

# Chapitre 3

## Groupement perceptuel en vision par ordinateur

---

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

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

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

### 3.1 Principes d'organisation perceptuelle

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

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

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

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

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

– *Champs perceptuels*

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

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

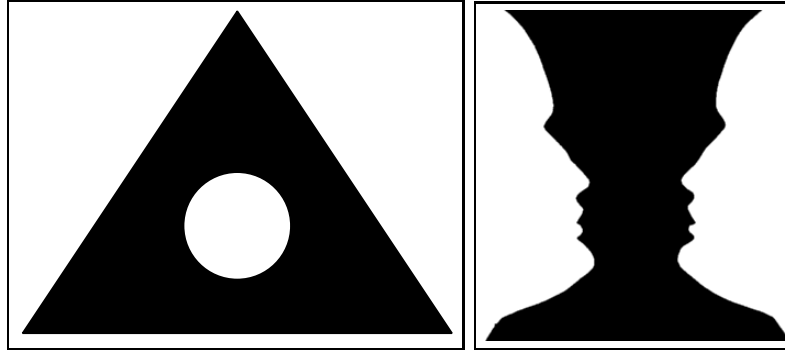
Köhler tenta de rendre compte des phénomènes observés par les Gestaltistes en des termes similaires. Il définit la notion de "champs perceptuel" présents dans le cerveau. Ces champs de force seraient responsables de l'impression de groupement spontané que donnent des éléments visuels présentant des propriétés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

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

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

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

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



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

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

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

– *Prägnanz et Gestaltqualität*

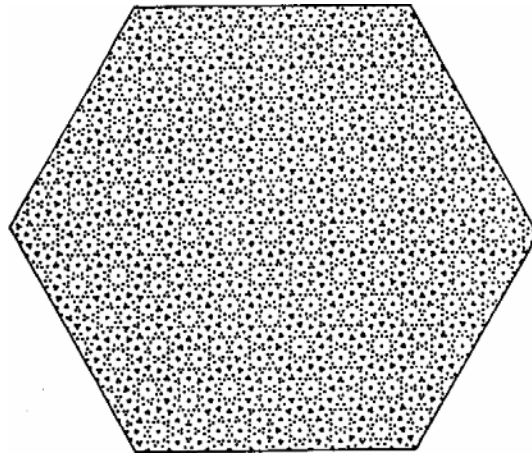
Deux termes ont été définis pour désigner différents aspects de ce principe. D’une part, la “prégnance” ou idée de “bonne forme”, de l’allemand *Prägnanz*. Ce terme rend compte d’une idée de saillance, de mise en valeur de certaines propriétés de façon immédiate, avant tout processus d’interprétation.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

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

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

## 3.2 Règles de groupement

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



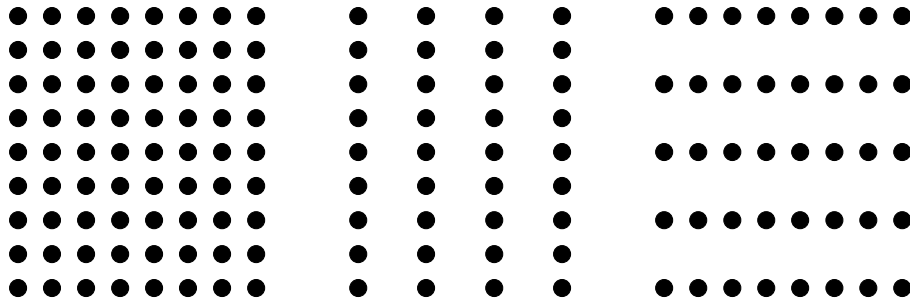
**Figure 3.2** - *Les motifs de Marroquin révèlent le côté dynamique et continu des processus de groupements perceptuels.*

