Département de formation doctorale en informatique **École doctorale IAEM Lorraine**
UFR STMIA

Modèle connexionniste neuromimétique pour la perception visuelle embarquée du mouvement

THÈSE

présentée et soutenue publiquement le 21 octobre 2005

pour l'obtention du

Doctorat de l'université Henri Poincaré – Nancy 1
(spécialité informatique)

par

Claudio CASTELLANOS SÁNCHEZ

Composition du jury

Président :	Didier GALMICHE	Professeur à l'Université Henry Poincaré, Nancy I
Rapporteurs :	Thierry VIÉVILLE	Directeur de Recherche à l'INRIA, Sophia Antipolis
	Mathias QUOY	Professeur à l'Université Cergy-Pontoise, Paris
Examinateur :	Marie-Dominique GIRAUDO	Maître de Conférence à l'Université de la Méditerranée
Directeurs :	Frédéric ALEXANDRE	Directeur de Recherche à l'INRIA, LORIA, Nancy
	Bernard GIRAU	Maître de Conférence à l'Université Nancy 2

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thloria.

Remerciements

Je tiens à remercier ici les personnes ayant contribué à l'aboutissement de cette thèse qu'il aurait été difficile d'aborder sans l'aide de spécialistes de différentes domaines abordés.

Je remercie tout d'abord à mes directeurs Frédéric ALEXANDRE et Bernard GIRAU qui m'ont accueillit et soutenu tout au long de mon parcours de recherche dans le cadre du projet d'Intelligence Neuromimétique (CORTEX), et aussi aux memebres du jury qui par la qualité de leurs remarques m'ont aidé à éclairer mes points faibles.

Je ne peut pas oublier les memebres de l'équipe CORTEX : permanents, stagiaires, invités et post-doctorants avec lequels j'ai partagé soit mes inquiétudes, soit mes loisirs, soit mes démarches autour de ma recherche. Merci à tous et à toutes pour vos suggestions, vos idées, vos encouragements.

Cette période a été si enrichissante, d'une part nourri par des multiples rencontres au sein du laboratoire d'accueille (LORIA), notamment dans l'équipe CORTEX, et d'autre part par les expériences vecues lors de congrès nationaux et internationaux.

Aujourd'hui je constate que la recherche ne se fait pas de façon isolée. Elle est le fruit de l'effort communautaire de propositions-oppossitions qui nous met en face l'un de l'autre et nous encourage à la création-conception de nouveaux outils pour continuer à la découverte.

Je dédie cette thèse
à mon épouse.
C'est elle qui m'a encouragé
jusqu'au bout de mes démarches.

Table des matières

Introduction

Le système visuel de l'être humain a été optimisé par des millions d'années de sélection naturelle. Il est composé d'au moins 10^{10} neurones et organisé en une série de couches successives de la rétine jusqu'aux aires corticales supérieures.

La quantité d'information fournie par le nerf optique (autour de 10^8 à 10^9 bits par seconde) excède largement la capacité du cerveau pour la traiter complètement et l'assimiler comme une expérience consciente. Naturellement, notre système nerveux en choisit certaines parties depuis l'entrée pour être traitées de façon préférentielle en déplaçant le processus de l'attention d'un endroit à l'autre de façon sérielle : c'est l'attention visuelle sélective. Mais l'estimation du mouvement est aussi une façon de réduire la redondance de l'information visuelle. La perception du mouvement visuel ne dépend pas a priori de l'interprétation ou de la reconnaissance de la forme. Elle est fondée sur la conjonction d'un processus d'intégration et d'un processus de segmentation. Le premier tend à combiner l'ensemble des signaux locaux des vitesses de façon à produire des zones de mouvement cohérent, et le deuxième tend, au contraire, à différencier les signaux locaux de manière à individualiser les objets les uns des autres.

Dans le monde du vivant, notamment chez les animaux, la perception du mouvement est utilisée pour la détection de la forme 3D des objets en mouvement, la profondeur, les jugements de coïncidence et l'estimation de la direction et de la vitesse parmi d'autres. Cette perception est souvent très importante à la survie. D'après la recherche en neurobiologie, le cerveau des primates prend autour d'un sixième de leur neurones pour les traitements visuels. Ainsi, les chercheurs en connexionnisme s'inspirent de plus en plus de la complexité des interactions neuronales et leur organisation dans le cerveau pour arriver à offrir des modèles neuromimétiques qui proposent des solutions aux problèmes que les systèmes classiques ont du mal à résoudre. Mais la conception, la réalisation et la mise en œuvre d'un système de vision artificielle ne se résument pas à savoir traiter une image numérique par un algorithme si sophistiqué et performant soit-il. Cette démarche fait de forts appels à différents domaines. Aujourd'hui, la psychophysique, l'informatique et la neurophysiologie travaillent de façon complémentaire dans le domaine de la perception visuelle du mouvement pour aboutir à la conception de systèmes de vision performants. Cependant, les systèmes de vision conçus sont réputés être très gourmands en calcul et font souvent appel à des architectures parallèles.

De plus en plus, l'émergence des modèles neuromimétiques gagne du terrain et attire le regard de la communauté scientifique. Leur façon de faire converger différents domaines de recherche et la pluridisciplinarité implicite pousse l'essor en robotique et en informatique avec une forte relation avec les neurosciences. Ainsi, le traitement systématique du problème de la perception visuelle débute dès les années soixante par l'application de l'analyse de Fourier en supposant que le système visuel ne fait qu'une transformée de Fourier locale. Donc, on pourrait commencer par un filtrage local temporel. Il va falloir attendre les années quatre-vingt pour voir la recherche connexionniste faire appel aux modèles informatiques bio-inspirés pour la détection du mouvement en faisant des détections locales par l'intégration de différentes directions à différentes échelles et pour aboutir à une réponse globale.

Enfin, mettre en relation la bio-inspiration et les modèles connexionnistes a fait émerger un vaste domaine de recherche où la modularité, la rétro-action et les fortes interactions latérales pallient les durs chemins classiques de la perception du mouvement, à savoir : la haute complexité des systèmes d'équations différentielles à résoudre, la relaxation inhérente à leur application, la forte sensibilité aux changements de l'environnement, l'application en temps réel et les contraintes de généralisation.

Dans cette thèse, je cherche à comprendre quelles sont les capacités de perception visuelle du mouvement dont on peut doter une machine en interaction physique avec son environnement pour qu'elle puisse y poursuivre des objets en mouvement. Nous sommes au croisement des chemins entre les neurosciences, la robotique et l'informatique. Il s'agit de dépasser les contraintes mathématiques des modèles classiques pour nous inspirer de la neurobiologie et proposer un modèle connexionniste neuromimétique issu des interactions entre le cortex visuel primaire (V1), le cortex temporal médian (MT) et le cortex temporal supérieur médian (MST) pour la perception visuelle du mouvement.

Actuellement, pour les tâches perceptives, notamment visuelles, les fortes contraintes de l'autonomie n'ont pas encore été satisfaites par les modèles neuronaux utilisés. C'est pourquoi, dans ce travail je définit un modèle connexionniste bio-inspiré permettant de traiter différentes tâches de perception visuelle dynamique liées à la robotique autonome. Ce travail implique la gestion de difficultés pour tenir compte du caractère modulaire des architectures neuronales tout en y incluant la perception visuelle et en identifiant les contraintes supplémentaires induites par cette modularité. Je cherche aussi à maintenir la satisfaction des contraintes au coeur de la définition des modèles connexionnistes, malgré la complexité des tâches en jeu. De plus, je prends en compte les différentes étapes menant à une perception visuelle dynamique embarquée en partant de la définition de modèles connexionnistes bio-inspirés traitant les aspects spatial et temporel et en manipulant ces modèles massivement distribués qui, à partir des traitements locaux, interagissent pour aboutir à une réponse globale du système pour faciliter le travail ultérieur d'implantation temps-réel embarquée sur FPGA. Les méthodes et les algorithmes obtenus visent à être intégrés aux robots mobiles autonomes travaillant en temps réel dans des environnements inconnus et dynamiques.

Je propose dans la première partie de mon mémoire l'état de l'art d'où sont extraites les principales motivations de mon modèle : tout d'abord les fondements neurobiologiques (§I.1), notamment les interactions dans le cortex visuel primaire et le cortex temporal moyen, suivi par une révision des approches classiques pour l'obtention du flot optique en mettant en valeur les approches avec une très forte relation avec la neurobiologie (§I.2). J'ai finalement choisi de m'inspirer des approches issues des modèles énergétiques. Les principaux modèles bio-inspirés seront ensuite résumés avant un bilan où je mettrai en valeur les principaux aspects retenus, à savoir : le pré-filtrage et les interactions latérales (§I.3). Ces aspects auront été mis en valeur pour les modèles bio-inspirés analysés.

Dans la deuxième partie de ce mémoire, je propose un modèle connexionniste neuromimétique constitué de trois modules pour la perception visuelle du mouvement (§II.1) : dans le premier module j'utilise des filtres spatio-temporels causaux inspirés des filtres de Gabor définis plus précisément dans §II.2. Le chapitre propose aussi une analyse des contraintes d'implantation matérielle sur des circuits FPGA. Dans le deuxième module je fais appel à de fortes inhibitions antagonistes locales. Le chapitre §II.3 les définit ainsi que la faisabilité d'implantation matérielle. Et dans le dernier module, j'applique les principes de la théorie des champs neuronaux continus pour la localisation spatiale des objets en mouvement (§II.4).

Première partie

La perception visuelle du mouvement

Chapitre 1

Fondements neurobiologiques

Sommaire

La perception du mouvement chez les animaux est utilisée pour, entre autres, la détection de la forme 3D des objets en mouvement, l'estimation de la profondeur, des jugements de coïncidence et l'estimation de la direction et de la vitesse. Selon Sekuler et ses collaborateurs [SWB02], lorsqu'on analyse cette perception, on y voit des interactions et des intégrations des mesures locales, de la structure et du mouvement propre (un bref parcours historique de la recherche en perception du mouvement est présenté dans l'annexe A). Mon travail a pour but de mettre en valeur ces interactions et ces intégrations locales. De plus, il apparaît que la détection et l'analyse du mouvement sont obtenues par une cascade d'opérations neuronales, à savoir [SWB02] :

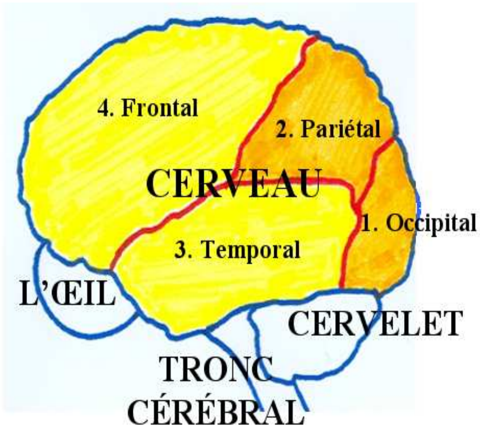
- L'enregistrement des signaux de mouvement local dans des régions restreintes du champ visuel.
- L'intégration de ces signaux pour décrire, de façon plus générale, la direction et la vitesse de l'objet en mouvement.

Selon Field et ses collaborateurs [FHH93], le système visuel intègre l'information de l'orientation spatiale locale pour la définition des contours spatiaux et, d'après Ledgeway et Hess [LH02], il intègre aussi la direction des signaux de mouvement local provenant des différentes parties du champ visuel si ces signaux sont alignés tout au long de l'axe du contour spatial.

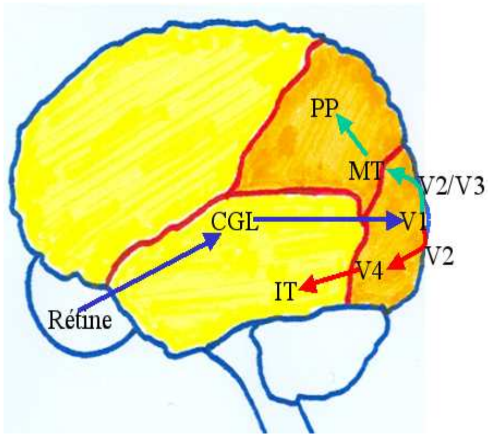
Cette intégration orientée locale est une partie fondamentale de mon modèle proposé et qui va être présenté dans la deuxième partie de ce mémoire.

Dans ce premier chapitre, je vais présenter les différents traitements appliqués aux signaux lumineux depuis la rétine et jusqu'aux aires corticales supérieures pour finalement inférer les principes de base que j'ai mis en place pour mon modèle connexionniste.

Tout d'abord je vais situer la partie du cerveau de l'être humain directement impliquée dans les processus visuels (§I.1.1). Puis je décrirai les principales étapes suivies par l'information lumineuse depuis la rétine et jusqu'aux aires corticales supérieures (§I.1.2). Ensuite, je vais montrer les principales caractéristiques des aires corticales clés dans la perception du mouvement (§I.1.3) pour finir avec les aspects les plus utiles à mettre en œuvre dans mon modèle connexionniste.



(a) Le cerveau, le cervelet, le tronc cérébral et l'organisation fonctionnelle du cortex en quatre lobes.



(b) Parcours des signaux lumineux acquis dans la rétine, puis CGL, V1, V2/V3, et soit vers V4 et IT soit vers MT et PP.

FIG. 1.1 – a) Une vue latérale de l'organisation anatomique de l'encéphale de l'être humain et b) le parcours suivi par les signaux lumineux depuis la rétine.

1.1 Le cerveau de l'être humain

D'après les travaux d'imagerie par résonance magnétique fonctionnelle (fMRI) de Sunaert et ses collaborateurs [SvHMO99], la détection du mouvement chez l'être humain met en fonction plusieurs aires corticales anatomiquement distinctes. Cette détection dépend de l'activité distribuée de ces différentes aires, chacune d'elles extrait des informations tirées de l'image acquise et pré-traitée dans la rétine.

D'un point de vue fonctionnel, le cortex de l'être humain peut être divisé en lobes (et en aires à l'intérieur de chaque lobe) qui traitent un ensemble de processus particuliers (voir figure I.1.1), à savoir :

1. Le lobe occipital pour les processus visuels.
2. Le lobe pariétal pour les processus sensitifs et de guidage.
3. Le lobe temporal pour les processus auditifs et ceux de la compréhension du langage.
4. Le lobe frontal pour les processus de l'action, la décision et l'abstraction.

La plupart des aires des deux premiers lobes (occipital et pariétal) sont impliquées dans les tâches de perception du mouvement. La partie supérieure du lobe pariétal des deux hémisphères cérébraux est impliquée dans la localisation spatiale [SvHMO99].

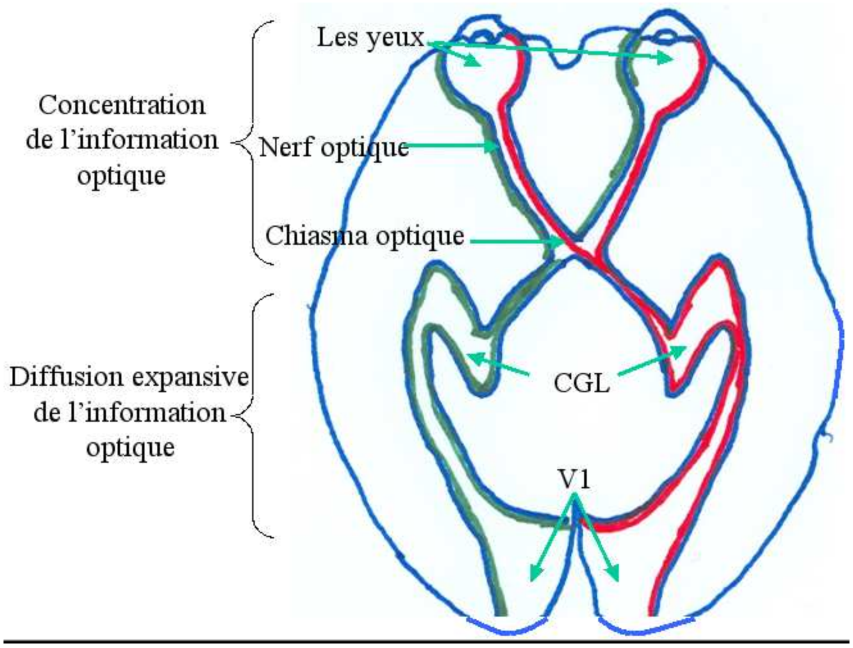


FIG. 1.2 – De la rétine jusqu'à l'aire visuelle primaire en faisant un croisement et un relais, l'information des signaux lumineux parcourt en parallèle deux voies établies : une par les cellules magnocellulaires et l'autre par les cellules parvocellulaires.

Dans l'hémisphère droit, le cortex [1] pré-frontal et le cortex pariétal supérieur sont impliqués dans le traitement des phénomènes attentifs tels que le suivi du mouvement.

L'aire motrice supplémentaire (lobe frontal) est impliquée dans l'initiation et la sélection du mouvement appris. Le cortex pré-frontal est engagé dans le mouvement à commande interne (mouvement librement choisi) et il est dépendant d'une décision personnelle non conditionnée par des stimuli extérieurs.

Pour leurs propriétés dans les tâches du mouvement, je vais prendre en charge les lobes qui le traitent (l'occipital et le pariétal) pour les voir plus en détail dans §I.1.3.

1.2 Le parcours des signaux lumineux chez l'être humain

D'après Davidoff [Dav03], nous pouvons décrire le chemin parcouru par les signaux lumineux chez l'être humain en quatre étapes (voir figure I.1.2) :

1. L'acquisition et la compression des signaux lumineux au niveau de la rétine.
2. Une intégration binoculaire et temporelle au niveau des corps genouillés latéraux.
3. Leur analyse corticale au niveau de l'aire visuelle primaire.
4. Leur traitement de plus haut niveau dans le cortex temporal inférieur et dans le cortex pariétal postérieur.

Au niveau rétinien, les bâtonnets et les cônes sont les toutes premières cellules à capturer l'information des signaux lumineux [2]. Puis les circuits de cellules bipolaires, horizontales, amacrines et ganglionnaires la traitent ensuite pour finalement la compresser.

[1] Le cortex est la partie, fortement plissée, la plus externe du cerveau.

[2] Les bâtonnets sont des éléments réceptifs particuliers de la rétine auxquels on attribue le rôle principal dans la perception des stimuli lumineux, lorsque l'œil est adapté à l'obscurité. Très sensibles aux changements du

Après des traitements locaux au niveau rétinien, toutes les informations concernant l'image sont envoyées au cerveau par deux voies parallèles [Rod98] : la parvocellulaire et la magnocellulaire. Dans la première l'information de la forme et de la couleur est amenée par les cellules ganglionnaires parvocellulaires et, dans la deuxième, c'est l'information du mouvement et de l'espace qui est amenée par les cellules ganglionnaires magnocellulaires. Ces deux chemins semblent être spécialisés pour deux sortes de traitement spatio-temporel [Mer91] : l'un avec une haute fréquence spatial et une basse fréquence temporelle (voie parvocellulaire) et vice-versa (voie magnocellulaire). Une réponse tonique et une réponse phasique, respectivement.

Ensuite, les signaux lumineux acquis et compressés traversent le corps genouillé du thalamus où un nombre aussi important de neurones relais contrôlent le flux d'information notamment par saccades.

Jusqu'ici, le traitement des signaux lumineux a été acheminé par deux voies parallèles indépendantes, en gardant une représentation topologique rétinienne.

Dans le cortex visuel primaire (V1) [3] il y a 500 fois plus de neurones que dans les corps genouillés latéraux (CGL). Ces neurones réalisent la première analyse corticale des signaux lumineux reçus depuis les CGL. Donc, l'image y est analysée en zones par des familles de neurones qui fournissent une description détaillée de chaque zone de l'image en termes d'orientation et de fréquence spatiale parmi d'autres.

Cela, autrement dit, génère différentes représentations orientées et fréquentielles pour chaque neurone de V1 à partir de son champ récepteur nourri des réponses de neurones provenant de l'ensemble CGL-rétine.

Un champ récepteur est, selon Smith [Smi03], la région dans l'espace et dans le temps sur laquelle un neurone répond à la lumière avec un changement de voltage. Il est composé en deux parties : une région centrale circulaire où le neurone est plus sensible et une autre concentrique, antagoniste et très large appelée la périphérie. Le champ récepteur d'un neurone est déterminé par les champs récepteurs des neurones pré-synaptiques, les propriétés du filtrage du signal de ses synapses d'entrée et ses propres signaux de traitement.

Après le traitement en V1 l'information corticale est ensuite distribuée vers plusieurs aires "extra-striées" (dehors V1) par deux grandes voies : une voie dite dorsale qui est impliquée dans l'analyse du mouvement et de l'espace, et la voie dite ventrale qui est impliquée dans la reconnaissance d'objets [RZZ+03, OK03]. Ces deux voies font des traitements fonctionnellement spécialisés et sont organisées hiérarchiquement. Elles aboutissent dans le cortex temporal inférieur pour la voie ventrale, et dans le cortex pariétal postérieur pour la voie dorsale [RZZ+03].

Il apparaît que la voie dorsale est pour la plupart la continuation au niveau cortical de la voie magnocellulaire et la ventrale de la voie parvocellulaire.

Selon Rao et ses collaborateurs [RZZ+03], les neurones de la voie dorsale (constituée pour la plupart par des cellules magnocellulaires) s'activent avant ceux de la voie ventrale (constituée pour la plupart par des cellules parvocellulaires). Il semble plus globalement que l'analyse du traitement du mouvement commence avant que celle pour la reconnaissance des motifs ("patterns" en anglais).

Je vais détailler maintenant les deux voies qui je viens d'énoncer.

contraste.

Les cônes sont des cellules en forme de bouteilles qui, garnissant avec les bâtonnets la surface de la rétine, servent à la vision précise ou à celle en couleur.

[3] On considère que le cortex visuel est la portion du cerveau qui traite les tâches de vision comprenant l'aire striée (ou 17 de Brodmann), lieu de projection et de réception des sensations visuelles, les aires péri-striée et para-striée (ou 18 et 19 de Brodmann) où s'effectuent les phénomènes d'intégration de la vision. C'est là que s'effectue la projection et le traitement des messages nerveux provenant de la rétine.

1.2.1 Les deux voies de cheminement des signaux lumineux

L'information lumineuse acquise au niveau rétinien par les cônes est concentrée en majorité par les neurones ganglionnaires parvocellulaires et de même pour les bâtonnets et la voie magnocellulaire. Donc, on a deux types de neurones qui ramènent leur information indépendamment et en parallèle vers les CGL et V1.

La voie parvocellulaire-ventrale

Pour la reconnaissance de formes, la couleur et la forme sont très importantes. Ce sont les cellules ganglionnaires parvocellulaires qui recueillent l'information de couleur et fournissent la plus grande contribution à celle de la forme. Au niveau cortical la voie ventrale poursuit l'analyse de ce type d'information.

D'après Merigan [Mer91], les neurones du chemin parvocellulaire transmettent l'information depuis la rétine et jusqu'à l'aire visuelle primaire avec une haute fréquence spatiale et une basse fréquence temporelle du stimulus. C'est-à-dire, ces neurones répondent aux caractéristiques de l'image plutôt dans l'espace que dans le temps.

La voie ventrale ou chemin cortical occipito-temporal est composée, chez le primate, d'environ quatre niveaux d'analyse. Après V1, on trouve V2, V4, puis le cortex inféro-temporal (IT). Ces deux dernières aires effectuent les opérations de traitement visuel de plus haut niveau comme le calcul d'éléments de texture par exemple. Leurs résultats sont tout de suite envoyés vers les structures de décision comme le cortex frontal, les structures motrices comme le striatum, et les structures limbiques comme l'hippocampe et l'amygdale.

Pour des tâches complexes de reconnaissance, il est fort probable que l'information soit obligée de passer par la voie rétine-CGL-V1-V2-V4-IT où on voit une augmentation progressive et systématique de la taille des champs récepteurs depuis la rétine jusqu'à IT. Dans cette dernière aire, on peut trouver des cellules avec des champs récepteurs quasiment aussi grands que le champ visuel lui-même.

La figure I.1.3 montre les relations de tailles des champs récepteurs des neurones issus de la voie ventrale. Par exemple, pour une excentricité de 50 ° dans les CGL il existe un ensemble des neurones dont leur champs récepteurs couvrent cet espace, par contre dans l'aire AIT, un seul neurone suffit pour le couvrir étant donnée que son champ récepteur a une taille d'environ 50 ° .

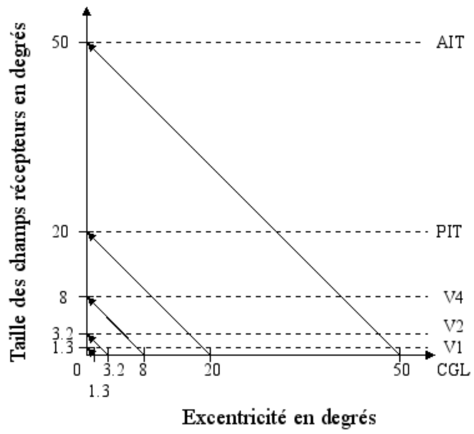


FIG. 1.3 – Changements de la taille des champs récepteurs dans la voie ventrale par rapport à leur excentricité à la fovéa, d'après Wallis et Rolls [WR97].

Étant données les propriétés de traitement spatial de cette voie, nous pouvons la considérer très schématiquement comme une voie de traitement statique.

La voie magnocellulaire-dorsale

Pour le traitement du mouvement et de la localisation spatiale, l'analyse du contraste et les changements spatio-temporels seront nécessaires. Les cellules ganglionnaires magnocellulaires ramènent l'information notamment du contraste qui a été acquise par les bâtonnets. Au niveau cortical, la voie dorsale poursuit l'analyse de cette information déjà traitée par V1.

Selon De Valois et Kooi [DVK91], les neurones du chemin magnocellulaire ont une basse sélectivité au contraste chromatique mais ont une haute sélectivité à la direction du mouvement. Leurs réponses dépendraient de la vitesse, du contraste et de la texture [KLP$^+$01]. D'après Merigan [Mer91], ces neurones transmettent l'information rétinienne vers V1 à hautes fréquences temporelles et à basses fréquences spatiales. Autrement dit, ils font un traitement spatial faible mais un traitement temporel fin.

Au niveau cortical, les signaux neuronaux ramenant l'information de mouvement passent de V1 vers V2-V3, puis vers le cortex temporal médian et le supérieur médian (MT/V5 et MST, respectivement), puis vers le cortex pariétal postérieur (PP) [SWB02, MVE83a].

