Utilisation des réseaux de neurones pour la reconnaissance de scènes complexes : simulation d'un système visuel comprenant plusieurs aires corticales

Neural Nets for Complex Scenes Understanding : Simulation of a Visual System with Several Cortical Areas

Philippe GAUSSIER
ENSEA ETIS
GDR 134 « Traitement du signal et Image »
Allée des chênes pourpres
95014 Cergy Pontoise Cedex

Jean-Pierre COCQUEREZ
ENSEA ETIS
GDR 134 « Traitement du signal et Image »
Allée des chênes pourpres
95014 Cergy Pontoise Cedex

RÉSUMÉ

Dans cet article, nous présentons un système général d'interprétation d'images, basé sur des concepts neurobiologiques et psychologiques. L'ensemble des traitements est réalisé à l'aide de réseaux de neurones. Ce système est une sorte de robot simulé capable d'agir dans son environnement afin de reconnaître des objets dépà appris. L'un de ses principaux attraits est qu'il permet une communication simple entre les traitements de haut et de bas niveau. Enfin et surtout, il a été conçu pour montrer que l'on n'a pas besoins d'avoir des régions bien fermées ou des contours parfaits pour réaliser une bonne interprétation. Notre robot est un exemple relativement simple de ce que les réseaux de neurones intégrés à une approche cybernétique permettront de réaliser. Malgré cela, il est déjà intrinsèquement capable de reconnaître plusieurs objets dans une scène complexe même s'ils sont bruités, déformés, ou en partie occultés. Notre travail se veut simplement être une base pour de futurs développements dans ce domaine encore peu exploré par la communauté du traitement des images et du signal.

MOTS CLÉS

Réseaux de neurones, reconnaissance des formes, détection de contours, fermeture de contours, extraction de points caractéristiques, modélisation de cartes corticales.

SUMMARY

Our study tries to combine scattered results in image processing, artificial intelligence, psychology, or neurobiology to improve our understanding of the cerebellar cortex and to realize systems that overstem the limitations of present systems. Our system emulates a little robot with a single eye. Its « brain » has several cortical areas. It is able to learn a given number of objects. We distinguish two sets of neural networks. The first one performs low level processing and extracts characteristic points. The second one processes a state space transformation of the input picture, tries to recognize the learning objects and proposes a reconstruction to confirm the recognition.

Un système de reconnaissance de scènes complexes

During the training, the robot extracts characteristic points of one object and makes a transformation from each of these points. During the interpretation, the robot focuses its eye on a characteristic point, processes a complex logarithmic transformation and performs a mental rotation to match the present transformation with the learned representation. To complete its interpretation or to remove ambiguity, the robot focuses on other characteristic points used during learning. Objects can be recognized in real scene even if they are partially occluded or if the picture is noisy. All the processes are performed by neural nets. The whole systems is made of less than ten layers, that is biologicaly possible.

KEY WORDS

Neural nets, pattern recognition, edge detection, contours closing, extraction of caracteristic points, modelling of cortical areas.

1. Introduction

Notre démarche est liée à la volonté de réunir des résultats épars obtenus en traitement des images, en intelligence artificielle (IA), en psychologie, ou neurobiologie afin de mieux comprendre le fonctionnement du cortex cérébral et de réaliser des systèmes capables de dépasser les limites des systèmes actuels.

Le système étudié est un petit robot doté d'un seul œil qu'il est capable de déplacer de façon à simuler des saccades oculaires. Son but est de reconnaître dans une scène complexe des objets précédemment appris même s'ils sont bruités, en partie occultés ou ont subit une rotation par rapport à la version apprise. Nous avons réalisé l'ensemble de cette chaîne de traitements sous la forme de réseaux de neurones. Elle comprend une extraction de contour, la recherche de points caractéristiques, l'interprétation, la commande des mouvements oculaires et la simulation de rotations mentales. Les différentes représentations d'un objet sont apprises par des réseaux simulant le fonctionnement des différentes aires corticales. Le réseau complet peut être entièrement réalisé avec moins de dix couches consécutives de neurones ce qui permet de considérer notre modèle comme plausible en terme de vitesse par rapport à des neurones réels [Thorpe 88].

Dans la première partie de cet article, nous présentons différents modèles de réseaux de neurones ainsi que leur architecture. Ces modèles sont les outils de base pour simuler notre robot.

Dans la deuxième partie, nous étudions comment des considérations psychologiques issues des travaux sur les illusions d'optiques permettent de déboucher sur un modèle plausible des premières étapes du traitement visuel chez l'homme : extraction de contours, fermeture des contours, et recherche de points d'intérêt.

Enfin nous montrons, dans la troisième partie comment réaliser un dispositif ayant un début de comportement intelligent. Pour cela, nous avons développé un système comprenant des modules de réseaux de neurones configurables : c'est-à-dire qu'il est possible de les dimensionner, de définir leur mode d'apprentissage etc... Ces modules peuvent être associés de différentes manières, ce qui offre une très grande souplesse pour la conception de systèmes complexes.

1.1. APPORT DE L'APPROCHE CONNEXIONNISTE EN ANALYSE D'IMAGES

Actuellement, beaucoup de systèmes d'interprétation d'images comportent des traitements de bas niveau et de haut niveau bien séparés. Schématiquement, les premiers utilisent des techniques de segmentation pour extraire des primitives dans l'image et élaborer une description de l'image, ils sont en général aveugles alors que les seconds emploient des systèmes à base de règles ou à propagation de contraintes pour fournir une interprétation. Dans ces dispositifs, la validité de l'interprétation est fortement tributaire de la qualité des résultats des traitements de bas niveau. Par exemple, un contour non extrait, ou une segmentation donnant localement un mauvais découpage en régions peut induire une mauvaise interprétation. Cela signifie, que les processus de haut niveau doivent pouvoir réagir sur les paramètres de réglage des processus bas niveau. Il se pose alors un problème non trivial de communication entre les processus de différents types. Ceci peut être résolu de différentes manières, par exemple, en utilisant un système bouclé qui segmente à nouveau les régions non interprétées [Gaussier 91] [Cocquerez 92] ou par un système multi-agent [Garbay & Pesty 89].

Ce problème de communication peut être occulté lorsque les dispositifs de bas niveau fournissent les résultats de traitements effectués pour différentes valeurs de paramètres et pour différentes échelles. La difficulté est alors de nature différente car on est souvent confronté à des problèmes de fusion de données. L'approche connexioniste en s'inspirant des modèles neurobiologiques apporte des solutions à ces problèmes :

— Premièrement, les réseaux de neurones, grâce au parallélisme inhérent à leur structure, permettent de briser la combinatoire des systèmes basés sur la programmation logique, la mise en correspondance de graphes, la prédiction et vérification d'hypothèses etc...

— Deuxièmement, les réseaux de neurones apportent une solution globale au problème posé. Ils ont par nature plusieurs niveaux qui communiquent entre eux résolvant structurellement les problèmes de communication évoqués précédemment.

— Troisièmement, ils peuvent fonctionner correctement dans un environnement bruité grâce à la non linéarité des fonctions d'activation et des synapses [Gaussier 92].

— Quatrièmement, dans les modèles neurobiologiques, le

capteur (l'œil) est intimement associé aux mécanismes d'interprétation dans la mesure où il se focalise sur des points d'intérêt grâce à des mouvements oculaires. Ce mécanisme nous paraît fondamental, car il est à l'origine du concept « **d'image mentale** », il permet d'aborder, différemment des méthodes actuelles, les problèmes de multirésolution, d'angle de vue, de perception 3D...

1.2. LES SYSTÈMES EXISTANTS À BASE DE RÉSEAUX DE NEURONES

Le principe généralement employé dans les systèmes actuels consiste à analyser un petit morceau d'image afin d'extraire des caractéristiques locales qui seront ensuite intégrées dans des classes plus générales. Les réseaux employés sont des réseaux à couches du type ascendant (« feed-forward ») avec des rétroactions possibles (« feed-back »), le niveau d'abstraction de chaque couche allant en augmentant de l'entrée (pixels) vers la sortie (symboles).

La reconnaissance de caractères est, sans nul doute, l'application la plus développée utilisant des réseaux de neurones. Différentes techniques ont été testées avec des succès relatifs pour résoudre ce problème. Le Cun a appliqué la rétropropagation de gradient (RPG) [Le Cun 89], [De Saint Pierre 87]. D'autres ont développé des modèles d'apprentissage particuliers et fabriqué toutes sortes de représentations invariantes par translation, rotation... de l'image de départ ; c'est le cas des études de [Omatu 90], [Lynch & Rayner 89], [Fukushima 88], [Li 90], [Khotanzad & Lu 89], [Widrow 88]. Un système a été développé pour la reconnaissance de caractères arabes manuscrits par [El-Sheiikh & El-Taweel 89], afin de tenir compte de la complexité des différents cas à reconnaître, ces auteurs ont préféré utiliser plusieurs réseaux modulaires allant du général vers le particulier.

Un autre problème très étudié est la reconnaissance d'un objet quelconque dans une scène particulière : chars, avions, yeux, visages, pièces industrielles [Allen 90], [Gupta 90], [Hines 89], [Cottrel & Fleming 90], [Herold 88]. Pour certain de ces travaux, la nature des images (infrarouges, radars) permet d'isoler les objets à identifier du fond grâce à un seuillage binaire utilisant un histogramme, par exemple. Dans tous les cas, à partir du moment où l'objet a été isolé, il est normalisé de façon à remplir au maximum la fenêtre de reconnaissance. Des procédures spéciales réalisent un changement d'échelle et une rotation de l'objet de façon à faciliter son identification. L'apprentissage utilise la RPG ([1]). Les exemples appris sont les différentes cibles possibles avec leur identification. Afin de rendre l'apprentissage tolérant au bruit qui peut être important, les auteurs font aussi apprendre des formes bruitées. La principale limitation de cette méthode est qu'elle nécessite que les objets à reconnaître soient parfaitement séparés du fond, ce qui n'est pas le cas des images complexes telles que les scènes naturelles ou les images aériennes.

Il existe encore beaucoup d'autres domaines d'application qui vont de la détection de textures à la vision stéréoscopique en passant par la segmentation d'objets en mouvement et l'estimation de ce mouvement. Toutes ces réalisations sont en fait des simulations de réseaux de neurones. Il n'existe à ce jour que très peu d'applications implémentées en microélectronique pour la segmentation d'images [Poggio 85], [Mead & Mahowald 88].

Dans ces approches, les réseaux de neurones sont souvent employés comme classifieurs. On rencontre alors le problème inhérent à toute classification : il faut trouver un jeu de paramètres discriminants. Après l'apprentissage de quelques exemples, les résultats produits par les réseaux de neurones sont en général satisfaisants. Cependant, il est difficile ensuite de les améliorer. En effet, face à l'augmentation du nombre des cas à classer, le caractère discriminant des paramètres est plus aigu et le bon dimensionnement du réseau devient crucial. En conclusion, pour ce type d'applications, les réseaux de neurones dégagent l'utilisateur de la partie fastidieuse de la programmation, mais ils ne le dispensent pas d'une bonne analyse du problème.

2. Structure générale d'un système d'interprétation d'images

Sous le terme de réseaux de neurones, on regroupe aujourd'hui un nombre important de modèles essayant d'imiter les fonctionnalités du cerveau en reproduisant certaines de ses structures de base. Le premier modèle proposé fut celui de McCulloch et Pitts en 1943 pour lequel ils étudièrent les opérations logiques effectuées par les neurones. Il existe un nombre important d'ouvrages généraux traitant des réseaux de neurones [McClellan 86], [Kohonen 89], [Khanna 89] et d'articles [Lippmann 87], [proceeding IEEE 90].

Une spécificité importante des R.N.([1]) est l'apprentissage qui consiste à les obliger à réagir d'une manière particulière à un ensemble donné de stimuli. Les algorithmes d'apprentissage sont répartis en deux grandes catégories qui comprennent les algorithmes supervisés (nécessité d'un professeur) et les algorithmes non supervisés (ou auto-organisés).

2.1. PRÉSENTATION DE QUELQUES ALGORITHMES D'APPRENTISSAGE NON SUPERVISÉ

Les algorithmes d'apprentissage présentés dans cette partie sont dit non supervisés [Rumelhart & Zipser 85] [Barlow 89], ils sont sensés être dotés d'un mécanisme d'auto adaptation comme les êtres vivants ayant un système nerveux. Ils nous permettront de définir les caractéristiques que devront posséder les neurones de notre système.

La plupart des algorithmes d'apprentissage sont basés sur la règle de Hebb [Hebb 49]. Il s'agit d'un des premiers mécanismes d'évolution proposé pour les synapses (modification des poids) [Hopfield 82]. Dans ces modèles, le renforcement des synapses a lieu lorsque les neurones pré et post synaptiques sont simultanément activés. Malheureu-

([1]) Rétropropagation du gradient.

([1]) Réseaux de neurones.

sement, la croissance des synapses excitatrices n'est pas limitée, ce qui provoque un risque d'instabilité. Différentes méthodes permettent de stabiliser l'apprentissage du réseau : les inhibitions récurrentes au travers de synapses modifiables [Easton & Gordon 84], le seuillage ou la normalisation des coefficients synaptiques [McClellan 86]. Différents types de structures de réseaux de neurones auto-organisés ont été étudiés [Grossberg 76], [Földiak 89] (réseaux à deux couches avec rétroactions).

Afin de permettre à un réseau utilisant la règle de Hebb de s'auto organiser, on ajoute des liaisions latérales inhibitrices qui permettent de n'avoir plus qu'un seul neurone actif après convergence [Lippman 87]. Si l'on présente un nombre de fois suffisant une forme en entrée du réseau, chaque neurone doit acquérir une sensibilité différente [Grossberg 76] [Grossberg 88] [Kohonen 89]. Ce type de réseaux est appelé « **Winner-take-all** » (le gagnant est le seul à s'activer). Pour éviter un mauvais apprentissage lorsque le réseau se trouve dans un environnement comportant des contradictions, Marshall dans [Marshall 90] [Marshall 89] propose de diminuer les inhibitions entre les neurones lorsqu'il y a ambiguïté afin de permettre à plusieurs sorties d'être actives simultanément. Il s'agit de rendre indépendantes ces sorties. Elles ne seront plus alors corrélées entre elles. Il existe un certain nombre de modèles de R.N. basés sur ce paradigme appelés **competitive learning model** tels que le système ART de Grossberg, les réseaux de Kohonen, ou le Néocognitron de Fukushima.

Les réseaux de Kohonen [Kohonen 89] tiennent compte de l'observation biologique selon laquelle certains neurones ont un rôle spécifique et que des neurones voisins réagissent à des entrées qui se ressemblent [Kohonen 72] a montré que la structure des connexions du réseau devient isomorphe à la structure de l'ensemble des stimuli dans le cas où l'on utilise des liaisons latérales inhibitrices. Il introduit le concept de **cartes topologiques auto adaptatives.** Dans ces cartes, chacun des neurones répond d'une manière spécifique à un type de stimuli mais l'implémentation qu'en a fait Kohonen n'est pas biologiquement plausible du fait que la convergence est longue et que les liaisons changent de type (excitatrices ou inhibitrices).

En revanche, l'idée de départ du modèle ART de Grossberg (Adaptative resonance theorie) [Carpenter & Grossberg 87] est de montrer comment une situation particulière peut accorder les détecteurs de caractéristiques afin de répondre à un ensemble convexe de formes spatiales imposées. Plus précisément, on veut que les détecteurs répondent automatiquement à des caractéristiques moyennes choisies dans un ensemble, même si les caractéristiques moyennes n'ont jamais été étudiées auparavant. La solution consiste en un système dans lequel l'apprentissage est du type descendant (du niveau le plus haut, le plus abstrait vers le niveau le plus bas qui correspond à la couche d'entrée) sur des informations de type ascendant. Ce système permet de protéger ce qui a été précédemment appris d'un effacement par les nouvelles données. Il autorise aussi l'incorporation automatique des nouveautés dans la base de connaissance du système en préservant la consistance de l'ensemble des événements déjà appris.

Pour des opérations de type « automatisme-asservissement », Sutton et Barton [Barto 83] [Barto 81a] [Barto 81b] proposent un modèle dans lequel un seul signal d'erreur améliore le comportement d'un ensemble de neurones. Cette idée est plausible car il existe dans le cerveau des cellules libérant des « médiateurs » chimiques qui influent sur le fonctionnement d'un grand nombre de neurones sans que les neurones soient directement connectés (par voie sanguine par exemple) ; ceci permet, en quelque sorte, au neurone d'avoir une information sur l'état global du système : douleur, plaisir, stress... Le réseau est mono couche mais bouclé. Il est sensé imiter une mémoire associative. La difficulté principale réside dans le fait qu'un neurone particulier ne peut pas savoir s'il est responsable ou non d'un mauvais résultat. L'idée de base est que les neurones vont « moyenner » les signaux d'erreur en fonction des différentes situations et qu'au bout d'un certain temps l'ensemble des neurones réagira correctement.

[Klopf 82] a proposé une théorie sur les bases des comportements adaptatifs intelligents. Dans son modèle, il introduit à la fois la règle de Hebb et le modèle de Sutton-Barto. En fait dans les réseaux de neurones adaptatifs, on distingue grossièrement deux types d'apprentissages. Le premier est du type associatif (règle de Hebb) alors que le deuxième est du type conditionnement Pavlovien, c'est-à-dire un apprentissage par renforcement (Sutton-Barto Klopf). Klopf part de l'idée que les neurones sont un peu comme des individus isolés qui ne cherchent qu'à maximiser leur plaisir (hédonisme). Pour un neurone, le plaisir sera le stade de dépolarisation et la peine le stade d'hyperpolarisation. Le mécanisme de base, mis alors en place dans un tel neurone, a pour but d'obtenir un maximum d'excitations et d'éviter un minimum d'inhibitions.

Le modèle neurone sigma-pi [Durbin 89] correspond a une modification plus fine de la réalité biologique que celle issue des travaux de Mc Cullock & Pitts ou Widrow & Hoff. Il comprend, comme les modèles précédents, une unité de sommation mais les entrées sont des produits d'entrées élémentaires ; d'où son nom neurone « sigma-pi ». Ainsi, le neurone ne réagit plus a un stimulus particulier mais à une combinaison de stimuli [Koch 85]. Ce neurone peut donc reconnaître une forme quelle que soit sa position (invariance par translation) [Giles 87]. Il peut réaliser des fonctions logiques complexes telles qu'un multiplexage ou un aiguillage (lorsque les entrées sont binaires le produit est équivalent à un « et » logique).

D'autres modèles, tels que le néocognitron ou le CMAC, présentent un intérêt particulier à cause de leur architecture. Le modèle du néocognitron proposé [Fukushima 82] est à apprentissage non supervisé. Il est capable de reconnaître correctement une forme donnée même si elle a subi une translation ou si elle a été déformée. Lorsque plusieurs formes sont présentées simultanément sur la même image, le néocognitron, grâce à un mécanisme d'**attention sélective** [Fukushima 88], reconnaît les différentes formes présentes, les unes après les autres.