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

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

### – Proximité

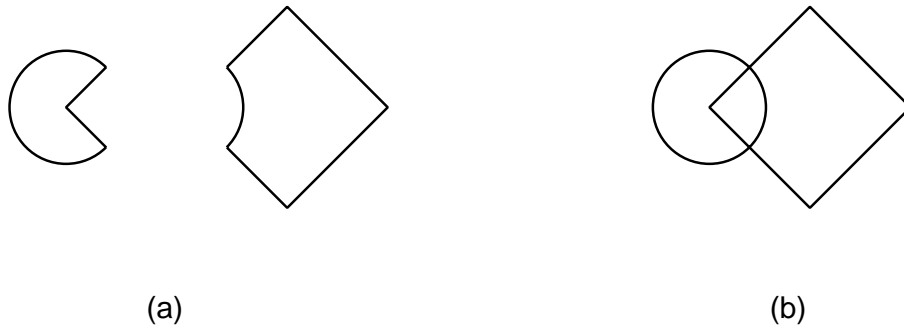
C'est la relation de groupement la plus naturelle. Plus deux éléments visuels sont proches, plus ils ont de chances d'appartenir à un même groupe ou motif. La notion de proximité est cependant soumise à la définition d'une échelle d'observation.



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

### – Continuité

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



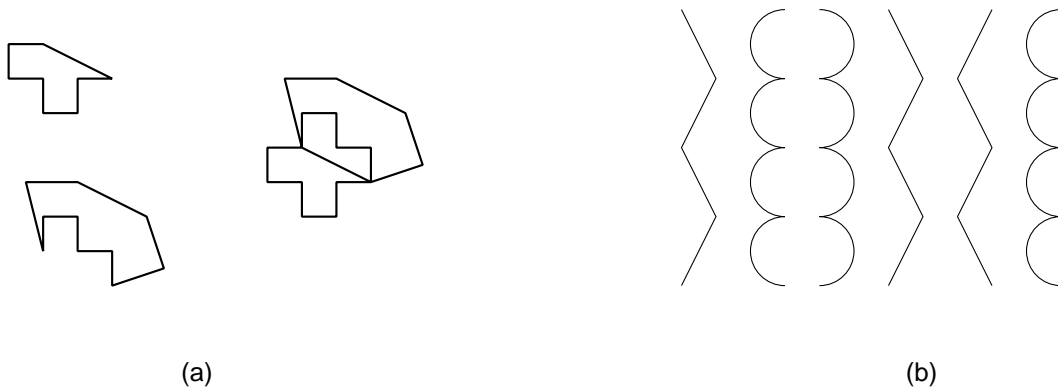
**Figure 3.4** - Groupement par **continuité** - Les figures distinctes de l'exemple (a) s'effacent au profit de figures plus continues dans l'exemple (b). Il est ainsi difficile de voir dans (b) autre chose qu'un cercle complet et un carré complet qui se superposent.

### – Similarité et symétrie

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

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

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

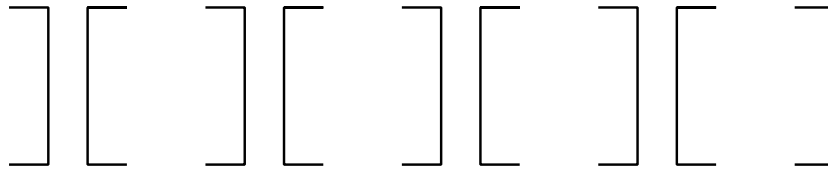


**Figure 3.5** - *Groupement par symétrie* - L'exemple (a) montre l'importance de la symétrie dans l'apparition de formes saillantes. L'exemple (b) montre comment l'importance de la proximité est atténuée par l'existence de symétries.