La voie dorsale ou chemin occipito-pariétal est alors composée d'environ cinq niveaux d'analyse : V1, V2, V3, MT-MST et PP. Dans la figure I.1.4, je vous propose un schéma très général du chemin suivi par les signaux lumineux dans le cerveau de l'être humain.

Vu le haut degré de réponses temporelles, nous pouvons considérer cette voie comme une voie de traitement dynamique.

Il faut souligner qu'il existe de nombreuses autres voies liées aux fonctions visuelles. Par exemple : la rétine envoie également des projections vers d'autres structures sous-corticales comme les tubercules quadrijumeaux supérieurs pour des fonctions oculo-motrices.

La suite du chapitre reprend pas à pas les différentes étapes du parcours des signaux lumineux en commençant par la rétine.

1.2.2 L'acquisition et la compression des signaux lumineux dans la rétine

L'acquisition des signaux lumineux dans la rétine parcourt une série d'étapes hiérarchiques dans une population neuronale différenciée : des neurones bâtonnets-cônes aux neurones ganglionnaires en passant par les neurones bipolaires et en les liant par les cellules horizontales et les cellules amacrines. Toute l'information recueillie est concentrée dans les cellules ganglionnaires qui vont la transporter vers les CGL, puis V1.

En résumé, la rétine traite l'intensité des signaux lumineux dans l'espace et le temps, puis elle les transmet vers le cerveau. La codification notamment du contraste et du mouvement y est déjà réalisée par une multitude des neurones rétiniens [4] vus comme des opérateurs élémentaires locaux non adaptatifs [Smi03].

Le traitement neuronal de l'image rétinienne a ce qu'on appelle un caractère "passe-bande" sur le plan spatial. Cela implique que le système visuel n'est pas très sensible aux gradients graduels de luminance à travers l'image : seule l'information concernant les gradients assez abrupts de luminance est gardée. C'est ce qui engendre le phénomène de "ganzfeld" : un champ de vision parfaitement homogène est vu comme étant gris, indépendamment de sa couleur ou de son intensité réelles [O'R90].

[4] Selon Sterling [Ste90], on peut trouver jusqu'à douze types différents de cellules dans la rétine.

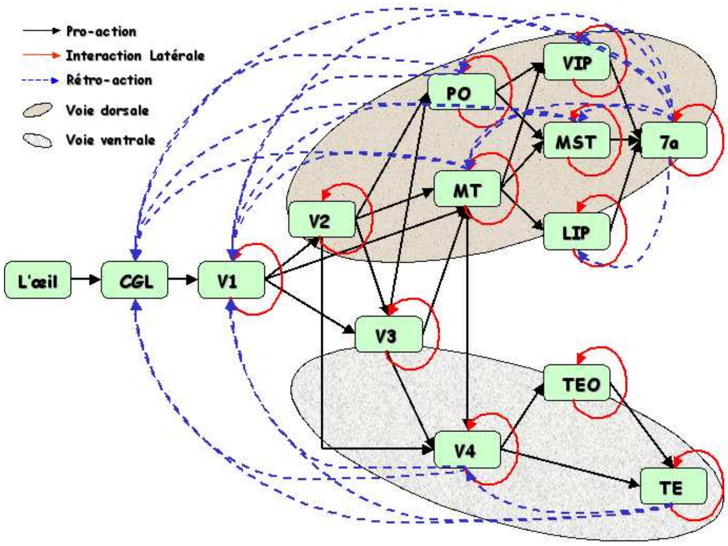


FIG. 1.4 – Un schéma très général du chemin des signaux lumineux dans les aires visuelles du cerveau de l'être humain.

Ainsi, les neurones qui déchargent le plus tôt correspondent aux endroits de l'image où le contraste est le plus élevé, alors que ceux qui déchargent plus tardivement correspondent à un contraste moindre.

D'autre part, la rétine et le thalamus répondent fortement aux mouvements des contours, aux points lumineux et aux grilles, mais leurs réponses ne sont sélectives ni à la direction ni à la vitesse de mouvement. Ces dernières caractéristiques apparaissent dans les neurones de V1 [5].

Trois niveaux de traitement

Pour l'acquisition et la compression des signaux lumineux, la rétine est organisée en trois niveaux de traitement : l'acquisition, l'intégration et la compression.

Pour l'acquisition, les cellules photo-réceptrices forment une mosaïque d'environ 160 millions d'unités (les cônes et les bâtonnets, dont 1/16 de cônes). Chaque unité donne une réponse croissante en fonction de l'intensité lumineuse reçue (voir la figure I.1.5).

Les cellules horizontales et bipolaires assurent l'intégration et le partage de l'information. Selon Hérault [H01], les premières forment un réseau fonctionnel de type *filtrage spatio-temporel passe-bas* (équivalent à l'obtention d'une moyenne locale) et assurent un phénomène d'inhibition latérale sur l'activité des cônes par rétroaction négative et, probablement, sur les cellules bipolaires par action directe.

Selon Imbert [Imb83], les cellules amacrines et ganglionnaires compressent spatialement l'information de 160 à 1 (des cellules photo-réceptrices aux cellules ganglionnaires). Cette compression est proportionnelle à l'excentricité : plus faible dans la fovéa et plus forte dans la périphérie.

De plus, cette compression économise l'activité des cellules ganglionnaires étant donné qu'elle ignore tout ce qui est uniforme et retient les discontinuités spatiales, les distributions de luminosités et les discontinuités temporelles. L'information résultante mettra en valeur l'apparition ou

[5] Selon Grzywacz et ses collaborateurs [GSA03], il n'y a pas d'évidence que la sélectivité rétinienne à la direction contribue à la perception du mouvement. Par contre, il y a évidence que cette sélectivité contribue aux réponses oculo-motrices.

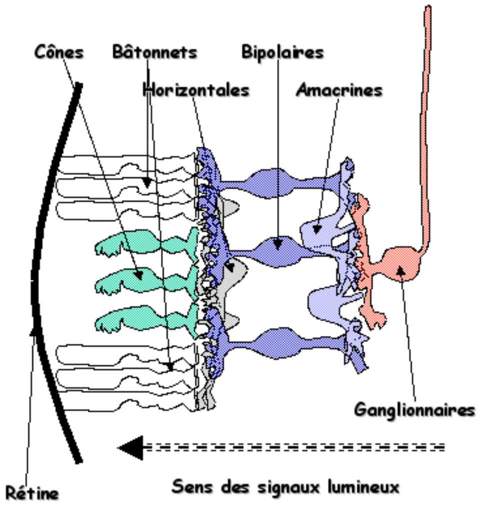


FIG. 1.5 – La rétine acquiert les signaux lumineux puis fait une forte compression spatiale. Dans l'image on peut voir la forte interaction dans la population neuronale rétinienne.

la disparition des contrastes locaux. Cette compression des éléments critiques du stimulus sera envoyée aux aires corticales par les cellules ganglionnaires.

Chaque cellule ganglionnaire répond chaque fois qu'il y a une concordance entre la forme et la taille codées dans son champ récepteur, d'une part, et entre la forme et la taille d'un stimulus visuel, d'autre part. Chacune a sa propre fonction de sensibilité au contraste, la fréquence spatiale qui produit la réponse maximale varie d'une cellule à l'autre. Elles sont reliées entre elles par les cellules amacrines.

Les cellules amacrines seraient responsables de la sélectivité directionnelle fondée sur des excitations et des inhibitions [Smi03]. L'inhibition de la direction nulle [6] pourrait agir pré-synaptiquement sur l'entrée excitatrice des cellules ganglionnaires sélectives à la direction dans un rayon local, et cette inhibition peut aussi agir post-synaptiquement dans les dendrites des cellules ganglionnaires probablement à travers un mécanisme non linéaire de forte inhibition [ZZ01].

Ces mécanismes d'excitation-inhibition au niveau rétinien sont aussi observés de façon plus élaborées au niveau de V1 (voir §I.1.3.1).

La concentration ganglionnaire de l'information

Les cellules ganglionnaires jouent le rôle d'un filtre qui détecte les différences d'illumination entre deux zones contiguës à l'intérieur de leur champ récepteur. Selon Zanker et Zeil [ZZ01], les champs récepteurs des cellules ganglionnaires correspondent à la somme des champs récepteurs des neurones pré-synaptiques.

Les cellules ganglionnaires parvocellulaires et magnocellulaires sont distribuées dans la rétine de la même façon que les cônes et les bâtonnets, respectivement (les premières sont plus concentrées dans la fovéa et les deuxièmes dehors de la fovéa). Les premières donnent une réponse tonique et transmettent lentement l'information, les deuxièmes donnent une réponse phasique et transmettent rapidement l'information.

Finalement, ces cellules sont assemblées dans le nerf optique et envoient leur information vers les CGL, en passant par le chiasme optique (voir la figure I.1.2).

[6] Direction opposée à la direction préférentielle du neurone (voir plus de détails dans §I.1.3.1).

1.2.3 Le relais binoculaire au niveau des CGL

En sortant de la rétine par le nerf optique et en parcourant les cellules magnocellulaires et parvocellulaires, les signaux lumineux compressés subissent les processus suivants :

1. Une sélection binoculaire au niveau du chiasme optique qui partage partiellement l'information venant par le nerf optique depuis les yeux : la moitié droite du champ visuel des deux yeux passe du côté gauche et la moitié gauche des champs visuels passe du côté droit (voir les deux couleurs différentes dans la figure I.1.2).
2. Une intégration temporelle au niveau des corps genouillés latéraux (CGL) servant de relais. Ils maintiennent séparées les cellules parvocellulaires (deux couches) et magnocellulaires (quatre couches) sans les mélanger et les acheminent en parallèle vers l'aire suivante. Chacune des six couches de ces corps voit la totalité de la moitié du champ visuel. Il y a donc six représentations cartographiques bien alignées et superposées les unes par rapport aux autres dans chaque corps genouillé latéral.

Ces derniers sont des relais assurant la distribution des fibres optiques dans le cerveau parce que les neurones des corps genouillés latéraux sont encore plus sensibles que les neurones ganglionnaires aux différences d'illumination entre le centre et la périphérie de leur champ récepteur, c'est-à-dire qu'ils sont plus sensibles au contraste local.

Selon Davidoff [Dav03], les cellules parvocellulaires dans les CGL sont capables de distinguer entre les informations de luminance et de longueur d'onde grâce aux différentes propriétés de leurs champs récepteurs étant donné qu'ils ont une grande résolution spatiale pour certains stimuli définis au niveau de la luminance.

Jusqu'ici les différences d'organisation des cellules rétiniennes se traduisent en des innervations de zones différentes dans les structures supérieures du cerveau. Une partie du traitement de l'information visuelle repose sur cette organisation parallèle qui maintient séparées, au niveau des structures visuelles supérieurs, les différentes classes fonctionnelles élaborées dès le niveau rétinien.

Selon Clifford et Ibbotson [CI03], la plupart des cellules des CGL ont des réponses similaires aux cellules ganglionnaires rétiniennes centre-périphérie mais ces cellules répondent plus faiblement à l'orientation et à la direction que dans le cortex.

1.2.4 Le cortex visuel

Tout de suite après les relais inter-cérébraux, l'information des signaux lumineux rejoint l'aire visuelle primaire qui effectue la première analyse corticale en assurant la sensibilité aux contrastes lumineux par des champs récepteurs allongés.

Selon Maunsell et Newsome [MN87], le calcul du mouvement est hiérarchique dans le cortex. Cette hiérarchie commence avec les cellules simples et les cellules complexes sélectives à la direction dans la couche IVb de V1.

Les neurones de V1 sélectifs à la direction et qui ont des champs récepteurs orientés et petits chez les mammifères, mesurent généralement de façon ambiguë le mouvement, c'est-à-dire, ces neurones réagissent aussi au bruit. Mais, selon Snowden et ses collaborateurs [STEA91, And97], lorsqu'un mouvement cohérent est donné les cellules de MT s'activent de mêmes que quelques cellules de V3A [TMH+97]. Par conséquent, un processus intégrateur plus sensible au mouvement cohérent commence à partir de V1.

Donc, V1 serait le meilleur endroit pour débuter le calcul du mouvement car ses cellules directionnelles sont des filtres d'énergie de mouvement local [AB85, Hee87, GY90, DVCM+00].

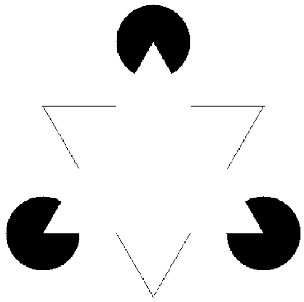


FIG. 1.6 – Le triangle de Kanizsa n'existe pas. Pourtant, certains neurones de V2 peuvent répondre à ce type de contour illusoire.

Les neurones de V1 fournissent l'entrée aux étapes de traitement ultérieur où les neurones ont des champs récepteurs larges qui répondent à la vitesse et à la direction globale du mouvement, en incluant les directions de mouvement spécifiées par la superposition des différents signaux de mouvement local [SWB02, BA98].

En plus des propriétés issues des neurones dans V1, les neurones dans V2 répondent aux contours construits par le système visuel alors qu'ils n'existent pas dans l'image même : il s'agit des contours illusoires [SWB02] (voir figure I.1.6).

L'aire V3 reçoit d'importantes connexions de l'aire V2 et envoie des projections vers les aires MT et V4 [SWB02]. Plusieurs neurones de l'aire V3 ont des propriétés semblables à l'aire V2, étant par exemple pour la plupart sélectifs à l'orientation. L'aire V3 possède aussi des cellules qui sont sensibles à la couleur et au mouvement.

Dans §I.1.3.1 je vais montrer les principales caractéristiques de V1. L'analyse des aires V2 et V3 n'est pas traitée dans ce mémoire car je ne m'occupe pas des phénomènes illusoires traités par V2 (voir figure I.1.6) et ni de l'aire V3.

1.2.5 Le traitement ultérieur dans le cortex pariétal

Finalement, dans les aires 20 et 21 du cortex temporal [7] et 7 du cortex pariétal [8] leurs cellules font un traitement secondaire des signaux lumineux et reçoivent leurs informations visuelles afférentes directement ou indirectement de l'aire visuelle primaire de façon non topographique, c'est-à-dire de façon distribuée, globale et différenciée.

L'aire MST

Selon Orban [Orb97], au niveau de l'aire temporale médiane (MT), les neurones sont sélectifs aux motifs de mouvement linéaire uniforme. Leur sélectivité à la direction est plus complexe à cause de leurs grands champs récepteurs [9]. Dans §I.1.3.2 je vais traiter cette aire plus en détail.

Au-delà de l'aire MT se trouvent d'autres régions impliquées dans l'analyse du mouvement comme l'aire MST par exemple. Les cellules qui s'y trouvent sont non seulement sensibles au

[7] Ces aires définies par Brodmann correspondent approximativement à l'aire inféro-temporale (IT) : antérieure et postérieure, respectivement.

[8] Cette aire définie par Brodmann correspond à l'aire pariétale postérieure, PP.

[9] Il apparaît que la taille des champs récepteurs des neurones de MT ont jusqu'à dix fois la taille de ceux de V1.

déplacement linéaire comme dans l'aire MT, mais également au déplacement radial (vers un point ou à partir d'un point) ou encore au mouvement circulaire (dans le sens horaire ou contre-horaire). Elles sont en fait sélectivement activées par des configurations complexes de mouvements correspondant au déplacement des éléments de notre environnement lorsqu'on se déplace dans celui-ci.

Selon Newsome et ses collaborateurs [NWK88], les cellules de MST traitent les signaux relatifs à l'objet mobile en utilisant l'information de la vitesse des yeux. Ils ont accès à la représentation du mouvement de la cible par rapport à la tête en éliminant le mouvement rétinien. Le mouvement propre est aussi utilisé à ce niveau. Ces cellules répondent de manière sélective à une direction donnée d'un stimulus mobile dans leur champs récepteur. Cependant selon Erickon et Thier [ET91], lorsque les yeux bougent en recevant les mêmes stimuli avec un environnement stationnaire, ces cellules perdent cette sélectivité. Le modèle proposé ne prendra pas en compte ces informations extra-rétiniennes comme celle du mouvement de la tête et celle du mouvement propre.

Selon Saito et ses collaborateurs [SYT$^+$86], la structure de MST a deux grandes subdivisions : la région latéro-ventrale (MSTv) et la région dorso-médiale (MSTd).

Les cellules de MSTd et MSTv répondent à un petit stimulus visuel qui bouge dans une direction donnée mais à la direction contraire si ce stimulus est grand.

Les neurones de MSTv répondent mieux aux petits points de lumière et, selon Dursteler et Wurtz [DW88], leur réponses aident à maintenir la fixation des yeux pendant la poursuite.

D'autre part, les neurones dans MSTd ont des grands champs récepteurs et répondent mieux au mouvement des grands champs (environ de 60°). Ces neurones sont sensibles aux stimuli visuels de mouvement de translation simple, de rotation, d'expansion/contraction et leurs combinaisons (radiales, circulaires et planaires) [DW91a, DW91b, DW97b, DW97a, GAS94]. Ainsi il semble que cette aire est impliquée dans l'analyse détaillée du flot optique.

Selon les travaux de Perrone [Per92, PS94] l'information relative au focus d'expansion [10] pour la navigation peut être représentée par les neurones de MSTd.

D'autres aires du cortex pariétal

Les neurones des aires inféro-pariétale médiane (MIP) et inféro-pariétale latérale (LIP) sont aussi sélectives à la direction du mouvement. De plus, selon Eskandar et Assad [EA02], les neurones de MIP et de LIP continuent encore à émettre des potentiels d'action de manière sélective à la direction si une partie de la source de mouvement disparaît (si le stimulus réapparaît celui-ci s'intègre avec la prédiction des réponses émises par ces neurones).

Selon Burr et ses collaborateurs [BMV98] les neurones de ces aires ont des zones d'intégration de plus de 70° et, d'après Duffy et Wurtz [DW91a], les neurones sensibles au mouvement ont des champs récepteurs autour de 100° [DW91a].

Finalement, il apparaît que les cellules de l'aire intra-pariétale ventrale (VIP), de la région antérieure de l'aire temporale supérieure poli-sensorielle antérieure (STPa) et de l'aire 7a exhibent des réponses préférentielles au mouvement complexe des composantes du flot optique comme celle liées au mouvement de formes radiales, circulaires et planaires [AS99, DW91a, DW91b, DW97b, DW97a, GAS94, PS00].

Enfin, le cortex pariétal pourrait être responsable de la conduite de nos propres actions et permettrait d'évaluer si elles atteignent nos buts. Le cerveau crée en quelque sorte un modèle, ou copie, des intentions du mouvement que l'on veut faire, au niveau du cortex pariétal, avant la

[10] Selon Gibson [Gib50], lorsque un observateur bouge à travers un environnement statique, une singularité dans le flot optique appelée focus d'expansion peut être utilisée pour déterminer la direction du mouvement propre.

commande d'exécution qui passe par le cortex moteur. Quand ce cortex pariétal est lésé, le mouvement est toujours possible, mais maladroit ; le sujet peut difficilement ajuster le mouvement, le corriger ou en apprendre un nouveau. Selon Rougier et ses collaborateurs [RV05, VRA05] le cortex pariétal joue donc un rôle essentiel dans l'anticipation et dans le contrôle en ligne de nos mouvements.

Le troisième module de mon modèle s'inspire du traitement dans le cortex pariétal pour maintenir fixée l'attention sur un objet en mouvement et le poursuivre (voir §II.1 et §II.4).

1.2.6 La rétro-action de l'information des aires supérieures vers les aires inférieures

Selon Miyauchi et ses collaborateurs [MSWM93], il existe des connexions en rétroaction sur le traitement du mouvement liées aux phénomènes attentionnels. Ils concluent que le système visuel a des projections spécialisées en rétroaction sur certaines aires corticales.

D'après les expériences faites par Beckers et Homberg [BH92] dans les aires V1 et MT et les études psychophysiques de l'intégration de la perception, la rétro-action des aires visuelles secondaires vers V1 est liée à la conscience visuelle [SWB02, PLW01].

La rétro-projection de MT vers V1 (qui est faite, selon Pascual et Walsh [PLW01], dans une période de temps très courte) semble être critique pour la conscience visuelle. De plus, la rupture de l'activité dans V1 due à l'arrivé de cette rétro-action apparaît interférer avec la perception des attributs ayant été codés par les aires extra-striées [PLW01].

D'autre part, il semble que dans MT existe un mécanisme descendant pour la reconstruction du mouvement [SBG$^+$00, KK00].

Malgré la grande influence de l'innervation en provenance de la rétine sur l'organisation du CGL, environ 80% des connexions excitatrices qui entrent dans le CGL ne proviennent pas de la rétine mais bien du cortex visuel primaire ! Il semble donc que le cortex visuel primaire exerce une rétroaction importante sur le CGL. En d'autres termes, le rôle principal du CGL pourrait être de modifier en retour les réponses visuelles de ce même CGL (voir figure I.1.4).

1.2.7 Bilan

Le plus important, pour la suite, sont les traitements suivis dans la voie magnocellulaire-dorsale par l'information du mouvement et de l'espace depuis la rétine et jusqu'aux aires corticales supérieures. Le traitement du contraste et les changements spatio-temporels seront donc mis en valeur (voir §I.2.3).

Bien qu'au niveau rétinien il existe déjà une détection de mouvement utile pour les commandes oculo-motrices opérées dans les cortex pariétal et pré-frontal, l'analyse de l'orientation et de la vitesse pour la localisation spatiale est plutôt opérée à partir de V1. Ainsi, le traitement rétinien a retenu les discontinuités spatiales, les distributions de luminosités et les discontinuités temporelles pour mettre en valeur l'apparition ou la disparition des contrastes locaux.

De cette façon, je me concentre dans les traitements opérés dès V1 où l'orientation et la vitesse sont nécessaires pour repérer les objets en mouvement et les poursuivre.

Par la suite nous allons nous plonger plus profondément dans les traitements pour la perception du mouvement faits par les aires V1 (en général) et MT (mouvement cohérent).

1.3 Le traitement cortical

Dans cette section je vais montrer plus en détail les principales caractéristiques des aires retenues pour concevoir mon modèle, à savoir : les aires V1 et MT.

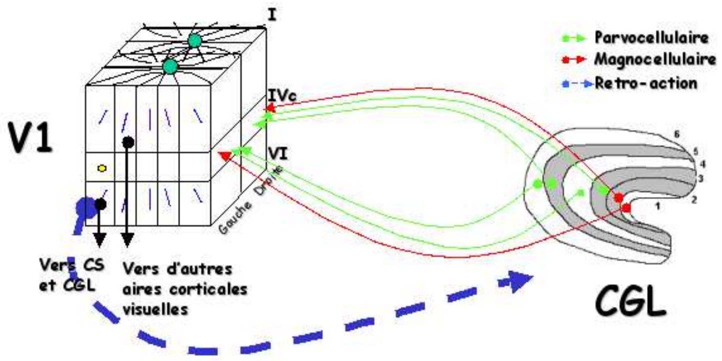


FIG. 1.7 – La relation entre les CGL et V1. Au niveau de V1 on perçoit en profondeur les hypercolonnes de dominance oculaire (issues des yeux gauche et droit). En face, les colonnes d'orientation (dans ce schéma, cinq) et, de haut en bas, ses différentes couches. Dans la couche quatre, V1 reçoit l'information provenant des CGL et elle envoie des réponses en rétro-action sortant de la couche cinq.

1.3.1 V1, la première analyse corticale

L'information acquise dans la rétine et intégrée dans les CGL, est reçue dans la couche IV de V1 où commencera sa première analyse corticale (voir figure I.1.7).

Selon Hubel et Weisel [HW72], les neurones de la couche IV du cortex visuel primaire sont la cible principale des axones sortant des CGL. À la différence des champs récepteurs de type centre-périphérie avec des centres circulaires (ON ou OFF) et des périphéries antagonistes dans les neurones des CGL, leurs neurones cibles (les cellules simples) de V1 ont des champs récepteurs allongés (jusqu'à 20 $^\circ$ dans leur longueur) se recouvrant partiellement [HW62].

Enfin, toutes les cellules de V1 sont sélectives à une direction préférentielle et à une fréquence spatiale. Elles sont en général peu sensibles à la couleur du stimulus de mouvement et limitées à une extraction locale du mouvement. Or, ces cellules ne "voient" qu'une partie locale de la scène et, en conséquence, leurs réponses sont partielles. Cette ambiguïté fut détaillée par Hans Wallach dans les années 40 [WSR96] et elle est connue comme le *problème d'ouverture* [Mar82, SKS00] et ses caractéristiques subséquentes : les différentes illusions de mouvement. Donc on aura autant de réponses que de directions différentes pour la même information locale reçue.

Maintenant je vais mettre en valeur les caractéristiques qui permettent aux neurones de V1 de percevoir le mouvement. Ensuite je ferai une description plus détaillée de V1 : tout d'abord sa topologie puis les neurones qui la composent et finalement les interactions que ces neurones provoquent.

La perception du mouvement

La détection du mouvement est déterminée par la réactivité du neurone le plus sensible à une direction [11]. Selon Grzywacz et Norcia [GN03], la perception du mouvement dépend de la

[11] Dans ce mémoire, l'orientation sera l'axe sur lequel on peut voir deux directions. Par exemple, dans l'axe horizontal, la direction à la gauche et la direction à la droite seront les deux directions opposées dans l'orientation

sélectivité corticale à la direction et contribue au contrôle du mouvement des yeux.

Je considère qu'un neurone est accordé à une direction (direction préférentielle) si sa réponse est fortement liée à une valeur particulière du stimulus (ou à un azimut des valeurs) et diminue de façon monotone pour les valeurs des stimuli qui s'éloignent de sa direction préférentielle (en référence à Perrone et Newsome [PN98]).

En outre, selon Grzywacz et Norcia [GN03], un neurone visuel est sélectif à une direction, si le mouvement correspondant provoque la plus grande réponse dans ce neurone et je l'appelle direction préférentielle. Si le mouvement est dans la direction opposée je l'appellerai la direction non préférée ou nulle.

Les différents neurones de V1 sont sélectifs à l'orientation des contours dans l'image. Selon Sterling [Ste02], ils répondent en premier, et de façon asymétrique, à la direction du mouvement. Ils ont une forte réponse au stimulus d'un objet qui bouge dans la direction préférentielle et une très faible réponse dans la direction contraire. Dans cette direction de mouvement préférentielle ou accordée, ils effectuent une analyse *locale* de l'énergie du mouvement [STEA91, MAGN85b, QAA94].

D'ailleurs, ces cellules peuvent être modélisées par des filtres spatio-temporels extrayant l'énergie du mouvement dans les régions voisines de l'image rétinienne [AB85, vSS84]. On peut modéliser le champ récepteur de ces cellules sélectives à l'orientation spatiale comme un produit des interactions inhibitrices et excitatrices dans l'espace et dans le temps.