Le CMAC est une architecture de R.N. issue des travaux d'Albus sur le cervelet. Il est composé de 2 couches d'association (mapping) combinant de façon aléatoire les résultats de la couche d'entrée binaire et d'une couche de

sortie sur laquelle s'effectue l'apprentissage [Miller 87], [Herold 88]. Ce réseau peut accepter des valeurs analogiques qui sont codées en binaire sur la couche d'entrée. La première couche cachée réalise un OU avec des entrées prises au hasard. Cette couche est théoriquement de taille infinie. Dans la pratique, elle est prise assez grande pour que l'ensemble des configurations à stocker puissent s'y retrouver (le choix des entrées est fait au hasard : c'est le concept de mémoire distribuée). La deuxième couche cachée réalise un ET avec des résultats de la première couche pris eux aussi au hasard. Enfin, la dernière couche apprend à retrouver les mêmes résultats grâce à un apprentissage supervisé de type Widrow-Hoff. Les résultats sont meilleurs que ceux obtenus avec un adressage associatif du type « Hash coding ». Mais le problème principal de ce réseau est qu'il est incapable de généraliser à des formes inconnues.
Dans son modèle de la colonne corticale, Burnod propose une architecture et des règles de fonctionnement pour chacun des quatre niveaux d'organisation du cortex cérébral (cellule, colonne, cartes et aires corticales). Ce modèle essaie à la fois de mieux représenter la complexité biologique du cortex cérébral : son organisation en aires et l'existence de mini structures élémentaires (colonnes corticales). Il offre des mécanismes permettant de simplifier les simulations en évitant de prendre en compte chaque neurone, ce qui serait trop long et qui n'est pas toujours bien modélisé, ni expliqué. Cette approche concilie l'aspect apprentissage et l'aspect génération de plans utilisée en IA : « Les apprentissages corticaux ne sont plus de simples associations mais des processus actifs... Chaque nouvel apprentissage produit des déséquilibres générateurs de nouveaux apprentissages. » Un déséquilibre dans le réseau correspond à la génération d'un but. La propagation du déséquilibre entre différentes colonnes a pour conséquence la recherche d'une solution au but proposé, donc elle conduit à la génération d'un plan d'action pour arriver au but choisi [Burnod 87], [Burnod 89], [Alexandre 87], [Dingeon 89], [Otto 90].

2.2. REMARQUES GÉNÉRALES SUR LES PROPRIÉTÉS D'UN NEURONE ET SUR L'ARCHITECTURE DES RÉSEAUX

Toutes les lois d'apprentissage vues précédemment sont intéressantes. L'architecture du réseau doit être bien étudiée, notamment, il faut disposer d'une structure particulière pour que l'apprentissage soit aisé. D'autre part, pour obtenir un comportement « intelligent », le réseau doit pouvoir mémoriser les événements, les dater et focaliser son attention. Enfin, pour contrôler l'apprentissage et fournir un but au système, il faut qu'il y ait un mécanisme d'attribution de récompenses ou de punitions (réalisé par un médiateur chimique). Un réseau de neurones artificiels doit pouvoir disposer de plusieurs lois d'apprentissage (ou d'une loi très générale) comme l'apprentissage associatif et l'apprentissage par renforcement. Il doit posséder une capacité de mémorisation :

— à cour terme (STM : Short Term Memory) à l'aide de liaisons de bouclage
— à long terme (LTM : Long Term Memory) obtenue par modification des poids synaptiques
— définitive quand les poids ne doivent plus être modifiés lorsqu'ils ont donné lieu à une certaine sûreté de fonctionnement ou lorsqu'ils correspondent au codage d'une information importante.

La nature des liaisons est donc aussi fondamentale. Chaque liaison doit pouvoir être « typée », au choix, selon les propriétés suivantes :

— excitatrice, inhibitrice
— poids modifiables suivant un degré de confiance, maturation de la liaison...
— dissociation entre des stimuli conditionnels (CS) et inconditionnels (US)
— simulation de l'accoutumance (dérivation) ou inversement de l'intégration des données dans le temps.

Par ailleurs, les connexions doivent pouvoir s'établir entre des neurones d'un même niveau (couche ou aire corticale) et des neurones de différents niveaux, ainsi :

— des connexions liant des neurones d'une couche de bas niveau à une couche de plus haut niveau autorisent l'existence d'un flux d'informations correspondant à l'extraction de caractéristiques de plus en plus complexes.
— des connexions liant des neurones d'une couche de haut niveau à une couche de bas niveau permettent la remise en cause de décisions prises sur les couches de niveau inférieur.
— des connexions entre neurones d'une même couche (ou groupe) sont nécessaires pour permettre l'auto organisation de la couche en y développant les principes de compétition et de coopération.

Une architecture typique de R.N. appliquée à la simulation de processus mentaux complexes devra comporter des caractéristiques que l'on retrouve dans le néocognitron, le modèle d'Albus, ou encore le modèle de la colonne corticale de Burnod.

En conclusion, l'intelligence d'un réseau est surtout fonction de son architecture. C'est pourquoi nous avons conçu un système composé de différents groupes de neurones comme le système DARWIN de l'équipe d'Edelman [Reeke 90]. Chaque groupe de neurones (cluster) peut obéir à un modèle différent. Au cours d'une simulation, plusieurs modèles peuvent donc cohabiter. Cela permet de rendre compte de la diversité biologique des structures neuronales. L'apprentissage se fait avec des règles locales de modification des poids. Pour différencier les périodes d'apprentissage des périodes de simple utilisation (sans apprentissage), il est nécessaire de disposer d'un mécanisme, agissant globalement sur un grand nombre de neurones, qui simulerait l'état de vigilance, d'attention, de peur, de douleur, ou de bonheur d'un être vivant (communication grâce à des médiateurs chimiques). Enfin, comme les mécanismes de base de la perception sont compliqués et pas toujours bien compris (système chaotique) [Freeman 91], nous serons conduit à utiliser des R.N. dont le comportement ne sera que globalement plausible par rapport au modèle biologique.

3. Application des réseaux de neurones aux traitements de bas-niveau : extraction de primitives ou caractéristiques de bases

3.1. Introduction

Il s'agit dans ce cas précis de simuler certains aspects de la vision naturelle (humaine et animale). Les recherches dans ce domaine utilisent des résultats obtenus aussi bien en psychologie qu'en neurobiologie. La rapidité voire le caractère quasi automatique des premières étapes du traitement visuel chez l'homme sont très difficiles à caractériser parce que les contours que nous percevons ne sont pas forcément corrélés avec les données optiques. Grossberg suggère ainsi que les « contours illusoires » ont un effet important dans les processus de reconnaissance. D'autre part, ces contours sont nécessaires pour expliquer la vision des contours continus alors que la rétine est parcourue par des veines qui ont pour effet d'altérer l'image reçue.

La séparation visuelle de régions dans une image est chez l'homme un phénomène complexe qui met en œuvre différentes techniques liées à des paramètres texturaux, de forme, de densité, de couleur... ou bien liées à la mise en correspondance d'éléments sur deux images stéréoscopiques et à l'utilisation de séquences d'images pour isoler correctement un objet. Les travaux de base pour ces aspects sont ceux de [Beck 83], [Beck 85], [Caelli 80], [Glünder 86], [Julesz 81a], [Julesz 81b], [Julesz 86].

On peut remarquer que l'extraction de contours fins est faite très rapidement chez l'homme, elle est dite « **préattentive** », c'est-à-dire faite localement et de manière massivement parallèle. La possibilité d'améliorer le contracte d'une image et de rendre indépendante des variations locales de luminosité (parties d'une image sur ou sous exposée) semble aussi être liée aux toutes premières couches de traitement de l'image. [Skrzypek 90] a ainsi implémenté un réseau du même type que celui de Grossberg afin d'améliorer la qualité des images avant de leurs faire subir d'autres traitements. [Ozawa 90] a simulé un modèle de l'optique de l'œil sur la perception des objets. Il considère que l'œil fixe son attention sur un point particulier $P(i, j)$ de l'image $\mathbf{I}(x, y)$. L'image perçue $F(P)$ est la convolution de l'image réelle par une fonction particulière $h(P)$ représentant la projection de l'image sur la rétine et tenant compte des caractéristiques de cette dernière. La fonction $h(P)$ est différente selon la distance au point de vue : c'est-à-dire si l'on se trouve ou non sur la fovéa. La fonction h est composée de plusieurs gaussiennes simulant les interactions latérales au niveau des premières étapes du traitement visuel.

Un problème important est que la reconnaissance d'un objet est très sensible au contexte dans lequel il se trouve. Depuis les travaux des Gestaltistes [Rock & Palmer 91], il a été démontré que les caractéristiques locales étaient perceptuellement ambiguës mais que leurs combinaisons étaient facilement identifiables (séparation entre les figures ainsi formées ou entre une figure et le fond). Le principe de recherche est différent de ceux d'IA ou du traitement d'images s'appuyant sur des grandeurs physiques : gradient, laplacien, etc... Le mécanisme est ici fondamentalement parallèle et fait intervenir des interactions hiérarchiques (avec un très grand nombre de neurones concernés). Il permet de résoudre d'une façon unique les problèmes liés aux textures, aux images bruitées et à la combinaison des deux.

3.2. SYSTÈMES DE SEGMENTATION PAR RÉSEAUX DE NEURONES

Des considérations biologiques montrent l'importance des contours pour la reconnaissance d'objets. Les opérateurs classiques de détection de contours utilisés en traitement des images sont des systèmes à seuil qu'il est difficile de régler de façon optimale sur l'ensemble de l'image. En conséquence, il subsiste souvent des lacunes dans les frontières d'objets ou bien de faux contours sont proposés. Pour interpréter l'image, on est amené à fermer les contours ou à associer différents contours entre eux. En général, ces traitements sont toujours de « bas niveau ». Comme nous l'avons déjà signalé dans l'introduction (cf. § 1.1), les décisions qui sont prises conditionnent fortement la qualité de l'interprétation. La difficulté majeure réside dans les problèmes de communication entre les processus de haut et de bas niveau.

L'approche neuronale que nous proposons et qui est inspirée des travaux de Grossberg présente les mêmes fonctionnalités qu'en traitement des images, c'est-à-dire qu'il existe des aspects bas niveau et haut niveau. Cependant, la différence fondamentale vient du fait que structurellement les différentes couches de neurones communiquent. Ainsi, il ne faut pas chercher à comparer directement les systèmes connexionnistes d'extraction de primitives à ceux du traitement d'images car l'extraction de caractéristiques n'est pas une fin en soi dans l'approche connexionniste. Elle n'est qu'un traitement opéré à un niveau donné et intégré dans une structure globale d'interprétation où la mise en cause est possible à chaque étape.

Grossberg propose un modèle général pour l'interprétation d'images. Il distingue d'abord un premier niveau de prétraitement des données lié aux neurones photo-sensibles. À partir de ces données l'extracteur de contours (BCS : boundary contour system) fabrique une image des contours qui sert à l'extracteur de primitives (FCS : feature contour system). Ces données sont exploitées par l'interprétation (ORS : objet recognition system) (cf. figure 1).

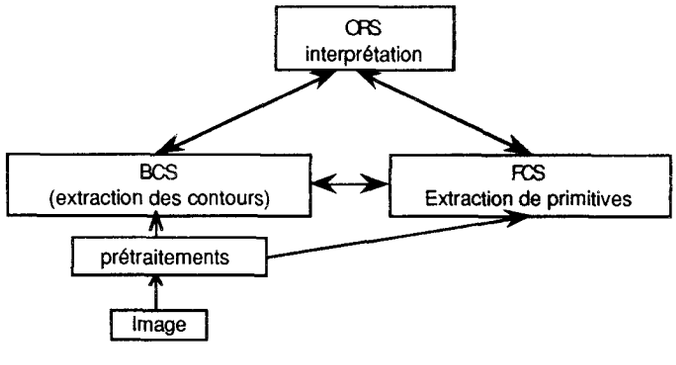


Figure 1. — Organisation du système complet.

Le système BCS comprend deux blocs de traitements. Le premier réalise une compétition entre différents contours possibles dans un voisinage donné alors que le deuxième fait coopérer les résultats fournis par le premier bloc afin de prolonger les contours. Les résultats obtenus sont alors réinjectés vers le premier bloc au travers d'une boucle de rétroaction pour qu'ils soient pris en compte au même titre que des contours vrais (utilisation pour la segmentation de nuage [Lehar 90]).

Les règles de fermeture des contours et le principe d'extraction de contours « fins » par compétition ont été imaginées pour expliquer des illusions d'optiques célèbres [Grossberg & Mingolla 85a], [Grossberg & Mingolla 85b], [Grossberg & Mingolla 87] (cf. figure 2).

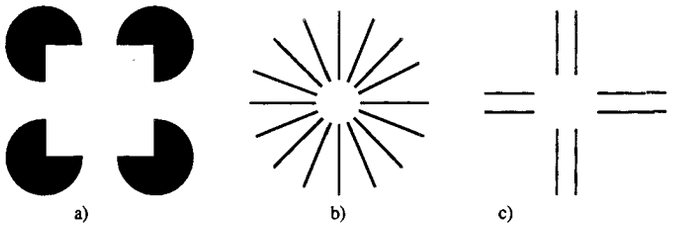


Figure 2 : Exemples d'illusions d'optique.

On voudrait donc que notre système réussisse à fermer les contours même dans les cas de contours illusoires. D'une part, on remarque que nous prolongeons les contours dans la direction perpendiculaire aux fins de lignes des contours réels. Ainsi, sur la figure 2 *b* nous devinons un cercle et sur la figure 2 *c* nous percevons un rectangle au centre de la figure. Les deux étages de compétitions que nous allons proposer créent ces contours illusoires perpendiculaires. D'autre part, dans la figure 2*a* les lignes de contours sont prolongées en ligne droite. Il y aura donc une compétition entre les deux possibilités de fermeture des contours : soit dans le prolongement soit perpendiculairement. La direction choisie pour prolonger les contours ne sera donc pas locale mais globale (localement les deux hypothèses sont formulées, mais ce sont les interactions avec le reste de l'image qui renforce l'une des deux possibilités). C'est donc un niveau supérieur qui favorisera l'une ou l'autre des possibilités de fermeture. La figure 3 représente le schéma général du système d'extraction de contours avec la fermeture préattentive que nous allons développer dans la suite.

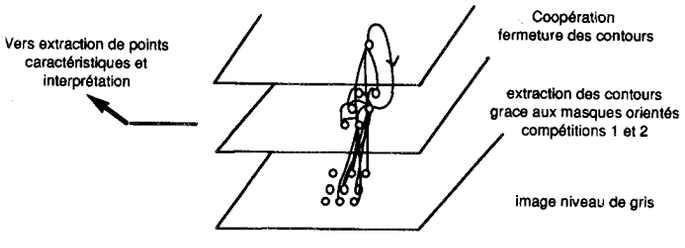


Figure 3. — Les différentes couches du système BCS.

3.2.1. Masques orientés pour l'extraction de contours

Les neurones responsables de la perception des contours ont été bien étudiés par les neurobiologistes [Hubel & Wiesel 68], [Blakemore & Campbell 69], [Sperling 89], [Watt & Morgan 83] et formalisés pour le traitement des images [Marr & Hildreth 80], [Daugman 88] (fonctions de Gabor). Les éléments d'entrée de cette première couche d'extraction de contours sont des masques directionnels simplifiés. Soient (i, j) une position donnée correspondant à un pixel et K une orientation particulière. Soit S_{pq} la valeur d'un pixel perçue après un prétraitement. On définit L_{ijk} et R_{ijk} la partie gauche et droite de la fenêtre F centrée sur (i, j).

$$U_{ijk} = \sum_{(p,\,q) \in L_{ijk}} S_{pq} \qquad V_{ijk} = \sum_{(p,\,q) \in R_{ijk}} S_{pq} .$$

La sortie de cette couche est définie par :

$$J_{ijk} = [U_{ijk} - \alpha \,.\, V_{ijk}]^+ + [V_{ijk} - \alpha \,.\, U_{ijk}]^+ \qquad (1)$$

avec :

$$[x]^+ = \text{Max}\,(x, 0) .$$

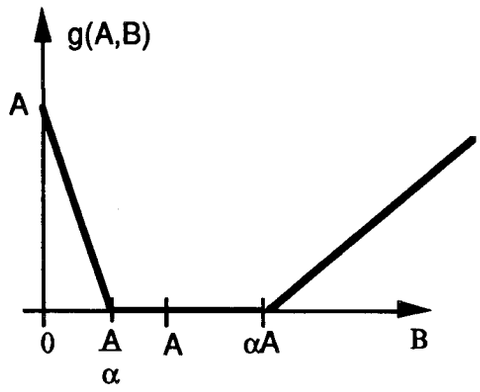


Figure 4. — Représentation de J = g (A, B) en fonction des variations de V = B pour un U = A fixé.

La valeur de α joue le rôle d'un seuil. Si α est grand, on réalise un seuillage important qui va se traduire par une perte d'information définitive.

L'expression (1) conduit à la propriété suivante :

$$J_{ijk} \geqslant 0 \quad \text{si} \quad \frac{\text{Max}\,(U_{ijk}, V_{ijk})}{\text{Min}\,(U_{ijk}, V_{ijk})} > \alpha .$$

Ces opérateurs ne sont pas à proprement parler des extracteurs de contours, mais plutôt des détecteurs de variations de contraste. Ils permettent d'initialiser le processus de formation des contours. Des implémentations ont été proposées par [Allen 90] [Lehar 90]. La méthode que nous avons retenue est celle de Grossberg. Le système a été prévu pour fonctionner en utilisant 4 directions : horizontale, verticale et les deux diagonales. Les 4 masques utilisés sont les suivants :

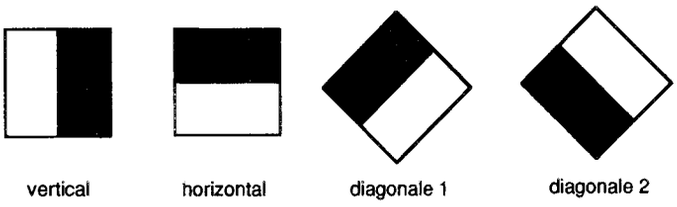


Figure 5. — Masques orientés.

Les dimensions de la fenêtre sont paramétrables. Plus la fenêtre est grande, plus le moyennage est important.