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

### – Fermeture

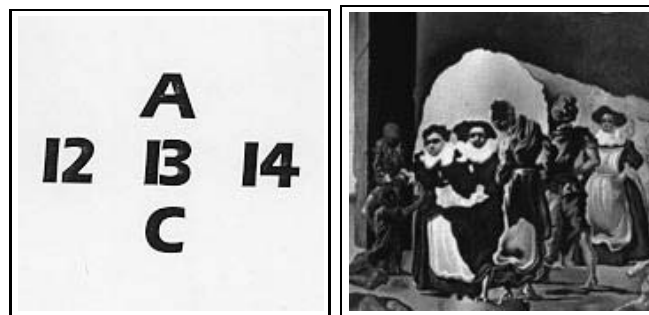
Une grande importance est attribuée à des formes aussi complètes que possible. Les groupements reflètent, de préférence, des formes fermées, convexes, symétriques. La fermeture est l'une des caractéristiques proposées par les Gestaltistes pour distinguer une figure, stable et structurée, par rapport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



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

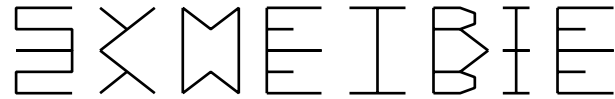
### – Familiarité et contexte

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



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

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

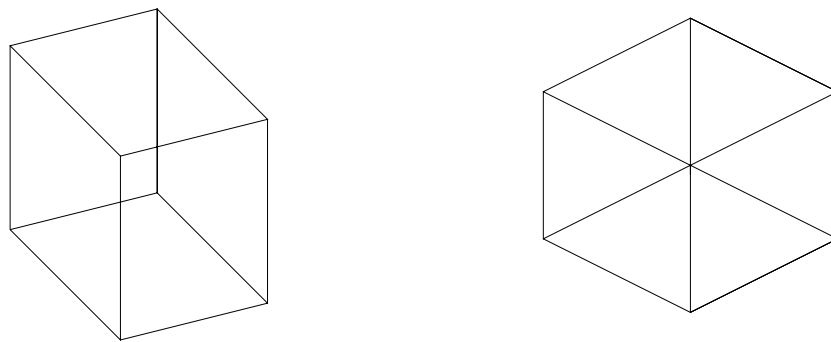


**Figure 3.8** - *Le sens du mot “SYMETRIE” disparaît devant l’influence du groupement par continuité et par symétrie.*

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

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

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



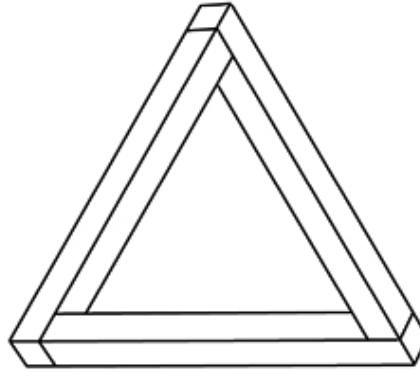
**Figure 3.9** - *Principe de simplicité - En l'absence d'autres indices, la figure de gauche apparaît comme la projection 2D d'un cube 3D alors que celle de droite apparaît comme un motif uniquement 2D.*