Enfin, le mouvement local de l'image rétinienne est extrait par les neurones de V1 ayant un champ récepteur semblable à une petite fenêtre spatialement délimitée. Ainsi, ses neurones peuvent y enregistrer la présence de mouvement dans une direction déterminée. Je prendrai en compte ces faits pour la conception de mon modèle.

Les colonnes d'orientation : une organisation topologique et spatiale

L'organisation en colonnes corticales est liée au traitement d'une caractéristique, soit le contraste, soit la couleur, soit l'orientation, soit le mouvement, etc., dans une région donnée du champ visuel.

Organisée en six couches en profondeur et en bandes dans sa surface [Ale90, FB99], les neurones de l'aire visuelle primaire sont aussi organisés en colonnes d'orientation et de dominance oculaire [12] (voir la figure I.1.7). L'orientation des colonnes tourne progressivement de 10 ° d'une colonne à l'autre. En assemblant tous ces neurones on peut retrouver la disposition complète des directions de 0 ° à 360 ° .

Pour chaque partie de la rétine, il existe une zone de V1 appelé hypercolonne [13] traitant l'information reçue comme la sélectivité à l'orientation [14], la disparité binoculaire et la sélectivité à la direction du mouvement des stimuli.

L'orientation, la fréquence spatiale, la présence de terminaisons, la disparité binoculaire, la direction du mouvement, la couleur, etc., sont des traitements locaux et distribués dans V1 qui émergent de l'activité partagée de chaque hypercolonne corticale et à l'intérieur d'elle-même. Un neurone isolé ne peut pas coder un seul paramètre. Donc, un processus d'auto-organisation serait une bonne représentation des hypercolonnes corticales dans V1.

horizontale.

[12] Ces bandes correspondent aux terminaisons nerveuses de l'œil gauche et de l'œil droit et elles alternent ainsi d'un œil à l'autre, se disposant un peu comme les rayures d'un zèbre.

[13] Une hypercolonne corticale contient au moins deux colonnes de dominance oculaire (cfr. [Mou79]).

[14] Entre deux hypercolonnes adjacentes il y a une différence d'environ 10 ° à 20 ° .

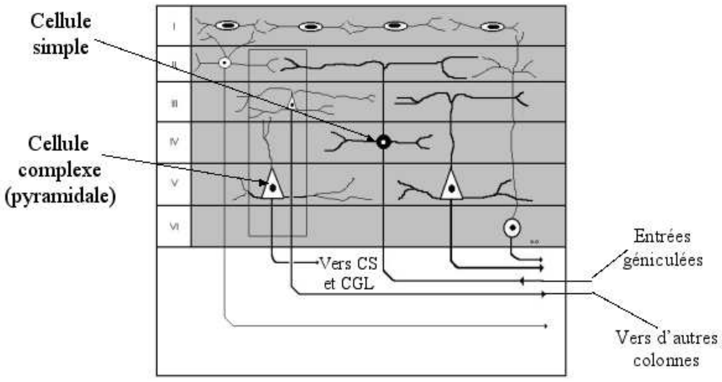


FIG. 1.8 – Les différentes couches dans V1 et ses différents neurones.

Selon Sato et ses collaborateurs [SKT+96], la couche $4C\alpha$ (magnocellulaire) a une grande sensibilité à l'orientation par rapport à la couche $4C\beta$ (parvocellulaire). De plus, il apparaît que dans cette couche, les neurones avec des orientations préférées similaires sont plus fortement liés que ceux ayant des orientations préférées orthogonales. Ce principe est mis en évidence dans le deuxième module de mon modèle connexionniste (voir §II.1 et §II.3).

Les différents neurones dans V1

Il semble que dans V1 les cellules simples voient l'axe d'orientation, les cellules complexes le mouvement de l'axe et les cellules hypercomplexes les bords et les angles. Je vais par la suite décrire ces trois types de cellules.

Il est établi que certains neurones du cortex visuel répondent sélectivement aux stimuli d'une certaine fréquence spatio-temporelle [AB85, WAJ85, NS94a]. Les filtres de Gabor, dont les principes seront décrits dans §I.2.3.1, semblent être un modèle approprié pour de tels types de neurones.

Selon Clifford et Ibbotson [CI03], on peut trouver des cellules directionnelles (comme les cellules simples accordées à l'orientation), d'autres cellules influencées directionnellement (non opposées, comme les cellules complexes) et des cellules opposées au mouvement (comme les cellules hypercomplexes). Les premières ne sont pas directionnelles mais sont fortement accordées à l'orientation. Selon Mazer et ses collaborateurs [MVM+02] ces cellules donneraient leur meilleure réponse lorsqu'une barre ou grille orientée bouge perpendiculairement à leur orientation préférée. Et d'après Henry et ses collaborateurs [HBD74] les deuxièmes tendent à être accordées à l'orientation. Les dernières, avec des propriétés opposées au mouvement, semblent répondre fortement dans la direction préférée et sont inhibées par le mouvement dans la direction opposée.

- Les cellules simples sont situées dans la couche IV, ont des petits champs récepteurs et sont sensibles à l'orientation et à la luminosité fixe. Elles possèdent des régions bien différenciées excitatrices et inhibitrices de telle sorte que le neurone ne répond bien que lorsqu'une barre orientée se situe dans une zone précise de son champ récepteur.

De plus, elles ont d'importantes non linéarités temporelles. Par exemple, plus le contraste d'un stimulus sera fort, plus la réponse de la cellule apparaîtra rapidement. Ce phénomène est connu comme "l'avance de phase". Et leur réponse à une fréquence temporelle dépendra du contraste du stimulus. En particulier, augmenter ce dernier augmente la sensibilité de la cellule aux hautes fréquences temporelles.

- Les cellules complexes n'ont pas de zone si clairement différenciée et peuvent répondre à un contour qui a la bonne orientation, quelle que soit sa position dans leur champ récepteur. Elles sont situées hors de la couche IV, ont des grands champs récepteurs et sont sensibles à l'orientation fixe et à la luminosité variable. Il y a deux types de ces cellules :

- Les cellules standards répondent selon le stimulus reçu.
- Les cellules spéciales ou pyramidales répondent de façon fixe quel que soit le stimulus reçu. Elles sont localisées dans les couches III et V. Ces cellules activent les tubercules quadrijumeaux antérieurs, structure essentielle dans l'élaboration et l'exécution des réactions visuo-motrices.

- Les cellules hypercomplexes présentent des zones d'inhibition telles qu'elles répondent optimalement lorsqu'un contour termine dans leur champ récepteur. Elles rehaussent les combinaisons des formes simples.

Movshon et Newsome [MN96] ont trouvé que les neurones de V1 qui projettent directement vers MT sont de type hypercomplexe, c'est-à-dire, ils répondent à une large gamme des fréquences spatiales et temporelles et sont aussi sensibles aux très bas contrastes. Il est possible que ces cellules reçoivent leurs entrées soit des cellules simples directionnelles, soit directement des fibres des CGL, soit des deux.

Les interactions latérales

Il existe un réseau très important de connexions horizontales au sein de V1 parmi les neurones des différentes colonnes corticales. Il est largement admis que ces connexions jouent un rôle primordial dans la structuration de l'information visuelle car les neurones corticaux sont fortement excitateurs et hautement inter-connectés [15] [LN04].

D'après Douglas et ses collaborateurs [DMW89, DKM$^+$95] et de Somers et les siens [SNS95], les circuits locaux de V1 réalisent un équilibre entre l'excitation et l'inhibition latérale, et, selon Nowlan et Sejnowski [NS95, NS94b], ces mécanismes font l'intégration du mouvement dans MT.

Selon Barlow [Bar81], au niveau rétinien, les inhibitions latérales éliminent l'information superflue générée par la corrélation de luminosité dans l'image. Par contre, selon Douglas et ses collaborateurs [DMM96], l'inhibition latérale augmente le bruit. D'après les résultats montrés dans le chapitre §II.3 je reste à mi-chemin entre ces deux points de vues.

- Les liens horizontaux. Selon Tucker et Katz [TK03], les neurones avec des propriétés similaires dans leurs champs récepteurs sont assemblés en modules fonctionnels et dans chaque module leurs cellules pyramidales sont densément inter-connectées dans un circuit neuronal local. Ces liens horizontaux avec un long rayon équilibreraient excitation et inhibition.

Les potentiels post-synaptiques excitateurs libérés par les connexions horizontales semblent s'amplifier par une addition temporelle [TD94] et réduire l'activation des neurones pyramidaux par une inhibition locale [TK03]. De plus, une stimulation sur les connexions horizontales distantes permet l'émergence des potentiels post-synaptiques inhibiteurs à différentes amplitudes.

- "Shunting inhibition". Une forte et dense inhibition arrive lorsque le potentiel dû aux synapses inhibitrices est près du potentiel de réponse de la cellule cible. Selon Koch et ses collaborateurs [KMS03], une telle inhibition en court-circuit annule localement l'excitation lorsqu'elle

[15] Selon Braintenberg et Schuz [BS91], un neurone typique du cortex visuel reçoit entre 5000 et 10000 entrées.

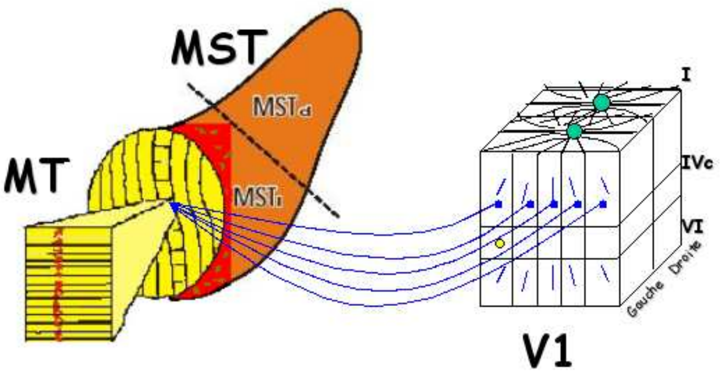


Fig. 1.9 – La relation entre les aires V1 et MT. Les neurones de MT récupèrent les sorties de neurones de V1 en différentes orientions pour les intégrer et aboutir à une réponse locale cohérente.

se trouve sur le chemin entre l'emplacement de la synapse excitatrice et le corps de la cellule cible.

Selon Grzywacz et Merwine [GM03], bien qu'un tel mécanisme soit théoriquement capable de produire les interactions localisées qui sont nécessaires pour expliquer beaucoup des propriétés sélectives de la direction, il n'a pas encore été possible de faire des essais pour confirmer ce mécanisme.

D'autre part, un mécanisme d'interaction latérale entre toutes les inhibitions localement fortes sur toutes les différentes directions de mouvement dans V1 apporterait une réponse émergente correspondant à la direction globale.

Ce mécanisme d'interaction latérale est très utilisé dans les modèles d'inspiration biologique (voir §I.3) alors que le mécanisme d'inhibition en court-circuit ("shunting inhibition") reste encore à modéliser. J'implanterai plutôt celui d'interactions latérales que j'appellerai Mécanisme de Fortes Interactions Antagonistes, MFIA (voir §II.1.3 et §II.3).

Selon Braddick et Quian [BQ01], l'effet de recouvrement local suggère que, dans n'importe quel petit secteur du champ visuel, les détecteurs de mouvement local accordés aux différentes directions ont une forte interaction. Par exemple, un signal directionnel simple est capable de perdurer à cet emplacement. Dans le cas de mouvements opposés, les signaux sont localement équilibrés et le résultat est une suppression de n'importe quel mouvement perçu.

1.3.2 MT, la détection du mouvement cohérent

La perception visuelle du mouvement n'est pas complètement déterminée par les réponses locales dans le champ récepteur des neurones, elles sont durement traitées pour obtenir la vitesse après les avoir collectées et combinées depuis V1 et en les intégrant dans MT. C'est cette combinaison de signaux qui résout l'ambiguïté des réponses individuelles locales des neurones de V1 (problème d'ouverture, voir figure I.1.9).

Par la suite je détaille comment les neurones de MT intègrent les réponses issues de V1 puis comment ils font la détection de la vitesse. Je présente ensuite l'organisation neuronale à l'intérieur de MT et, finalement, les différences principales entre les neurones de V1 et de MT.

L'intégration des réponses issues de V1

MT reçoit les réponses des neurones de V1 et construit une représentation rétinotopiquement organisée de l'espace visuel. Et, comme une grande partie des neurones de MT répondent à une direction de mouvement donnée indépendamment de l'orientation du stimulus, cette aire élimine l'ambiguïté de la direction du mouvement.

Il y a trois hypothèses différentes qui expliquent le processus d'intégration des réponses de V1 dans MT : par la combinaison de différentes sorties de V1, par les inhibitions mutuelles des neurones antagonistes [16] et par le renforcement de la direction dans un mouvement cohérent.

Dans la première hypothèse selon Movshon et Newsome [MN96], les neurones de MT fonctionnent comme si leurs champs récepteurs étaient une combinaison de multiples sorties des neurones de V1.

Dans la deuxième hypothèse selon Heeger et ses collaborateurs [HBD+99], la perception du mouvement est basée sur des mécanismes opposés correspondant à des inhibitions mutuelles entre neurones sensibles aux mouvements dans des directions opposées.

Et dans la troisième hypothèse selon Britten et ses collaborateurs [BSNM93], les neurones de MT répondent aux signaux de mouvement stochastique et la force de leurs réponses augmente avec le nombre de points en mouvement cohérent.

Même si nous voyons cette diversité d'explications pour l'intégration de l'information de V1 dans MT, il apparaît que presque tous les neurones de MT sont sélectifs à la direction et à la vitesse d'un stimulus de mouvement qu'ils calculent de façon locale, explicite et non-ambiguë [PT01].

Dans ma démarche j'ai choisit la deuxième hypothèse pour la conception du mécanisme des fortes interactions antagonistes (MFIA, voir §II.1.3 et §II.3). De plus, il est couramment accepté qu'il existe un conflit entre la réduction du bruit et la représentation des mouvements multiples. La suppression locale de la direction dans MT diminue les réponses au bruit mais elle rend presque impossible le fait que ces cellules peuvent représenter plus d'un vecteur de mouvement dans une petit aire. La suppression directionnelle dans MT est très forte lorsque deux directions du mouvement diffèrent de plus de 90 ° et faible si la différence angulaire est mineure [STEA91, QAA94]. Par exemple, à 45 ° on pourrait voir un mouvement cohérent résultat de la moyenne des directions de deux mouvements composants proches.

La détection de la vitesse

Dans MT presque tous les neurones sont sélectifs à la direction et à la vitesse d'un stimulus de mouvement en laissant dehors les attributs de la couleur et de la forme. Par contre, selon Perrone et Thiele [PT01] et Adelson et Movshon [AM82], les neurones de V1 ne peuvent pas estimer la vélocité car ils confondent les changements des formes de stimulus (telles que l'orientation) avec les changements dans la vélocité du stimulus.

- Une intégration sélective et pondérée. D'après Hess et Ledgeway [HL03], les neurones proches correspondant à des vitesses similaires créent une association significative des champs des vitesses. En conséquence, la détectabilité des contours de mouvement est détectée lorsqu'il y a une différence de vitesses claire entre les éléments définissant le contour et ceux du fond.

L'adaptation du mouvement sélectif à la direction semble produire un déséquilibre des réponses des neurones sélectifs à la direction dans MT et engendre une asymétrie correspondant

[16] J'appelle neurone antagoniste au neurone répondant préférentiellement à la direction opposée d'un autre neurone.

aux seuils de discrimination de la vitesse [HRH01].

Donc, la vitesse est construite par la somme des sorties pondérées de V1 accordées aux bandes de fréquences spatio-temporelles appartenant à un plan commun [Hee87, GY90, SH98]. Et, selon Schrater et ses collaborateurs [SKS00], la réponse d'une famille de ces détecteurs peut déterminer la présence et la vitesse des formes en translation locale.

Il est possible que l'information de la vitesse locale ne soit pas simplement intégrée de la même manière que les signaux locaux de la direction [LH02]. Selon Perrone et Thiele [PT01], quelques neurones de MT combinent leurs entrées temporelles et spatiales pour le calcul de la vitesse indépendamment de la structure spatiale et temporelle du stimulus.

Selon Todd et Perotti [TP99], les neurones accordés à une direction préférentielle ont des interactions latérales modulatrices et activent un processus compétitif pour aboutir à la sélection non ambiguë d'une vitesse particulière. Ce processus compétitif est dépendant de la fréquence spatiale, de la forme, du contraste et de la disparité des objets [AM82, MAGN85a, BQA95].

- L'influence du contraste. Finalement, il apparaît que la vitesse et le contraste ne sont pas codés indépendemment [MG94]. De plus, il est possible que le stimulus du contraste affecte la vitesse [Tho82, MG94] sinon, en bas contraste, il y aurait une discrimination des seuils de vitesses [MG94].

De plus, il est possible que la perception de la vitesse puisse fortement dépendre de la composition du contraste, la couleur et la fréquence spatiale d'un stimulus [DPW99, Tho82, MSN86], mais la perception de la direction est complètement indépendante de ces facteurs. Elle est aussi modifiée par la distance du sujet à l'objet (effet de vitesse constante) en suggérant que son calcul est contextuel et, par inférence, moins liée aux motifs élémentaires d'activité du cortex.

En conclusion, les neurones de MT calculent et représentent la vélocité locale de l'image rétinienne. De plus, ils sont sélectifs à la vitesse.

La construction de la sélectivité à la vélocité dans MT à partir des cellules afférentes de V1 serait analogue à la construction de la sélectivité à l'orientation dans V1 à partir de cellules afférentes du CGL [HW68, HW62].

L'organisation neuronale dans MT

Le mécanisme permettant de lier les directions locales adjacentes tout au long d'un contour spatial est non sensible aux différences de vitesse des éléments individuels du contour [BSD01].

Le système visuel utilise la même forme spécialisée d'intégration spatiale pour extraire des contours définis par la vitesse comme pour extraire des contours définis par la direction du mouvement, mais avec un certain degré de préférence pour des vitesses similaires. Selon Hess et Ledgeway [HL03], les liens correspondant à la vitesse sont beaucoup plus faibles que ceux correspondant à la direction et à l'orientation.

Les neurones de MT sont sélectifs à la vitesse [LRO93, MVE83b, PT01] sous forme de groupements locaux [DAN98, MVE83b] plutôt qu'en colonnes [LN03], mais par rapport à la direction et à la disparité binoculaire, ils sont organisés en colonnes [Alb84, DAN98, ADG84].

Étant donnée la complexité de la perception de la vitesse, une organisation en colonne ne permettrait pas une meilleure performance de calcul [LN03]. Donc, il est possible que MT ne soit pas organisé en colonnes pour le traitement de la vitesse.

Les différences entre les neurones de V1 et de MT

Les cellules de MT sont, pour la plupart, sélectives à la direction, mais il ne semble pas qu'elles extraient l'information du mouvement local par elles-mêmes. Au contraire elles reçoivent les signaux directionnels des étapes primaires du système visuel [LPB01].

La différence des réponses des neurones de MT et de V1 est que certains neurones de MT reçoivent en entrée les réponses des neurones de V1 avec différentes directions préférentielles. D'autre part, selon Heeger et ses collaborateurs [HBD+99], alors que les neurones de V1 répondent également aux différents stimuli de mouvement pair et impair [17], les neurones de MT répondent mieux aux stimuli impairs. Et, selon Braddick et Quian [BQ01], les cellules de MT ont de grands champs récepteurs [18] et ont une plage spatiale des interactions directionnelles trois fois plus grande que celles de V1.

Dans MT, il y a un très haut gain aux bas contrastes et une saturation complète aux hauts contrastes [TRK+95].

En outre, un neurone de MT répondra mieux si les signaux du mouvement sont renforcés par la luminosité, la couleur ou la texture.

Selon Pack et Born [PB01], la dynamique non-ambiguë des neurones de MT chez le singe fait que ses neurones répondent tout d'abord à la composante perpendiculaire du mouvement et ensuite à l'orientation d'un contour en mouvement. Donc, ils font une estimation initiale de la direction du mouvement local par intégration des signaux ambigus et non ambigus sur une surface spatiale uniforme et, au fur et à mesure, ils améliorent leurs réponses.

1.3.3 Bilan

Ce que je retiens de cette section est le grand ensemble de filtrages orientés dédiés pour chaque point de l'espace visuel représenté dans V1 d'une part, et la sélectivité à la direction et l'extraction locale du mouvement réalisée par ces capteurs d'autre part. L'extraction locale du mouvement effectuée dans le champ récepteur des cellules de V1 est comparable à un filtrage spatio-temporel que je simule pour modéliser les interactions inhibitrices et excitatrices dans l'espace et dans le temps au sein de mon modèle connexionniste (voir §II.1.2 et §II.2).

D'autre part, la détection locale du mouvement n'est pas complètement déterminée par les réponses locales des neurones de V1, il va falloir les intégrer dans MT où il est possible d'éliminer l'ambiguïté de la direction du mouvement et de rendre plus cohérentes les zones de mouvement par de fortes interactions antagonistes (voir §II.1.3 et §II.3).

Dans le chapitre suivant nous allons faire un bref survol sur les différentes approches pour la modélisation du flot optique et je mettrai en valeur celles d'inspiration biologique comme le sont notamment les approches énergétiques.

[17] Un mouvement provoqué par des couples d'objets dans une même direction s'appelle mouvement pair. Il sera impair s'il existe un objet en mouvement différent.

[18] Selon Duffy et Wurtz [DW91a, GAS94, TFS89], de 6 ° à 10 ° de diamètre, c'est-à-dire, environ dix fois celui des neurones de V1.

Chapitre 2

La modélisation classique du flot optique

Sommaire

Le chapitre §I.1 a permis de faire un rapide survol de la façon dont les signaux lumineux sont traités dans le cerveau de l'être humain, plus particulièrement la voie magnocellulaire-dorsale. Ici, je vais présenter les fondements mathématiques pour le traitement des signaux lumineux du mouvement, à savoir le flot optique.

L'observation et l'analyse des changements temporels dans les scènes conduisent à des informations essentielles pour la compression de données d'une séquences d'images (MPEG), l'évitement d'obstacles, l'enregistrement des images et leur composition, la segmentation fondée sur le mouvement et sa détection, la structure, la profondeur et la trajectoire d'un objet en mouvement qu'on doit suivre ou qu'on doit éviter.

Pour tout cela une première étape de traitement des distributions des signaux du mouvement local inclut généralement les opérations fondamentales suivantes [Bra97] (voir figure I.2.1) :

1. *La détection du mouvement* où le résultat est une image binaire indiquant pour chaque pixel [19] s'il y a eu du mouvement ou pas.

[19] Élément le plus petit d'une surface d'affichage auquel on puisse associer individuellement une couleur (ou un niveau de gris) et une intensité.
Le pixel étant la plus petite surface homogène d'une image, il est en quelque sorte l'équivalent d'un point. Dans

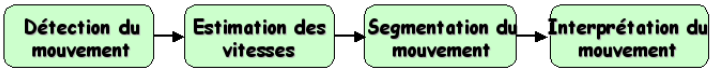


FIG. 2.1 – Les principales étapes classiques pour le traitement du mouvement.

2. *L'estimation du champ des vitesses apparent (le flot optique)*, fournissant une mesure de la vitesse pour chaque région de l'image.
3. *La segmentation du mouvement apparent* où l'image est segmentée en régions de mouvement uniforme.
4. *L'interprétation du mouvement* fournissant les paramètres globaux sur le mouvement des divers objets de la scène.

La détection et l'estimation sont des analyses du mouvement de bas niveau. La segmentation et l'interprétation du mouvement sont des analyses de haut niveau. Dans ma démarche, je privilégierai le traitement de bas niveau.

D'autre part, le mouvement fournit une information plus riche sur la forme 3D des objets en mouvement car l'information émise par ces objets crée des changements spatio-temporels dans la distribution de la lumière à l'intérieur de la rétine d'un observateur (en considérant le cas le plus simple où l'observateur est statique). Ces changements sont à l'origine du flot optique, lequel fournit des informations sur la vitesse, la direction, la trajectoire des mouvements de l'observateur et la structure 3D de l'environnement [Koe86].

La recherche sur le flot optique se développe, entre autres, en deux branches :
- Le guidage des mouvements de l'observateur et la planification des trajectoires où la précision des jugements, le guidage dans la locomotion et le support contre les collisions sont importants.
- L'obtention de la forme et de la structure 3D des objets en mouvement.

Dans ce chapitre je resterai dans la première de ces deux branches et je vais faire un rappel notamment sur la deuxième opération fondamentale pour le traitement des signaux du mouvement : le flot optique. Tout d'abord, nous allons voir le concept, puis les modèles classiques pour sa détection et son estimation. Je finirai ce chapitre en mettant en valeur l'approche retenue : les modèles énergétiques. Ce rapide survol me fournira les principes choisis pour m'inspirer dans la conception du premier module de mon modèle (voir §II.1).

2.1 La détermination du flot optique

Le flot optique est une distribution spatio-temporelle des vecteurs de mouvement de l'image qu'estime la distribution obtenue par une projection des vecteurs 3D de l'objet en mouvement sur la rétine [GN03]. Il ne peut pas être calculé en un point de l'image de façon indépendante des points voisins sans introduire des contraintes additionnelles [HS80] et son estimation a été proposée comme une étape de pré-traitement dans plusieurs algorithmes de vision de haut niveau [GMN+99].

le cas d'un écran monochrome, le pixel s'identifie avec le point et, dans le cas d'un écran couleur, il est constitué de trois points de couleurs différentes (rouge, vert, bleu). En variant l'intensité de chacun des points, on peut faire apparaître des milliers de couleurs différentes. La quantité de pixels composant l'écran détermine la résolution.

Le terme pixel est formé de la juxtaposition de pix et el qui sont les formes abrégées de *picture element*.

Le terme élément d'image est moins couramment utilisé que le terme pixel. Les termes eldim et élim ont été proposés pour traduire le terme anglais pixel, mais ils ne se sont pas implantés dans l'usage.