3.2.2. Premier niveau de compétition

Nous illustrons le principe de compétition à partir des masques verticaux et horizontaux. Il en est de même pour les masques en diagonale. La première compétition concerne les neurones représentant une même orientation. Dans un voisinage donné, il ne doit rester qu'un neurone actif lié à une direction donnée, les autres doivent être inhibés. Il s'agit donc d'une compétition entre un neurone et ses voisins. Cette opération porte le nom de « On center-Off surround interaction » et peut être matérialisée par deux neurones : l'un actif lorsque la région centrale est activée et l'autre dans le cas opposé. Chaque J_{ijk} excite un neurone W_{ijk} de la couche supérieure et cherche à inhiber les neurones W_{pqk} de sa couche lorsqu'ils appartiennent à son voisinage. Ce dernier peut être défini par :

$$|p - i|^2 + |q - j|^2 < \text{Seuil} \qquad (2)$$

(cette relation définit un voisinage circulaire).
L'équation différentielle des neurones de la couche W s'écrit :

$$\frac{d}{dt} W_{ijk} = - W_{ijk} + I + f(J_{ijk}) - W_{ijk} \sum_{p,q} f(J_{pqk})\, A_{pqij} \,. \qquad (3)$$

Où :
$\begin{cases} A_{pqij} \text{ est le poids de l'inhibition entre } (p, q) \text{ et } (i, j) \\ f(J_{pqk}) = B \,.\, J_{pqk} \text{ avec B = constante, par exemple .} \end{cases}$

I est une entrée constante permettant la suppression de l'inhibition (désinhibition) lors de la rétroaction (feed-back). Lorsqu'on atteind l'équilibre, on démontre [Grossberg & Mingolla 85b] que les sorties W tendent vers la valeur :

$$W_{ijk} = \frac{I + f(J_{ijk})}{1 + \sum_{p,q} f(J_{pqk})\, A_{pqij}} \,. \qquad (4)$$

On peut prendre $A_{pqij} = A_0$ si (p, q) et (i, j) vérifient (2) et $A_{pqij} = 0$ sinon ; ce qui simplifie l'écriture.
Cette première compétition tend à fournir des contours fins directement à partir d'une image de type « gradient ». Il faut que le voisinage de compétition soit de taille supérieure ou égale à la taille des masques orientés. Par ailleurs, la normalisation des valeurs est nécessaire pour éviter, d'une part une croissance démesurée qui perturberait la compétition et d'autre part une diminution excessive qui conduirait à la disparition de la totalité des contours. La sortie est nulle si elle est inférieure à un seuil. Lorsqu'une direction est désactivée par l'action d'un neurone correspondant à un contour voisin ou à cause de la rétroaction, la direction perpendiculaire pour le même pixel est activée grâce à l'entrée « désinhibitrice » I.

En résumé, dans un voisinage de compétition donné, ayant les mêmes dimensions que la fenêtre, on ne garde que le plus grand des contours horizontaux pris sur la même verticale et réciproquement pour les contours verticaux (*fig.* 6). En pratique, nous avons seulement utilisé les masques horizontaux et verticaux. Les contours sont alors construits en quatre connexité et peuvent être épais lorsqu'ils sont en diagonale. Ce problème n'est pas gênant pour l'interprétation.

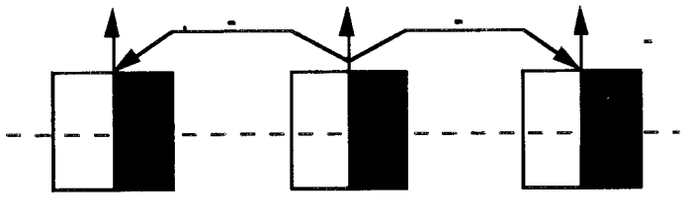


Figure 6. — Compétition de premier niveau pour contours verticaux.

3.2.3. Deuxième compétition

Il s'agit d'une compétition entre des neurones représentant différentes orientations possibles d'un pixel de contour donné (*fig.* 7). Si un neurone attaché à une direction donnée est inhibé, le neurone de la direction perpendiculaire pourra alors être activé faiblement (utile pour la fermeture des contours). Les équations de fonctionnement sont les suivantes :

On notera k une direction donnée et K sa direction perpendiculaire. On fabrique ainsi :

$$X_{ijk} = W_{ijk} - W_{ijK}$$
$$X_{ijK} = W_{ijK} - W_{ijk}$$

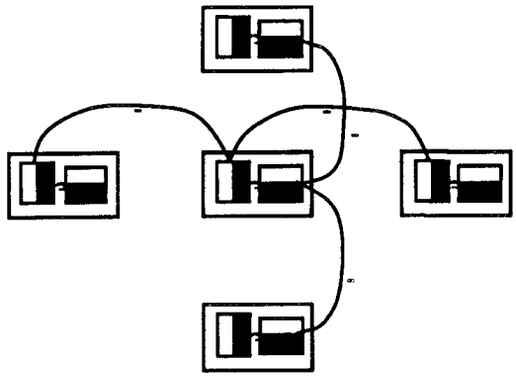


Figure 7. — Schéma général de la compétition entre voisins et entree différentes directions pour un même point.

La sortie du neurone avant normalisation est :

$$O_{ijk} = C \cdot [W_{ijk} - W_{ijK}]^+ \qquad C \text{ constante positive}$$
$$O_{ijK} = C \cdot [W_{ijK} - W_{ijk}]^+ .$$

Toutes ces sorties interagissent pour chaque position au travers de liaisons du type « On center-Off surround », leur potentiel Y_{ijk} est donné par l'équation différentielle :

$$\frac{d}{dt} Y_{ijk} = - DY_{ijk} + I + (E - Y_{ijk}) O_{ijk} - Y_{ijk} \sum_{m \neq k} O_{ijm} .$$

A l'équilibre, on a $\frac{d}{dt} Y_{ijk} = 0$, donc on trouve :

$$W_{ijk} = \frac{E \cdot O_{ijk}}{D + \sum_{m=1} O_{ijm}} .$$

Si D est petit par rapport à $\sum_{m=1} O_{ijm}$ alors : $\sum_{m=1} Y_{ijm} = E$.

On a donc réalisé une opération équivalente à une normalisation de la sortie. Si Y_{ijk} est excité alors Y_{ijK} est bien inhibé.

3.2.4. Coopération orientée

La mise en place de pixels de fermeture se fait grâce à une boucle de rétroaction (*fig.* 8). Elle utilise les contours fins produits par le deuxième niveau de compétition afin de fournir des « possibilités » de fermeture.

Pratiquement, nous avons choisi d'utiliser un neurone pour assurer la fermeture de la manière suivante : si l'on note Z_{ijk} le potentiel d'un neurone de l'étage de coopération. Un de ces neurones sera actif si et seulement s'il reçoit assez d'excitations positives d'entrées alignées du deuxième étage de compétition. C'est-à-dire qu'une lacune pourra être comblée si dans une direction donnée il existe des morceaux de contours qui pourraient se raccorder en formant une droite et si cette droite n'est pas coupée par des points de contours ayant une direction perpendiculaire. Cela peut être traduit par l'équation différentielle suivante :

$$\frac{d}{dt} Z_{ijk} = - Z_{ijk} + g\left(\sum_{p,q,r} [f(Y_{p,q,r}) - f(Y_{p,q,R})] F^{r,k}_{p,q,i,j} \right) + g\left(\sum_{p,q,r} [f(Y_{p,q,r}) - f(Y_{p,q,R})] G^{r,k}_{p,q,i,j} \right)$$

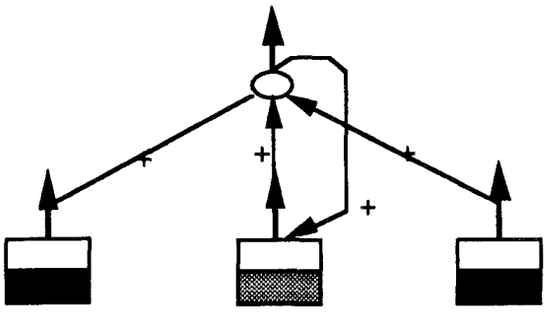


Figure 8. — Boucle de coopération, schéma de principe.

— k est la direction associée au neurone Z_{ijk} étudié sur la couche de coopération

— r est la direction du neurone $Y_{p,q,r}$ de la deuxième couche de compétition

— g est une fonction de seuillage, par exemple : $g(s) = \frac{H \cdot [S]^+}{K + [S]^+}$

— R est la direction perpendiculaire à r

— $F^{r,k}_{p,q,i,j}$ et $G^{r,k}_{p,q,i,j}$ sont des fonctions donnant les poids des liaisons entre le neurone lié au pixel candidat à la fermeture et les neurones de son voisinage dans la deuxième couche de compétition

— f est la fonction de sortie des neurones du deuxième étage de compétition.

La prise en compte de contours alignés dans une certaine direction peut être faite en utilisant les sorties de neurones pris dans le voisinage dessiné figure 9.

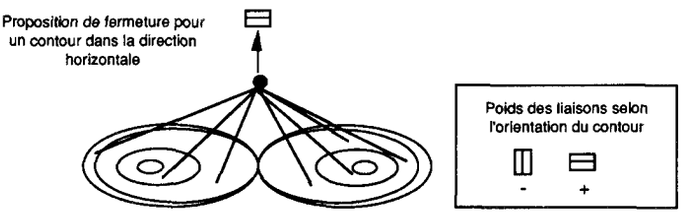


Figure 9. — Forme du masque pour le prolongement (lignes de niveaux d'intensité des poids synaptiques).

Les neurones correspondant à la même direction que le point de contour proposé par le neurone de la couche de coopération ont une action excitatrice. Inversement, les neurones correspondant à la direction perpendiculaire ont une action inhibitrice maximale. On peut représenter les coefficients d'entrée des neurones de la couche de coopération par les équations suivantes :

$$F^{r,k}_{p,q,i,j} =$$
$$= \left[e^{-2\left(\frac{N_{pqij}}{P} - 1\right)^2} \cdot \left|\cos (Q_{pqij} - r)\right|^R \times \right.$$
$$\left. \times [\cos (Q_{pqij} - k)]^T \right]^+$$
$$G^{r,k}_{p,q,i,j} =$$
$$= \left[- e^{-2\left(\frac{N_{pqij}}{P} - 1\right)^2} \cdot \left|\cos (Q_{pqij} - r)\right|^R \times \right.$$
$$\left. \times [\cos (Q_{pqij} - k)]^T \right]^+ .$$

Avec : R et T sont des entiers positifs impairs, et

$N_{pqij} = \sqrt{(p-i)^2 + (q-j)^2}$:

distance par rapport au point d'influence

$Q_{pqij} = \arctan\left(\frac{q-j}{p-i}\right)$:

direction pour contour possible.

La largeur et la longueur de la zone de coopération effective sont fixées par les paramètres R, T et P.

3.2.5. Rétroaction

Les sorties des neurones de fermeture sont renvoyées en entrée des neurones du premier niveau de compétition. Elles renforcent l'activité des neurones de la première couche attachés aux pixels candidats à la fermeture. Cette influence se fait dans le cadre de chaque orientation et met en œuvre un mécanisme du type « On-Center, Off-surround ». Les potentiels résultant V_{ijk} vérifient l'équation :

$$\frac{d}{dt} V_{ijk} = - V_{ijk} + h(Z_{ijk}) - V_{ijk} \sum_{p,q} h(Z_{pqk}) \cdot W_{pqij}$$

avec $h(z) = L\,[z - M]^+$ une fonction de seuillage : $h(z) = \begin{cases} 0 \text{ si } z < M \\ z - M \text{ sinon .} \end{cases}$

L est une constante.

Les W_{pqij} servent de constantes de la même manière que les A_{pqij}. Le signal V_{ijk} sert alors d'entrée à W_{ijk}. L'expression de W_{ijk} se transforme alors en :

$$W_{ijk} = \frac{I + B \cdot J_{ijk} + V_{ijk}}{1 + B \cdot \sum_{p,q} J_{pqk} A_{pqij}} .$$

3.2.6. Résultats

Nous avons testé les opérations de bas niveau sur une image aérienne de zone urbaine sur laquelle nous essayons différentes approches pour l'interprétation. Sa taille est de 256 × 256. Sur la figure 10 *b*, nous présentons les résultats de la deuxième couche de compétition (les masques orientés ont une dimension de 2 × 2). L'existence de petits contours perpendiculaires résulte de la présence de contours d'intensité très élevée qui désinhibent les contours perpendiculaires situés dans le voisinage. En prenant un seuil plus important, on obtient figure 10 *c* des contours proches de ceux que donnerait un détecteur classique ; les contours illusoires ont alors disparu. La figure 10 *d*, présente les résultats de la couche de coopération : contours + propositions de fermeture. On voit que beaucoup de lacunes pourraient être fermées. La figure 10 *e* montre la modification des contours après la prise en compte des propositions de la couche de coopération. On constate que des contours qui n'étaient pas visibles (car on les avait seuillés) le sont maintenant : ils ont donc augmenté d'intensité. Ce mécanisme permet donc d'avoir des seuils locaux. En revanche, certaines des propositions de fermeture n'ont pas abouti car elles se sont retrouvées en concurrence avec d'autres possibilités. La distance pour prolonger les contours était ici assez faible (8 pixels au maximum). Compte tenu de la résolution de l'image, l'utilisation d'une distance plus grande risquerait de créer des prolongements non significatifs. Enfin, la figure 11 *b* et *c*, montrent les résultats obtenus avec une fenêtre plus grande (10 × 10 et 20 × 20). Il ne reste sur cette image que les structures de grandes tailles (bâtiments, bosquets). Il faut noter que prendre un masque de grande taille n'est pas équivalent à

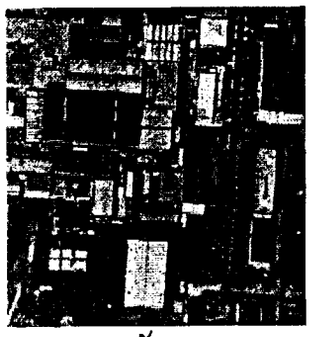
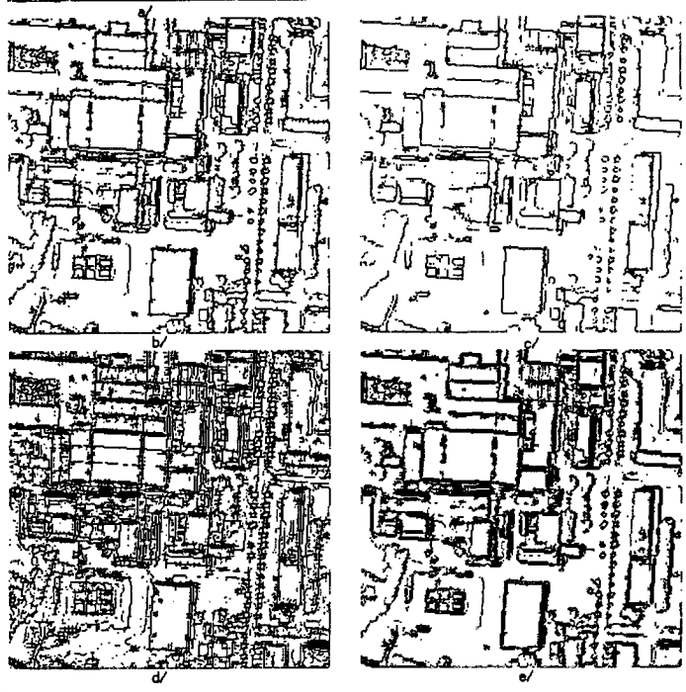


Figure 10. — Résultats du bas niveau.

Image de départ	
Contours	Contours seuillés
Coopération	Contours fermés une itération



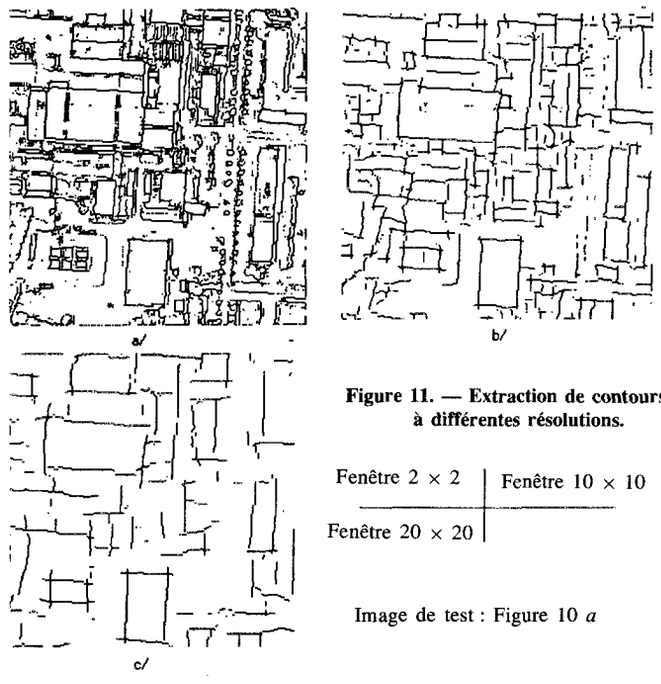


Figure 11. — Extraction de contours à différentes résolutions.

Fenêtre 2 × 2	Fenêtre 10 × 10
Fenêtre 20 × 20	

Image de test : Figure 10 *a*

effectuer un filtrage sur un voisinage important puis à utiliser un masque de petite taille. De façon classique, nous ne risquons pas de délocaliser les contours tant que la dimension des détails est supérieure à la taille de la fenêtre.

3.2.7. Conclusion

L'un des problèmes majeurs de la méthode Grossberg est lié au choix des constantes (coefficient d'amplification pour le retour, taille des fenêtres...) et au grand nombre d'itérations nécessaire à l'obtention de contours réellement fins et à la fermeture des contours (ce nombre ne correspond pas à une réalité biologique). Une solution pourrait être de faire gérer par le système d'interprétation plusieurs résolutions en même temps. Il faudrait donc introduire plus de souplesse dans l'architecture du réseau (le choix des liaisons) et les valeurs des coefficients synaptiques. L'un des intérêts majeurs du bouclage est surtout de permettre un seuillage local adaptatif de l'image (c'est comme si l'on avait un seuil fonction de l'éclairement moyen sur un voisinage de l'image). Enfin, il ne faut pas oublier qu'il ne s'agit que d'une fermeture préattentive des contours donc ne mettant pas en œuvre de mécanisme d'interprétation. Il n'était pas question pour nous de vouloir concurrencer d'une manière quelconque des extracteurs de contours du type Canny, Deriche [Canny 86] [Deriche 90]. Nous voulions seulement proposer une structure à laquelle puisse facilement se raccorder une interprétation et où la remise en cause des décisions prises puisse s'effectuer de manière élégante et naturelle. Le système d'interprétation devra donc être capable de travailler sur des images de contours non fermées et bruitées. Mais la structure du système adopté permettra, une fois l'interprétation faite, de modifier les contours avec la même facilité que la couche de coopération pour la fermeture des contours.

3.3. EXTRACTION DE POINTS CARACTÉRISTIQUES

Il existe de nombreuses méthodes d'extraction de points caractéristiques. On peut détecter les angles en utilisant des masques particuliers ou les apprendre grâce à un apprentissage par compétition [Rumelhart & Zipser 85] fonctionnant sur une fenêtre de l'image et qui apprendrait à reconnaître les différentes configurations possibles (angle, droites...). Notre approche est différente. Elle utilise les conclusions de différents travaux sur la localisation des points anguleux prenant en compte des études faites sur les illusions d'optique [Grossberg 87], [Seibert 89] et [Ozawa 90]. Par exemple, sur le dessin *a* de la figure 12, on a l'impression que les deux flèches ont des longueurs différentes, on perçoit les extrémités des deux segments verticaux à des endroits différents. De la même manière, pour le dessin *b* la ligne oblique qui coupe les deux segments verticaux paraît être composée de deux segments qui ne seraient pas dans le même alignement. On peut ici aussi supposer que la mauvaise localisation des extrémités des segments explique notre erreur.