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

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

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



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

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

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

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

### 3.3 Application à la vision par ordinateur

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

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

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

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

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

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

### 3.3.1 Principe de “non-accidentalité”

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

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

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

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

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

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

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

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

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

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

–  $P(\text{Causalité})$

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

–  $P(\text{Organisation})$

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

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

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

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

### 3.3.2 Choix des règles de groupements

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

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

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

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

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

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

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

### 3.3.3 Choix des primitives à grouper

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

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

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

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

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

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

### 3.4 Techniques de groupement de contours

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

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

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

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

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

### 3.4.1 Approches algorithmiques

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

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

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

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

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

### 3.4.2 Méthodes d’optimisation

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

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

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

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

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

- Étiquetage et relaxation

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



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

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

– Programmation dynamique

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

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

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

– Optimisation de réseaux de neurones

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

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

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

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

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

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

### 3.4.3 Théorie des graphes

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

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

### 3.4.4 Autres approches

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

– *Classification et indexation*

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

– *Théorie de transformations*

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

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

– *Simplicité de codage*

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

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

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

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

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

### 3.5 Principes de notre méthode

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

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

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

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

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

– *Groupements élémentaires*

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

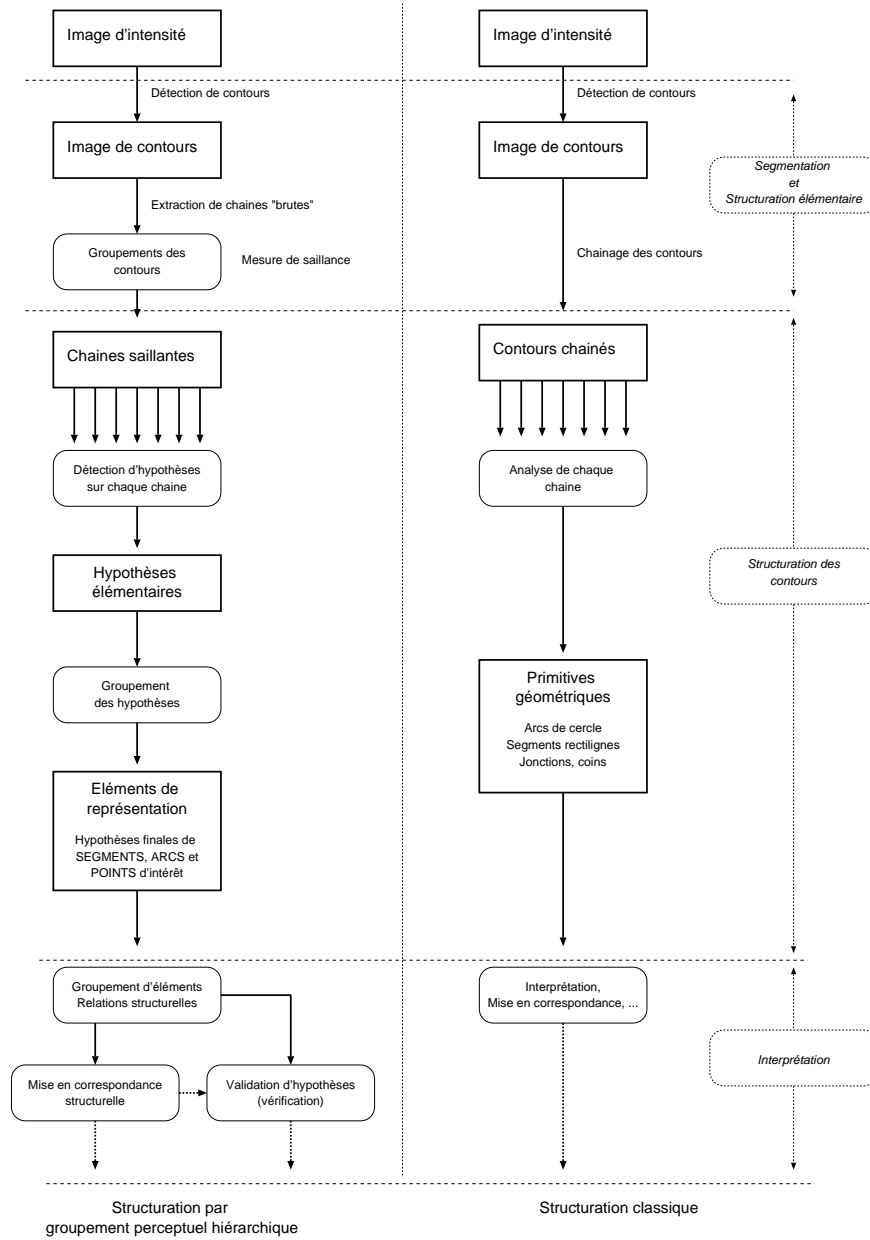
A cette fin, nous choisissons une méthode de groupement par optimisation. Une mesure de saillance structurelle, inspirée des réseaux de saillance de Shashua et Ullman, nous permet en effet de définir des structures d'intérêt pour les niveaux suivants, un peu à la manière des précurseurs sémantiques de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

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

– *Groupements intermédiaires*

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

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



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

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

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

– *Groupements de haut niveau*

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

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

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

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

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

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

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