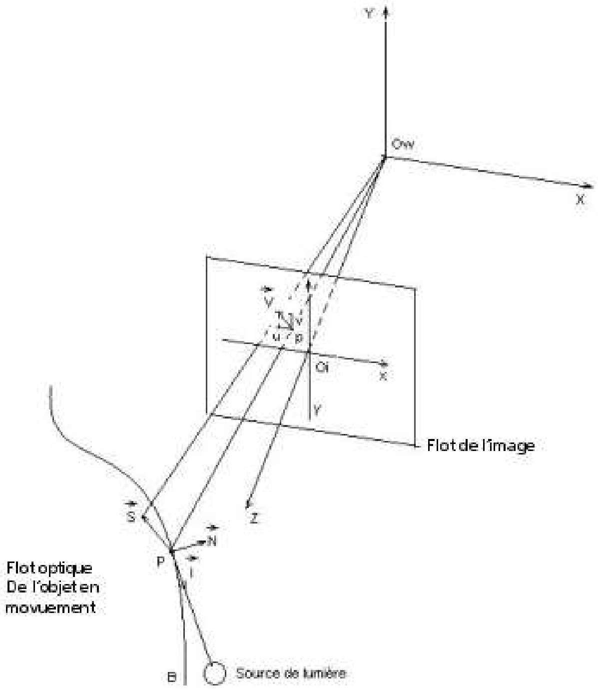


FIG. 2.2 – Le flot de l'image (plan Oi) est la projection 2D des vecteurs de mouvement 3D (le flot optique réel). La vitesse $\vec{V}$ est perpendiculaire à la normale $\vec{N}$ au point P de l'objet B qui bouge dans la direction $\vec{S}$ et qui a une intensité $\vec{I}$.

Pour distinguer le flot optique réel de celui qu'on perçoit, j'appellerai le deuxième flot de l'image ou flot optique perçu (voir figure I.2.2).

En effet, on estime le flot optique, on ne le mesure pas. On l'estime en mesurant l'intensité de l'image.

2.1.1 L'équation de contrainte du flot optique

Pour obtenir l'équation de contrainte du flot optique (ECFO), on mettra en relation les changements de luminosité dans un pixel de l'image et ses mouvements.

Soit la luminosité d'une image en un point (x, y) du plan de l'image au temps t dénoté par $E(x, y, t)$. Si la luminosité reste constante et si la forme de la luminosité se déplace d'une distance Δx dans la direction x et d'une distance Δy dans la direction y au temps $t + \Delta t$, on obtiendra alors

$$E(x, y, t) = E(x + \Delta x, y + \Delta y, t + \Delta t) \tag{2.1}$$

Dans la limite de $\Delta t \to 0$ on aura

$$\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \tag{2.2}$$

C'est *l'équation de contrainte du flot optique* (ECFO) qui définit une simple contrainte locale du mouvement de l'image [BB95a].

Si on écrit

$$u = \tfrac{dx}{dt} \quad et \quad v = \tfrac{dy}{dt} \tag{2.3}$$

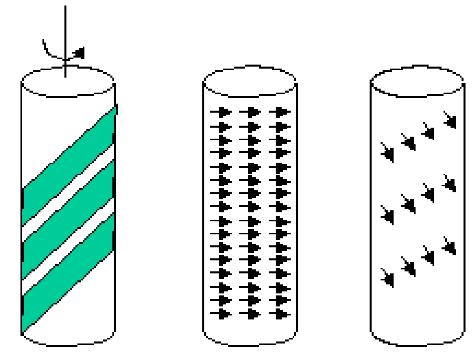


FIG. 2.3 – Le champ de mouvement et le flot optique du poteau tournant d'un coiffeur.

on voit alors qu'on a une équation linéaire avec deux variables inconnues u et v,

$$E_x u + E_y v + E_t = 0 \tag{2.4}$$

soit

$$(E_x, E_y) \cdot (u, v) = -E_t \tag{2.5}$$

où E_x, E_y et E_t sont les dérivées partielles de la luminosité selon x, y et t, respectivement.

Comme la vitesse du flot (u, v) ne peut pas être calculée localement sans introduire des contraintes additionnelles, par exemple, en limitant la différence entre la vitesse du flot dans un point et la moyenne de vitesse sur un petit voisinage autour de ce point. On pourra minimiser la somme des carrés du laplacien des composantes du flot x et y, où les laplaciens alors de u et v sont définis comme

$$\Delta^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad et \quad \Delta^2 v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \tag{2.6}$$

Dans les situations simples, les deux laplaciens sont égaux à zéro. Si l'observateur se déplace soit de façon parallèle à un objet plat, soit tourne autour d'une ligne perpendiculaire à la surface, soit traverse orthogonalement vers la surface (en supposant une projection en perspective), alors la deuxième dérivée partielle de u et v disparaîtra [HS80].

Mais cette proposition a plusieurs restrictions. Dans les régions de l'image où la luminosité est nulle, la vitesse estimée pourrait être prise par défaut comme la moyenne des vitesses estimées dans un petit voisinage. C'est un des principes de régularisation. Lorsque la luminosité dans une région est une fonction linéaire des coordonnées de l'image on ne peut qu'obtenir la composante du flot optique dans la direction du gradient. Plus généralement, calculer le mouvement génère le problème d'ouverture, quelque soit les contours [Mar82, MU81].

2.1.2 Le problème d'ouverture

La figure 2.3a montre le poteau tournant d'un coiffeur. Le champ de mouvement (2.3b) diffère du flot optique (2.3c). Cette différence est due au problème d'ouverture. La figure 2.4 montre ce problème : on ne détecte pas le mouvement de la barre se déplaçant vers le nord, la seule direction détectée est celle colinéaire au gradient de l'intensité : la composante b dans la figure 2.4, une seule composante parmi les deux nécessaires pour l'estimation correcte du mouvement de la barre.

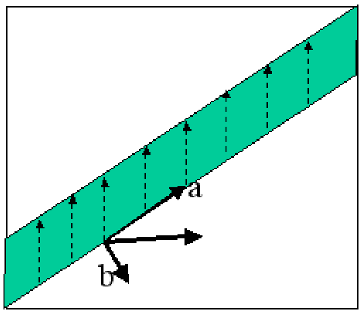


FIG. 2.4 – Le problème d'ouverture. On ne peut estimer que la composante b (celle de la direction du gradient de l'intensité).

Pour résoudre ce problème, il a été proposé de considérer :

1. Plusieurs mesures dans chaque pixel.
2. La définition de contraintes de régularité en restreignant l'espace des solutions.

De plus, d'après la définition du flot optique de Horn et Schunck [HS93], on voit que les effets spéculaires [20], les ombres, les textures insuffisantes et l'occlusion [21] rendent fausse l'ECFO. Dans des séquences d'images artificielles on peut connaître le flot optique réel et faire les hypothèses nécessaires à une régularisation pertinente. Mais son application aux séquences d'images naturelles reste difficile car il est plus difficile d'avoir des hypothèses réalistes sur certaines caractéristiques de leur flot optique réel.

2.2 Méthodes d'estimation du flot optique

Pour que le flot optique réel et celui perçu ou estimé coïncident, il faut une surface réfléchissante lambertienne [22] et une translation pure parallèle au plan de l'image [VP89]. Ce sont des conditions très restrictives à respecter, donc l'utilisation de cette hypothèse dans le monde réel reste peu applicable.

En général, pour obtenir le flot optique d'une séquence d'images on réalise les étapes suivantes (voir la figure I.2.5) :

1. La récupération de l'information de vitesse localement disponible (en utilisant un petit voisinage spatio-temporel) :
   (a) Pré-filtrage (passe-bas ou passe-bande) en vue d'extraire les éléments pertinents du signal et de rehausser le rapport signal-bruit.
   (b) Extraction des mesures des dérivées spatio-temporelles, par des méthodes de corrélation, par exemple.
   (c) Intégration des mesures soit par régularisation, soit par mise en correspondance, soit par la minimisation d'une autre fonctionnelle.
2. La propagation de la vitesse des régions ayant une information complète vers celles ayant une information partielle ou nulle de façon à récupérer l'estimation du champ de flot optique en chaque point.

[20] Qui réfléchit la lumière comme un miroir.
[21] Lorsqu'un objet en mouvement cache un autre aussi en mouvement. Voir sa définition dans §I.2.4.
[22] Les surfaces naturelles provoquent des réflexions diffuses, réparties dans de nombreuses directions. Si l'énergie renvoyée est exactement la même dans toutes les directions, la surface est dite lambertienne.

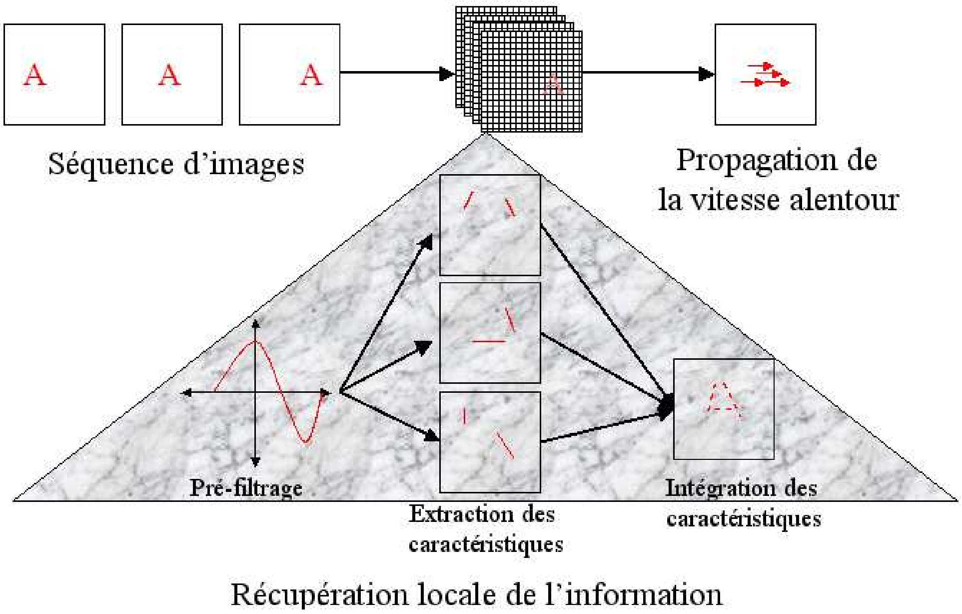


FIG. 2.5 – Les principales étapes dans l'obtention du flot optique.

La façon de récupérer l'information localement disponible, le pré-filtrage choisi, l'extraction des mesures et leur propagation peut avoir des solutions diverses. Par la suite, nous allons survoler les principales approches pour l'estimation du flot optique (voir §I.2.2.1 à §I.2.2.5) et je vais mettre en valeur les approches fréquentielles pour leur forte inspiration biologique (§I.2.3).

2.2.1 Les principales approches pour l'estimation du flot optique

En accord avec la conservation de l'intensité lumineuse, différentes approches proposent comment estimer le flot optique (on les développera plus loin), à savoir : les approches différentielles, de mise en correspondance, stochastiques et fréquentielles (voir tableau I.2.1).

- Dans les approches fondées sur le gradient (différentielles), on suppose, tout d'abord, la conservation de l'intensité de l'image dans un petit intervalle de temps. Ensuite, on utilise le résultat d'une fonction de l'intensité de l'image plutôt que l'intensité de l'image elle-même. Finalement, on calcule la vitesse des dérivées spatio-temporelles des intensités de l'image en supposant que le domaine de l'image est continu (différentiel) dans l'espace et dans le temps [LK81, Nag89]. L'estimation est très locale (ponctuelle), donc génère l'ambiguïté caractéristique : le problème d'ouverture.
- Dans les approches de mise en correspondance (*block matching*), comme celle de Barnard et Thompson [BT80], on suppose la conservation locale de la distribution d'intensité, laquelle est filtrée, tout d'abord, par un filtre passe-bande, puis traitée selon un schéma hiérarchique. Pour cela, on fait un appariement entre deux images consécutives d'une séquence donnée. Cette approche définit un déplacement (lequel est une approximation de la vitesse) comme le meilleur ajustement parmi les régions voisines variant dans le temps [Ana89, OS92]. Elles partagent et utilisent les corrélations entre les zones des deux images successives ou entre leurs éléments caractéristiques (segments de contour, coins, etc., appelés *tokens*). Bref, on suit les caractéristiques marquantes d'un objet tout au long d'une séquence d'images où les mouvements des objets sont détectés en faisant une recherche exhaustive des caractéristiques similaires dans deux images successives.

- Les approches stochastiques (comme celles de relaxation) sont fondées sur une contrainte concernant les discontinuités du mouvement [KD88, HB93]. Elles manipulent les occlusions en modélisant les intensités discontinues par des processus stochastiques [HB93] ou en utilisant les filtres de Kalman [Bla92].
- Dans les approches fréquentielles, les fréquences spatio-temporelles sont mises en relation avec la vitesse du stimulus de mouvement, et le flot optique devient l'identification d'un plan d'énergie dans l'espace de la fréquence spatio-temporelle. En général, les mécanismes sensibles au mouvement fondés sur l'énergie orientée spatio-temporelle dans l'espace de Fourier peuvent estimer le mouvement dans les endroits où les autres approches échouent. Ainsi, la détection du mouvement dans l'image revient ici à extraire l'orientation spatio-temporelle [AB85, Hee87, GY90]. On connaît deux sous-approches : fondée sur la phase et fondée sur l'énergie.
  - Dans le premier cas, on définit les composantes de la vitesse en termes de mouvement instantané au niveau de la phase des contours de la sortie des filtres passe-bande de Gabor accordés à la vitesse [FJ90, WWB88]. Ces filtres seront décrits dans §I.2.3.1.
  - D'autre part, les approches fondées sur l'énergie sont fondées sur une caractérisation du mouvement dans le domaine de la fréquence [AB85, WAJ85]. Elles utilisent les réponses énergétiques de filtres spatio-temporels orientés qui sont sensibles à des stimuli en mouvement. Ces filtres peuvent être des filtres séparables, i.e les filtrages spatiaux et temporels peuvent être appliqués séquentiellement sur l'image d'entrée [AB85, Hee87]. Pour l'obtention de la vitesse, il existe un schéma de codification de la population où la vitesse est exprimée comme le rapport des réponses des filtres d'énergie accordées aux différentes vitesses. En supposant que la vitesse a un comportement linéaire, les réponses de chaque filtre changeront avec les différents contrastes mais leur rapport restera constant.

Les deux premières approches, différentielles et de mise en correspondance, sont les plus anciennes, mais, selon Kearney et Thompson [KT87], avec des images naturelles, elles conduisent généralement à des erreurs d'estimation assez grandes. De plus, les modèles de mise en correspondance ont au moins deux problèmes sérieux : le premier est la détermination des régions et des caractéristiques pour la poursuite et le deuxième, le choix d'un critère de correspondance.

Selon Spinei [Spi98], une analyse de l'erreur des méthodes de mise en correspondance confirme que le problème majeur de cette approche est que les grandes erreurs sont faites là où l'image est hautement texturisée. Mais c'est dans ces points qu'il existe la plupart de l'information de mouvement.

Les approches stochastiques fondées sur les champs de Markov sont plus performantes dans la mesure où elles utilisent plusieurs contraintes apportant de l'information pour l'estimation : cohérence spatiale et temporelle, prise en charge des régions de discontinuités du mouvement, etc. Les processus de prise en compte des contraintes dans le temps sont assez lourds à mettre en œuvre avec un algorithme séquentiel et sont plus adaptés pour des architectures parallèles ou pour une implantation sur silicium.

Le principe de la dernière approche consiste à localiser le plan d'énergie non nulle ou *le plan de vitesse*. Ce plan correspond à un mouvement de translation 2D apparent sur un support spatial délimité par le biais de filtres spatio-temporels orientés. Plusieurs transformations ont été utilisées pour construire de tels filtres, à savoir : les transformées de Fourier à court terme [Wen90], de Wigener-Ville [VH93] et des approches tensorielles [BHKG91]. Les filtres de Gabor sont les plus utilisés grâce à leurs propriétés et à leur ressemblance avec certains "filtres" du cortex visuel. En effet, toutes les théories ayant comme point de départ les expériences physiologiques sont fondées sur l'existence de filtres spatio-temporels orientés. Heeger [Hee87], par exemple, a proposé et implanté une méthode de calcul du flot optique en combinant les réponses fournies par une famille de tels filtres.

Il existe évidemment d'autres approches et de plus, de nombreuses combinaisons de celles vues ici. Les principaux aspects des approches présentés dans cette section vont être développés un peu plus en détail pour mettre en valeur leurs principes.

<table>
<tr><th>Approche</th><th>Sous Approche</th><th>Proposition classique</th><th>Autres propositions</th></tr>
<tr><td rowspan="3">Différentielles</td><td>Globale</td><td>Fennema et Thompson [FT79]</td><td>Horn et Schunck [HS80]
Nagel [Nag87]</td></tr>
<tr><td>Locale</td><td>Lucas et Kanade [LK81]</td><td>Uras et al. [UGVT88]
Proesmans et al. [PvGPO94]</td></tr>
<tr><td>Autres</td><td></td><td>Limb et Murphy [LM75]
Cafforio et Rocca [CR76]</td></tr>
<tr><td>Mise en correspondance</td><td></td><td>Reichardt [Rei61]</td><td>Thompson et Barnard [TB81]
Prager et Arbib [PA83]
Anandan [Ana89]
Ogata et Sato [OS92]
Singh [Sin91]
Camus [Cam97]</td></tr>
<tr><td>Stochastiques</td><td></td><td>Konrad et Dubois [KD88]</td><td>Ascher et Grzywacz [AG00]
Heeger et Jepson [HJ92]
Heitz et Bouthemy [HB93]
Black [Bla92]
Geman et Geman [GG84]</td></tr>
<tr><td rowspan="2">Fréquentielles</td><td>Énergétiques</td><td>Adelson et Bergen [AB85]</td><td>Adelson et Movshon [AM82]
Welch [Wel89]
Watson et Ahumada [WAJ85]
Heeger [Hee87]
Van Santen et Sperling [vSS85]
Simoncelli [Sim93]
Nowlan et Sejnowski [NS95]</td></tr>
<tr><td>Fondées sur la phase</td><td>Waxman et al. [WWB88]</td><td>Fleet et Jepson [FJ90]
Grzywacz et Yuille [GY90]</td></tr>
</table>

TAB. 2.1 – Principales approches pour l'estimation du flot optique.

2.2.2 Approches différentielles

Dans les approches différentielles ou fondées sur le gradient on applique une combinaison des sorties des filtres des dérivées spatiales et temporelles pour estimer la vitesse [FT79, HS80, JC95]. Par conséquent, le mouvement est calculé comme le rapport des sorties accordées à la fréquence temporelle et spatiale.

Les accords à la direction et à la vitesse du modèle de gradient apparaissent dans une étape postérieure à l'accord de la fréquence temporelle.

Cette approche calcule la vitesse par dérivées spatio-temporelles des intensités de l'image en supposant que le domaine de l'image est continu (différentiel) dans l'espace et dans le temps.

On peut percevoir deux types de traitement : l'un global et l'autre local. Si le traitement d'une séquence d'images utilise une régularisation sur des grandes régions on dit que l'approche est globale, mais si celui-ci applique une minimisation sur un voisinage local, alors l'approche est locale.

Les méthodes globales

Les méthodes globales incluent un terme de régularisation du lissage et font un calcul intensif du flot optique sur de grandes régions, par exemple pour la plus rudimentaire :

$$\int_D ((\nabla I \cdot V + \frac{\partial I}{\partial t})^2 + \lambda^2((\nabla V)^T \nabla V))dx \tag{2.7}$$

où D est le domaine d'intérêt, $\nabla I \cdot V + \frac{\partial I}{\partial t}$ une autre façon d'exprimer l'ECFO, $\lambda^2((\nabla V)^T \nabla V)$ le terme de régularisation, où λ est une constante à fixer, et $V = (u, v)$.

La solution pour V est donnée par les équations d'Euler-Lagrange annoncées et résolues de façon itérative. Il faut avoir des surfaces lambertiennes, une projection orthogonale et un mouvement de translation pure des objets parallèlement à la scène pour que l'ECFO soit vérifiée.

Pour palier un peu ces contraintes, Schunck [Sch85] a proposé d'utiliser un terme additionnel contenant le flot de la divergence qui exprime l'expansion/compression dans un voisinage de l'image mais en restant avec des dérivées de premier ordre :

$$\nabla I \cdot V + I(u_x + v_y) = -\frac{\partial I}{\partial t} \tag{2.8}$$

Les méthodes locales

Les méthodes locales utilisent l'information de la vitesse normale dans un voisinage local pour aboutir à la meilleure vitesse en utilisant une minimisation par les moindres carrés.

Lucas et Kanade [LK81], par exemple, ont estimé la vitesse en cherchant la minimisation sur V de

$$\sum_R W^2(x, y)(\nabla I \cdot V + \frac{\partial I}{\partial t})^2 \tag{2.9}$$

où $W(x, y)$ est une fonction imagette et R est un voisinage spatial.

2.2.3 Approches par mise en correspondance

Si le support temporel est petit ou si on a une faible proportion signal-bruit, la différentiation numérique dans les approches différentielles et fréquentielles est difficile à réaliser. Donc, on utilise les méthodes de mise en correspondance car cette approche ne dépend pas des caractéristiques significatives de l'image et une taille variable des imagettes de corrélation peut être utilisée près des bornes pour manipuler différents mouvements. Cette approche définit un déplacement (lequel est une approximation de la vitesse) comme le meilleur ajustement parmi les régions voisines variant dans le temps.

Le coefficient de corrélation de deux fonctions est défini comme :

$$C_{\Delta x}(f, g) = \int_D f(x + \Delta x)g(x)dx \tag{2.10}$$

En cherchant Δx on maximisera cette intégrale pour obtenir le déplacement de f sur g si $f(x + \Delta x) = g(x)$.

En présence de grandes disparités, les algorithmes locaux sont sensibles aux fausses correspondances dues à l'augmentation requise de recherche dans l'espace pour maîtriser les mouvements rapides.

L'algorithme d'Ogata et Sato [OS92] calcule la corrélation avec une estimation grossière du mouvement, obtenue par des filtres de Gabor accordés à la vitesse. Une telle estimation restreint la taille des zones de recherche et réduit le nombre de calculs nécessaires pour obtenir les disparités.

Mais, selon Srinivasan et ses collaborateurs [SPK99], un système de détection du mouvement par mise en correspondance fournit une indication ambiguë de la vitesse de l'image car il confond la vitesse de l'image avec sa structure spatiale, i.e la corrélation dans la zone de recherche ne sera pas nécessairement obtenue en relation à la forme en mouvement car il y aura d'autres corrélations issues de la structure de l'image.

2.2.4 Approches stochastiques

Ascher et Grzywacz [AG00] ont présenté un modèle bayésien pour mesurer la vitesse visuelle en permettant d'estimer la vitesse rétinienne avec des hypothèses plus réalistes sur la forme des filtres spatiaux et temporels. Ce modèle est consistant avec les observations de la perception de la vitesse telles que la dépendance de la vitesse perçue au contraste.

D'autre part, les processus stochastiques [GG84] ou la mise en place de filtres de Kalman [Bla92] ont été utilisés pour modéliser les intensités discontinues et manipuler plus facilement les occlusions.

D'ailleurs, les filtres de Kalman sont typiquement utilisés pour l'actualisation de l'estimation de la vélocité spatio-temporelle en temps réel (navigation et poursuite). Ils sont utilisés pour le calcul de mouvement de premier, deuxième et troisième ordre. De plus, ils réduisent l'incertitude de l'estimation sur le temps.

2.2.5 Approches fondées sur la fréquence

Jusqu'ici, les approches énumérées ci-dessus utilisent un point de vue fortement mathématique. D'autre part, les approches fréquentielles ont de forts fondements neurobiologiques et travaillent dans le domaine de la fréquence spatiale et temporelle plutôt que dans le domaine de l'espace-temps.

Approches fondées sur la phase

Les filtres fondés sur la phase sont plus robustes à la variation d'illumination de la scène que ceux fondés sur la dérivée de l'intensité ou ceux fondés sur l'énergie. Ils semblent capables de répondre aux composantes de la vitesse de multiples mouvements dus à l'occlusion ou à la transparence.

La méthode de Fleet et Jepson [FJ90] définit les composantes de la vitesse en termes de mouvement instantané au niveau de la phase des contours issue de la sortie des filtres passe-bande de Gabor accordés à la vitesse. Chaque sortie du filtre de Gabor est une valeur complexe à savoir,

$$R(x, y, t) = \rho(x, y, t)e^{i\phi(x,y,t)} \tag{2.11}$$

où $\rho(x, y, t)$ et $\phi(x, y, t)$ sont l'amplitude et la phase de la signal de sortie, respectivement.

La composante 2D de la vitesse dans la direction normale au niveau de la phase de contour est

$$v_n = \frac{-\frac{\partial \phi}{\partial t}(x, y, t)\Delta\phi(x, y, t)}{\|\Delta\phi(x, y, t)\|^2} \tag{2.12}$$

où $\frac{\partial \phi}{\partial t}(x, y, t)$ est la dérivée temporelle de la phase et $\Delta\phi(x, y, t)$ est son gradient spatial. Les dérivées de la phase sont calculées en utilisant l'identité

$$\Delta\phi(x, y, t) = \frac{Im[R^*(x, y, t)\Delta R(x, y, t)]}{|R(x, y, t)|^2} \tag{2.13}$$

où R^* est le complexe conjugué de $R(x, y, t)$, $\Delta R(x, y, t)$ est le gradient de $R(x, y, t)$ et Im est la partie imaginaire du nombre complexe.

Fleet et Jepson mettent en relation la vitesse et l'information locale de la phase car la phase du signal est relativement insensible aux variations de l'amplitude dues aux changements d'illumination de la scène.

Voilà l'utilisation de l'ECFO définie dans le domaine de la fréquence, à savoir :

$$\phi_t(x, y, t) + \Delta\phi(x, y, t) \cdot V = 0 \tag{2.14}$$

où la vitesse est $V = (u, v)$.

Malheureusement, on a besoin d'un ensemble de filtres pour couvrir l'espace de la fréquence [23] car les fonctions d'amplitude et de phase ($\rho(x, y, t)$ et $\phi(x, y, t)$, respectivement) doivent se fixer dans une plage du spectre à analyser.

Caractéristique	Différentielles	Matching	Relaxation	Fréquentielles
Ressources	+	+	+++	+++
Robustesse	++	+	++	+++
Bio-inspiration	+	-	++	+++
Performance	++	+	+++	+++
Images Naturelles	+	+	++	+++
Implantation	++	+++	+	+

TAB. 2.2 – Tableau comparatif des principales approches pour l'estimation du flot optique. De rien $-$ à très bien $+++$

Approches fondées sur l'énergie

Les approches énergétiques utilisent des filtres linéaires orientés dans l'espace-temps et accordés à la direction et à la fréquence temporelle [EBA92, Hee87, SH98].