A partir de cette constatation, plusieurs modèles ont été imaginés pour expliquer ces illusions. Le principe consiste à faire diffuser les contours et à ne garder que les maxima locaux. L'équation différentielle de la diffusion a pour solution une Gaussienne. Un neurone, ayant ses poids fixés

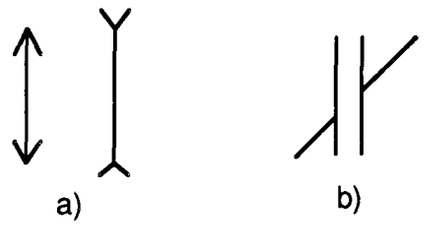


Figure 12. — Illusions d'optique illustrant le mécanisme d'extraction de points caractéristiques.

pour correspondre à une Gaussienne en trois dimensions, donnera ainsi le résultat de la diffusion en un point particulier (masque d'entrée). Les points appartenant à un contour sont les entrées du réseaux mises à 1. Pour tous les autres pixels de l'image l'entrée associée est 0. Si l'image est une droite isolée, les neurones correspondant à cette droite auront tous la même valeur. Après la compétition, tous les points correspondant à cette droite auront tous la même valeur. Après la compétition, tous les points correspondant à la droite seront supprimés sauf les 2 extrémités. Lorsque l'image est un angle formé par 2 segments, la partie aiguë de l'angle voit se superposer l'influence de la diffusion des contours des 2 segments. Il apparaît un maximum qui est supérieur aux résultats obtenus pour les points des segments (*fig.* 13). Après la compétition, ce point sera donc conservé. Plus l'angle est aigu, plus sa valeur sera grande. Il sera utilisé comme point de focalisa-

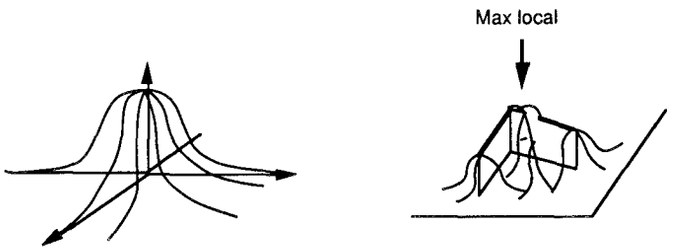


Figure 13. — Illustration de la diffusion avec une Gaussienne.

tion de la rétine dans la quatrième partie de l'article consacrée aux représentations mentales des objets appris et à leur interprétation. La position et la magnitude de ces points d'intérêt dépendent de la taille du domaine de diffusion (plus la diffusion sera importante plus le point caractéristique sera éloigné de l'angle) et de la taille des masques orientés (plus il y aura de contours, plus il y aura de points anguleux et moins ces points seront pertinents).

Il est possible de sélectionner uniquement les points caractéristiques présents à partir d'une résolution spatiale donnée grâce à une modification de la forme du masque précédent (une Gaussienne simple) de manière à ce que jusqu'à une certaine distance dans le voisinage des points de contours servant de support à la diffusion, aucun point caractéristique ne puisse s'activer. Cela revient donc à introduire une zone d'inhibition. Le nouveau masque pourra donc être facilement réalisé à partir de la différence

de 2 Gaussiennes de même centre (ou Laplacien d'une Gaussienne). La forme du masque sera donc celle d'une cellule « center-OFF » ce qui est loin d'être un hasard. Il s'agit d'un chapeau mexicain inversé : inhibition au centre, excitation pour un voisinage plus large.

Enfin, il est plus intéressant d'extraire les points caractéristiques à partir d'une segmentation sur un grand voisinage, donc de faible résolution (ne laissant subsister que les contours les plus importants). On effectue ensuite la reconnaissance sur une image de contours de meilleure résolution (voisinage plus petit) de façon à ce que la reconnaissance soit plus précise. Il nous semble que dans le cerveau plusieurs résolutions doivent être utilisées en même temps d'une part pour la focalisation de l'attention (d'abord vers une zone intéressante, puis vraiment sur le point le mieux placé) et d'autre part pour la reconnaisance (reconnaissance avec des contours obtenus pour des résolutions différentes).

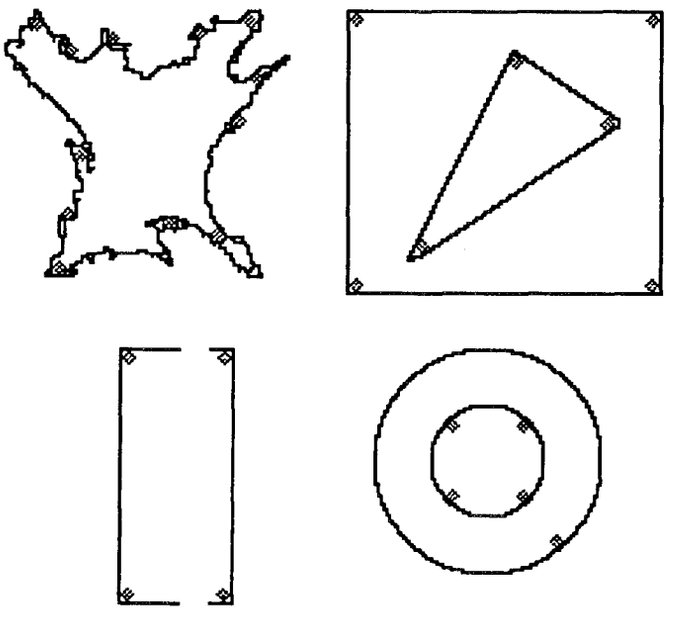


Figure 14. — Exemple de contours et de points caractéristiques.

4. Vision haut niveau, interprétation

4.1. Introduction

Il existe une bibliographie importante d'articles essayant d'expliquer l'organisation ou la fonction des aires corticales [Van Essen & Maunsell 83], [Anderson 83], [Zeki & Shipp 88], [Ballard 86], [Burnod 89] et plus généralement la façon dont le cerveau fabrique un « code » ou une représentation mentale afin de mémoriser ses expériences et résoudre des problèmes particuliers [Marr 69], [Grossberg 80], [Ballard 86]. Nous ne rentrerons pas dans le détail des explications relatives aux différentes aires. Nous supposerons seulement l'existence de deux représentations visuelles comme dans [Zipser & Andersen 88].

Notre approche s'inspire de la « Gestalt théorie ». Nous considérons que l'information pertinente se situe dans les variations d'intensité et que la perception des caractéristiques locales présente des ambiguïtés. La première étape de l'interprétation consiste à faciliter la reconnaissance en utilisant une représentation possédant des propriétés d'invariance. Ensuite, il faut définir les entrées et les sorties nécessaires pour son bon fonctionnement ; elles sont évidemment fonction de la représentation adoptée pour les données. Nous proposons ensuite différentes architectures de complexité croissante autorisant un comportement de plus en plus évolué.

4.2. GÉNÉRALISATION DES MÉTHODES TRADITIONNELLES, FABRICATION D'INVARIANTS

La fabrication d'invariants est un problème clé en reconnaissance des formes. Il s'agit de trouver des primitives peu dépendantes de l'orientation et de l'échelle des objets dans la scène. En effet, lors d'une tentative de reconnaissance, il faut pouvoir isoler l'objet et le remettre, si possible, dans l'orientation utilisée pour l'apprentissage. La plupart des méthodes existantes résolvent le problème d'échelle en normalisant l'image de l'objet pour le faire entrer dans une fenêtre de taille fixe. Dans les cas les plus simples, l'image normalisée subit une rotation pour rendre sa direction principale parallèle à l'axe des ordonnées, par exemple. On peut alors utiliser directement un réseau pour l'identification.

Dans [Lynch & Rayner 89], les contours de l'objet, une fois isolés de la scène, sont utilisés par une transformation rendant le résultat invariant par translation, rotation et changement d'échelle. Le résultat de cette opération est une courbe qui représente la distance du contour au centre de gravité de l'objet pour une position angulaire donnée. L'origine de la courbe est recalée par rapport à la valeur maximale de la courbe. Cette méthode est difficilement utilisable dans le cas où la segmentation est mauvaise. En effet, la position du centre de gravité peut être largement modifiée par la présence de lacunes dans les contours de l'objet. De même, la présence de bruit peut modifier la position du maximum servant au recalage de la courbe. La comparaison avec une courbe de référence peut se révéler alors très difficile.

Il existe plusieurs réalisations permettant de fabriquer des invariants, on peut dénombrer quatre approches de base :

1) Transformer l'image d'entrée jusqu'à ce qu'une correspondance avec une forme apprise soit trouvée. Cette méthode est très coûteuse en temps de calcul mais une combinaison des opérations permet de la rendre rapide [Austin 89].

2) Fabriquer une image invariante par translation, rotation... qui sera alors apprise par le système d'interprétation [Zetzche & Caelli 89]. Cela nécessite de pouvoir bien isoler l'objet à analyser. La transformation peut être faite par un réseau de neurone [Widrow 88] ou par une transformation de Fourier de l'image (invariance par rotation) combinée à une transformation logarithmique (le changement d'échelle devient une translation), ou par une transformation dans l'espace $\{\ln(\rho), \theta\}$ (ρ et θ étant les

coordonnées polaires par rapport à une origine fixée) [Wechsler & Zimmerman 88] [Seibert 89]. La transformation logarithmique accorde une plus grande importance au centre de l'image qu'à la périphérie.

3) Stocker l'ensemble des représentations possibles de l'image pour ensuite les analyser. C'est une méthode très rapide lors de l'exécution mais trop coûteuse en mémoire.

4) Utiliser une technique de reconnaissance par mise en correspondance avec un modèle enregistré. Ces méthodes font appel à la théorie des graphes (recherche d'isomorphismes ou de cliques) ou utilisent la prédiction et vérification d'hypothèses ou encore la relaxation. Cette dernière peut être réalisée avec un R.N. [Hinton & Lang 85], [Feldman 85], [Bienenstock & Van der Malsburg 87] (fabrication d'un modèle invariant par translation, rotation... et mise en correspondance avec des caractéristiques de l'image). La solution employée est la création d'un graphe associant les caractéristiques simples d'un objet à des étiquettes symboliques (voir [Hinton & Lang 85]). Ceci ne correspond pas à un schéma biologique possible car trop de cellules, codant une même notion sur des neurones ayant une localisation différente, seraient nécessaire (de plus leur nombre serait variable suivant le cas traité).

Une autre possibilité consiste à permettre une certaine latitude sur la position de l'objet et à intégrer les caractéristiques dans une structure de plus en plus complexe tout en veillant à ce qu'elle ne soit pas trop contraignante au niveau de la position. C'est ce que fait Fukushima avec son Néocognitron [Fukushima 88]. Cependant, sa solution n'est pas biologiquement acceptable car elle comprend trop de couches (le principe n'en reste pas moins intéressant). On peut aussi essayer d'utiliser la transformée de Hough [Ballard 81] pour reconnaître des formes quelconques. Dans le cas de la reconnaissance des droites, elle a même été implémentée sous forme de réseaux de neurones [Costa & Sandler 89].

4.3. TRANSFORMATION POLAIRE $\{\log(\rho), \theta\}$

Des études sur le nerf optique et la projection des images visuelles dans le cortex cérébral [Schwartz 77], [Schwartz 80] ont montré que le nerf optique effectuait une représentation (mapping) de l'image rétinienne dans un espace $\{\log(\rho), \theta\}$ (ρ et θ en coordonnées polaires par rapport au centre de la fovéa). Elle ne nécessite même pas une couche de neurone pour sa réalisation. Cette méthode a été décrite en détail par [Messner 85], [Caelli & Nagendran 87] et [Wechsler & Zimmerman 88] [Cartenter & Grossberg 87b] pour la reconnaissance des formes.

Les photo-récepteurs de la rétine ne sont pas répartis de façon homogène et cartésienne comme c'est le cas dans les matrices CCD. Il s'ensuit quelques problèmes. L'origine de l'image est choisie en son milieu. La représentation $\{\ln(\rho), \theta\}$ est définie par :

$$\rho = \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

$$\theta = \arctan\left(\frac{y - y_0}{x - x_0}\right)$$

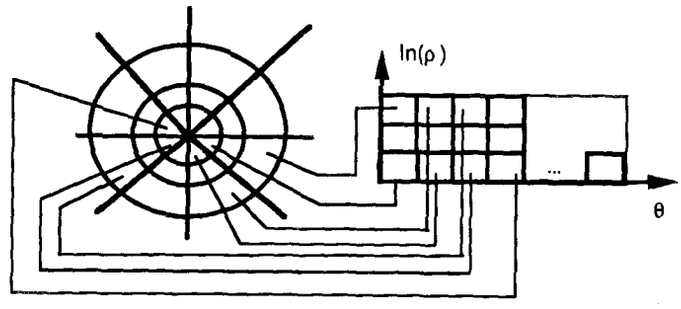


Figure 15. — Transformation $\{\log(\rho), \theta\}$ à partir du point de focalisation.

x_0, y_0 coordonnées de l'origine choisie dans le repère de l'image.

Un exemple de transformation est donné figure 16.

Figure 16. — Transformée log polaire.

a) Transformée de l'image des contours de la clé (origine pt caract.) de la figure 27.

b) Transformée de l'image des contours du cube (origine pt caract.) de la figure 27.

Cette transformation est très sensible aux translations opérées dans l'image de départ. Si l'origine de la transformation est placée en un point particulier par rapport à l'objet à reconnaître (son centre de gravité par exemple), le fait de tourner l'objet par rapport à cette origine ou d'effectuer une homothétie par rapport à ce même centre a pour conséquence de translater l'image résultat en $\{\theta\}$ ou en $\{\rho\}$ selon qu'il s'agit d'une rotation ou d'un changement d'échelle. Si l'origine, qui correspond au point de focalisation, change par rapport à celle utilisée pour l'apprentissage, la reconnaissance n'est plus possible. En effet, l'image résultat est complètement déformée à cause de la nature même de la transformation. Ainsi, le choix du centre de gravité de l'objet comme point de référence n'est pas très judicieux car ce point change de place lorsque l'objet est bruité ou en parti occulté. Comme de plus la transformée $\{\ln(\rho), \theta\}$ donne plus d'importance au voisinage du point de focalisation qu'aux zones plus éloignées, l'erreur est encore amplifiée.

La solution retenue consiste à prendre plusieurs points de focalisation au lieu d'un seul pour permettre la reconnaissance de l'objet lorsqu'il est en partie occulté ou lorsqu'il correspond à plusieurs régions dans l'image. Ces derniers

sont choisis parmi les points caractéristiques extraits avec la méthode du 3.3. Ainsi, si un point caractéristique n'est pas trouvé à cause du bruit ou simplement parce qu'il n'est pas visible, les points caractéristiques présents pourront suffirent à reconnaître l'objet (s'ils sont assez nombreux). Le choix des points caractéristiques résulte d'un algorithme plausible établi à partir des observations des déplacements de l'œil sur une image. Nous avons choisi les points anguleux de la scène en considérant qu'ils étaient riches en information.

Pour limiter la dépendance vis-à-vis de la position du point de focalisation, nous empêchons le système de voir le voisinage immédiat du point de focalisation (zone aveugle au centre de la vision, voir *fig.* 17). On utilise une translation $\{\ln(\rho/a), \theta\}$ où « a » est un facteur d'échelle. Si $a > 1$, le centre de l'image est alors perdu, car l'espace de représentation ne comprend que des coordonnées positives et $\ln(\rho/a) < 0$ pour $\rho < a$. Ce procédé présente l'avantage de réduire l'amplification des détails provoquée par la transformation polaire au voisinage de l'origine.

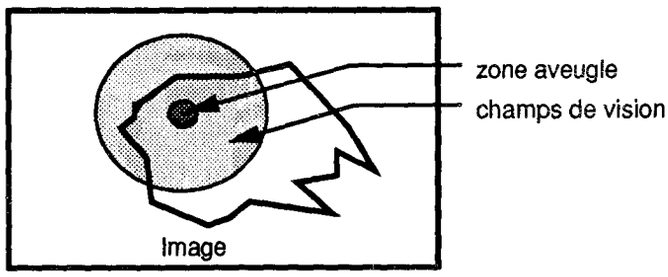


Figure 17. — Champs visuels de notre système.

Remarquons que la focalisation sur un point anguleux supprime l'invariance par changement d'échelle. En réalité, l'image transformée est relativement insensible aux changements d'échelles. Par exemple : soit X le facteur de multiplication d'un objet (rapport d'homothétie), soit α le pas de quantification pour la dimension $\{\log(\rho)\}$, si $\alpha \cdot \log(\rho) < 1$, le déplacement dans l'espace de représentation sera inférieur au pixel. L'image obtenue sera la même que sans le changement d'échelle (la discrétisation est dans ce cas responsable de l'invariance).

Enfin, pour limiter l'influence du bruit, l'image étudiée ne comprend que les contours situés dans le voisinage des points caractéristiques. On peut ainsi chercher les points caractéristiques avec une basse résolution pour ne garder que les plus significatifs par rapport à la taille des objets dans la scène. Les contours situés dans leur voisinage (obtenus avec une meilleure résolution) sont les seuls conservés. Cela empêche l'apprentissage des parties non significatives des objets. De même, lors de l'interprétation, on évite ainsi que l'image en entrée du réseau comporte trop de contours. On filtre donc les contours intéressants. Ce mécanisme peut être réalisé simplement avec des neurones Sigma-PI. Chaque neurone de la couche image résultante a une entrée vers un point de l'image contour de départ et des entrées vers les points caractéristiques de son voisinage. Si le neurone correspondant à l'entrée contour et l'un des neurones correspondant à l'image des points caractéristiques sont actifs alors la sortie du neurone considéré est active (schéma de principe *fig.* 18).

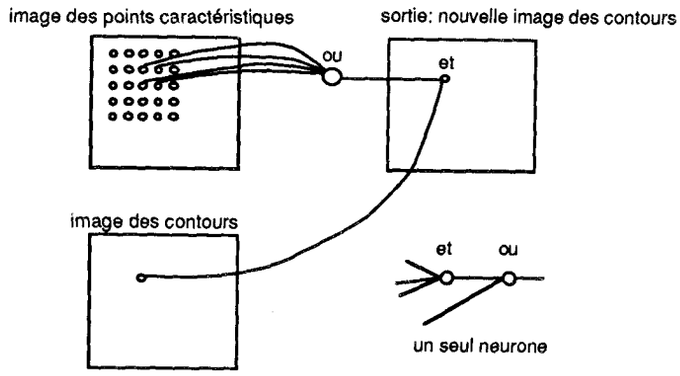


Figure 18. — « Filtrage » de l'image des contours par l'image des points caractéristiques.

4.4. IMAGES MENTALES ET SYSTÈME D'INTERPRÉTATION

Différents travaux mettent en évidence l'existence de plusieurs mécanismes de vision et au moins deux systèmes de représentations visuelles : pariétal et temporal (What & Where) [Jeannerod 74] [Thorpe 88] [Treisman 87] [Burnod 89]. Ils font ressortir l'importance des **saccades oculaires** pour la reconnaissance de scènes complexes [Norton & Stark 71], [Andersen 89]. Les mouvements oculaires se font entre des points caractéristiques choisis en raison de leur contenu informatif. La **focalisation de l'attention** sur un point caractéristique puis son changement vers un autre point caractéristique a été modélisée par [Koch 85] en utilisant des groupes de neurones à apprentissage compétitif (du type Winner-take-all).

