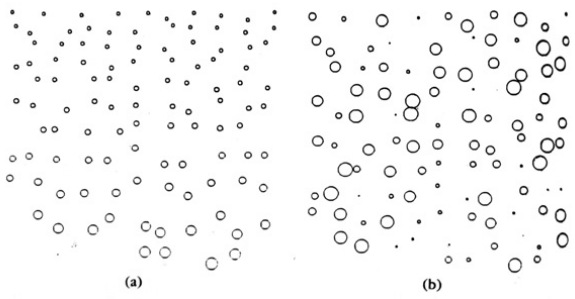

Figure 1.6 - *La présence d'un gradient de texture est un indice immédiatement utilisable concernant l'orientation de la surface. L'exemple (a) semble s'éloigner alors que (b) semble parallèle à l'observateur.*

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

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

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

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

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

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

1.2.3 Théories neuro-physiologistes

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

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

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

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

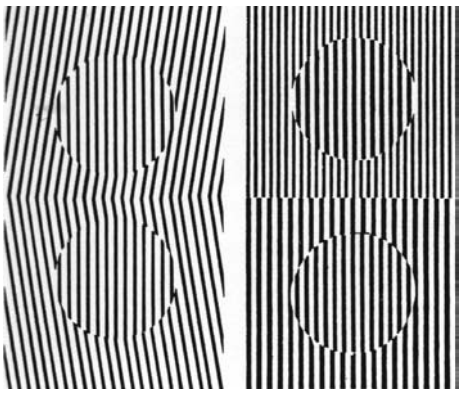

Figure 1.7 - *Exemple de modification apparente de la fréquence spatiale par le système visuel. Dans l'exemple de gauche, les disques ont la même orientation verticale - l'orientation de l'arrière plan semble dévier les disques dans le sens opposé. Dans le cas de droite, les disques ont la même fréquence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

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

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

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

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

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

d'attention visuelle de l'école Empiriste.

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

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

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

1.3 Méthodologies de la vision par ordinateur

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

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

1.3.1 Paradigme reconstructif

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

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

1.3.1.1 Processus descriptif et paradigme de Marr

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

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

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

- *Niveau calculatoire*

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

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

- *Représentation et algorithme*

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

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

– *Implémentation physique*

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

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

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

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

– *L'image*

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

– *Ebauche primaire - "Primal sketch"*

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

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

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

– *Représentation 3-D*

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

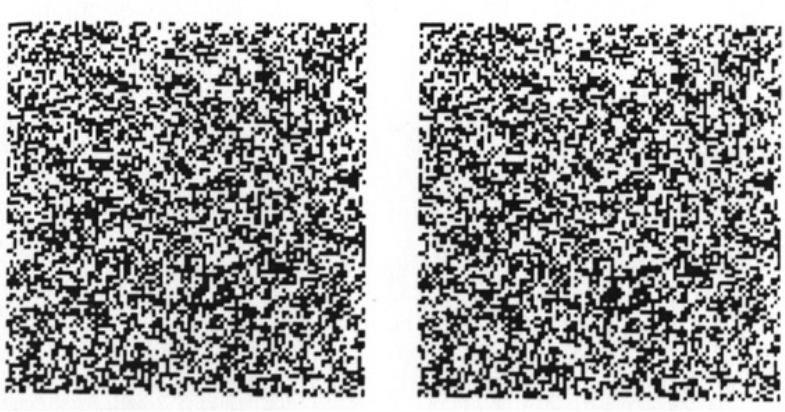

Figure 1.8 - *Deux stéréo-grammes de points aléatoires. La superposition des deux images, à l'aide d'un stéréoscope par exemple, fait apparaitre la forme d'un carré dont la profondeur est différente du reste de l'image.*

1.3.1.2 Représentation et reconstruction

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

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

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

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

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

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

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

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

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

1.3.1.3 Limites et perspectives

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

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

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

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

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

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

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

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

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

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

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

– L'approche reconstructive est trop figée.

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

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

1.3.2 Approche "intentionnelle"

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

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

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

1.3.2.1 Processus dynamique et interactif

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

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

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

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

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

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

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

1.3.2.2 Vision "dirigée"

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

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

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

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

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

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

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

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

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

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

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

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

1.3.2.3 Intérêt et limites

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

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

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

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

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

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

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

1.3.3 Approche "système"

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

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

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

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

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

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

1.3.3.1 Reconstruction intentionnelle

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

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

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

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

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

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

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

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

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

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

1.3.3.2 Approche "système" de la vision par ordinateur

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

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

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

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

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

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

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

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

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

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

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

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

1.4 Conclusion

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

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

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

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

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

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