Elles utilisent des filtres accordés à la vitesse et sensibles à l'orientation dans le domaine de Fourier et dans des images variant dans le temps. Ce sont des mécanismes sensibles au mouvement fondés sur l'énergie orientée spatiale et temporelle dans l'espace de Fourier. Ils pourraient estimer le mouvement dans les endroits où les autres approches échouent [Spi98].

Si l'image varie très faiblement dans une petite période de temps

$$I(x, y, t) \approx I(x + \Delta x, y + \Delta y, t + \Delta t), \tag{2.15}$$

la transformée de Fourier de la translation 2D du signal d'intensité sera

$$\hat{I}(k, w) = \hat{I}_0(k)\delta(kV + w) \tag{2.16}$$

[23]Typiquement on propose un ensemble des filtres directionnels pour des différentes fréquences spatiales et temporelles. Ces ensembles parfois dépassent la soixantaine [NHU04, GY90] juste pour l'aspect spatial.

où $\hat{I}_0(k)$ est la transformée de Fourier de $I(x, y, 0)$, et (x, y) est la position spatiale. δ est la fonction delta de Dirac et (k, w) représente la fréquence spatio-temporelle. Par conséquent, l'ECFO dans l'espace de la fréquence est

$$kV + w = 0 \tag{2.17}$$

laquelle montre que la vitesse d'une forme de translation 2D est une fonction de sa fréquence spatio-temporelle et forme un plan traversant l'origine de l'espace de Fourier.

Adelson et Bergen d'un côté et van Santen et Sperling d'un autre côté [AB85, vSS85] ont explicité que la détection du mouvement dans l'image est équivalent à extraire l'orientation spatio-temporelle. Le filtrage de Gabor est alors présenté comme une technique pour l'extraction de l'énergie spatio-temporelle. Adelson et Bergen ont noté que ces filtres sont sensibles au contraste du signal, c'est-à-dire, un stimulus avec un bas contraste génère une basse réponse d'amplitude et inversement.

Le tableau I.2.2 résume grossièrement les principaux atouts des approches vues ci-dessus.

Jusqu'ici j'ai fait un rappel des fondements des principales approches pour l'estimation du flot optique. Parmi elles, les approches fréquentielles semblent les plus proches de la neurobiologie. Par la suite, je vais décrire plus en détail la sous-approche fréquentielle fondée sur l'énergie (les modèles énergétiques) et leurs fortes relations biologiques.

2.3 La bio-inspiration des modèles fréquentiels énergétiques

Deux des modèles fondamentaux en psychophysique pour la perception du mouvement sont le modèle d'énergie et le modèle élaboré par Reichardt [HR56, vSS84, AB85, WAJ85] (voir aussi [AM82, Wel89, NS95]) où la détection du mouvement est limitée à un détecteur de mouvement local simple et les signaux de mouvement de détecteurs de mouvement local sont combinés dans le temps [WMG94]. Et, en suivant les propriétés des cellules ganglionnaires magnocellulaires (voir §I.1.2.1), cette combinaison temporelle semble être plus performante que la spatiale [GWM95].

Les approches énergétiques utilisent une distribution spatiale asymétrique utile pour modéliser les différentes tailles des champs récepteurs des neurones. De cette façon, ils semblent répondre assez vite lorsqu'un objet bouge dans leur direction préférée.

Avec Grzywacz et Yuille [GY90], un modèle qui dépend dans l'espace et dans le temps de l'information du mouvement en utilisant un champ récepteur fondé sur les filtres de Gabor a été proposé.

Faisons un rappel des fondements mathématiques des filtres de Gabor, puis décrire leur relation avec la physiologie cérébrale et finalement leur utilisation dans les approches énergétiques où je vais mettre en valeur la détection locale du mouvement issue des champs récepteurs des neurones de V1 et leur manipulation d'après les principes de l'approche énergétique.

Les caractéristiques que je vais mettre en valeur m'aideront pour la conception du premier module du modèle connexionniste proposé dans §II.1.

2.3.1 Les filtres de Gabor

Un filtre de Gabor est un filtre passe-bande local avec certaines propriétés de localisation optimales dans les deux domaines : spatial et de fréquence [Dau85]. Une des propriétés caractéristiques la plus importante est qu'ils ont la même expression mathématique dans les deux domaines. Cette propriété permet des implantations rapides dans les domaines spatial ou de fréquence avec leur localisation optimale [Gab46]. De plus, un filtre énergétique 3D de Gabor est un filtre spatio-temporel orienté [Hee87].

D'autre part, les "ondelettes" de Gabor sont une famille de fonctions tirées d'une gaussienne déplacée dans le temps et dans la fréquence réalisant une analyse locale de Fourier [IA02]. Cette famille de filtres est utilisée dans le traitement de signaux, en mathématiques, en physique quantique, et dans le génie électrique, comme une approximation à la transformée de Fourier à court terme. Les "ondelettes" de Gabor sont les ancêtres des ondelettes actuelles, bien qu'elles ne partagent pas leurs propriétés.

Formulation mathématique

La formulation mathématique de ces "ondelettes" a été pour la première fois proposée par Gabor [Gab46], à savoir : l'expansion d'une fonction f dans une série de fonctions élémentaires construites par blocs simples de translations et de modulations (c'est-à-dire, une translation dans le domaine de la fréquence). On représente f comme la série

$$f(t) = \sum_{n,m \in Z} c_{m,n} g_{m,n}(t) \tag{2.18}$$

où la fonction élémentaire $g_{m,n}$ est donnée par

$$g_{m,n}(t) = g(t - na)e^{2\pi imbt}, m, n \in Z \tag{2.19}$$

avec une fonction fixe g et des paramètres de déplacement dans le temps et la fréquence $a, b > 0$.

Cette définition a été largement utilisée puis adaptée pour la proposition des filtres de Gabor. Ces filtres sont utilisés pour l'analyse des signaux de la parole, l'estimation spectrale variant dans le temps, la représentation et l'identification des systèmes linéaires, la communication digitale et la représentation des images, l'analyse de la reconnaissance des formes, la vision par ordinateur, la compression vidéo, le traitement des signaux médicaux et la vision biologique. Dans ce dernier domaine, ce sont ces filtres qui modélisent la réponse des cellules simples du cortex visuel [Dau80, Mar80].

D'après Gabor [Gab46], un filtre 1D est

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma_x} exp\left(-\frac{x^2}{2\sigma_x^2}\right) exp(2\pi i f_{x_0} x) \tag{2.20}$$

où le deuxième facteur est l'enveloppe gaussienne, le troisième l'onde de modulation et f_{x_0} est le point où $g(x)$ a sa meilleure réponse fréquentielle. En développant le troisième facteur :

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma_x} exp\left(-\frac{x^2}{2\sigma_x^2}\right) cos(2\pi f_{x_0} x) + i\frac{1}{\sqrt{2\pi}\sigma_x} exp\left(-\frac{x^2}{2\sigma_x^2}\right) sin(2\pi f_{x_0} x) \tag{2.21}$$

où le premier terme est la partie réelle ou la phase cosinus et le deuxième, la partie imaginaire ou phase sinus.

La formulation de la transformée du Fourier de l'équation I.2.20 est

$$G(f_x) = A \int_{-\infty}^{\infty} g(x) exp(Bif_x x)dx \tag{2.22}$$

et, en suivant les travaux de Bracewell [Bra65] je fixe $A = 1$ et $B = -2\pi$ pour rendre plus facile la manipulation différentielle de l'équation 2.22. Alors l'intégrale devient

$$\begin{aligned} G(f_x) &= \int_{-\infty}^{\infty} g(x) exp(-2\pi i f_x x)dx \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{\infty} exp\left(-\frac{x^2}{2\sigma_x^2}\right) exp(2\pi i f_x x) exp(-2\pi i f_x x)dx \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{\infty} exp\left(-\frac{x^2}{2\sigma_x^2}\right) dx \\ &= exp(-2(\pi\sigma_x f_x)^2) \end{aligned} \tag{2.23}$$

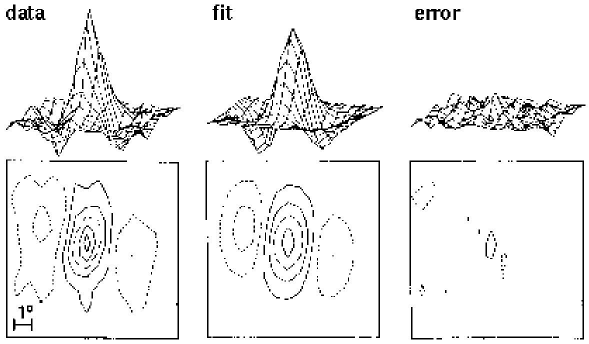


FIG. 2.6 – Adaptation d'un filtre de Gabor aux réponses d'un neurone simple quelconque de V1 (d'après Jones et Palmer [JP87]). À gauche, les données expérimentales dans le domaine spatial. Au milieu la réponse du filtre de Gabor adapté. À droite, la différence entre les deux premiers.

En 2D

$$\begin{aligned} g(x,y) &= \tfrac{1}{2\pi\sigma_x\sigma_y} exp\left(-\tfrac{x^2}{2\sigma_x^2} - \tfrac{y^2}{2\sigma_y^2}\right) exp(2\pi i(f_x x + f_y y)) \\ &= \tfrac{1}{2\pi\sigma_x\sigma_y} exp\left(-\tfrac{x^2}{2\sigma_x^2} - \tfrac{y^2}{2\sigma_y^2}\right) cos(2\pi(f_x x + f_y y)) \\ &+ i\tfrac{1}{2\pi\sigma_x\sigma_y} exp\left(-\tfrac{x^2}{2\sigma_x^2} - \tfrac{y^2}{2\sigma_y^2}\right) sin(2\pi(f_{x0} x + f_{y0} y)) \end{aligned} \tag{2.24}$$

et sa transformée de Fourier

$$\begin{aligned} G(f_x, f_y) &= \textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) exp[-2\pi i(f_x x + f_y y)]dxdy \\ &= \left[\textstyle\int_{-\infty}^{\infty} g(x) exp(-2\pi i f_{x0} x)dx\right]\left[\textstyle\int_{-\infty}^{\infty} g(y) exp(-2\pi i f_{y0} y)dy\right] \\ &= [exp(-2(\pi\sigma_x f_x)^2)] \cdot [exp(-2(\pi\sigma_y f_y)^2)] \\ &= exp[-2\pi^2((\sigma_x f_x)^2 + (\sigma_y f_y)^2)] \end{aligned} \tag{2.25}$$

Finalement, en 3D

$$g(x,y,z) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_z} exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right) exp[2\pi i(f_x x + f_y y + f_z z)] \tag{2.26}$$

où son spectre d'amplitude (transformée de Fourier) est

$$G(f_x, f_y, f_z) = exp[-2\pi^2((\sigma_x f_x)^2 + (\sigma_y f_y)^2 + (\sigma_z f_z)^2)] \tag{2.27}$$

Voilà les équations couramment utilisées dans la modélisation des réponses des cellules simples du cortex visuel primaire [Dau80, Mar80, AB85, Dau85, vSS85, DD03, GY90]. Elles sont utilisées plutôt dans le domaine de la fréquence en modulant leur amplitude et leur phase.

D'autre part, la somme des réponses de deux filtres de Gabor en quadrature (l'un en phase cosinus et l'autre en phase sinus) et avec la même orientation et largeur de bande est égale à un filtre d'énergie de Gabor [Mog00, Hee87], à savoir,

$$\begin{aligned} \hat{G}(f_x, f_y, f_z) &= \tfrac{1}{4} exp(4\pi^2[(\sigma_x^2(f_{x0} - f_x)^2 + (\sigma_y^2(f_{y0} - f_y)^2 + (\sigma_z^2(f_{z0} - f_z)^2]) \\ &+ \tfrac{1}{4} exp(4\pi^2[(\sigma_x^2(f_{x0} + f_x)^2 + (\sigma_y^2(f_{y0} + f_y)^2 + (\sigma_z^2(f_{z0} + f_z)^2]) \end{aligned} \tag{2.28}$$

où (f_{x0}, f_{y0}, f_{z0}) est l'endroit où le filtre de Gabor atteint son énergie maximale. Cette équation signifie que la réponse d'un filtre sera d'autant plus importante que le plan d'énergie non nulle qui passe à la proximité de son centre (f_{x0}, f_{y0}, f_{t0}).

Cet outil est utilisé dans [Hee87, SH92, Sim93, Mog00] pour l'estimation du flot optique. Mais il a quelques inconvénients car son manque d'orthogonalité rend le calcul très difficile et, dans la vision biologique, ce problème produit une redondance d'autant plus forte que l'échantillonnage spatial et fréquentiel est très resserré. C'est le prix de la robustesse.

L'utilisation d'un haut échantillonnage permet la représentation des filtres de Gabor simples, robustes et d'implantation rapide et fournit des reconstructions de l'image avec un haut rapport signal-bruit et une haute qualité visuelle [NT91]. Un inconvénient mineur des fonctions de Gabor est qu'elles ne sont pas passe-bande pures pour être considérées comme des ondelettes.

Il y a deux catégories de filtres de Gabor : une banque de filtres [JF91] qui a besoin d'un large ensemble de filtres avec des paramètres prédéterminés qui couvrent le plan de la fréquence (par exemple dans [Spi98, SPH98, Mog00]), d'une part, et de simples filtres de Gabor [TPH95] où leur réponses sont mises en correspondance directement avec les caractéristiques des textures et utilisées pour la segmentation et la classification, d'autre part. Avec ces derniers, le dimensionnement est très réduit mais très dépendant de l'application [BBL02].

Pour le filtrage spatial je reste dans le deuxième cas pour mon modèle connexionniste. Pour le traitement temporel je vais travailler autrement (voir §II.1.2 et §II.2).

Relation avec la physiologie cérébrale

L'utilisation de filtres spatio-temporels orientés sensibles à des stimuli en mouvement a pour origine la recherche concernant la vision des mammifères, qui avait mis en évidence la présence de cellules simples dans l'aire corticale V1 [Mar80, Dau80, MPJP85] et de cellules complexes dans l'aire MT [NGM83] se comportant comme des filtres passe-bande spatio-temporels. La réponse de ces cellules dépend de la vitesse d'un objet en mouvement ainsi que de sa forme et de ses caractéristiques fréquentielles [HR81].

D'après les études électro-physiologiques, les filtres de Gabor 3D ont été utilisés pour modéliser la structure des champs récepteurs spatio-temporels des cellules simples [WAJ85, AB85], et les cellules complexes par leur énergie [ECVK86].

Même si Canny [Can86] a montré que les dérivées gaussiennes sont optimales pour la détection de contours dans des environnements bruités, Spinei [Spi98] a montré que les filtres de Gabor ont une bonne performance dans la plupart des applications. Ils ont été largement utilisés comme outils pour des tests visuels dans des études psychophysiques et physiologiques.

2.3.2 Les approches énergétiques

La découverte que l'énergie du mouvement pourrait être calculée par les caractéristiques de premier ordre [24] d'une image a inspiré les modèles énergétiques spatio-temporels du mouvement. Ces modèles travaillent par décomposition des motifs de mouvement en ensembles 1D des variations des composantes sinusoïdales avec différentes orientations, fréquences spatiales, contrastes, et fréquences temporelles (composantes de première ordre) [AB85, vSS85, WAJ85].

Mais les modèles spatio-temporels utilisant l'information de mouvement de premier ordre ne peuvent pas prédire toute la perception du mouvement chez l'être humain et on a besoin de l'information du mouvement de second ordre [CM89], par exemple, la mise au carré [DBH92].

[24] Smith [SS94] décrit les caractéristiques de premier ordre d'une image comme : "Une description des variations spatio-temporelles en intensité et en longueur d'onde contenues dans une image...".

Le principe des méthodes énergétiques

Comme nous avons vu dans §I.2.2.5, les approches fréquentielles sont plus robustes mais très gourmandes en temps de calcul. Voici leur principe.

Pour une fréquence spatiale exprimée en cycles/pixel et pour une fréquence temporelle exprimée en cycles/image, la vitesse (distance parcourue par unité de temps ou en pixels/image) est donnée par

$$v = \frac{f_t}{f_x} \tag{2.29}$$

Si l'objet en mouvement est suffisamment texturé, son spectre est situé sur une droite de pente v. Cette pente, et donc la vitesse, peut être mesurée en utilisant les réponses de plusieurs filtres placés à des fréquences spatio-temporelles différentes.

Soit une séquence d'images décrite par la fonction de luminosité $i(x, y, t)$ et sa transformée de Fourier $I(f_x, f_y, f_t)$. Soit $i(x, y)$ la même image mais fixe et $I(f_x, f_y)$ sa transformée de Fourier. Lorsque toute l'image est en translation à vitesse uniforme $\vec{v} = (v_x, v_y)$, on peut écrire la relation

$$i(x, y, t) = i[x + \Delta_x(t), y + \Delta_y(t)] = i(x - ut, y - vt) \tag{2.30}$$

d'où l'on déduit par la transformée de Fourier

$$I(f_x, f_y, f_t) = I(f_x, f_y)\delta(f_t + v_x f_x + v_y f_y) \tag{2.31}$$

où δ représente la distribution de Dirac. Cette expression montre que le spectre d'énergie pour une image en translation a des valeurs nulles partout sauf dans un plan passant par l'origine, dont l'équation est

$$f_t + v_x f_x + v_y f_y = 0 \tag{2.32}$$

Cette équation n'est autre que la transformée de Fourier de l'ECFO (voir §I.2.1.1).

En 2D, estimer u et v revient à déterminer la pente du plan d'énergie non nulle de l'équation I.2.32 à l'aide de filtres spatio-temporels. Ces filtres ne sont pas sélectifs à une vitesse particulière mais plutôt réglés sur des fréquences spatio-temporelles. Ils ne permettent donc pas à eux seuls l'estimation de la vitesse.

La première méthode énergétique

La première méthode énergétique d'estimation du mouvement a été la méthode de Heeger [Hee87] consistant à échantillonner l'espace des fréquences spatio-temporelles au moyen de filtres de Gabor 3D puis, par une technique de minimisation, déterminer le plan de vitesse correspondant le mieux aux réponses fournies par l'ensemble des filtres. Cette méthode suppose que l'image est suffisamment texturée pour que le spectre associé soit distribué sur une large gamme de fréquences spatio-temporelles.

L'ensemble de cette méthode repose sur la connaissance de l'énergie locale liée au mouvement : cela impose donc un filtrage par convolution spatiale plutôt qu'une implantation dans le domaine fréquentiel utilisant des transformées rapides de Fourier.

Selon Spinei [Spi98], l'estimation de la vitesse est formulée comme une minimisation de l'écart quadratique entre les réponses mesurées des filtres pour une image quelconque et les réponses prédites pour une image de bruit blanc en translation.

Face au problème d'ouverture

Les modèles énergétiques résolvent le problème d'ouverture en différents pas :

1. Un ensemble de filtres de Gabor sont accordés à différentes parties de l'espace 3D des fréquences spatio-temporelles.
2. L'image est transformée du domaine espace-temps au domaine des fréquences spatio-temporelles.
3. La fréquence spatio-temporelle de l'image déclenche une activité croisée de l'ensemble des filtres de Gabor.
4. Une estimation par moindres carrés détermine le plan par lequel la forme d'activité distribuée est la meilleure.
5. Ce plan définit la vitesse de l'image (ou d'une partie de l'image).

Pour les lecteurs que voudraient aller un peu plus loin dans l'étude des méthodes énergétiques, je vous donne, dans l'annexe B, un brève parcours, son utilisation par des banques des filtres de Gabor et quelques équivalences.

Ce que nous pouvons retenir jusqu'ici est que l'ensemble de cette méthode repose sur la connaissance de l'énergie locale liée au mouvement. Par conséquent le travail se fait dans le domaine fréquentielle plutôt que dans le domaine espace-temps. L'équation I.2.32 nous propose donc de chercher un ensemble de fréquences spatio-temporelles adéquates pour proposer un système d'équations linéaires qui résoudraient cette équation et, en conséquence, le problème d'ouverture en différents pas comme nous venons de le voir (pour la définition du problème d'ouverture voir §I.2.1.2).

2.4 Bilan

Différentes propositions ont été suggérées pour aborder le problème de la perception visuelle du mouvement mais la plupart d'entre elles reviennent au même principe de base : le pré-filtrage spatio-temporel.

Un exemple est le travail de Watson et Turano [WT95]. Ils ont mesuré le contraste minimal pour discriminer la direction du mouvement en le modélisant par une famille de filtres de Gabor [25], mais en *confondant* le mouvement avec l'orientation.

D'après Derrington et Henning [DH93], il suffit de faire la détection du contraste pour l'identification de la direction du mouvement, de telle façon que les mécanismes visuels extrayant le mouvement sont étiquetés par la direction.

D'autre part, Grzywacz et ses collaborateurs [GWM95] ont proposé un modèle en utilisant plusieurs observations pour la détection des trajectoires. Dans ces trajectoires les connexions locales renforcent les réponses dans une direction définie. Ces connexions sont les analogues spatio-temporels des champs d'associations spatiaux utilisés pour la détection de contours.

Le détecteur de mouvement biologiquement plausible le plus connu est celui par mise en correspondance ou modèle Hassenstein-Reichardt ou modèle de Reichardt [HR56]. Il fait des calculs par comparaison de deux signaux lumineux consécutifs mesurés par un photo-récepteur (un filtre passe-bas). La comparaison est sur un voisinage local et il fournit une réponse complètement sélective à la direction.

[25]Selon Sekuler et ses collaborateurs [SWB02], les filtres de Gabor sont souvent utilisés comme des stimuli dans la recherche en vision. En outre, ils sont des bons approximateurs de beaucoup des champs récepteurs spatio-temporels des neurones visuels.

L'interprétation de la variation de l'intensité comme un mouvement pur relatif est restrictive car la vélocité est une quantité géométrique indépendante des conditions d'illumination. En conséquence, une illumination uniforme de la scène et une surface lambertienne sont explicitement ou implicitement supposées [BB95a].

Pour avoir une meilleure estimation du flot optique il faut avoir une acquisition robuste d'images dans le temps et l'espace en vue de mieux calculer les dérivées spatio-temporelles par des méthodes différentielles, réduire les zones de recherche par des méthodes de mise en correspondance et limiter les quantités de lissage par des méthodes fréquentielles. Nous pourrions augmenter la proportion d'acquisition des images, améliorer le pré-filtrage de l'image ou mettre un processus hiérarchique pour réduire les effets du lissage.

Les effets d'illumination sont aussi un problème dans les séquences d'images. Woodham a travaillé avec de multiples sources de lumière [Woo90], Markandey et Flinchbaugh ont travaillé avec un ensemble des contraintes multi spectrales [MF90] et Mukawa a bien modélisé les effets d'ombres [Muk93].

D'après les résultats de Barron et ses collaborateurs [BFB94] les techniques pour l'estimation du flot optique de meilleure performance sont (par ordre décroissant) :

1. Celle de Fleet et Jepson [FJ90], fondée sur la fréquence et la phase.
2. Celle de Lucas et Kanade [LK81], fondée sur le gradient avec une méthode de moindres carrés.
3. Celle d'Anandan [Ana89], fondée sur la mise en correspondance en suivant une approche multirésolution (pour plus de détails de ce type d'approche voir l'annexe B.2).
4. Celle d'Uras et al. [UGVT88], fondée sur le gradient local de deuxième ordre.
5. Celle de Singh [Sin91], fondée sur la mise en correspondance.

D'autre part, McCane et ses collaborateurs [MNGG01] concluent que la complexité des objets en mouvement pourrait aider à vérifier si un algorithme peut estimer de forts mouvements dans de petites aires, et la complexité de mouvement de la caméra pourrait vérifier si un algorithme peut estimer le flot optique dans une image.

Généralement, si nous considérons les articles revus ici, les techniques fréquentielles sont meilleures que les différentielles et elles-mêmes meilleures que celles de mise en correspondance. Même si je regrette la gourmandise en calcul des approches fréquentielles, leur proximité aux principes d'inspirations neurobiologiques reste très intéressante.

Dernièrement, il semble que les chercheurs dans ce domaine s'inclinent de plus en plus vers les approches stochastiques à cause de l'essor en probabilité et son application aux modèles bio-inspirés.

Il faut garder à l'esprit que toutes ces méthodes doivent faire face à :

1. *La perte d'information due à l'échantillonnage.* Un échantillonnage très faible dans une image peut entraîner des effets forts gênants de repliement du spectre spatial. De même, des problèmes de repliement temporel peuvent apparaître, avec des effets néfastes surtout pour les techniques différentielles fondées sur l'ECFO proposée par Horn et Schunck [HS80].
2. *Le problème des régions uniformes.* Dans des régions homogènes non-texturées en mouvement, il est impossible de percevoir des changements de luminosité entre deux images successives et donc de faire l'estimation des vecteurs de vitesse. Il faut prendre en compte l'information de contour et de mouvement dans l'image.
3. *Le problème d'occlusion.* Lorsque deux objets se cachent partiellement au cours de leur mouvement, il est très difficile de savoir à quel objet correspond le mouvement estimé au

niveau de la frontière entre ces deux objets. De plus, quand une région se déplace, elle en découvre une autre située en-dessous. Les pixels apparus ne peuvent être corrélés entre deux trames puisqu'ils n'existent que dans une seule. Donc, l'hypothèse de la conservation de l'intensité des points entre deux images est alors mise en échec. Les occlusions sont la cause des mouvements discontinus et les méthodes de mise en correspondance y sont sensibles. Les occlusions ont été bien décrites analytiquement dans l'espace de Fourier [BB95b, FL95].

4. *Le problème d'ouverture.* Il est lié à la taille spatio-temporelle des filtres utilisés dans la détection du mouvement. Lorsque cette ouverture est réduite, on dispose d'une excellente spécificité positionnelle, mais d'une très mauvaise spécificité fréquentielle. On va donc seulement obtenir une estimation correcte pour la composante de vitesse perpendiculaire au contour, alors que la composante tangentielle reste totalement indéterminée et apparaît nulle dans la zone d'observation. Le problème d'ouverture est moins fort dans les méthodes fréquentielles que dans les méthodes différentielles grâce au lissage intrinsèque introduit par le filtrage spatio-temporel.

D'autre part, il y a des modèles qui interprètent l'image en mouvement comme la projection de la surface des objets en mouvement rigide dans le temps et essaient de récupérer leurs paramètres de mouvement : trois pour la translation et trois pour la rotation en permettant la représentation d'un mouvement rigide 3D général entre un observateur et la surface d'un objet [HR03].

Dans ce rapide parcours des approches pour l'estimation du flot optique, nous voyons que la robustesse a un prix excessif en calcul, mais elle se fonde sur des propositions tirées de la recherche en neurobiologie.