D'autre part, nous savons reconnaître des objets ayant subi de faibles translations ou rotations par rapport à leur orientation habituelle mais il ne nous est pas facile d'identifier, par exemple, une lettre ayant subi une rotation trop importante. Cela veut dire que nous sommes obligés d'effectuer une **rotation mentale** qui est un processus beaucoup plus long que l'identification directe [Shepard & Cooper 82], [Farha & Hammond 88]. Lorsque l'on cherche un objet qui est compris dans une liste, il est très difficile de le localiser directement [Elliman 88]. Il nous faut en général parcourir toute la liste. Par exemple, la recherche des lettres *k* dans la chaîne suivante :

*lopilop*M*lzerrhjkcvbera*P*dsdjypeflm*L*nvckht*O*ugxdf*

nous oblige à parcourir toute la liste. Pour reconnaître les détails d'un objet, il est donc nécessaire de focaliser son image sur la fovéa (centre de l'œil). Cette région a une plus grande densité de cellules nerveuses que le reste de la rétine (ce qui implique une meilleure résolution). On peut donc supposer que pour assurer l'invariance en rotation et

en translation il suffise que la zone de focalisation corresponde à la fovéa. L'invariance par changement d'échelle est alors liée à l'optique de l'œil (grossissement variable, variation de la distance focale). Si un objet à analyser est situé hors de la fovéa, le cadrage est obtenu par un mouvement oculaire. Il faut donc un mécanisme de **focalisation de l'attention** sur un endroit particulier.

Enfin, il ne faut pas oublier l'importance des **mouvements oculaires** qui permettent de translater un objet pour le mettre dans une position favorisant sa reconnaissance (celle-ci correspond à la position lors de l'apprentissage). Différents auteurs se sont intéressés au contrôle des mouvements oculaires afin de suivre un objet en mouvement [Marshall 89], [Grossberg 88], [Johnson & Grogan 89], pour centrer un objet dans une fenêtre particulière [Elliman 89] ou enfin pour expliquer certains phénomènes de mémorisation [Grossberg & Marshall 89].

Plusieurs indications d'ordre psychologique permettent de justifier cette approche de la vision. Il apparaît, en effet, que la qualité de l'apprentissage d'une image est proportionnelle au temps de présentation de l'image devant l'œil. Ce temps correspond au parcours d'un nombre de points caractéristiques plus ou moins grand. Lors de la reconnaissance, l'observateur parcourt les points caractéristiques de la même manière que lors de l'apprentissage : le trajet oculaire est donc important pour la reconnaissance de l'objet [Baron 85], [Norton 71].

Ce mécanisme vient se superposer au mécanisme plus connu mais pas forcément mieux compris de l'apprentissage par cœur de l'objet (reconnaissance par corrélation entre l'image apprise et l'image observée). Dans ce cas, l'objet est mémorisé pour un point de vue particulier. L'invariance par translation est assurée par les mouvements oculaires qui doivent ramener le point de focalisation sur l'objet au centre de la fovéa et par des neurones du type Sigma-Pi pour l'invariance par un petit décalage. Une solution pour éliminer les problèmes liés à la rotation est de mémoriser un objet pour une orientation particulière, toutes les autres orientations seront reconnues par un mécanisme de rotation qui tentera de faire correspondre l'image vue à l'image apprise (le mécanisme cherchera à translater l'image en θ dans la représentation polaire).

Si les points caractéristiques sont utilisés comme points de focalisation, la représentation des mouvements oculaires sous forme d'image mentale présentera des phénomènes d'illusions d'optique comme ceux de la figure 12 *a*.

Selon la taille du voisinage utilisé pour l'extraction de points caractéristiques, un cercle peut apparaître comme un objet n'ayant pas de points caractéristiques car la diffusion des contours du cercle se fera de manière uniforme ; il n'y aura pas de maximum local. L'idée pour résoudre ce problème consiste à extraire les points caractéristiques sur l'image résultat (après la transformation log polaire). Quelle que soit la position du cercle par rapport au point de focalisation, le cercle va apparaître comme une ellipse (une droite si l'on focalise au centre du cercle). On peut donc distinguer sur cette ellipse deux points caractéristiques (les deux endroits où la variation de courbure est la plus importante) appartenant à un diamètre du cercle. Il est alors possible de focaliser sur l'un de ces deux points, l'image se transforme alors en un paraboloïde. Ainsi, il serait plus intéressant, pour trouver d'une manière générale des points caractéristiques sur des courbes, de réaliser la recherche de points caractéristiques après avoir effectué la transformée log polaire. Nous n'avons pas choisi cette méthode car, sur les exemples étudiés, ce problème n'apparaît pas et parce que l'image log polaire dégrade beaucoup l'image des contours ; il serait alors nécessaire de travailler avec une précision plus grande (d'où un système plus lent).

4.5. UN MACRO-RÉSEAU DE NEURONES POUR LA RECONNAISSANCE DE FORMES

Notre robot utilise les contours et les points caractéristiques produits par le bas niveau pour reconnaître un objet dans une scène. Les points caractéristiques servent à diriger les saccades visuelles vers des zones intéressantes à mémoriser ou à reconnaître. La transformée $\{\log(\rho), \theta\}$ calculée en un point caractéristique donné est utilisée conjointement avec un dispositif réalisant une translation en θ (équivalent à une rotation). Ainsi, un objet appris peut être reconnu quelle que soit son orientation. Le robot se focalise donc sur chacun des points caractéristiques et opère une transformation $\{\log(\rho), \theta\}$ qu'il cherche à mettre en correspondance avec les vues apprises. S'il échoue pour un point caractéristique donné, il se focalise sur le point caractéristique non encore utilisé qu'il juge le plus intéressant. Quand le robot commence à reconnaître un objet, il utilise le trajet oculaire appris pour confirmer la reconnaissance sur d'autres points de vue. Il y a compétition et coopération entre l'utilisation des points caractéristiques extraits pour diriger les saccades visuelles automatiquement et les sollicitations résultant de l'exploitation des trajets oculaires appris.

Le schéma général du macro-réseau est fourni figure 19. Les différentes parties de ce schéma sont détaillées dans les paragraphes suivants.

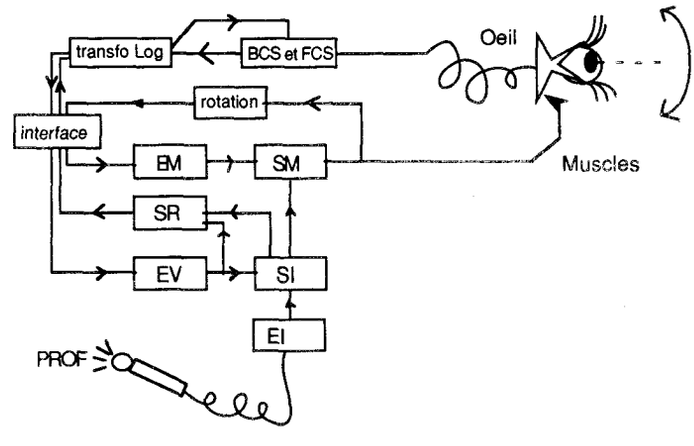


Figure 19. — Schéma général du « robot ».
EM : entrée musculaire, EV : entrée visuelle (image log des contours)
SR : sortie reconstruction pour reconstruire l'image des contours théorique
SI : sortie interprétation d'une vue, SM : sortie musculaire
PROF : utilisateur donnant un numéro à un objet lors de l'apprentissage
BCS et FCS : extraction des contours et des points caractéristiques
l'interface gère les rotations mentales.

4.5.1. Reconnaissance à partir d'un point de focalisation

Dans cette partie, nous mettons en œuvre un système d'interprétation très simple destiné surtout à expliquer le fonctionnement des groupes de neurones servant d'entrées et de sorties. Les groupes de neurones utilisés (à part les groupes d'entrées) fonctionnent selon le principe de compétition mais, il peut arriver qu'aucun neurone dans un groupe ne soit gagnant si l'activité de chaque neurone est inférieure à un certain seuil. Ce seuil dépend d'une constante fixée par l'utilisateur et du niveau d'attention du système. Lors de l'apprentissage le niveau d'attention du système est élevé, le seuil sera donc toujours plus faible que lors de la phase d'utilisation où l'attention est moins importante. Le neurone gagnant modifie ses poids de façon à reconnaître la forme d'entrée selon le même principe que celui des cartes toplogiques de Kohonen. Ici la modification est brutale pour éviter d'une part une convergence lente et d'autre part parce que l'on sait que les déclenchements contraires au but recherché sont impossibles. En effet, l'architecture du réseau et les liaisons fixes assurent la stabilité de l'apprentissage par rapport aux entrées que l'on peut considérer comme une sorte de corrigé.

Cette partie du système (*fig.* 20) est capable de reconnaître N objets à partir d'un certain nombre de points de vue. Chaque objet O_i a été appris à partir de p_i points caractéristiques. L'image des contours de la scène observée subit une transformation $\{\log(\rho), \theta\}$ dont le résultat est quantifié dans un espace discret borné possédant $T_\rho \cdot T_\theta$ cellules associées aux neurones du groupe « entrée vision » (**EV**). En fait, l'image est réduite de façon à correspondre au nombre de neurones de la matrice d'entrée **EV.** Chaque neurone *i* du groupe « sortie interprétation » (**SI**) reçoit en entrée, chacune des $T_\rho \cdot T_\theta$ cellules de **EV** et une entrée du groupe « entrée interprétation » (**EI**) correspondant à l'interprétation associée au point de vue présenté.

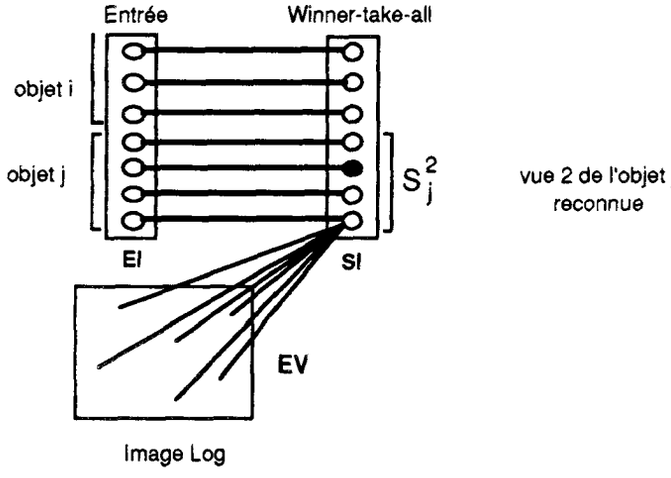


Figure 20. — Schéma synoptique de la partie « vision » du R.N. de la figure 19.

Il y a autant de neurones dans les groupes **EI** et **SI** que de vues différentes à mémoriser. Afin de simplifier la simulation, on impose que chaque objet soit appris de n_v points de vue différents (cf. § 4.5.3). De même, on définit à l'avance le nombre d'objets à apprendre N. Lors de la création du réseau, **EI** et **SI** contiendrons donc $N \cdot n_v$ neurones. Chaque neurone du groupe **EI** est relié à un neurone du groupe **SI** au moyen d'une liaison synaptique non modifiable et représente un stimulus inconditionnel. Le poids de cette liaison est suffisamment grand pour faire toujours gagner la sortie associée lorsque l'utilisateur le désire. Par exemple, le poids d'une liaison inconditionnelle est de 1.1 alors que tous les coefficients modifiables des synapses liées à **EV** sont normalisés de sorte que leur somme fasse 1. Cette façon de procéder évite un apprentissage supervisé du type Widrow-Hoff ou RPG [1]. En effet, chaque groupe peut forcer l'apprentissage d'un autre groupe. Cette architecture de réseau assure donc la stabilité de l'apprentissage. Lorsqu'un neurone de SI est sélectionné, il modifie ses poids de façon à se spécialiser sur l'entrée présentée. L'ensemble de ses poids modifiables va donc mémoriser l'image d'entrée (masque). On réalise ainsi un apprentissage associatif très simple et très rapide. Dans le modèle le plus « brutal » une seule présentation peut suffire pour que l'objet à reconnaître soit mémorisé.

En résumé, le groupe de neurones pour la « sortie interprétation » (**SI**) apprend à associer une interprétation à une configuration visuelle particulière. Les poids synaptiques des liaisons vers l'image $\{\log(\rho), \theta\}$, représentent une vue « mémorisée ». Lors de la reconnaissance, c'est une corrélation entre l'image présentée et l'image « mémorisée » qui s'opère. La sortie la plus active du groupe est alors validée si elle est supérieure à un certain seuil de tolérance pour la reconnaissance.

Un programme d'interface fait le lien entre le simulateur de réseaux de neurones et l'application développée. Il charge les groupes d'entrées du réseau de neurones avec les données fournies en les formattant si nécessaire. Par exemple, pour l'entrée visuelle, l'interface réduit l'image log polaire des contours pour la faire correspondre avec le nombre de neurones dans **EV**(NB_{ev}). (NB_{ev}) est une puissance de 2 d'un entier : $NB_{ev} = \dim 2^2$ (entier). Le rapport de réduction de l'image de départ, supposée elle aussi carrée de dimension dim 1, est défini par dim 2/dim 1. Inversement, pour les sorties du réseau qui représentent des images (§ 4.5.7), une mise à l'échelle en sens inverse est effectuée. De même, l'interface effectue une rotation en sens inverse de celle de l'interprétation afin que les résultats du réseau correspondent à la situation étudiée.

4.5.2. Gestion des mouvements oculaires

Comme nous l'avons indiqué précédemment, un dispositif doit gérer les mouvements oculaires en fonction des positions des points caractéristiques. On définit donc une entrée musculaire (**EM**) qui est une grille bidimensionnelle représentant une quantification de l'espace $\{\log(\rho), \theta\}$ dans lequel on a placé les points caractéristiques [Ballard 86]. Le groupe de sorties musculaires (**SM**) comprend autant de neurones que de cellules dans **EM.** Chaque neurone de **SM** est uniquement relié dans un premier temps

[1] Rétropropagation du gradient.

à la cellule correspondante dans **EV.** Le groupe de sortie **SM** opère une compétition. Le neurone gagnant fournit donc une information topologique sur le prochain point de focalisation. Cette information est traitée par l'interface qui provoque un mouvement oculaire. L'environnement se trouve alors modifié.

Avec le système présenté figure 21, les mouvements oculaires sont non contrôlés. Ils sont effectués en direction du point caractéristique ayant la plus forte intensité. Lorsqu'un point caractéristique a été étudié, il est inhibé pour le reste de l'interprétation (afin de simplifier). La seule sortie active correspond ainsi à un mouvement vers le point caractéristique le plus intéressant qui n'ait pas encore été exploré.

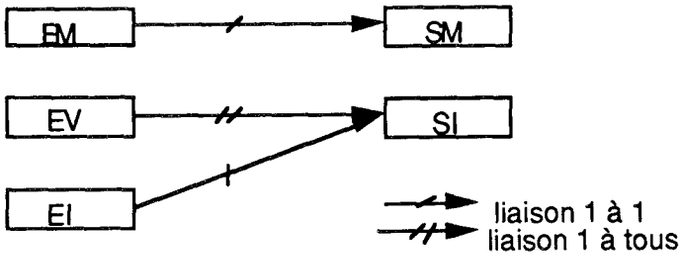


Figure 21. — Schéma du système le plus simple.

4.5.3. Modes de fonctionnement du système : apprentissage et utilisation

Pour l'apprentissage, l'objet à apprendre doit se trouver sur fond uniforme. L'objet peut être une entité composée de plusieurs morceaux disjoints, cela ne gène ni l'apprentissage, ni la reconnaissance. Le nombre de points caractéristiques à apprendre par objet est fixé a priori. On peut interpréter cette contrainte comme une limitation du temps de présentation de chaque objet. Le simulateur fonctionne exactement de la même manière que lors de la phase d'utilisation sauf que l'on suppose que le robot est plus attentif, c'est-à-dire que les seuils de déclenchement des neurones sont abaissés d'un certain facteur. Pour rendre le système plus plausible biologiquement, on pourrait imaginer d'avoir une seule entrée d'interprétation par objet. Le choix du neurone pour l'apprentissage d'un prototype relatif à une vue particulière serait alors commandé par la sortie musculaire. Elle servirait de commande à un démultiplexeur (neurones Sigma-Pi) qui auraient en entrée : l'entrée interprétation générique et forcerait en sortie l'apprentissage d'un neurone particulier.

Pour la phase d'utilisation, on suppose que toutes les liaisons sont définitivement fixées. L'interface charge les différents groupes d'entrées du réseau avec les données relatives à la situation présente. Pour un point de focalisation donné, il fabrique en fait n situations d'entrées qui correspondent aux différentes rotations possibles. Les résultats sont récupérés à chaque fois. Lorsque toutes les possibilités de rotations ont été testées, l'interface effectue le mouvement correspondant au neurone de **SI** ayant obtenu le meilleur taux de succès. Il prend en compte l'éventuelle rotation effectuée. De plus, si une des sorties seuillées de **SI** est active, il inscrit pour le point caractéristique étudié (sur l'image d'interprétation) le numéro de l'objet reconnu. Ce numéro est facilement accessible, il est lié à l'organisation de **SI.** Soit i le numéro du neurone actif. Le numéro de l'objet est la partie entière de (i/n_v).

Un autre mode de fonctionnement consiste à fixer un seuil élevé sur la sortie musculaire de façon à ne déclencher son action qu'en cas de réussite de l'interprétation. Dès qu'un neurone de **SM** est actif, l'interface effectue le mouvement sans essayer les autres possibilités de rotations. Ce procédé a l'inconvénient d'être très sensible aux différents seuils choisis. Dans la pratique, il y a toujours des cas où le mouvement est déclenché alors que l'interprétation est mauvaise, ainsi que l'angle de rotation. Pour que l'algorithme que nous avons utilisé et qui semble être le plus simple, donne de bons résultats, il faut que les différentes possibilités de rotation soient testées de manière séquentielle avec un seuil évoluant dans le temps. Ce dernier est d'abord très grand pour déclencher le mouvement uniquement si la reconnaissance est sûre à 100 %, puis il devient de plus en plus faible pour effectuer malgré tout un mouvement même si aucune interprétation n'est vraiment sûre. Un tel mécanisme impliquerait de recommencer plusieurs fois l'ensemble des rotations possibles ce qui serait très long. Une solution plus réaliste serait d'avoir un réseau mémorisant au fur et à mesure les résultats pour prendre une décision en fonction du temps écoulé.