Je vais insérer les filtres de Gabor dans mon modèle connexionniste pour la modélisation des champs récepteurs de neurones simples de V1 et j'en ferai une manipulation différente pour le traitement temporel (voir §II.1.2 et §II.2).

Par la suite, nous allons faire un survol parmi les modèles bio-inspirés dédiés à la perception du mouvement où nous retrouverons l'application de certaines des approches vues dans ce chapitre.

Chapitre 3

La modélisation bio-inspirée

Sommaire

Dans ma démarche nous avons déjà vu les principes neurobiologiques de la perception visuelle du mouvement (§I.1) et nous avons parcouru les différentes approches pour sa modélisation (§I.2). Nous allons explorer dans ce chapitre les modèles bio-inspirés qui mettront en valeur ces principes et feront des appels à l'une ou l'autre approche.

Dans les systèmes de vision biologiques et artificiels, l'analyse du mouvement commence avec des mesures locales telles que les sorties des cellules sélectives à la direction dans V1 [Nak85] ou des opérateurs de dérivées temporelles et spatiales dans les systèmes artificiels [HS80, LK81]. Puis, elles sont intégrées pour générer des descriptions plus larges et globales du mouvement. Cette intégration est essentielle car, comment nous l'avons vu dans le chapitre précédant, les mesures du mouvement local sont ambiguës.

Nous avons aussi vu que les aires V1 et MT/MST sont les premières actrices du traitement du mouvement et de la localisation spatiale (voir §I.1.3). Donc, de manière simplifiée V1 traite plutôt l'orientation du mouvement, MT sa direction et sa vitesse et MST le mouvement propre et le mouvement complexe (§I.1.2.5).

Les premiers efforts dans la modélisation bio-inspirée ont été focalisés sur le comportement des neurones simples dans V1 puis, sur le partage de leur information locale et leurs interactions latérales inter- et intra-colonnes d'orientation (voir §I.1.3.1). Ils ont continué par les neurones complexes et hypercomplexes pour aboutir à l'intégration des sorties de V1 dans MT, où un traitement différent des colonnes d'orientation fait émerger la vitesse des zones de mouvement cohérent (voir §I.1.3.2).

Ce chapitre effectue un survol des différents modèles pour la perception du mouvement inspirés de la neurobiologie. En commençant par la modélisation des neurones simples de V1 jusqu'à la modélisation des aires corticales supérieures comme MST, je me concentrerai sur les aires V1 et MT et je choisirai deux modèles représentatifs pour les présenter plus en détail.

3.1 Les principales aires cérébrales de la voie dorsale

Il est couramment accepté que les filtres de Gabor spatio-temporels sont fortement mis en relation avec la fonction des cellules simples dans V1 chez les primates et leur énergie avec la réponse des cellules complexes (voir §I.1.3.1 et §I.2.3 [KB91, Dau80, Mar80]).

De la modélisation des cellules simples par filtres de Gabor à leur interaction locale, différentes propositions ont été faites. D'une part, Daugman [Dau88] a proposé un réseau neuronal avec des champs récepteurs inspirés des cellules simples de V1 en les paramétrant par une famille de filtres de Gabor 2D. D'autre part, les modèles énergétiques (voir §I.2.3) calculent l'énergie d'une famille de ces filtres qui permet de modéliser de nombreuses propriétés des cellules complexes [PR83, EBA92, Hee92b] et, donc, de la perception du mouvement. Ce travail a abouti à des modèles pour le calcul de la vélocité soit par l'énergie d'une famille de filtres de Gabor [Hee87], soit par la combinaison des sorties énergétiques d'un grand nombre de filtres spatio-temporels [GY90], soit par différentes familles en fréquence de filtres de Gabor [NHU04].

Les mécanismes d'énergie spatio-temporels sélectifs à un intervalle de différentes fréquences spatiales et temporelles donnent des réponses maximales lorsque leur fréquences préférées sont localisées dans un plan. Heeger [Hee87] et Grzywacz et Yuille [GY90] font la détection du dit plan par la maximisation de l'énergie du mouvement. Par contre, Nagano et ses collaborateurs [NHU04] la font par l'intégration des différentes familles de filtres de Gabor à différentes fréquences.

Ces travaux ont donné une façon de voir et d'interpréter les interactions latérales des neurones simples et complexes issus de V1 et leur contribution respective aux réponses des neurones de MT (voir §I.1.3). Le tableau I.3.1 résume de façon générale la modélisation des cellules corticales et les premiers modèles informatiques pour l'estimation de la vitesse.

Modélisation bio-inspirée (I)		
Cellules Corticales	Simples	Kulikowski et Bishop [KB91]
		Daugman [Dau80]
		Marcelja [Mar80]
	Complexes	Pollen et Ronner [PR83]
		Emerson et al. [EBA92]
		Heeger [Hee92b]
Premiers modèles informatiques	Réseaux neuronaux de type Gabor	Heeger [Hee87]
		Daugman [Dau88]
		Grzywacz et Yuille [GY90]
		Nagano et al. [NHU04]

TAB. 3.1 – Quelques premiers modèles bio-inspirés

Dans cette section je ferai un rappel des divers modèles inspirés des aires de la voie dorsale : V1, MT et MST. Puis, dans la section suivante (§I.3.2), je vais en choisir deux pour les décrire plus en détail.

3.1.1 Le groupe V1-MT-MST

La recherche sur la perception du mouvement au niveau cortical a mis en valeur les aires V1, MT et MST (voir §I.1.2.1). Par contre, la participation des aires V2 et V3 à la perception du mouvement reste à étudier.

V1

Les interactions dans V1 peuvent être modélisées soit en intégrant un processus de diffusion et un processus d'inhibition localement connecté [WWS90, Mog00], soit par de simples interactions excitatrices-inhibitrices [Gro73, GL75, Gro76, Gro88] et avec de rétro-actions [LN04], soit par l'intégration orientée du mouvement dans un petit voisinage [RDSW92, RSD$^+$00, FT02, Gil85].

- Waxman et ses collaborateurs[WWS90] proposent un modèle à deux niveaux : un de diffusion, l'autre de rehaussement. Le premier étale les entrées dans le temps et transmet ses réponses discrétisées au deuxième niveaux. Celui-ci rehausse le contraste dans les zones les plus actives puis la réponse est rétro-propagée vers le premier niveau selon les principes de la mémoire à court terme proposé par Grossberg [Gro73, GL75, Gro76, Gro88]. Moga [Mog00] a proposé une variation de ce dernier principe.
- Les principes de la mémoire à court terme [Gro73, GL75, Gro76, Gro88] ont été appliqués à la modélisation des interactions latérales au niveau d'un neurone cortical (voir §I.1.3.1). Ces fortes interactions locales d'excitation et surtout d'inhibition pourraient affaiblir l'émission d'un potentiel d'action. En outre, Latham et Nirenberg[LN04] ont proposé que la rétro-action des interactions excitatrices et particulièrement les inhibitrices créent un environnement stabilisateur qui permet aux ensembles de neurones plus activés de créer une zone d'activité forte.
- D'autre part, les interactions entre les colonnes corticales semblent servir à distribuer de façon compétitive les sorties d'une colonne vers une autre colonne [RDSW92, RSD$^+$00]. À cause de cette interaction inter-colonne, les neurones participants parviennent à une sorte d'adaptation corticale [FT02], i.e. à une étape de stabilisation.
  En outre, l'activation directe (excitation) d'une petite région du cortex produirait des péristimuli inhibiteurs [Gil85], fait que Kohonen a utilisé pour focaliser les zones d'activité [Koh84]. Enfin, le traitement des colonnes d'orientation et leur interaction font émerger différentes propriétés auto-organisatrices. Il est aussi couramment accepté que les interactions ON-OFF des cellules ganglionnaires contribuent à une telle dynamique (voir §I.1.3.1).

Bref, ce que nous pouvons retenir ici est la mise en œuvre de fortes interactions neuronales au niveau de V1 dans les modèles présentés ci-dessus. Ces interactions coopératives-compétitives parmi les différentes colonnes d'orientation permettant la convergence d'une auto-organisation locale, distribuée et orientée sont notamment mises en valeur par mon modèle connexionniste où je prends en considération ces idées (voir §II.1).

V1-MT

L'interaction entre V1 et MT peut aussi se voir soit comme un simple processus en pro-action [SGLF02] (défini ci-dessous), soit comme un processus compétitif-coopératif dans V1 [SH98], soit comme l'intégration de ses sorties et de leur segmentation au niveau de MT [Min03].

Dans un processus en pro-action, un long champ récepteur avec recouvrement partiel [SGLF02] permet de modéliser les propriétés dynamiques basiques des neurones de V1 et leur sélectivité à la vitesse par la lecture des réponses spatio-temporelles. Puis, dans l'étape de type MT, les sorties de V1 sont passées par un filtrage passe-haut et rectifiées en rehaussant les réponses éphémères de V1 et en supprimant les réponses qui varient plus lentement dans le temps.

Mais Simoncelli et Heeger [SH98] montrent qu'un modèle séquentiel pondéré, rectifié et normalisé des aires V1 et MT en aval pourrait mettre en valeur un processus d'inhibition par division. L'interaction entre V1 et MT y extrait la structure du mouvement à travers un calcul neuronal local, coopératif et compétitif.

Cette proposition autorise donc une inhibition compétitive qui permettrait aux neurones plus activés de coder la direction du mouvement.

Un autre point de vue est de considérer les interactions entre V1 et MT comme un processus d'intégration et de segmentation du mouvement [Min03] où l'intégration est vue comme le résultat de différentes étapes de traitement : un filtrage orienté à petit rayon et des interactions compétitives qui mettent en valeur les zones de mouvement et qu'on repère pour maintenir la poursuite. Puis un filtrage à long rayon d'action qui augmente la réponse des cellules directionnelles et, finalement, une population distribuée codant la vitesse préférée.

Ensuite, une compétition spatiale et des inhibitions à la direction opposée suivies d'une rétro-action coopérative/compétitive à long rayon d'action (identifiée avec les aires MT et MST) réorganiserait les signaux directionnels ambigus.

C'est donc un processus en pro-action hiérarchique qui s'impose d'après les premiers résultats de la recherche en neurophysiologie cérébrale. Mais la découverte de fortes interactions coopératives-compétitives dans les aires corticales nous invite à chercher d'autres chemins où les interactions locales, distribuées, coopératives et compétitives dans V1 semblent avoir un rayon d'action aussi local. Ainsi, les neurones corticaux supportent ces interactions et leur réponse est intégrée dans MT. Ce sont les principes que j'ai retenus (voir §II.1).

MT

Les neurones dans MT sont accordés aux vélocités des caractéristiques extraites principalement dans V1 et répondent mieux aux zones de mouvement cohérent (voir §I.1.3.2). On peut les modéliser soit simplement par des différences de gaussiennes [FWQ02, TP99, TAAH95] soit, de plus, par une combinaison des sorties excitatrices issues de V1 et une interaction latérale excitatrice-inhibitrice au niveau de MT [Koe96]. Cette dernière approche fait que les interactions excitatrices-inhibitrices propagent spatialement la vitesse en respectant les indices visuels de segmentation (comme les contrastes). Puis, elle met en interaction latérale les neurones de MT pour aboutir à une auto-organisation.

Le traitement au niveau de MT diffère de celui au niveau de V1. Dans le dernier les interactions latérales sont inter- et intra-colonnes d'orientation, dans le premier elles intègrent les vitesses plutôt que l'orientation (voir §I.1.3.2). Le modèle proposé intègre ces deux types d'interactions dans le deuxième module (voir §II.1.3 et §II.3).

MST

Au niveau de MST la segmentation du mouvement cohérent [ZS98] et du mouvement propre se ferait dans l'interaction de ses deux voies en auto-récurrence : la ventrale (MSTv) et la dorsale (MSTd) [PGM00].

Les interactions entre MSTv et MSTd semblent traiter le mouvement de l'objet et du fond, respectivement. Leurs cellules répondraient mieux soit aux stimuli des petits mouvements en éliminant les gros mouvements (dans MSTv), soit aux gros mouvements en éliminant les petits mouvements (dans MSTd) [PGM00]. Bref, on a une répartition des cellules de MT et de MST en unités inhibées ou excitées avec une structure auto-récurrente de MST en aval.

Modélisation bio-inspirée (II)			
Aire	*Type*	*Auteurs*	*Simulé/Validé*
V1	Diffusion-Inhibition	Waxman et al. [WWS90]	√/√
		Moga [Mog00]	√/×
	Exc-Inh-Rétro-action	Latham et Nirenberg [LN04]	×/′
	Interaction Orientée	Reggia et al. [RDSW92, RSD$^+$00]	√/×
		Fellez et Taylor [FT02]	√/×
		Gilbert [Gil85]	√/×
V1-MT	Pro-action	Sèries et al. [SGLF02]	×/′
	Compétitif-coopératif	Simoncelli et Heeger [SH98]	√/′
	Intégration-Segmentation	Mingolla [Min03]	√/×
MT	Différences Gaussiennes	Fernández et al. [FWQ02]	√/√
		Todd et Perotti [TP99]	√/×
		Treue et al. [TAAH95]	√/×
	Intégration Exc-Inh	Koechlin [Koe96]	√/′
MST	Mouvement cohérent	Zemel et Sejnowski [ZS98]	√/×
	Mouvement propre	Pack et al. [PGM00]	√/′

TAB. 3.2 – Modèles bio-inspirés énoncés dans ce chapitre

D'après ces idées, MST serait responsable du traitement du mouvement propre par la mise en corrélation de l'information rétinienne avec celle issue de V1 (voir §I.1.2.5). Ce traitement envoie ses réponses en retour vers les aires inférieures (MT et V1) pour mettre en valeur les mouvements des objets dans la scène et en inhibant le mouvement propre.

Mais, si on suppose que les neurones de MST segmentent partiellement les champs de mouvement, une technique d'optimisation non supervisée fournirait des unités accordées à des formes signalant un mouvement cohérent [ZS98]. Alors, un modèle d'une mémoire bi-directionnelle pour MSTd peut tout d'abord segmenter les champs de mouvement de différents objets mobiles selon les formes cohérentes du mouvement de chaque objet puis récupérer l'information de cet objet.

Dans le modèle proposé, je m'inspire des interactions entre V1 et MT, mais il restera dans mes perspectives la démarche du traitement du mouvement propre (au niveau de MST) et la rétro-action des sorties de MST vers les aires V1 et MT, et celle de MT vers V1.

Le tableau I.3.2 résume les modèles bio-inspirés cités ci-dessus. Dans la dernière colonne, appelée *Simulé/Validé*, j'ai mis un √ dans le côté *Simulé* pour les modèles rapportés qui n'ont été testés qu'avec des images synthétiques ou des cinématogrammes [26], et un × sinon. Dans le coté *Validé* on a mis un √ pour les modèles rapportés et qui ont été validés avec des images réelles depuis leur acquisition, un ′ pour ceux validés seulement pour les données psychophysiques et × pour ceux qui ne sont testés que pour la simulation.

Dans la sous-section suivante nous allons faire un rappel très général de comment les modèles connexionnistes classiques ont été adaptés pour percevoir le mouvement.

[26] En psychophysique, pour faire de tests visuels on génère des points lumineux aléatoires ayant des mouvements réguliers sur un écran d'ordinateur. On peut les gérer en vitesse, en orientation et en quantité.

3.1.2 La modélisation en connexionnisme classique

Le problème de la perception du mouvement a été aussi abordé par l'adaptation des modèles neuronaux classiques parmi lesquels je vais citer quelques modèles utilisant soit le perceptron multi-couches, soit les cartes auto-organisatrices, soit des réseaux complètement connectés.

Modèles issus des réseaux de neurones en pro-action

L'utilisation d'un perceptron à deux couches comme un simple classificateur a été décrite soit pour estimer la vélocité d'un objet en mouvement [JD03], soit pour modéliser MT-MSTd [BV98, BWV03].

Dans le premier cas, les sorties normalisées des cellules complexes issues des filtres énergétiques de Gabor sont l'entrée du perceptron. Le filtrage est fait en trois fréquences spatiales et cinq fréquences temporelles, toutes les huit différentes. Son unique sortie donne la vitesse de l'objet en mouvement.

Dans le deuxième cas, les entrées du réseau sont déjà les réponses de MT dans seize directions différentes et la couche cachée représente MSTd et sa sortie les propriétés de MST.

D'autre part, il y a des modèles de réseaux de neurones en pro-action [Ser93] qui s'inspirent du perceptron multi-couches où on cherche à stabiliser les interactions latérales dans chaque couche [HN93, FCMMFGD03].

Un modèle simple en pro-action fait une correspondance entre les données de l'image et les propriétés de l'objet d'intérêt [Ser93]. La vélocité de cet objet y est une fonction linéaire des composantes de la vitesse (pour une orientation constante) mais pas des composantes de la direction.

Enfin, notons deux propositions inspirées du perceptron multi-couches, une avec un réseau en deux couches connectées dans la direction préférentielle du mouvement et dans un voisinage de rayon 1 pour chaque neurone de la couche cachée [HN93], ou encore une autre avec une interaction latérale forte et cumulative, i.e les sorties de la couche supérieure et celles des interactions latérales de la périphérie sont accumulées-intégrées au centre et redistribuées vers la périphérie [FCMMDFG03].

Modèles issus des réseaux de neurones auto-organisateurs

Dans le contexte auto-organisateur, un réseau de type Kohonen fournit une représentation d'une orientation en particulier [RV03], ou encore, considère l'ECFO comme un problème d'optimisation [Sto01].

Dans la première proposition le réglage des compétitions est fondé sur un mécanisme de type Kohonen sur des champs récepteurs de type ON-OFF et un autre mécanisme identique pour les compétitions de ces neurones permet d'extraire une représentation dans une orientation particulière.

D'autre part, deux réseaux de neurones, un du style WTA [27] et un autre entièrement connecté, résolvent l'ECFO en faisant une optimisation [Sto01]. Le premier cherche les discontinuités du mouvement et le deuxième fournit une perception sélective du mouvement visuel en accord avec un mouvement préférentiel donné.

Enfin, les modèles cités dans cette section reprennent certaines caractéristiques du comportement des neurones corticaux : leur localité, leur interaction latérale, leur intégration hiérarchique,

[27] Winner-Take-All (le gagnant prend tout) : Le WTA simule les mécanismes de compétition existant entre neurones ou populations de neurones liés par des connexions inhibitrices latérales. Après convergence, seul le neurone ayant la plus grande activité reste actif et inhibe tous les autres.

etc. Je vais par la suite détailler deux modèles bio-inspirés pour mettre en valeur ces caractéristiques.

3.2 Deux modèles bio-inspirés

Après un rapide parcours des modèles connexionnistes bio-inspirés, je vais détailler deux modèles retenus parmi eux. Ceux-ci montrent les interactions de V1, MT et MST en général. Le premier présente les interactions des neurones de V1 et de MT sous un schéma identique pour les deux aires, où une intégration hiérarchique nous fournit la réponse à la vitesse. Par contre, le deuxième met plus en valeur les relations neuronales intra-corticales au niveau de MST, en faisant une première discrimination entre le mouvement propre tiré de l'analyse des signaux extra-rétiniens (comme les mouvements des yeux) et les réponses fournies par MT.

Après la présentation de ces deux modèles je vais mettre en valeur, dans §I.3.3, les principales caractéristiques retenues des trois premiers chapitres de la première partie de ce mémoire.

3.2.1 Modèle Simoncelli-Heeger

Simoncelli et Heeger [SH98] proposent une inhibition particulière par division dans V1 où la réponse de chaque neurone est divisée par une quantité proportionnelle à l'activité sommée d'un voisinage spatio-temporel, puis une inhibition soustractive dans MT où la réponse d'un neurone sélectif à la vitesse somme les réponses afférentes compatibles en direction et en vitesse provenant de V1 et soustrait les réponses de V1 non compatibles en direction et en vitesse.

Architecture du modèle

L'interaction entre V1 et MT extrait la structure du mouvement à travers un calcul neuronal local et compétitif caractérisé par (voir la figure I.3.1) :

1. Un calcul local du stimulus de contraste (niveau 0).
2. Une modélisation des neurones simples et complexes de V1 en utilisant la troisième dérivée gaussienne spatio-temporelle et une mise en commun spatiale (niveau 1 et 2).
3. Une modélisation des neurones de MT en additionnant les réponses des cellules de V1 (niveau 3 et 4).
4. Un calcul de l'élément gagnant pour chaque pixel du champ visuel.

L'étape V1 (niveaux 0, 1 et 2) et l'étape MT (niveaux 3 et 4) calculent de façon indépendante la somme linéaire pondérée des entrées suivie d'une rectification et d'une normalisation par division.

La forme du calcul est identique dans chaque étape. Les poids des champs récepteurs de V1 sont fixés par la sélectivité à l'orientation et à la direction, pendant que les champs récepteurs de MT sont fixés par la sélectivité à la vélocité (direction et vitesse).

L'entrée du modèle est un stimulus visuel variant dans le temps et sa sortie correspond à l'équilibre dynamique du taux de déclenchement des neurones de MT. Ces valeurs forment une représentation distribuée (codification de la population) de la vélocité de l'image pour chaque région spatiale du stimulus visuel.

Donc, la réponse neuronale est une sorte de somme pondérée sur tout l'espace local et un passé récent du contraste dans ce stimulus.

Pour prendre en compte les importantes non-linéarités des cellules simples comme la saturation et l'inhibition aux orientations croisées, les auteurs ont normalisé la réponse de leur champs

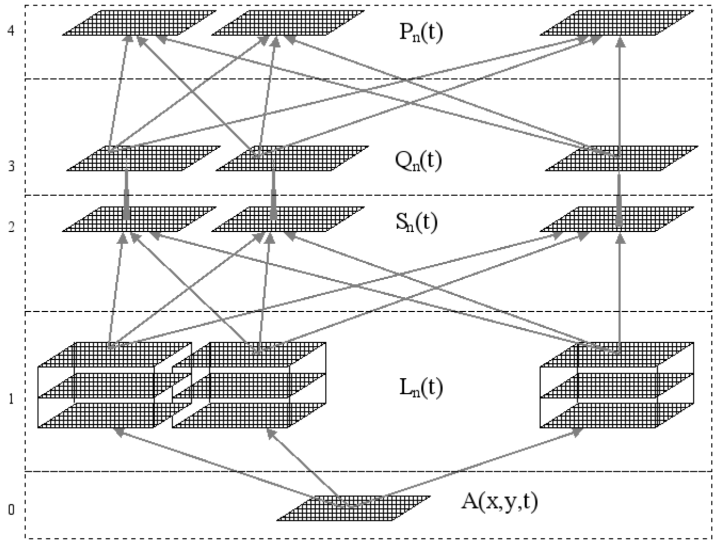


FIG. 3.1 – Architecture du modèle Simoncelli-Heeger. Niveau 0 : Acquisition de l'image. Niveau 1 : Réponse linéaire de la n-ième cellule simple. Niveau 2 : Réponse de la n-ième cellule simple. Niveau 3 : Réponse linéaire sélective à la vitesse sur un moyenne locale des réponses de V1. Niveau 4 : Réponse du n-ième neurone en MT.

récepteurs [Hee93, Hee92a] en les divisant par l'opération de normalisation de chaque neurone dans un petit voisinage l'entourant (voir l'équation I.3.4).

D'autre part, les cellules complexes font la somme pondérée des cellules simples afférentes et distribuées sur une région spatiale locale où chaque cellule a la même orientation et la même phase dans l'espace-temps.

Équations de Simoncelli-Heeger

Dans une vision monoculaire et en échelle de gris, un stimulus visuel $I(x, y, t)$ est représenté par le contraste local (cfr. [SEC84])

$$A(x,y,t) = \frac{I(x,y,t) - \bar{I}}{\bar{I}} \tag{3.1}$$

où $\bar{I}$ est l'intensité moyenne du stimulus dans l'espace et le temps.

La réponse linéaire de la n-ième cellule simple $L_n(t)$ est la somme pondérée du contraste local plus une valeur constante α_1

$$L_n(t) = \int\int\int s_n(x,y,T)A(x,y,t-T)dxdydT + \alpha_1 \tag{3.2}$$

Les fonctions de pondération $s_n(x, y, T)$ sont un ensemble de dérivées directionnelles de troisième ordre d'une gaussienne avec 28 orientations différentes dans l'espace-temps pour chaque imagette.

Puis, la réponse de la n-ième cellule simple $S_n(t)$ est

$$S_n(t) = \frac{K_1 \lfloor L_n(t) \rfloor^2}{\sum_m \lfloor L_m(t) \rfloor^2 + \sigma_1^2} \tag{3.3}$$

où $\sigma_1 \neq 0$ est la demi-saturation constante de la normalisation, K_1 est la réponse maximale et $\lfloor\cdot\rfloor$ est l'opération de normalisation définie comme

$$\lfloor L_n(t)\rfloor^2 \equiv max[0, L_n(t)]^2 \tag{3.4}$$

La réponse normalisée aura toujours une valeur entre 0 et K_1, K_1 étant la saturation pour les hauts contrastes. L'équation I.3.3 modélise directement ce qui passe chez l'être vivant par un mécanisme inhibiteur de rétro-action [Hee93].

La réponse des cellules complexes de V1 est calculée par la moyenne locale des réponses des cellules simples

$$C_n(t) = \sum_m c_{nm} S_m(t) \tag{3.5}$$

sur une région spatiale locale avec la même orientation spatio-temporelle et la même phase et avec tous les poids $c_{nm} > 0$.

La réponse linéaire sélective à la vélocité $Q_n(t)$ d'un ensemble de cellules de MT est

$$Q_n(t) = \sum_m p_{nm} C_m(t) + \alpha_2 \tag{3.6}$$

où p_{nm} est l'ensemble des poids et α_2 le niveau de réponse minimale. L'indice n caractérise la position spatiale et la sélectivité à la vélocité de la cellule.

Finalement, la réponse d'un neurone dans MT $P_n(t)$ est

$$P_n(t) = \frac{K_2\lfloor Q_n(t)\rfloor^2}{\sum_m \lfloor Q_m(t)\rfloor^2 + \sigma_2^2} \tag{3.7}$$

où σ_2 et K_2 déterminent le niveau de demi-saturation et la valeur maximale de réponse d'une cellule dans MT, respectivement.

Donc, la partie centrale du modèle a été donnée par les équations I.3.3 et I.3.7. La rectification exprimée dans l'équation I.3.4 combinée à la normalisation réalise une sorte d'inhibition soustractive-divisive. Les différentes orientations impliquées convergent vers les cellules complexes qui calculent une énergie de mouvement. De plus, elles ont été pondérées (voir les équation I.3.5 et I.3.6) pour les intégrer dans l'étape MT [28].

D'après les équation I.3.3 et I.3.7, l'inhibition compétitive du modèle permet aux neurones les plus activés de coder la direction du mouvement.

Ce modèle a été confronté aux données expérimentales des réponses des cellules de V1 et MT avec des résultats proches.

3.2.2 Modèle Pack-Grossberg-Mingolla

Le modèle de Pack et ses collaborateurs [PGM00] se focalise sur les interactions entre MSTv et MSTd supposées traiter le mouvement de l'objet et du fond, respectivement. Leur cellules peuvent être classées en deux types : celles répondant mieux aux petits stimuli de mouvement et en éliminant les gros mouvements (dans MSTv), et celles répondant mieux aux gros mouvements (dans MSTd).

[28] Les valeurs des paramètres pour les expériences ont été publiées par Simoncelli et Heeger [SH98], à savoir : $\sigma_1 = 0.2$, $\alpha_1 = 0.07$, $K_1 = 4$, $\sigma_2 = 1$, $\alpha_2 = 0.8$ et $K_2 = 1.2$.

Architecture du modèle

Le modèle a deux niveaux. Le premier a 200 cellules de type MT accordées à la direction et à la vitesse du mouvement de l'image rétinienne perçue dans leurs champs récepteurs. La moitié de ces cellules a un voisinage inhibiteur de telle façon que soient éliminés les gros stimuli de mouvement dans une direction cohérente. L'autre moitié a un voisinage excitateur en exhibant des fortes réponses à des grosses tailles de stimuli.

Chaque neurone de MST peut être vu comme appartenant à un canal de manipulation du mouvement des yeux dans une direction préférée où il y a trois types de connexions excitatrices traitant : la vitesse de la cible dans une direction préférée du mouvement des yeux (calculée par MSTv), une copie afférente de la vitesse des yeux dans la direction préférée du mouvement des yeux (signal en rétro-action de MSTv à MSTd) et le mouvement du fond dans la direction opposée à la direction préférée du mouvement des yeux (calculé dans MSTd).

Le mélange des signaux provenant du mouvement de la cible et de la rotation des yeux est résolu par le calcul du mouvement du fond fait par les cellules de MSTd, lesquelles excitent celles de MSTv accordées à la direction opposée (voir la figure I.3.2).

L'interaction entre MSTv et MSTd peut être interprétée comme les liens des cellules représentant la fixation du fond (MSTv) et de la profondeur (MSTd) relatifs à un objet.

Le mouvement provoqué (de l'ordre de pas plus de 20 ° par seconde [PM98]) est le mouvement perçu d'une cible dans la direction opposée au mouvement du fond.

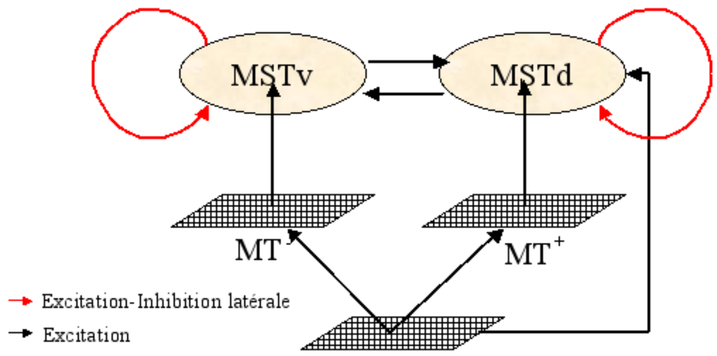


FIG. 3.2 – Architecture du modèle Pack-Grossberg-Mingolla.

Le modèle capture les interactions entre MT et MST dans des petits intervalles de temps en utilisant une dynamique de premier ordre pour un système de poursuite simple.

Équations de Pack-Grossberg-Mingolla

La vitesse de la poursuite p est

$$\frac{dp}{dt} = -p + S(x_4 - x3) \tag{3.8}$$

où x_3 et x_4 sont les sorties de MSTv pour la poursuite gauche et droite, respectivement. S est la corrélation neuronale de la décision consciente de la poursuite d'une cible [GKL92] : $S = 1$, poursuite ; $S = 0$, fixation. $p > 0$, poursuite à droite, $p < 0$ poursuite à gauche.

L'entrée du système est le champ de vecteurs de mouvement où $v(x, y)$ est sa vitesse au point (x, y) de la rétine. Les valeurs x et y sont normalisées entre $[-1, 1]$. Cette entrée est représentée par des carrés de taille r bougeant horizontalement dans le champ visuel. L'objet est distingué du fond par

$$O(x,y) = \begin{cases} 1 & si|x - x_0| \leq \frac{r}{2} \\ 0 & si|x - x_0| > \frac{r}{2} \end{cases} \tag{3.9}$$

et

$$\hat{v}(x,y) = O(x,y)v_0 - p \tag{3.10}$$

où $\hat{v}$ est la vitesse de l'image rétinienne, x_0 le centre de l'entrée carrée et v_0 sa vitesse.

Soit (i,j) le centre du champ récepteur avec une largeur h comme fonction de la position du champ récepteur dans le champ visuel

$$h(i,j) = \frac{H}{i^2 + j^2} \tag{3.11}$$

où H contrôle l'échelle du champ récepteur par rapport à l'excentricité.

Pour chaque cellule de MT caractérisée par sa structure centre-périphérie, sa réponse au centre α_{ij}^c est

$$\alpha_{ij}^c = \sum_{x,y} exp(-G(s_{ij} - \hat{v}(x,y))^2)exp(-h(i,j)((i-x)^2 + (j-y)^2)) \tag{3.12}$$

et pour la périphérie

$$\alpha_{ij}^p = \sum_{x,y} exp(-G(s_{ij} - \hat{v}(x,y))^2)exp\left(-\frac{h(i,j)}{25}((i-x)^2 + (j-y)^2)\right) \tag{3.13}$$

pour $H = 25$ (laquelle nous donne une excentricité de 70%) et G la différence gaussienne.

Un neurone de MT^+ sera

$$\beta_{ij}^+ = \alpha_{ij}^c + \alpha_{ij}^p \tag{3.14}$$

et un neurone de MT^- sera

$$\beta_{ij}^- = \alpha_{ij}^c - \alpha_{ij}^p \tag{3.15}$$

Il y a la même quantité de cellules MT^+ et MT^- [BT92]. Chaque cellule a été mise à une distance d de la fovéa avec une probabilité de $exp(-Pd^2)$, où P est la distribution des positions.

Pack et ses collaborateurs [PGM00] ont mis en correspondance les neurones de MT^+ avec les neurones de MSTd et ceux de MT^- avec MSTv.

L'activation des cellules MSTd est [Hod64]

$$\begin{aligned} \tfrac{dx_1}{dt} &= -x_1 + (1 - x_1)\left(x_3 + C\textstyle\sum_{i,j}\beta_{ij}^+ - p\right) - Fx_1x_2 \\ \tfrac{dx_2}{dt} &= -x_2 + (1 - x_2)\left(x_4 + C\textstyle\sum_{i,j}\beta_{ij}^+ + p\right) - Fx_1x_2 \end{aligned} \tag{3.16}$$

où C et F sont les paramètres montrant la force relative des connexions d'excitation et d'inhibition respectivement. x_1 et x_2 décrivent le déclin de l'équilibre potentiel. Le dernier terme décrit l'inhibition du deuxième composant par le premier.

L'activation des cellules MSTv est

$$\begin{aligned} \tfrac{dx_3}{dt} &= -x_3 + (1 - x_3)\left(x_1 + M\textstyle\sum_{i,j}s_{ij}\beta_{ij}^- - p\right) - Jx_3x_4 \\ \tfrac{dx_4}{dt} &= -x_4 + (1 - x_4)\left(x_2 + M\textstyle\sum_{i,j}s_{ij}\beta_{ij}^- + p\right) - Jx_3x_4 \end{aligned} \tag{3.17}$$

où s_{ij} est la vitesse de mouvement.

Les connexions MT/MSTv sont définies de telle façon qu'elles permettent à MSTv de reconstruire la vitesse du stimulus de mouvement.

Les paramètres type ont été : $C = 0.05$, $F = 1.0$, $J = 20.0$, $M = 0.5$.

Le modèle montre que le traitement du mouvement fait par MT et MSTd optimise le calcul des motifs de mouvement de position invariante sans sacrifier la capacité à calculer l'information du mouvement propre.

Le modèle prétend que le traitement du mouvement fait par MT et MSTd optimise le calcul des motifs de mouvement sans déformation sans sacrifier la capacité à calculer l'information du mouvement propre.

Bref, ce modèle met en valeur les interactions à l'intérieur de MST afin de discriminer le mouvement du fond de celui des objets, donc, l'extraction du mouvement propre mais avec l'intégration des signaux extra-corticaux (comme les mouvement des yeux). De plus, ce modèle suppose déjà que les réponses des neurones dans MT nous fournissent leur direction et leur vitesse préférentielle.

Pour ma part, je resterai au niveau du traitement des signaux corticaux et à partir de l'acquisition de l'image même. L'intégration des signaux extra-corticaux est une autre perspective à développer.

3.3 Bilan des éléments de bio-inspiration

Après ce survol sur les fondements neurobiologiques pour la perception du mouvement et sa modélisation classique, j'ai mis en évidence deux modèles bio-inspirés dans ce chapitre.

Ici, la perception du mouvement visuel ne dépend pas a priori de l'interprétation ou de la reconnaissance de la forme, mais elle est fondée sur la conjonction d'un processus d'intégration tendant à combiner les signaux de vitesses locales de façon à produire des zones de mouvement cohérent, et d'un processus de segmentation tendant, au contraire, à differencier les signaux locaux de manière à individualiser les objets les uns des autres [Bra93].

Selon ce que nous avons vu dans le chapitre §I.1, depuis la rétine et jusqu'aux aires corticales supérieures, le traitement de l'information lumineuse a subi des compressions, des intégrations temporelles et binoculaires, de fortes interactions latérales excitatrices-inhibitrices et de grandes rétro-actions. Aujourd'hui, la plupart les modèles existants pour l'estimation de la vitesse de mouvement et le traitement de la direction se fondent sur l'utilisation d'un réseau en rétro-action, multi-échelle et compétitif pour saisir les principales caractéristiques des aires de la voie dorsale du système visuel du primate (V1-V2-MT-MST).

Je vais mettre en valeur quelques aspects retenus pour la mise en place de mon modèle bio-inspiré, à savoir : le filtrage spatio-temporel et les fortes interactions latérales.

3.3.1 Les principales étapes

Tout d'abord, nous pouvons remarquer la fréquente application d'un développement hiérarchique pour la perception du mouvement fondé sur des traitements locaux (voir §I.2). Ces traitements appliquent un filtrage spatio-temporel et progressent vers de fortes interactions à chaque niveau du développement hiérarchique pour finir dans la segmentation du mouvement obtenue par l'intégration des signaux afférents puis, l'envoi de sa réponse par de grandes rétro-actions vers les aires inférieures.

Le filtrage spatio-temporel

Les signaux lumineux subissent, dans leur parcours depuis la rétine jusqu'aux aires corticales supérieures, plusieurs filtrages spatio-temporels. Leur application fournit l'information utile pour les traitements ultérieurs de la perception du mouvement dont notamment le contraste et la texture (voir §I.1.3.2).

Nous avons vu que les endroits de l'image où il y a une structure suffisante peuvent apporter une information importante pour la perception du mouvement (voir §I.2.2). De plus, leur détection est équivalente à extraire l'orientation spatio-temporelle. Par conséquent, seule l'information concernant les gradients assez abrupts de luminance est gardée, i.e. le squelette des objets suffira pour le traitement cortical. On élimine tout ce qui est uniforme dans l'image en ne retenant que les discontinuités spatiales, les distributions de luminosités et les discontinuités temporelles.

Des modèles classiques pour l'estimation du flot optique on retire l'application constante d'un filtrage spatio-temporel local et sa hiérarchie pyramidale progressive pour aboutir à une image filtrée réduite en taille et avec une forte rétro-action vers les couches inférieures de la pyramide (voir §I.2).

Étant donné que les champs récepteurs des différentes cellules dans le parcours visuel des signaux lumineux font une somme excitatrice des champs récepteurs des cellules pré-synaptiques, un filtre spatio-temporel peut modéliser leur champ récepteur (voir §I.1.3.1). Parmi les filtres les plus utilisés pour cette modélisation, les filtres de Gabor ont été couramment appliqués pour modéliser notamment les réponses de cellules simples de V1 (ibidem).

Cette démarche m'a encouragé à proposer un filtrage causal issu des filtres de Gabor. Dans §II.1.2 je vais décrire sa formulation dans mon modèle.

Les fortes interactions latérales

Le traitement cortical dans V1 est local, distribué, appliqué en parallèle, compétitif et coopératif (voir §I.1.3.1). Les différentes colonnes d'orientation traitent une seule orientation et partagent leurs réponses avec leurs voisines, mais inhibent les colonnes opposés à leur direction préférentielle (ibidem).

Dans une colonne d'orientation, l'excitation dans un voisinage proche contraste avec les fortes inhibitions provenant d'autres colonnes. Donc, les neurones avec des propriétés similaires dans leur champ récepteur seraient assemblés en modules fonctionnels pour réaliser une balance entre l'excitation et l'inhibition latérale.

La distribution compétitive-coopérative des réponses de chaque colonne d'orientation vers les autres colonnes (interaction densément latérale) fera que leurs neurones arrivent à s'auto-organiser. Ce fait a été bien pris en compte par Kohonen dans son modèle [Koh84].

Dans mon modèle (voir §II.1.3) j'applique ces idées dans la formulation de mon mécanisme de fortes interactions latérales inspiré des interactions des colonnes corticales dans V1. Simoncelli et Heeger (voir §I.3.2.1) ont appliqué une inhibition soustractive-divisive (voir les équations I.3.3 et I.3.7), je vais mettre de façon plus explicite ces interactions au niveau de V1.

Enfin, la forte interrelation entre les couches visuelles supérieures (MT-MST) et leur rétro-action vers les couches visuelles primaires renforcent l'information nécessaire pour la poursuite. Le module de focalisation de l'attention (voir §II.1.4 et §II.4) dans mon modèle connexionniste fait une sorte de coopération-compétition parmi les interactions latérales de chacune des cartes proposées inspirées des interactions de MT et de MST.

3.3.2 Avantages des approches connexionnistes bio-inspirées

Dans un contexte évolutif, fortement modulaire et intégrateur, les modèles bio-inspirés fournissent une approche qui essaie de mettre en place les différents aspects du traitement cortical pour aller au-delà des contraintes, parfois prohibitives, de la modélisation classique.

D'une part, la neurobiologie nous dit que dans notre cerveau il existe un grand réseau de coopération-compétition distingué en modules différents qui interagissent en pro-action, en rétro-action et par auto-organisation. D'autre part, les essais pour mettre en place ces idées demandent des ressources matérielles assez importantes.

Les modèles classiques nous fournissent déjà les premières approches, mais une intégration de très haut niveau, puis des couplages avec d'autres systèmes font appel à une modélisation plus robuste et ouverte.

Donc, la bio-inspiration fournit des idées de modularité fonctionnelle et de leurs interrelations qu'un modèle informatique pourrait mettre en place pour nous aider à faire émerger des comportements autonomes. Voilà un des principaux axes dans la recherche en robotique et informatique liée par les neurosciences. Et c'est dans cette optique que je me situe.

Ma proposition est alors inspirée des hypothèses et des résultats en neurosciences où les fortes interactions entre le cortex visuel primaire (V1), le cortex temporal médian (MT) et le cortex temporal supérieur médian (MST) m'amènent à concevoir un modèle pour la perception visuelle du mouvement qui prend en compte le caractère modulaire des architectures neuronales, qui maintient la localité, la distribution de l'information et les fortes interactions latérales et qui prend en compte des différentes étapes menant à une perception visuelle dynamique embarquée.

Je vais proposer dans la deuxième partie, un modèle connexionniste neuromimétique qui prend en charge les idées mises en valeur dans §I.3.3.1 et en respectant les contraintes fixées que je viens d'énoncer.

Deuxième partie

Une approche connexionniste bio-inspirée

Chapitre 1

Présentation globale du modèle

Sommaire

Nous avons vu que les aires corticales dans la voie dorsale se spécialisent dans la perception du mouvement, mais la rétine participe aussi à la détection et a été une source d'inspiration pour de nombreux modèles de perception du mouvement. Cependant l'analyse, la segmentation et l'intégration dans les couches supérieures du cerveau nous permettent de concevoir des modèles plus robustes et modulaires qui traitent le parcours du flot optique depuis la rétine.

Après l'analyse des divers traitements que reçoit le flot optique dans le chemin magnocellulaire-dorsal (voir §I.1.3), j'ai notamment retenu les aires corticales V1 et MT, lesquelles se focalisent vers l'obtention de l'orientation (V1), de la direction (V1/MT) et de la vitesse (MT). L'aire MST aura un rôle prépondérant pour la perception du mouvement propre, un aspect qui sera une des perspectives de ce travail.

Par la suite, dans §II.1.1 je vais présenter de façon globale la structuration en trois modules de mon modèle, puis je décrirai globalement chacun d'eux (dans §II.1.2, §II.1.3 et §II.1.4, respectivement).

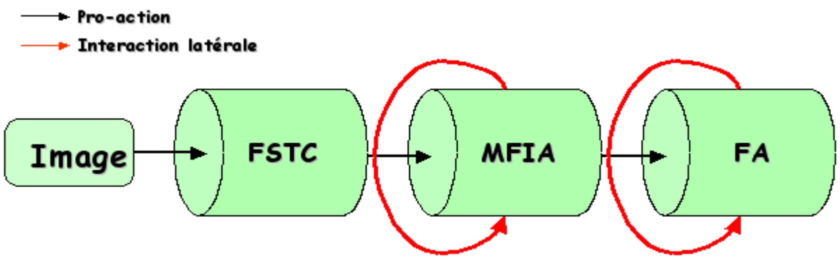
Les chapitres qui suivent dans cette deuxième partie vont préciser chaque module, leur implantation, le couplage avec les modules antérieurs et décriront leur application.

1.1 Les trois modules du modèle

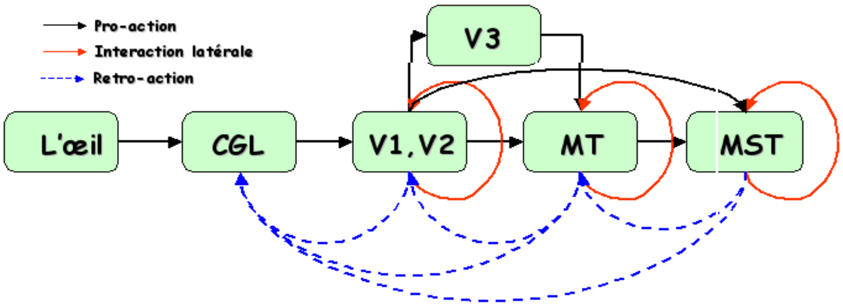
Selon §I.1.3.1, l'information du mouvement local de l'image rétinienne est extraite par les neurones de V1 ayant des champs récepteurs restreints à des petites aires d'interaction spatiale. Ces neurones sont densément interconnectés par des interactions excitatrices-inhibitrices.

En vue d'une réalisation bio-inspirée, connexionniste, fortement locale et sa faisabilité d'implantation embarquée sur un robot mobile, je vais décrire dans ce chapitre mon modèle connexionniste composé de trois modules. Le premier opère un filtrage spatio-temporel causal issu de filtres de Gabor (FSTC). Le deuxième effectue un mécanisme distribué de fortes interactions localisées fondé sur un principe antagoniste (MFIA). Finalement, le troisième dit de focalisation de l'attention (FA) permet de faire émerger un seul objet en mouvement à travers l'évolution d'une population neuronale densément interconnectée élaborée sur la théorie des champs neuronaux continus (CNFT).

En se référant à la figure II.1.1, mon modèle ne modélise pas le chemin magnocellulaire-dorsal que suit le flot optique chez l'être humain, par contre, je m'inspire de ce chemin pour proposer mon modèle. Mon filtrage spatio-temporel causal (FSTC) et mon mécanisme de fortes interactions antagonistes (MFIA) s'inspirent des réponses des cellules simples de V1 et de leur organisation en colonnes d'orientation, respectivement. L'intégration des réponses du MFIA dans le module de focalisation de l'attention (FA) prend les principes des neurones issus de MT et MST où ces neurones ont des grands champs récepteurs. Les interactions latérales et leur traitement ultérieur dans les sous-étapes de ce module feront émerger l'attention sur un seul objet en mouvement et permettront donc de le suivre.



(a) Architecture générale de mon modèle



(b) Chemin Magnocellulaire-Dorsal du flot optique

FIG. 1.1 – Vue globale des étapes suivies par mon modèle pour la perception du mouvement (a) et sa relation avec les différentes étapes de traitement dans le chemin magnocellulaire-dorsal (b). Les modules FSTC et MFIA s'inspirent des propriétés issues de V1 pendant que le module FA s'inspire de celles de MT et MST. Il faut noter que mon modèle n'incorpore pas la rétro-action des aires supérieures sur les aires inférieures montré en (b).

1.2 Filtrage spatio-temporel causal (FSTC)

Après l'étude des principaux algorithmes pour l'estimation du flot optique (voir §I.2.2 et §I.2.3), j'ai décidé de m'inspirer des approches fréquentielles, notamment de celles fondées sur

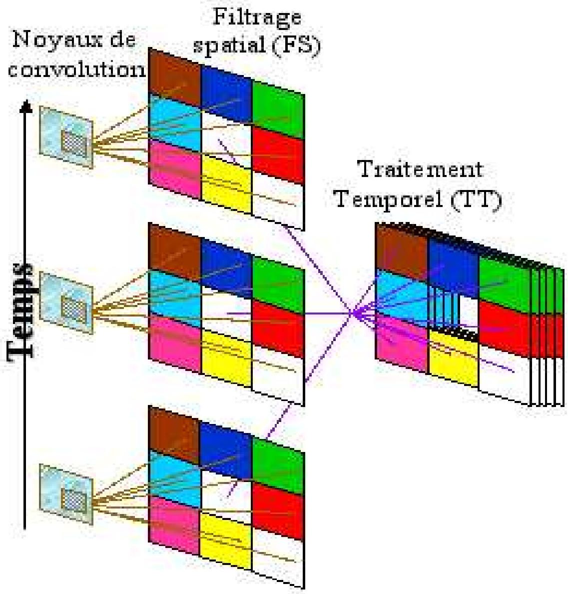


FIG. 1.2 – Architecture pour le premier module de mon modèle avec $\mathcal{D} = 8$ différentes directions et $\tau = 6$ le nombre d'images de la sous-séquence S (voir §II.1.2).

l'énergie, pour leurs propriétés proches des modèles biologiques et leurs caractéristiques de lissage et de robustesse issues de l'application des filtres spatio-temporels. De plus, comme nous avons vu dans §I.1.3.1, §I.2.2, et §I.2.3, les filtres les plus utilisés pour modéliser les réponses des cellules simples dans V1 sont bien les filtres spatio-temporels de Gabor.

Pour mon modèle, je propose une adaptation de ces filtres en les voyant comme des filtres spatio-temporels causaux, c'est-à-dire des filtres orientés spatio-temporels spécifiques à une vitesse supposée et fondés sur l'hypothèse d'une haute fréquence d'échantillonnage des images. Ce haut échantillonnage induirait une vitesse localement constante. De plus, ces filtres font, d'une part, un filtrage spatial orienté proche de la détection de contours (filtres de Gabor spatiaux) et, d'autre part, un traitement temporel adapté à chaque vitesse locale supposée.

Le filtrage est donc développé en deux étapes : un filtrage spatial et un traitement temporel causal.

1.2.1 Filtrage spatial

Pour le filtrage spatial, les filtres de Gabor sont implantés comme des noyaux de convolution de l'image dans D directions différentes (voir la figure II.1.2).

Étant donné que je veux faire la détection des vitesses au niveau local, j'élimine les mouvements à grande amplitude par un échantillonnage rapide. Cela me permettra d'avoir une vitesse instantanée, linéaire, locale et constante pour une direction et une vitesse de mouvement données [29].

Donc, la réponse d'une cellule simple $S_\theta(x, y, \tilde{t})$ dans V1 avec une direction préférentielle θ est exprimée dans mon modèle comme

$$S_\theta(x, y, \tilde{t}) = \frac{\mathcal{R} \cdot E_\theta(x, y, \tilde{t})}{\mathcal{P} + E_\theta(x, y, \tilde{t})} \tag{1.1}$$

[29] Cette hypothèse est aussi faite dans d'autres travaux avec une détection locale de la vitesse [Pir98].

où $\mathcal{R}$ représente le niveau de réponse maximal de la cellule et $\mathcal{P}$ la constante de demi-saturation. De plus, avec l'application d'une moyenne gaussienne sur la région Ω_G qui repère son champ récepteur, je veux rendre $E_\theta(x, y, \tilde{t})$ indépendante de l'intensité de telle façon que

$$E_\theta(x, y, \tilde{t}) = \frac{G_\theta(x, y, \tilde{t})}{M(x, y, \tilde{t})} \tag{1.2}$$

où

$$G_\theta(x, y, \tilde{t}) = \int_{\Omega_G} I(x - \xi_g, y - \eta_g, \tilde{t}) g_\theta(\xi_g, \eta_g) d\Omega_G \tag{1.3}$$

est la convolution de l'image avec le filtre de Gabor orienté g_θ sur la région Ω_G centré sur (x, y) avec $(\xi_g, \eta_g) \in \Omega_G$, et

$$M(x, y, \tilde{t}) = \int_{\Omega_G} I(x - \xi_g, y - \eta_g, \tilde{t}) m(\xi_g, \eta_g) d\Omega_G \tag{1.4}$$

est la pondération de l'image par sa moyenne gaussienne sur la région Ω_G centrée sur (x, y) où $m(\xi_g, \eta_g)$ est la gaussienne.

Dans §II.2 je décrirai la fonction de Gabor utilisée et le critère appliqué pour la moyenne gaussienne.

1.2.2 Traitement temporel causal

Le traitement temporel causal fait le calcul de la moyenne temporelle des filtres de Gabor spatiaux pour chaque direction et pour un ensemble de zones de recherche qui correspondent à V vitesses différentes supposées pour chaque région Ω_G filtrée. C'est-à-dire, pour chaque direction et vitesse supposées, je renforce le mouvement local avec la moyenne des filtres de Gabor spatiaux appliqués aux images passées dans les zones passées supposées. Ce principe est valable avec une forte hypothèse d'un haut échantillonnage de capture d'images pour assurer une détection locale du mouvement et qui induit une vitesse locale instantanée constante.