4.5.4. Réalisation des « rotations mentales »

Le choix de la représentation $\{\log(\rho), \theta\}$ pour la reconnaissance visuelle et pour la gestion des mouvements oculaires permet une mémorisation indépendante de la rotation de l'objet aussi bien pour sa reconnaissance que pour la génération du mouvement oculaire dans la bonne direction une fois l'interprétation faite. En effet, il suffit d'imaginer un système d'aiguillage ou de multiplexage qui effectuerait le décalage de l'ensemble des données musculaires et visuelles à la fois en entrée et en sortie (de manière bidirectionnelle) pour que l'on puisse simuler une rotation mentale et effectuer correctement la reconnaissance et le mouvement musculaire associé (cf. *fig.* 22). Le mécanisme

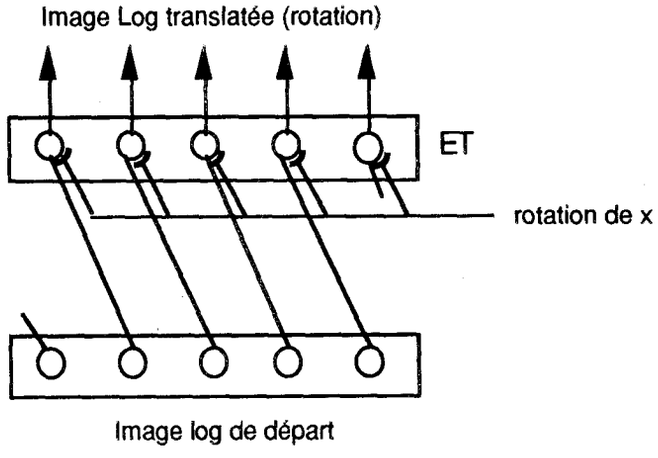


Figure 22. — Système d'aiguillage pour la rotation mentale.

de multiplexage n'est pas implémenté dans notre application sous forme de réseaux de neurones pour des raisons de rapidité et de place mémoire. Néanmoins, il est probable qu'une telle structure puisse exister dans le cerveau [Anderson & Van Essen 87]. On peut facilement la simuler avec des neurones du type sigma-pi (cf. § 2.1).

La figure 23 présente un schéma de l'architecture générale du système. On peut remarquer que l'ensemble des images d'entrée passe par le système de rotation mentale alors que la sortie musculaire passe en sens inverse pour retourner vers le monde extérieur. De même, si l'on introduit une rétroaction pour revenir sur la segmentation, il faut prendre soin que l'image résultat ait bien la même orientation que l'image réelle. Elle doit donc aussi subir une rotation en sens inverse. Le mécanisme de rotation mentale doit donc bien être bidirectionnel. Pour cela, il suffit de dupliquer le mécanisme de la figure 22 et de le mettre en sens inverse.

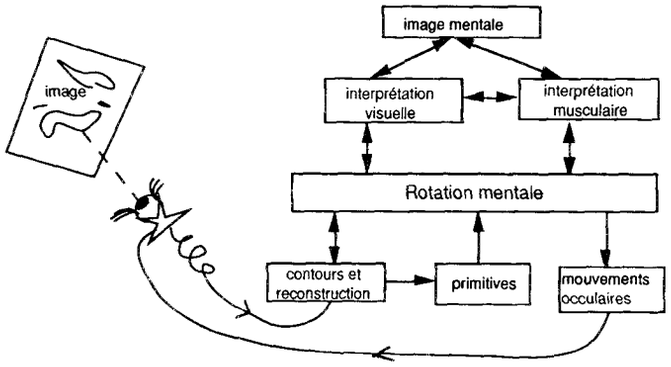


Figure 23. — Découpage hiérarchique du système.

4.5.5. Mécanisme de correction des saccades visuelles

Lorsqu'un mouvement oculaire est proposé par la sortie musculaire, il y a peu de chances pour qu'une fois le mouvement réellement effectué, l'œil se focalise exactement sur le point caractéristique choisi. En effet, le robot peut imaginer trouver un point caractéristique à un endroit donné et se tromper parce qu'il a fait une erreur d'interprétation ou parce que la partie de l'objet cherché est occultée. Une autre cause d'erreur est due à l'imprécision résultant de la quantification de l'espace $\{\log(\rho), \theta\}$ des cartes corticales **EM, SM...** Ce type d'asservissement a été étudié dans de nombreux articles [Grossberg 88], [Elliman 89] (nous ne l'avons pas implémenté avec un réseau de neurones).

Illustrons ces problèmes à l'aide de la figure 24. Si l'on part du point P_1 et que la sortie musculaire propose le mouvement $V(\ln(\rho), \theta)$, l'œil va se retrouver en P_2 qui n'est pas un point caractéristique. La procédure de correction va alors choisir le point caractéristique le plus proche de P_2 pour y amener l'œil, il s'agit de P_3. Si la définition de la carte musculaire n'est pas assez fine, le robot risque de se tromper de point caractéristique et de glisser d'un objet sur un autre sans pouvoir finir la reconnaissance d'un objet. On pourrait ici imaginer ajouter un réseau de neurone qui se chargerait de focaliser sur le point caractéristique le plus

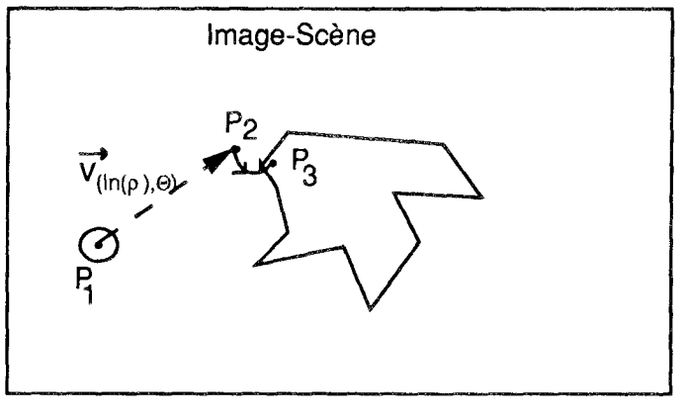


Figure 24. — Tracking visuel vers le point caractéristique choisi.

intéressant situé dans le voisinage du point choisi. Il s'agirait d'un mécanisme de contrôle et de rectification des saccades visuelles.

4.5.6. Vers un système complexe plus proche du modèle biologique

Une première amélioration du dispositif précédent consiste à commander les mouvements oculaires en utilisant les résultats de l'interprétation. Pour cela, il suffit de connecter chaque neurone de **SM** à toutes les sorties de **SI** (liaisons du type 1 vers tous dont les poids sont variables) (cf. *fig.* 25).

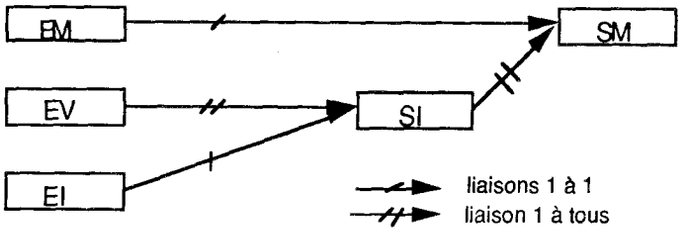


Figure 25. — Système commandant les mouvements oculaires.

Au temps t, on présente au réseau une image **EV**, et on lui propose d'effectuer le mouvement oculaire contenu dans **EM**. **EI** force la sortie et l'apprentissage d'un neurone particulier dans **SI**. **SM** est activé par le neurone actif de **SI** et va donc apprendre à associer le mouvement qu'il devra faire pour l'interprétation présentée sur ses entrées variables. Il faut que le simulateur itère au moins deux fois sur cette configuration pour pouvoir faire les deux associations. Le nombre d'itérations minimal pour avoir une sortie significative correspond au temps minimal pour que les stimuli d'entrée traversent l'ensemble des groupes du réseau. Il est donc au moins égal au nombre de groupes en cascades (ou couches). Dans le cas où l'on met des liaisons récursives, il faut fixer ce temps de façon à ce que l'information ait eu le temps de se propager dans les boucles, il faut alors éviter toute réinitialisation du réseau entre deux exemples à apprendre. En effet, le bouclage peut servir de mémoire à court terme STM (1).

(1) Short Time Memory.

4.5.7. Reconstruction de la forme mémorisée

Une autre amélioration consiste à ajouter au système précédent un nouveau groupe de neurones (**SR**) (cf. *fig.* 26) qui reproduirait l'image idéale apprise. Lorsqu'un objet est reconnu à partir d'un point caractéristique, le groupe **SR** (sortie reconstruction) active les neurones correspondant aux pixels actifs dans l'image de départ. Il réalise ainsi une reconstruction. L'apprentissage de **SR,** dont les neurones obéissent simplement à la règle de Hebb, est dirigé par l'entrée visuelle (log polaire) présente au moment de l'apprentissage. Les poids des connections entre **EV** et **SR** sont fixés définitivement lors de la création du réseau avec une valeur assez basse (ex. : 0,09) de façon à ce qu'ils ne perturbent pas l'utilisation (seuil utilisation = 0,1).

Le seuil vaut 0 pendant la phase d'apprentissage (attention maximale) et augmente pendant l'utilisation (interprétation d'autres images). La variation du seuil est commandée par le paramètre de vigilance qui est grand lors de l'apprentissage et faible lors de l'utilisation.

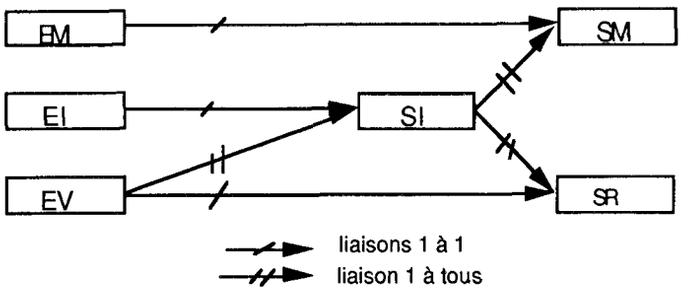


Figure 26. — Système avec reconstruction de l'objet reconnu.

4.5.8. Implémentation

L'ensemble des indications sur les groupes de neurones utilisés (type, nombre de neurones seuils spécifiques), la nature de leur règle d'apprentissage et leur mode de connexion (variables ou non, valeur ..., voisinage ...) est simplement décrit au moyen d'un logiciel graphique de création de macro R.N. développé au laboratoire E.T.I.S.(1). Il suffit de réaliser les schémas des figures 21, 25 et 26 pour que le programme se charge automatiquement de générer l'ensemble du réseau.

L'idée de base est de simuler par des réseaux de neurones tout ce qui serait difficile ou contraignant d'implémenter directement en langage C, toutes les autres fonctions, dont on a déjà démontré qu'elles étaient biologiquement plausibles, sont écrites en langage C chaque fois que l'on peut gagner du temps par rapport à un R.N. Il en est ainsi de la compétition du Winner-take-all, de l'extraction des contours, des points caractéristiques et de la transformation {log (ρ), θ}. De même, une procédure s'occupe de focaliser « l'œil » sur le point caractéristique le plus proche du point proposé puisque cet asservissement a été modélisé et est indépendant du reste de l'application.

(1) Equipe Traitement des Images et du Signal de l'ENSEA.

4.5.9. Résultats

Nous avons expérimenté notre dispositif sur une scène comportant 5 objets (une clé, une cigarette, un cube, un rectangle, et un capuchon de stylo). Seuls les 3 premiers ont été appris (*fig.* 27 *a*, *b*, *c*). Les images sont prises par une caméra CCD et numérisées sous le format 256 × 256 × 8 bits. Ensuite, les deux images de résultats sont réduites au format 128 × 128 en veillant à ce qu'il n'y ait pas de perte de points de contours. Les contours obtenus sont en général assez bruités du fait des conditions d'éclairage (*fig.* 27 *a*1, *b*1, *c*1).

Nous avons limité le nombre de points caractéristiques d'apprentissage (n_v) à 4. Sur les figures 27 *a*1, *b*1, *c*1 nous présentons le parcours de l'œil sur les points caractéristiques sélectionnés pour l'apprentissage. Les figures 27 *a*2, *b*2, *c*2 montrent les vues reconnues lors d'interprétation, une fois les 3 objets appris. On constate que le trajet oculaire est exactement le même et qu'il n'y a pas d'ambiguité. Le robot est bien capable de discerner les 3 × 4 vues apprises. Par ailleurs, notre espace de représentation {log (ρ), θ} est de dimension 32 × 32 (1 024 neurones) pour **EM** et **EV**. Ce qui fait qu'au total l'image a été réduite d'un facteur 8. On constate que ceci n'est pas préjudiciable pour l'interprétation. Comme notre système examine, pour chaque point de focalisation, toutes les possibilités de rotation, on peut considérer qu'il est capable de reconnaître les 12 vues apprises parmi les 12 × 32 vues possibles prenant en compte les différentes valeurs d'angle de rotation. Un rotation élémentaire représente ici un angle de 360/32 = 11.25°. Pour les problèmes d'implémentation et de rapidité, nous insistons sur l'importance de disposer d'opérateurs performants (transformation {log (ρ), θ} avec une zone aveugle au centre de l'image et un champ de vision limité, filtrage de l'image des contours par l'image des points caractéristiques).

Il faut signaler que pour le cube, il conviendrait de l'apprendre pour tous les cas de rotation dans l'espace 3D. Nous ne l'avons pas fait, mais nous avons constaté que de légers décalages ne perturbent pas la reconnaissance. Notre système pourra donc reconnaître des objets en 3D.

Après le test de reconnaissance des objets isolés, le système a été expérimenté sur la scène présentée figure 27 *d*. La clé et la cigarette ont subi une rotation importante par rapport à l'apprentissage. Parmi les deux objets inconnus, le capuchon de stylo peut être pris pour une cigarette. La figure 27 *d*1 représente le trajet oculaire pendant la phase de reconnaissance et la figure 27 *d*2 les interprétations des différents points de vue reconnues. Dans ce cas, il y a un étiquetage du point de vue avec le numéro de l'objet.

Lors de l'interprétation on constate que le rectangle non appris est ignoré (*fig.* 28 *a*). Le capuchon de stylo sur le cube est pris pour une cigarette mais un observateur non averti se trompe aussi (*fig.* 28 *b*). Sur la figure 28 *c*, on observe que les saccades sont bien dirigées de façon à explorer tous les points appris d'un objet avant de passer à un autre. Le cube est reconnu (*fig.* 28*d*) et les 3 objets sont reconnus (*fig.* 28 *e*).

On remarque figure 28 *b* que le robot exploite un point caractéristique qu'il interprète comme faisant parti de la

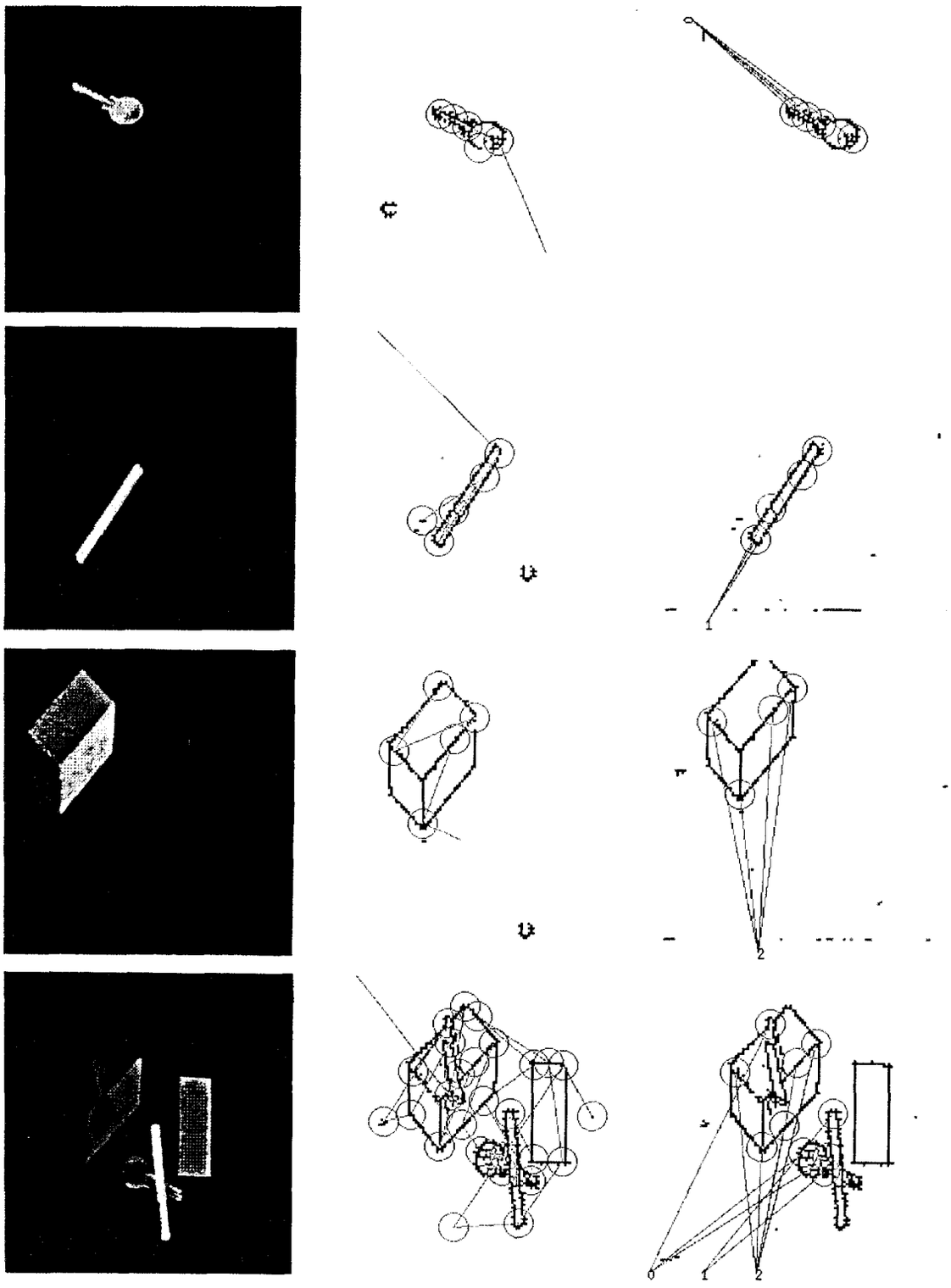


Figure 27. — Visualisation interprétation et chemin musculaire.

Image niveau de gris	Parcours sur les points caractéristiques	Vues apprises et vue reconnue (dernière ligne)
a	a_1	a_2
b	b_1	b_2
c	c_1	c_2
d	d_1	d_2

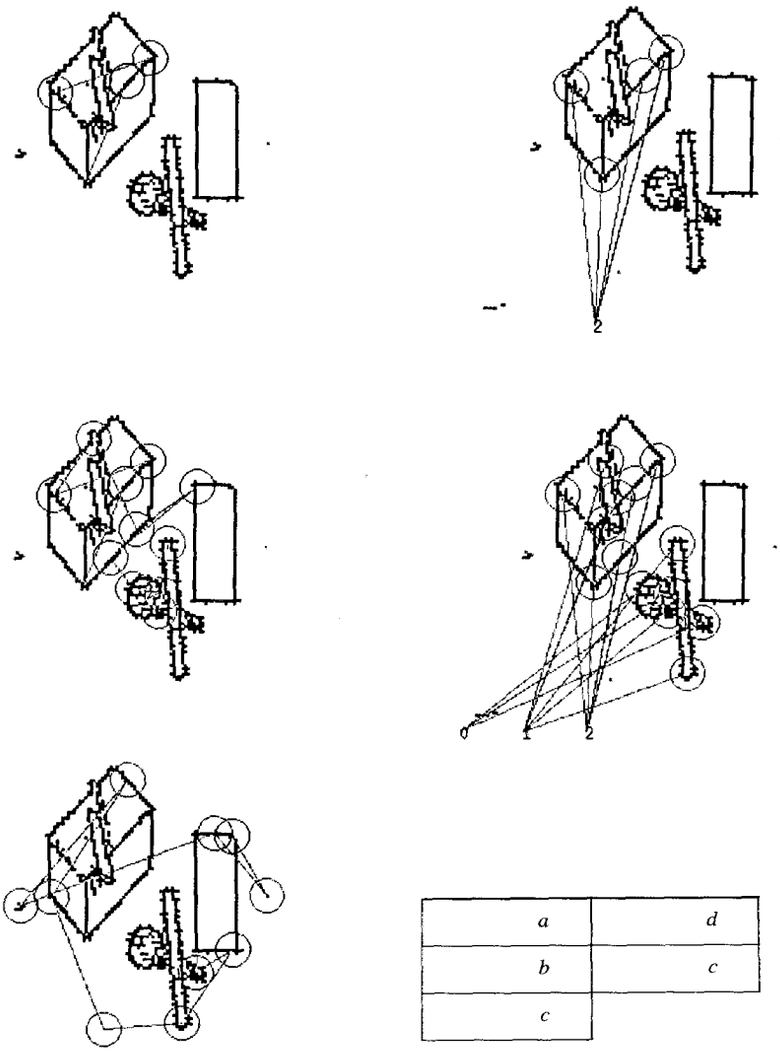
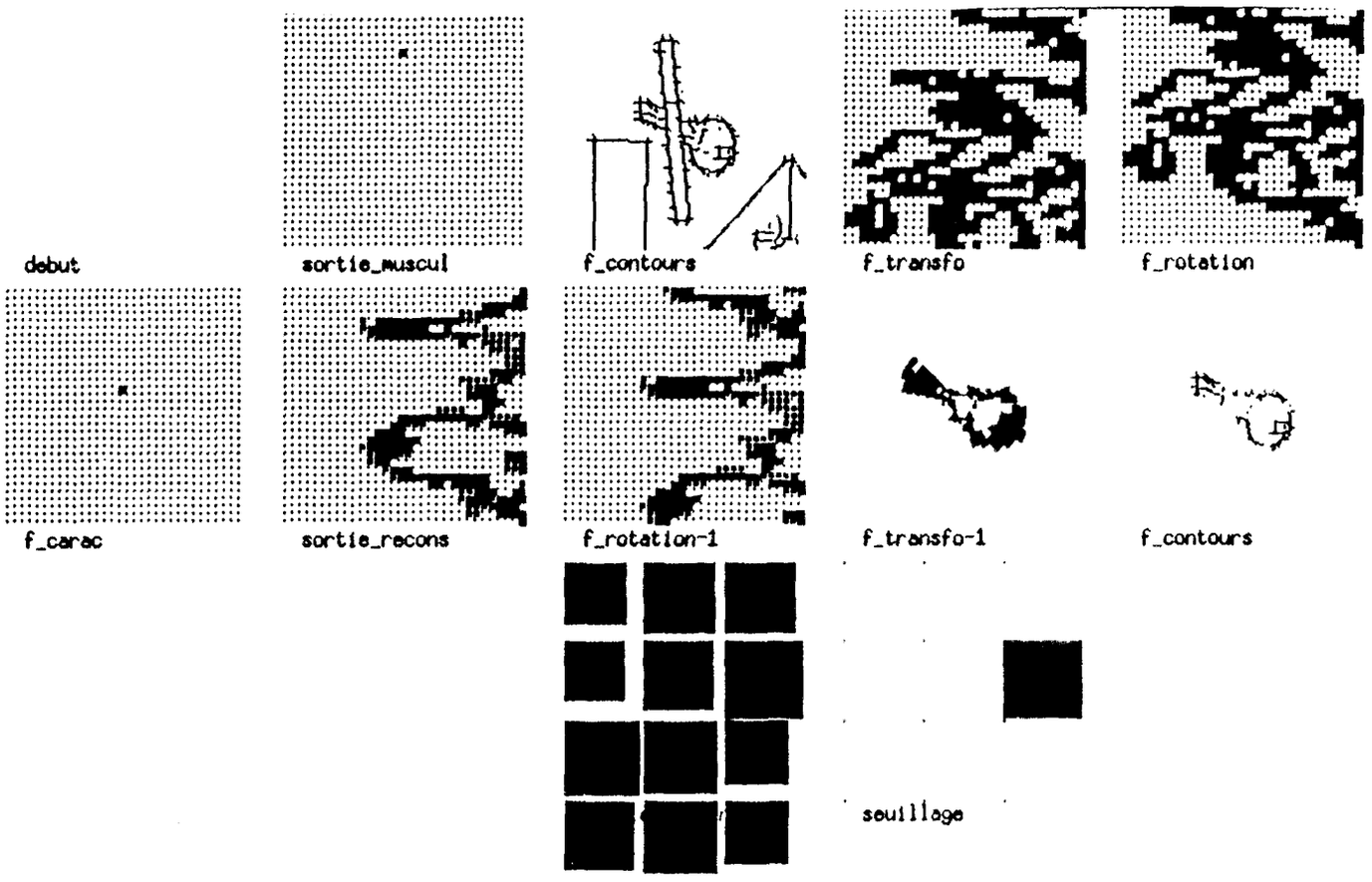


Figure 28. — Différentes étapes de l'exploration de la scène (*fig. 27 d*).
***a*) Le rectangle non appris est ignoré. *b*) Le capuchon de stylo sur le cube est pris pour une cigarette mais un observateur non averti se trompe aussi. *c*) Les saccades sont bien dirigées de façon à explorer tous les points appris d'un objet avant de passer à un autre. *d*) Le cube est reconnu. *e*) 3 objets sont reconnus.**

cigarette. Il continue alors son interprétation sur la cigarette (*fig.* 28 *c*) et finit d'interpréter la clé à la fin. Ce type d'erreur est inévitable, du fait de la simplicité du réseau. Il aurait fallu ajouter une mémorisation de l'objet à partir des mouvements oculaires (voir les futurs développements § 4.5.5). Malgré tout, ces performances sont très intéressantes car le système est capable de rejeter un objet non appris (cas du rectangle). Il est aussi capable de généraliser (cas du capuchon de stylo sur le cube). Enfin, on constate bien que la reconnaissance est toujours possible lorsque des objets sont superposés, ou lorsque les contours sont entachés d'un bruit important.

Les figures 29 et 30 représentent respectivement les différents états des cartes corticales au cours de l'analyse de la clé et du cube dans la scène de la figure 27. Pour la reconstruction, l'un des mécanismes de retroaction permet la sélection des neurones actifs au niveau de la carte contours. Ce procédé est très intéressant car il illustre très bien une collaboration entre les processus de bas et de haut niveau pour la segmentation d'images.

Notre système est donc un bon compromis entre la précision pour obtenir une bonne reconnaissance et la compression de l'information pour avoir un réseau matériellement réalisable et capable de généraliser. Si on analyse figure 29 et 30 les images $\{\log(\rho), \theta\}$ fabriquées, on se rend mieux compte de la difficulté pour obtenir une bonne interprétation et donc de la capacité intrinsèque du modèle de réseau utilisé. Sur cette figure, qui est exploitée pour la mise au point, on peut comparer la sortie musculaire SM proposée par l'entrée musculaire avec celle obtenue après interprétation. La différence est due aux sollicitations des sorties interprétation même si celles ci n'étaient pas assez grande pour activer une interprétation. Si les sorties de SI n'avaient pas réagi sur SM, les mouvements auraient été

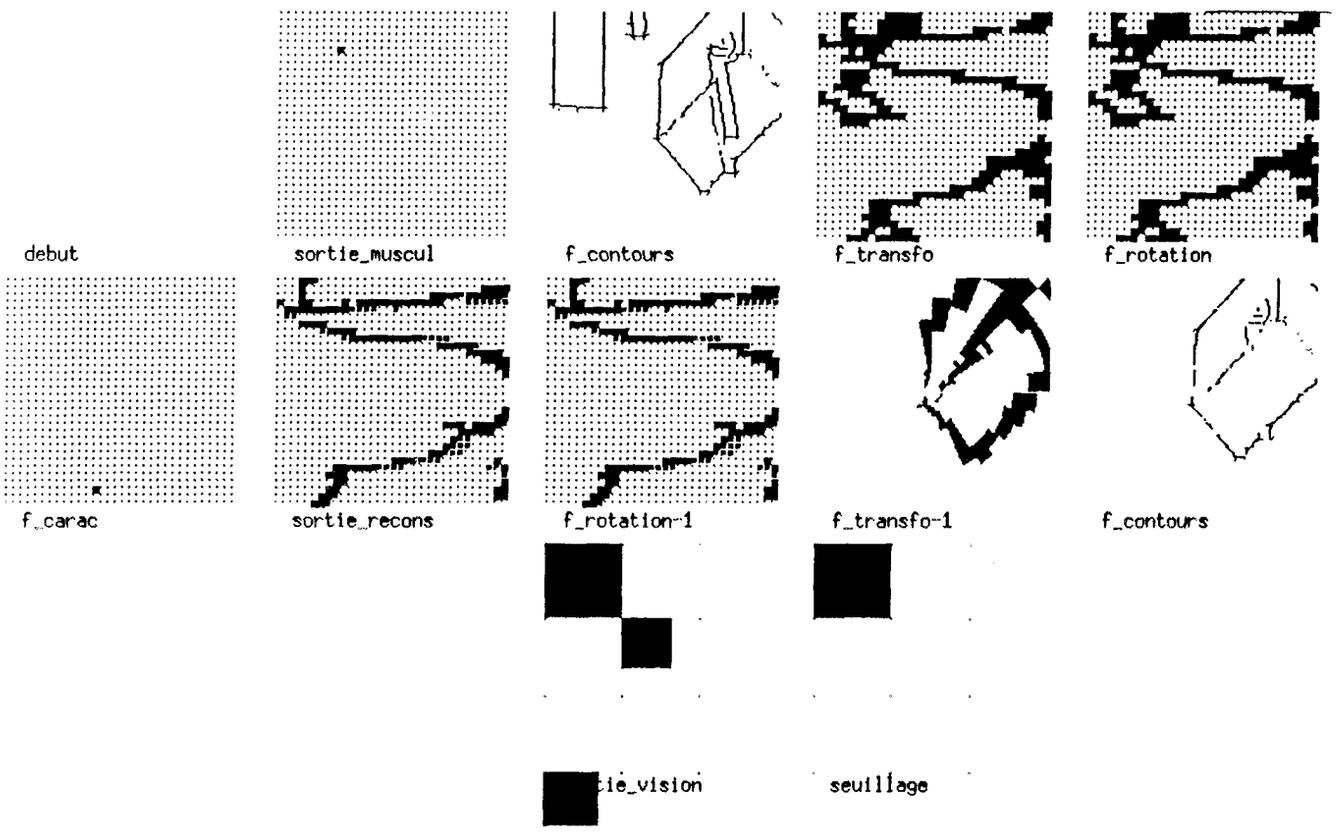


	a	b	c	d
e	f	g	h	i
		j	k	

Figure 29. — Visualisation des différentes cartes corticales du système pour la reconnaissance de la clé de la scène de la figure 27.
***a*) Sortie musculaire EM. *b*) Image des contours vus. *c*) Transformée (log ρ, θ) de l'image *b*). *d*) Entrée visuelle : translation en θ de l'image *c*). *e*) Entrée musculaire. *f*) Image reconstruite en coordonnées (log ρ, θ). *g*) Translation inverse sur l'image *f* pour retrouver l'orientation réelle. *h*) Transformation (log ρ, θ) inverse. *i*) Rehaussement de contraste par rétroaction sur la première couche. *j*) Activité des neurones dans la carte topologique de sortie ; le carré noir le plus grand correspond au neurone gagnant de la compétition. *k*) Neurone gagnant dans la couche de sortie.**

plus désordonnés. Pour déclencher une sortie interprétation à partir de SI, l'objet doit être reconnu au moins à 70 % par rapport à la forme mémorisée.

En conclusion, on constate que le système réussit bien à apprendre chaque vue. Il peut arriver qu'il confonde des vues lorsqu'elles se ressemblent car la transformée {log (ρ), θ} que nous avons implémentée limite le champ de vision à un voisinage assez proche (on a choisi un rayon de 60 pixels pour une image 128 × 128, ce voisinage correspond à peu près à la taille des objets à reconnaître). Elle supprime aussi le centre de l'image vue et elle perd beaucoup d'informations à cause du pas de quantification utilisé (lié au nombre de neurones de connection disponibles). Les mouvements oculaires effectués lors de l'apprentissage ressemblent à ce que les neurobiologistes ont pu observer chez l'homme (il serait facile d'ajouter un sens prioritaire et de parcourir ainsi l'image dans un sens particulier comme le pensait [Norton 71]). Le seul problème est qu'il faille trouver un bon compromis entre la résolution et une vitesse d'exécution acceptable, notamment pour tester toutes les possibilités de rotations. Nous avons choisi de travailler avec des cartes visuelles et musculaires de 1 024 neurones (32*32). Le fait qu'une partie de l'objet soit occultée n'est pas un problème tant que suffisamment de points caractéristiques sont conservés.



	a	b	c	d
e	f	g	h	i
		j	k	

Figure 30. — Visualisation des différentes cartes corticales du système pour la reconnaissance du cube de la scène de la figure 27.
***a*) Sortie musculaire EM. *b*) Image des contours vus. *c*) Transformée (log ρ, θ) de l'image *b*). *d*) Entrée visuelle : translation en θ de l'image *c*). *e*) Entrée musculaire. *f*) Image reconstruite en coordonnées (log ρ, θ). *g*) Translation inverse sur l'image *f* pour retrouver l'orientation réelle. *h*) Transformation (log ρ, θ) inverse. *i*) Rehaussement de contraste par rétroaction sur la première couche. *j*) Activité des neurones dans la carte topologique de sortie ; le carré noir le plus grand correspond au neurone gagnant de la compétition. *k*) Neurone gagnant dans la couche de sortie.**

5. Conclusion

Nous avons réalisé un système de simulation très souple et très évolutif dans lequel chacune des différentes entrées simule en fait une aire corticale. Le système de R.N. doit au moins contenir les groupes de neurones spécialement dédiés à une utilisation précise (**EV, EI, EM, SM, SI**) puisque ces groupes participent à l'interface entre le simulateur de R.N. et l'environnement. On peut de plus créer autant de groupes de neurones que l'on veut obéissant à une règle d'apprentissage particulière. Ce sont ces groupes de neurones et leurs règles d'interconnexion qui représentent la partie intelligente du système.

Ce système comporte aujourd'hui un nombre important de variables libres sur lesquelles il est possible de jouer pour améliorer les performances globales. L'une des voies les plus prometteuses semble être l'introduction de la multirésolution pour reconnaître des objets de tailles différentes grâce à l'existence d'une représentation dans laquelle l'objet sera le mieux codé. Nous avons montré qu'il n'est pas nécessaire d'avoir un système d'extraction de contours ni de fermeture de contours performant pour réaliser une interprétation de qualité. Une bonne segmentation sera surtout le résultat d'une bonne reconnaissance de l'objet étudié. Par ailleurs, on peut noter qu'en termes de temps simulé notre réseau est biologiquement plausible puisqu'il ne fait intervenir que 4 couches de neurones entre l'image perçue et la commance musculaire. Enfin, notre système montre que l'on peut se passer de la rétropropagation du gradient pour l'apprentissage de réseaux multicouches lorsque l'architecture est bien définie et si l'on suppose que

les couches sont apprises successivement (d'abord le bas niveau puis l'interprétation) [Burnod 89]. Pour aller vers un comportement réellement intelligent qui comprendrait la reconnaissance de formes génériques telles que des rectangles ou des bâtiments quelconques, il faudrait ajouter à notre robot d'autres modules gardant une trace de l'expérience en cours de façon à pouvoir interpréter dans le temps et propager des contraintes.

Par rapport à un système expert, on passe de la simulation d'un comportement à la simulation du mécanisme. Dans les S.E., on est limité par la capacité de l'expert à comprendre son propre raisonnement (qualité de l'expertise). Le R.N. permet de passer à la « mimétique » des chemins neuronaux qui engendrent le raisonnement. En utilisant un S.E. uniquement pour la partie « intelligente » du traitement, on se trouve confronté aux problèmes liés à la pertinence des connaissances utilisées et à leur codage.

Enfin, il faut noter que le fait d'apprendre un objet à partir de points de vue particuliers vérifie la Gestalt théorie dans le sens où ce ne sont pas des caractéristiques isolées qui permettent la reconnaissance, mais une forme globale qui est à chaque fois identifiée, même si certains de ses attributs changent (couleur, partie occultée, taille changée, bruit...). Focaliser l'attention sur des points caractéristiques ne fait que diminuer la mémoire nécessaire pour apprendre une forme. On n'est pas obligé d'avoir une information complètement distribuée comme dans un hologramme par exemple. Seules les informations pertinentes sont enregistrées. Cette façon d'interpréter et de mémoriser s'oppose aux méthodes classiques qui cherchent à reconnaître des caractéristiques isolées pour les associer en un objet (théorie structuraliste) et dont les « Gestaltistes » avaient démontré qu'elle ne correspond pas à notre façon de procéder [Rock & Palmer 91].

Manuscrit reçu le 12 juin 1991.

BIBLIOGRAPHIE

[Alexandre 87] F. Alexandre, Y. Burnod, F. Guyot, J. P. Haton, « La colonne corticale, nouvelle unité de base pour des réseaux multicouches », *C. R. Acad. Sci.*, Paris, t. 309, Série III, 1989, p. 259-264.

[Andersen 89] R. A. Andersen, « Visual and eye movement functions of the posterior parietal cortex », *Ann. Rev. Neurosci.*, 12, 1989, p. 377-403.

[Anderson 83] J. Anderson, « Cognitive and Psychological Computation with Neural Models », *IEEE Trans. on Syst., Man, and Cybernetics* 13(5), Sept./Oct. 1983, p. 799-815.

[Allen 90] E. Allen, M. Menon, P. Dicaprio, « A Modular Architecture for Object Recognition Using Neural Networks », *INNC 90*, Paris, July 90, p. 35-37.

[Anderson & Van Essen 87] C. H. Anderson & D. C. Van Essen, « Shifter circuits : A computational strategy for dynamic aspects of visual processing« , *Proc. Natl. Acad. Sci. USA*, Neurobiologie, vol. 84, Sept. 1987, p. 6297-6301.

[Austin 89] J. Austin, « High Speed Invariant Pattern Recognition Using Adaptative Neural Networks », *third international conference on Image Processing IEE*, n° 307, July 1989, p. 28-32.

[Ballard 81] D. H. Ballard, « Generalizing the Hough transform to detect arbitrary shapes », *Pattern Recognition*, vol. 13, n° 2, 1981, p. 111-112.

[Ballard 86] D. H. Ballard, « Cortical connections and parallel processing : Structure and function », *the behavioral & brain sciences* (1986) 9, p. 67-120.

[Barlow 89] H. B. Barlow, « Unsupervised Learning », *Neural Computation 1*, 1989, p. 295-311.

[Baron 85] R. J. Baron, « Visual memories and mental images », *Int. J. Man-Machine Studies*, n° 23, 1985, p. 275-311.

[Barto 81a] A. G. Barto, R. S. Sutton, P. S. Brouwer, « Associative Search Network : A Reinforcement Learning Associative Memory », *Biol. Cybern.*, n° 40, 1981, p. 201-211.

[Barto 81b] A. G. Barto, R. S. Sutton, « Landmark Learning : An Illustration of Associative Search », *Biol. Cubern.*, n° 42, 1981, p. 1-8.

[Barto 83] A. G. Barto, R. S. Sutton, C. W. Anderson, « Neuronlike Adaptative Elements that Can Solve Difficult Learning Control Problems », *IEEE trans on systems, man and cybernetics*, vol. SMC-13, n° 5, September/October 1983, p. 834-846.

[Beck 83] J. Beck, « Textural Segmentation, Second-Order Statistics, and Textural Elements », *Biol. Cybern.*, n° 48, 1983, p. 125-130.

[Beck 85] J. Beck, « Perception of transparency in man and Machine », *CVGIP*, n° 31, p. 127-138, 1985.

[Bienenstock & Von der Malsburg 87] E. Bienenstock & C. Von der Malsburg 87, « A Neural Network for Invariant Pattern Recognition », *Europhisics letters*, 4(1), July 1987, p. 121-126.

[Blakemore & Campbell 69] C. Blakemore, F. Campbell, « On the Existence of neurones in the Human Visual System Selectively Sensitive to the Orientation and Size of Retinal Images » ; *J. Physiol.*, n° 203, 1969, p. 237-260.

[Burnod 87] Y. Burnod, « Architecture par Niveaux du Cortex Cérébral : un Mécanisme possible », *Cognitiva 87*, May 1987, p. 268-274.

[Burnod 89] Y. Burnod, « An adaptive neural network : the cerebral cortex », *Collection Biologie théorique*, Masson 1989.

[Caelli 80] T. M. Caelli, « Facultative and Inhibitory Factors in Visual Texture Discrimination », *Biol. Cybern.*, n° 39, 1980, p. 21-26.

[Canny 86] J. F. Canny, « A computational approach to edge detection », *IEEE transactions PAMI8*, n° 6, 1986, p. 679-698.

[Caelli & Nagendran 87] T. Caelli, S. Nagendran, « Fast Edge-Only Matching Techniques for Robot Pattern Recognition », *GVGIP*, n° 39, 1987, p. 131-143.

[Carpenter & Grossberg 87] G. A. Carpenter & S. Grossberg « A Massively Parallel architecture for a Self-Organizing neural Pattern recognition Machine », *GVGIP 37*, 1987, p. 54-115.

[Carpenter & Grossberg 87b] G. A. Carpenter & S. Grossberg « Invariant Pattern recognition and recall by an attentive self organizing ART architecture in a nonstationary world », *Proceeding of Neural Network*, vol. 2, 1987, p. 737-745.

[Cocquerez 92] J. P. Cocquerez, P. Gaussier, S. Philipp, « Système d'interprétation mixte : réseau de neurones/système-expert appliqué aux images aérienne », *Traitement du signal*, vol. 8, n° 6, numéro spécial IA, 1991.

[Costa & Sandler 89] L. Costa, M. Sandler, « Neural Networks and Hough Transform for Pattern Recognition » 1989.

[Cottrel & Felming 90] G. W. Cottrel, M. Fleming, « Face Recognition using Unsupervised feature Extraction », *INNC 90 Paris*, July 90, p. 322-325.

[Daugman 88] J. Daugman, « Complete Discrete 2-D Gabor Transform by neural Networks for Image Analysis and Compression », *IEEE Trans on acoustics, speech, and signal processing*, vol. 36, n° 7, July 1988.

[Deriche 90] R. Deriche, « fast algorithms for low level vision », *IEEE transactions PAMI 12*, n° 1, 1990, p. 78-87.

[De Saint Pierre 87] T. Desaint Pierre, « Codification et apprentissage connexionniste de caractères multipolices », *Cognitiva 87*, Paris, mai 87, p. 284-289.

[Dingeon 89] C. Dingeon, F. Alexandre, F. Guyot, J. P. Halton, « Un autre apprentissage cortical : Différencier pour généraliser », *proc. Neuro-Nîmes*, Nîmes 1989, p. 305-315.

[Durbin 89] R. Durbin, D. E. Rumelhart, « Product Units : A Computationally Powerfull and Biologically Plausible Extension to Backpropagation Networks », *Neural Computation 1*, 1989, p. 133-142.

[Easton & Gordon 84] P. Easton & P. E. Gordon (1984), « Stabilization of Hebbian Neural Nets by Inhibitory Learning », *Biological Cybernetics*, 29, 127-136.

[Elliman 89] D. G. Elliman, R. N. Banks, J. A. Thompson, « Position Independent Recognition with a Neural Network », 1989.

[El-Sheikh & El-Taweel 89] T. S. El-Taweel, « Real-time Arabic Handwritten Character Recognition », *third international conference on Image Processing IEE*, n° 307, July 1989, p. 212-216.

[Farah & Hammond 88] M. J. Farah, K. M. Hammond, « Mental rotation and orientation-invariant object recognition : Dissociable processes », *Cognition*, n° 29, 1988, p. 29-46.

[Feldmann 85] J. A. Feldman, « Connectionist models and parallelism in high level Vision », *CVGIP 31*, p. 178-200, 1985.

[Földiak 89] P. Földiak (1989), « Adaptative Network for Optimal Linear Feature Extraction », *Proceedings of the International Joint Conference on Neural Networks*, Washington, DC, June 1989, I, 401-105.

[Freeman 91] W. Freeman, « La physiologie de la perception », *Pour la science*, n° 162, avril 1991, p. 70-78.

[Fikushima 82] K. Fukushima, « Neocognition : a new algorithm for pattern recognition tolerant of defaults and shifts in position », *Pattern Recognition*, vol. 15, n° 6, 1982, p. 455-469.

[Fukushima 88] K. Fukushima, « A neural Network for Visual Pattern Recognition », *Computer*, March 1988, p. 65-74.

[Garbay & Pesty 89] C. Carbay, S. Pesty, « Un système Multi-Agents pour la Résolution de problèmes », *AFCET 1989*, p. 355-368.

[Gaussier 91] P. Gaussier, J. P. Cocquerez, S. Philipp, « Un système d'interprétation mixte : réseaux de neurones/Système-expert appliqué à l'interprétation d'images aériennes », *AFCET*, Lyon 1991.

[Gaussier 92] « Simulation d'un système visuel comprenant plusieurs aires corticales : Application à l'analyse de scène », *Thèse de Doctorat*, Paris XI^{e}, novembre 1992.

[Giles 87] C. L. Giles, T. Maxwell, « Learning, invariance, and generalisation in high-order neural networks », *Applied optics*, vol. 26, 1987, p. 4972-4978.

[Glünder 86] H. Glunder, « Neural Computation of Inner Geometric Pattern Relations », *Biol. Cybern.*, n° 55, p. 239-251.

[Grossberg 76] S. Grossberg (1976), « Adaptative Pattern Classification and Universal Recoding : I. Parallel Development and Coding of Neural Feature Detectors », *Biological Cybernetics*, 23, 121-134.

[Grossberg 80] S. Grossberg, « How does the Brain Build a Cognitive Code ? », *Psychological Review*, vol. 87, n° 1, Janv. 1980, p. 1-51.

[Grossberg 87] S. Grossberg, « Cortical dynamics of three-dimentional form, color, and brightness perception : I. Monocular theory », *Perception & Psychophysics*, n° 41(2), 1987, p. 87-116.

[Grossberg 88] Grossberg S., « Nonlinear Neural Networks : Principles, Mechanisms, and Architectures », *Neural Networks*, vol. 1, 1988, p. 17-61.

[Grossberg & Marshall 89] S. Grossberg, J. Marshall, « Stereo Boundary Fusion by Cortical Complex Cells : A System of Maps, Filters, and Feedback Networks for Multiplexing Distributed Data », *Neural Networks*, vol. 2, 1989, p. 29-51.

[Grossberg & Mingolla 85a] S. Grossberg, E. Mingolla, « Neural Dynamics of Form Perception : Boundary Completion, Illusory Figures, and Neon Color Spreading », *Psychological Review*, vol. 92, n° 2, 1985, p. 173-211.

[Grossberg & Mingolla 85b] S. Grossberg, E. Mingolla, « Neural Dynamics of per ceptual grouping : Textures, boundaries, and emergent segmentations », *Perception & Psychophysics*, n° 38(2), 141-171.

[Grossberg & Mingolla 87] S. Grossberg, E. Mingolla, « Neural Dynamics of Surface Perception : Boundary Webs, Illuminants, and Shape-from-Shading », *CVGIP*, n° 37, 1987, p. 116-165.

[Gupta 90] I. Gupta, M. Sayeh, R. Tammara, « A Neural Network Approch to Robust Shape Classification », *Pattern Recognition*, vol. 23, n° 9, p. 563-568, 1990.

[Herold 88] D. J. Herold, W. T. Miller, L. G. Kraft, F. H. Glanz, « Pattern Recognition using a CMAC Based Learning System », *SPIE*, vol. 1004, 1988, p. 84-90.

[Hines & Hutchinson 89] E. L. Hines, R. A. Hutchinson, « Application of Multi-Layer Perceptrons to Facial Feature Location », *IEE image processing*, 1989, p. 39-43.

[Hinton & Lang 85] G. E. Hinton, K. J. Lang, « Shape Recognition and Illusory Conjunctions », *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 1985, pp. 252-260.

[Hebb 49] D. O. Hebb (1949). The organization of Behavior, New York, Wiley.

[Hopfield 82] J. J. Hopfield (1982), « neural Networks and Physical Systems with Emergent Collective Computational Abilities », *Proceeding Natl. Acad. Sci. USA*, vol. 79, 2554-2558, April 1982.

[Hubel & Wiesel 68] Hubel, D. H. Wiesel, « Receptive fields and functional architecture of monkey striate cortex », *J. Physiol*, 195, 1968, p. 215-243.

[Jeannerod 74] M. Jeannerod, « Les deux mécanismes de la vision », *La Recherche*, n° 41, janvier 1974, p. 23-32.

[Johnson & Grogan 89] J. D. Johnson, T. A. Grogan, « Neural Network Controlled Visual Saccades », *SPIE*, vol. 1076, Image understanding and Man-Machine Interface II, 1989, p. 44-49.

[Julesz 81a] B. Julesz, « A Theory of Preattentive Texture Discrimination Based on First-Order Statistics of Textons », *Biol. Cybern.*, n° 41, 1981, p. 131-138.

[Julesz 81b] B. Julesz, « Textons, the elements of texture perception, and their interactions », *Nature*, n° 290, March 1981, p. 91-97.

[Julesz 86] B. Julesz, « Texton Gradients : The Texton Theory Revised », *Biol. Cybern.*, n° 54, 1986, p. 245-251.

[Khotanzad & Lu 89] A. Khotanzad, J. H. Lu, « Object Recognition Using a Neural Network and Invariant Zerkine Features », *IEEE*, 1989, p. 200-205.

[Koch 85] C. Koch, S. Ullman, « Shifts in selective visual attention : towards the underlying neural circuity », *Human neurobiol*, n° 4, 1985, p. 219-227.

[Kohonen 72] T. Kohonen, « Correlation Matrix Memories », *IEEE trans. on computers*, vol. C-21, n° 4, April 1972, p. 353-359.

[Kohonen 89] T. Kohonen, Self-Organization and Associative Memory. New York : Springer-Verlag, 1989.

[Khanna 89] T. Khanna, « Foundations of Neural Networks », *Addison-Wesley Publishing Compagny*, 1989.

[Klopf 82] A. H. Klopf, « The hedonist neuron : A theory of memory, Learning and intelligence » New York, Hemisphere publishing corporation.

[LeCun 89] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, « Backpropagation applied to handwritten zip code recognition », *Neural Computation*, vol. 1, n° 4, 1989, p. 541-551.

[Lehar 90] S. Lehar, T. Howells, I. Smotroff, « Application of Grossberg and Mingolla Neural Vision Model to Satellite Weather Imagery », *INNC 90 Paris*, July 90, p. 805-808.

[Li 90] D. Li, M. R. Spicer, W. G. Wee, « Training a Neocognitron Neural Net for Gray level Images », *INNC 90 Paris*, July 90, p. 111-114.

[Linsker 88] R. Linsker, « Self Organization in a Perceptual Network », *Computer*, Marsch 1988, p. 105-116.

[Lippmann 87] R. Lippmann, « An Introduction to Computing with Neural Nets », *IEEE ASSP*, Magazine, April 1987, p. 4-22.

[Lynch & Rayner 89] M. R. Lynch, P. J. Rayner, « Optical Character Recognition using a New Connectionist Model », *third international conference on Image Processing IEE*, n° 307, July 1989, p. 63-67.

[Marr 69] D. Marr, « A Theory of Cerebellar Cortex », *J. Physiol.*, n° 202, 1969, p. 437-470.

[Marr & Hilderth 80] D. Marr & E. Hilderth, « Theory of edge detection », *Proc. R. Soc. London B.*, 207, p. 187-217, 1980.

[Marshall 89] J. A. Marshall (1989), « Self-Organizing Neural Network Architectures for Computing Visual Depth from Motion Parallax », *proceeding of the International Joint Coferences on Neural Networks*, Washington DC, June 1989, II, 227-234.

[Marshall 90] J. A. Marshall (1990), « Representation of uncertainty in self-Organizing Neural Network », *proceeding of INNC Paris*, July 1990, 809-812.

[McClelland 86] J. L. McClleland, D. E. Rumelhart, G. E. Hinton, « Parallel distributed processing, Exploration in microstructure of cognition », vol. 1, vol. 2, Cambridge, MIT press.

[Mead & Mahowald 88] C. A. Mead, M. A. Mahowald, « Asilicon Model of Early Visual Processing, Neural Networks, vol. 1, 1988, p. 91-97.

[Messner 85] R. A. Messner, « An Image Processing Architecture for Real Time Generation of Scale and Rotation Invariant Patterns », *CVGIP*, n° 31, 1985, p. 50-66.

[Miller 87] W. T. Miller, « Sensor-Based Control of Robotic Manipulators Using a General Learning Algorithm », *IEEE Journal of robotics & automation*, vol. RA-3, n° 2, April 1987, p. 157-165.

[Norton 71] D. Norton, L. Stark, « Eye Movements and Visual Perception », *Scientific American*, vol. 224(6), 1971, p. 34-43.

[Omatu 90] S. Omatu M. Fukumi, M. Teranisi, « Neural Network Model for Alphabetic Letter Recognition », *INNC 90 Paris*, July 90, p. 19-22.

[Otto 90] I. Otto, E. Guigon, Y. Burnod, « Coopération between temporal and parietal networks for invariant recognition », *INNC90 Paris*, p. 480-483.

[Ozawa 90] K. Ozawa, « Simulation Studies on Optical Illusions Based on a Position-Dependent Point Spread Function », *Pattern Recognition*, vol. 23, n° 12, 1990. p. 1361-1366.

[Poggio 85] T. Poggio, « Early Vision : from Computational Structure to Algorithms and Parallel Hardware », *GVGIP*, n° 31, 1985, p. 139-155.

[proceeding IEEE90] Proceeding of the IEEE, vol. 78, n° 9 et 10, Sept./Oct. 1990, p. 1536-1543.

[Reeke 90] G. E. Reeke, Jr. O. Sporns and G. Edelman, « Synthetic Neural Modeling : the Darwin Series of Recognition Automata », *Proceeding of the IEEE*, vol. 78, n° 9, Sept. 1990, p. 1498-1530.

[Rock & Palmer 91] I. Rock & S. Palmer, « L'héritage du gestaltisme », *Pour la science*, n° 160, février 1991. p. 64-70.

[Rumelhard & Zipser 85] D. E. Rumelhart & D. Zipser, « Feature Discovery ty Competitive Learning », *Cognitive Science 9*, 1985, p. 75-112.

[Schwartz 77] E. L. Schartz, « Spacial Mapping in the Primate Sensory Projection : Analytic Structure and Relevance to Perception », *Biol. Cybernetics*, n° 25, 1977, p. 181-194.

[Schwartz 80] E. L. Schartz, « Computational anatomy and fonctional architecture of striate cortex : a spacial mapping approach to perceptual coding », *Vision Research*, vol. 20, 1980, p. 645-669.

[Seibert 89] M. Seibert, A. Waxman, « Spreading Activation Layers, Visual Saccades, and Invariant Representations for Neural Pattern Recognition Systems », *Neural Networks*, vol. 2, 1989, p. 9-27.

[Shepard & Cooper] R. N. Shepard, L. A. Cooper, « Mental Images and their Transformations », *MIT Press*, Cambridge MA 1982.

[Skrzypek 90] J. Skrzypek, « Lightness Constancy : Connectionist Architecture for Controlling Sensitive », *IEEE Trans. on syst. Man and Cybernetics*, vol. 20, n° 5, Sept./Oct. 1990.

[Sperling 89] G. Sperling, « Three stages and two systems of visual processing », *Spacial Vision*, vol. 4, n°s 2/3, p. 183-207, 1989.

[Thorpe 88] S. J. Thorpe, « Traitement d'images chez l'homme », TSI 1987, p. 517-525.

[Treisman 87] Treisman, « L'identification des objets visuels », *Pour la science*, janvier 1987, p. 50-59.

[Van Essen & Maunsell 83] D. C. Van Essen & J. H. R. Maunsell, « Hierarchical organization and functional streams in the visual cortex », *TINS*, September 1983, p. 370-375.

[Wang & Arbib 90] D. Wang & M. A. Arbib, « Complex Temporal Sequence Learning Based on Short-Term Memory », *Proceeding of the IEEE*, vol. 78, n° 9, p. 1536-1543, Sept. 1990.

[Watt & Morgan 83] R. J. Watt & M. J. Morgan, « The recognition and representation of edge blur : evidence for spacial primitives in human vision », *Vision Res*, vol. 23, n° 12, 1983, p. 1465-1477.

[Wechsler & Zimmerman 88] H. Wechsler, G. L. Zimmerman, « 2-D Invariant Object Recognition Using Distributed Associative Memory », *Trans on Pattern Analysis and Machine Intelligence*, vol. 10, n° 6, 1988, p. 811-821.

[Widrow 88] B. Widrow, R. G. Baxter R.A., « Layered neural Nets for Pattern Recognition », *IEEE Trans on Acoustic, Speech, and signal Processing*, vol. 36, n° 7, July 1988.

[Zeki & Shipp 88] S. Zeki & S. Shipp, « The functional logic of cortical connections », *Nature*, vol. 225, 22 sept. 1988, p. 311-317.

[Zetzsche & Caelli 89] C. Zetzsche, T. Caelli, « Invariant Pattern Recognition Using Multiple Filter Image Representations », *Comptuteur Vision, Graphics, and Image Processing*, 45, 1989, p. 251-262.

[Zipser & Andersen 88] D. Zipser & R. A. Andersen « A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons », *Nature*, vol. 331, 25 February 1988, p. 679-684.