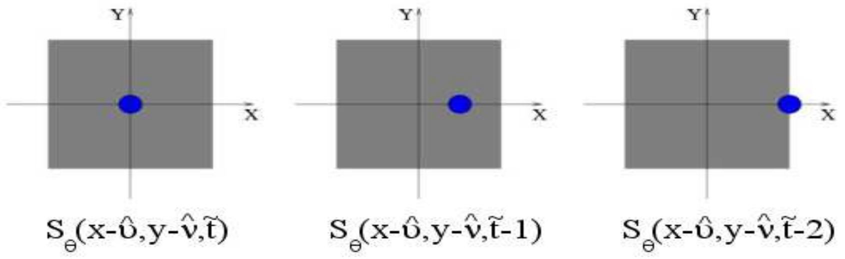


FIG. 1.3 – Recherche du mouvement supposé à gauche pour une vitesse de $v = 1$ orientée à $\theta = 180^\circ$ et pour une taille de $\tau = 3$ pour la sous-séquence S dans la sous-étape de traitement temporel causal (voir §II.1.2.2).

J'espère identifier ce mouvement local en cherchant un contraste spatial dans les endroits et les moments attendus. C'est-à-dire, je suppose que chaque pixel de l'image $I(x, y, t)$ a un mouvement linéaire constant dans la sous-séquence de taille τ, S_τ, avec une vitesse $v \in \Lambda = \{v_i = (\upsilon_i, \nu_i) \mid \|i\| \leq \mathcal{V},\, \mathcal{V} \in \mathcal{N}^*\}$ où $\mathcal{V}$ est le nombre de vitesses différentes supposées, et dans une direction $\theta \in \Theta = \{\theta_i \mid 0 \leq \theta_i < 2\pi,\, |\theta_i - \theta_{i+1}| = \theta_\epsilon = \frac{2\pi}{\mathcal{D}}\}$ où $\mathcal{D}$ est le nombre de directions différentes supposées et θ_ϵ la séparation entre deux directions consécutives.

Donc, la réponse de mon filtre spatio-temporel causal serait

$$F(x,y,\tilde{t},\theta,v) = \int S_\theta(x-\hat{\upsilon}, y-\hat{\nu}, \tilde{t})d\tilde{t} \tag{1.5}$$

où les vitesses sont exprimées en fonction du temps et de l'orientation

$$\hat{\upsilon} = \frac{\tilde{t}}{\tau-1}\upsilon cos\theta, \quad \hat{\nu} = \frac{\tilde{t}}{\tau-1}\nu sin\theta \tag{1.6}$$

La figure II.1.3 montre le cas d'un mouvement horizontal vers la gauche avec la vitesse $v = 1$.

La mesure obtenue par l'équation II.1.5 n'est pas capable de déterminer le vecteur de mouvement 2D. J'ai besoin de mettre en compétition un ensemble de filtres ne différant qu'en orientation et en vitesse. Ils seront donc rassemblés dans un seul vecteur appelé *vecteur des capteurs de mouvement* dont chaque réponse obtenue dans une orientation donnée est un capteur de mouvement.

Bien que l'utilisation de ces filtres nous permette de travailler dans le monde réel, la sensibilité au contraste et à la texture restera importante. En profitant de la grande quantité d'information d'une séquence d'images hautement texturisées, j'extrairai la réponse de mon filtre spatio-temporel causal pour chaque vitesse supposée $v \in \Lambda$ dans chaque région du champ visuel par l'application de l'équation II.1.5. D'autre part, la détection du contraste est suffisante pour l'identification de la direction du mouvement, de telle façon que les mécanismes visuels extrayant le mouvement sont étiquetés par la direction [DH93].

Dans §II.2 je décrirai les détails d'implantation du premier module de mon modèle.

1.3 Mécanisme des fortes interactions antagonistes (MFIA)

Dans la section précédente, j'ai modélisé les neurones de V1 sélectifs à l'orientation spatiale par mes filtres spatio-temporels causaux. Dans ce premier module je suis passé de 1 à $\mathcal{D} * \mathcal{V}$ valeurs différentes pour chaque pixel de l'image. Le deuxième module travaillera donc $\mathcal{D} * \mathcal{V}$ fois plus que le filtrage appliqué sur l'image d'entrée. Cette expansion de l'information est similaire à celle suivie par les signaux lumineux de CGL à V1 (voir §I.1.2).

Nous avons vu que la première analyse corticale est faite au niveau de V1 (voir §I.1.3) où les fortes interactions latérales excitatrices-inhibitrices inter- et intra-colonnes feraient émerger une auto-organisation dépendant des différentes colonnes d'orientation. Leurs neurones font d'abord un "filtrage" spatio-temporel et partagent leurs réponses avec une forte interaction locale.

Dans cette section je vais modéliser les champs récepteurs des neurones simples de V1 comme un produit des interactions inhibitrices et excitatrices dans l'espace et dans le temps à travers un mécanisme d'interactions antagonistes bio-inspiré.

1.3.1 Sens du mécanisme MFIA

En m'inspirant des interactions inter- et intra-colonnes d'orientation dans V1 (voir §I.1.3.1), le modèle proposé est hautement connexionniste et fortement local et distribué. J'y ai inséré en entrée mes filtres spatio-temporels causaux.

Dans ce module, le mécanisme d'inhibition antagoniste réalise des interactions locales d'excitation et d'inhibition dans les différentes colonnes d'orientation issues du cortex visuel primaire (voir la figure II.1.4). Le but de ce mécanisme est de renforcer la cohérence dans les zones de mouvement.

Mon mécanisme met en place cet ensemble d'interactions excitatrices et inhibitrices et utilise aussi un fort schéma local de connectivité avec un rayon d'influence dynamique dépendant de la

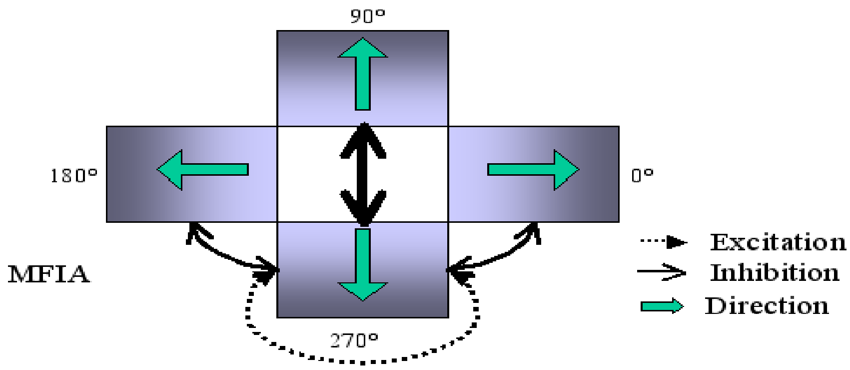


FIG. 1.4 – Le MFIA avec quatre directions préférentielles vu depuis les neurones à 270 °. L'épaisseur des flèches d'excitation et d'inhibition indique la force de l'interaction.

réponse de chaque neurone (voir la section §.3.1). Ce rayon est inspiré des modèles d'interaction latérale [FCMMFGD03, FJ90] où le concept d'un voisinage adaptatif choisit une région de travail pour chaque point de l'image et pour chaque niveau de calcul fondamental.

La connectivité locale et le rayon d'influence de chaque neurone joueront un rôle central pour l'élimination de la direction non préférée. Heeger [Hee93] propose que l'élimination de la direction non préférée pourrait être due à l'inhibition corticale du réseau générant des réponses normalisées. Dans mon cas, le MFIA mettra en valeur les interactions d'inhibition corticale.

Dans mon mécanisme, chaque neurone reçoit les deux types de signaux des neurones (excitateur et inhibiteur) inclus dans un voisinage proche du neurone considéré. Ce voisinage dépend du rayon d'influence du neurone considéré et règle son activité.

La figure II.1.5 montre la forte interaction entre les différentes colonnes d'orientation. Les interactions d'excitation et d'inhibition font interagir les neurones se trouvant dans le voisinage centré autour du neurone en considération. L'épaisseur des flèches indique la force de l'interaction.

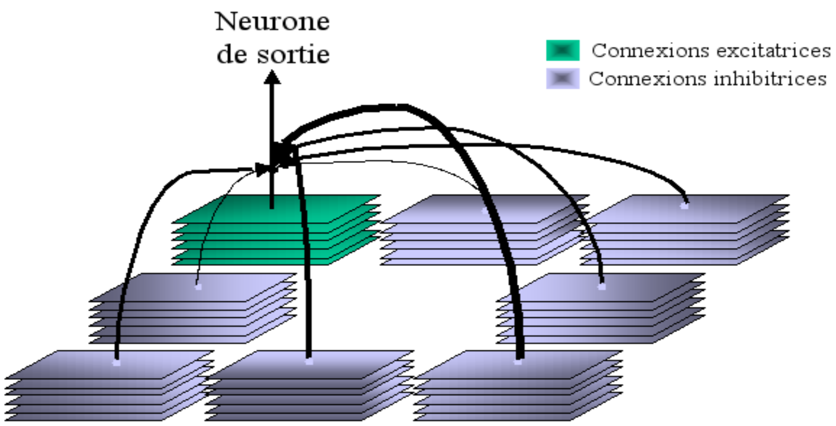


FIG. 1.5 – Les fortes interactions proposées par le MFIA changent l'état interne du neurone de l'ensemble à 135 ° appelé ici "neurone de sortie". Dans cette figure je montre huit directions préférentielles. L'épaisseur des flèches indique la force de l'interaction.

J'appelle mon mécanisme MFIA car la grande quantité des connexions d'inhibition diminue l'activité des neurones avec une direction préférentielle opposée ou antagoniste, c'est-à-dire les neurones qui ne partagent pas soit une direction soit une vitesse commune ou similaire. Par contre, les connexions excitatrices font augmenter l'activité des neurones vers l'émergence de réponses cohérentes, c'est-à-dire en assemblant les réponses des neurones aux vitesses et aux orientations similaires à travers un processus interactif.

1.3.2 Règle d'actualisation de l'état interne des neurones

Les fortes interactions de mon mécanisme font changer l'état interne du neurone et, par conséquent, son rayon d'influence. Dit d'une autre façon, cette actualisation de l'état interne du neurone génère un processus adaptatif dynamique.

En m'inspirant des idées tirées de Grossberg [Gro88], Moga [Mog00] et Fernández et ses collaborateurs [FCMMFGD03], l'équation de mise à jour de l'état interne du neurone localisé en (x, y) est

$$\begin{aligned} \eta\frac{\delta H(x,y,T)}{\delta T} &= -\mathcal{A} \cdot H(x,y,T) \\ &+(\mathcal{B} - H(x,y,T)) \cdot Exc(x,y,T) \\ &-(\mathcal{C} + H(x,y,T)) \cdot Inh(x,y,T) \end{aligned} \tag{1.7}$$

avec $H(x, y, T)$ son potentiel à l'instant T, $Exc(x, y, T)$ l'activité due aux contributions des interactions d'excitation dans les voisinages $\Omega^{\Theta_E}_{(x,y)}$ et $Inh(x, y, T)$ l'activité due à la contribution des interactions d'inhibition dans les voisinages $\Omega^{\Theta_I}_{(x,y)}$. Les voisinages d'excitation et d'inhibition dépendent du rayon d'influence de chaque neurone dans chaque direction. $\mathcal{A}$, $\mathcal{B}$ et $\mathcal{C}$ sont des constantes réelles et η le coefficient d'apprentissage.

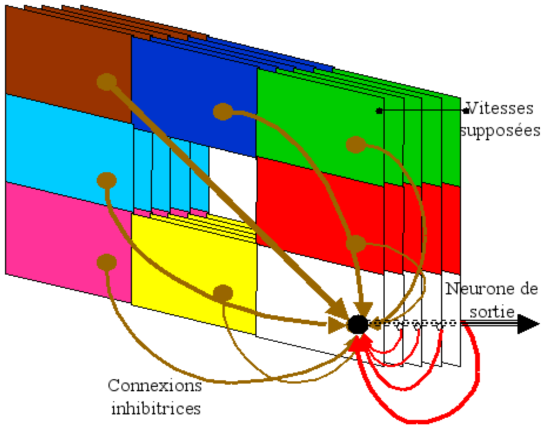


FIG. 1.6 – Vue générale de mon mécanisme MFIA simplifié à quatre directions préférentielles. Les connexions excitatrices interagissent dans la même direction pendant que celles inhibitrices interagissent avec les autres directions.

Soit ρ le rayon d'influence du neurone localisé en (x, y) au temps T. Ce neurone reçoit alors au moins ρ^2 connexions excitatrices des neurones dans la même direction et la même vitesse et, au moins, $(\mathcal{V} \cdot \mathcal{D} - 1) \cdot \rho^2$ connexions inhibitrices des autres neurones de son voisinage.

Enfin, j'ai $\mathcal{D} \cdot \mathcal{V}$ neurones différents qui codent l'information des directions et des vitesses pour un pixel. La figure II.1.6 montre de façon générale les interactions des quatre ensembles de neurones avec leur direction préférentielle correspondante.

D'après les résultats physiologiques, l'activation des neurones est déterminée par une balance précise entre les interactions latérales excitatrices sélectives à leur direction préférentielle et les interactions latérales inhibitrices non sélectives à leur direction préférentielle [SNS95, SUN95] se manifestant par un mécanisme de forte inhibition [YG89]. Dans mon cas, la topologie des interactions entre les colonnes d'orientation excitatrices et inhibitrices est fixe mais je peux la régler (voir §II.3).

Dans §II.3 je vais décrire plus précisément mon mécanisme MFIA, son équation d'actualisation, son paramétrage et sa faisabilité d'implantation matérielle sur FPGA.

1.4 Focalisation de l'attention (FA)

Le troisième module de mon modèle met en interaction une population de neurones pour les focaliser sur un seul objet mobile (focalisation de l'attention).

Ce module se fonde sur les interactions issues des aires visuelles supérieures (MT, MST, PP et le champ frontal oculaire -FEF-) et d'autres comme le colliculus supérieur et les CGL (voir §I.1.2.5 et §I.1.2.3, respectivement).

D'après les travaux de Rougier et ses collaborateurs [RV05, VRA05], les interactions dans chaque aire sont modélisées par l'application de la théorie des champs neuronaux continus en 2D [Tay99]. Cette théorie a été appliquée à la modélisation du contrôle par saccades fait par le colliculus supérieur, la formation de détecteurs de l'orientation, la représentation de l'orientation spatiale et les changements des champs récepteurs corticaux somatosensoriels avec apprentissage.

Dans mon cas, ce troisième module me fournit une réponse globale pour repérer l'objet mobile dans la scène, puisque mon mécanisme MFIA reste toujours local et distribué sans aboutir pour lui-même à une réponse globale.

Dans la suite, je vais résumer les sources neurobiologiques dont s'inspire mon troisième module, puis je décrirai succinctement son architecture et je finirai par montrer ses équations d'actualisation.

1.4.1 Les interactions dans les aires visuelles supérieures

Selon Sekuler et ses collaborateurs [SWB02], MT et MST sont spécialisées pour la détection du mouvement cohérent. Ils mettent en relation leur activité neuronale par l'exécution des mouvements oculaires de poursuite(SPEM) [NWK88, KSWY94, IT03].

MST est considéré comme un substrat pour la perception du flot optique en incluant les stimuli de la tête (les motifs du flot optique simulant des trajectoires tirés des expériences du mouvement propre [30] dans le monde [BMA+96, DW95]).

La sélectivité de la direction des neurones de MST serait dominée par la direction du mouvement visuel et ces neurones continuent à émettre des potentiels d'action pendant la saccade oculaire si la cible suivie est toujours dans la rétine [NWK88].

D'autre part, selon Movshon et ses collaborateurs [MAGN85b], les neurones sélectifs au mouvement dans V1 et les premières étapes de traitement dans MT (couche 4 et 6) ne répondent qu'aux composants de la vitesse dans la direction perpendiculaire au mouvement. Il a aussi proposé que cette population peut servir à intégrer les mesures des composantes du mouvement pour résoudre le problème d'ouverture.

Culham et ses collaborateurs [CBC+98] ont identifié les aires du cerveau humain participant le plus dans la poursuite de plusieurs objets en mouvement. La poursuite attentive de plusieurs objets indépendants et en mouvement fait interagir l'aire pariétale et l'aire frontale (responsables des saccades oculaires et de la focalisation de l'attention), l'aire MT et celles alentour.

[30] Lorsqu'on bouge, le flot optique acquis contient l'information de l'environnement et de nous-mêmes combinée très fortement [KLP+01].

La perception visuelle du mouvement émerge avec le déplacement associé aux influences exogènes (représentées par les propriétés des stimuli) et aux influences endogènes (dont l'expectative, l'attention, la mémoire et l'apprentissage) [SWB02].

Si nous déplaçons notre attention entre les deux composantes mobiles d'un stimulus de mouvement par transparence, notre attention peut affecter l'induction du MAEC [31] et augmenter ce mouvement dans la direction opposée [BQ01].

L'activation de l'aire visuelle V5/MT chez l'être humain a été considérée à la base du MAEC [Ree01].

Le MAEC a été attribué à la fatigue ou à l'adaptation des unités du cortex visuel sélectives aux directions particulières de mouvement. L'attention a été attribuée à l'activité de MT pour la grande quantité des cellules sélectives à la direction, action ayant été interprétée comme une corrélation neuronale de la perception du MAEC [Ree01].

1.4.2 Architecture du troisième module

La figure II.1.7 décrit les cinq cartes retenues du modèle proposé par Rougier et ses collaborateurs [RV05, VRA05] correspondant à ses deux premiers niveaux des trois possibles, à savoir : l'émergence de l'attention et la fixation de l'attention. Ces cinq cartes pour mon troisième module sont inspirées des interactions des aires MT, MST, PP et FEF.

Dans le niveau de l'émergence de l'attention (premier niveau), il n'y a qu'une seule bulle d'activité [32] qui émerge parmi les neurones de la population des cartes d'entrée, de visualisation et de focalisation. Dans le niveau de la focalisation de l'attention (deuxième niveau), le FEF et la mémoire de travail soutiennent la réponse des neurones gagnants (bulle d'activité émergente) et leur localisation dans la carte de focalisation de l'attention. Cette mémoire garde sa trace lorsque d'autres bulles d'activité essaient d'émerger en éliminant la première.

La plupart des cartes dans ce module ont de denses interactions latérales où chaque neurone est complètement connecté à tous les autres neurones de la même carte. De plus, la communication entre ces différentes cartes est fondée sur le principe des champs récepteurs locaux.

Dans §II.4 je vais montrer plus en détails ce troisième module.

1.4.3 Règle générale d'actualisation

L'état interne d'un neurone dans la carte $\mathcal{M}$ adjacente à la carte $\tilde{\mathcal{M}}$ est actualisé comme

$$\kappa \frac{\partial u(x,y,t)}{\partial t} = -u(x,y,t) \quad \begin{aligned} &+ \textstyle\int_{\hat{\mathcal{M}} \subset \tilde{\mathcal{M}}} w_{\hat{\mathcal{M}}}((x,y),(\hat{x},\hat{y})) I(\hat{x},\hat{y},t)\, d\hat{x} d\hat{y} \\ &+ \textstyle\int_{\mathcal{M}} w_{\mathcal{M}}((x,y),(\bar{x},\bar{y})) f(u(\bar{x},\bar{y},t))\, d\bar{x} d\bar{y} + h \end{aligned} \tag{1.8}$$

où $u(x,y,t)$ est le potentiel d'action du neurone dans la position (x,y) au temps t. $f(\cdot)$ représente la moyenne du taux de déclenchement du neurone [33], $I(\hat{x},\hat{y},t) \in \hat{\mathcal{M}}$ est le neurone localisé en

[31] Lorsqu'on fixe son regard sur un mouvement dans une direction pendant une certaine période de temps et ensuite on regarde une scène statique, cette scène semble se déplacer dans la direction opposée à celle qu'on a précédemment vu, mais avec un petit changement dans la position perçue [NJ99, Sno98]. Une longue exposition de notre regard au mouvement provoque un changement des seuils dans la détection de la direction et une altération de la perception de la vitesse. Les mouvements illusoires définis à la post-exposition s'appellent des *mouvements à effets consécutifs* (MAEC) [PGP00].

[32] Une bulle d'activité est générée par les interactions latérales de type centre-périphérie (excitation-inhibition) à l'intérieure d'une carte où tous ses neurones sont entièrement connectés. Donc, un groupe de neurones proches ont un état d'activation élevé tandis que les neurones plus éloignés ont un état d'activation quasi-nul.

[33] En considérant que l'activité des neurones est approximativement leur moyenne de taux de déclenchement [Tay99].

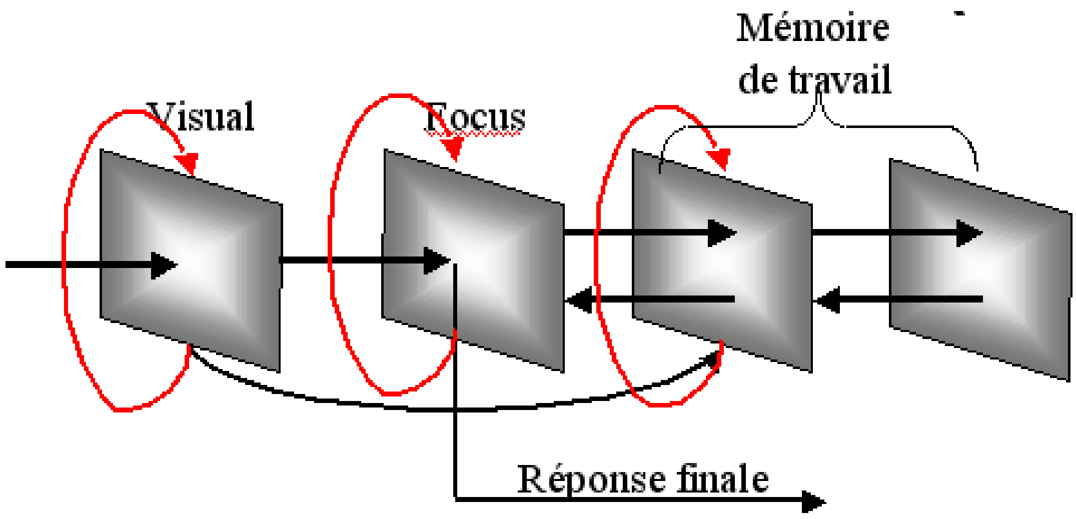


FIG. 1.7 – Architecture générale des sous-étapes de mon troisième module : l'émergence de l'attention (deux premières cartes) et la fixation de l'attention (Mémoire de travail).

$(\hat{x}, \hat{y})$ au temps t dans la carte $\tilde{\mathcal{M}}$ qui enverra sa réponse pondérée vers la carte $\mathcal{M}$. Finalement, h est le seuil du neurone.

Le deuxième terme de l'équation II.1.8 décrit l'interaction inter-cartes qui est fondée sur les champs récepteurs de la carte source $\tilde{\mathcal{M}}$ vers la carte cible $\mathcal{M}$. Alors, $w_{\hat{\mathcal{M}}}((x,y),(\hat{x},\hat{y}))$ est la fonction de pondération des connexions adjacentes du neurone $(\hat{x},\hat{y}) \in \hat{\mathcal{M}}$ vers le neurone $(x,y) \in \mathcal{M}$. Elle est définie par

$$w_{\hat{\mathcal{M}}}((x,y),(\hat{x},\hat{y})) = \mathcal{D}exp\left(\frac{|(x,y)-(\hat{x},\hat{y})|^2}{d^2}\right) \tag{1.9}$$

D'autre part, le comportement d'interaction latérale intra-carte est défini dans le troisième terme de l'équation II.1.8 où $w_{\mathcal{M}}((x,y),(\bar{x},\bar{y}))$ est la fonction de pondération des connexions latérales entre les neurones localisés en (x,y), $(\bar{x},\bar{y})$, respectivement dans la carte $\mathcal{M}$ exprimée comme

$$w_{\mathcal{M}}((x,y),(\bar{x},\bar{y})) = \mathcal{D}exp\left(\frac{|(x,y)-(\bar{x},\bar{y})|^2}{d^2}\right) - \mathcal{E}exp\left(\frac{|(x,y)-(\bar{x},\bar{y})|^2}{e^2}\right) \tag{1.10}$$

avec $\mathcal{D}$, $\mathcal{E}$, d, $e \in \Re_+^*$.

La description plus en détail de ce module, le couplage du deuxième module (MFIA) avec celui-ci (FA) et l'intégration des trois modules de mon architecture connexionniste sont montrés dans §II.4 de cette deuxième partie.

Jusqu'ici j'ai fait un rapide survol de mon modèle connexionniste neuromimétique pour la perception visuelle du mouvement. Dans les chapitres qui suivent, je vais exposer plus en détail chacun des modules le composant.

D'une façon plus particulière, dans §II.2 et §II.3 je montrerai aussi la faisabilité d'une implantation matérielle sur des circuits FPGA.

Chapitre 2

Filtrage spatio-temporel causal

Sommaire

Dans le chapitre précédent j'ai présenté de façon générale mon modèle connexionniste. Dans ce chapitre, je vais préciser son premier module : le filtrage spatio-temporel causal (FSTC).

Issue de la modélisation des neurones simples du cortex visuel, la première étape de mon filtrage spatio-temporel causal est le filtrage spatial (voir §II.1.2.1). Les paramètres spécifiques de cette étape sont discutés dans §II.2.1.1. Puis, un traitement temporel causal suit la première étape (voir §II.1.2.2). C'est ici, et je vais le préciser dans §II.2.1.2, que je fais l'interprétation des vitesses locales sous l'hypothèse d'un échantillonnage rapide des images.

Je vais aussi montrer la faisabilité de l'implantation matérielle de mon filtrage FSTC sur des circuits FPGA (voir §II.2.2). Finalement, je montre quelques exemples d'application (§II.2.3).

2.1 Présentation détaillée du filtrage FSTC

Comme nous avons vu dans §I.1.3.1, le mécanisme pour mesurer la direction du mouvement, le plus basique utilisé dans le système visuel, est la sélectivité directionnelle. Un neurone a une direction préférentielle si les mouvements symétriques d'avance et de recul captés par son champ récepteur provoquent différentes réponses : la réponse maximale indiquera la direction préférentielle et la minimale la direction nulle ou non préférentielle.

J'ai choisi d'estimer ce comportement par des filtres de Gabor (voir §I.2.3.1), étant donnée que le côté spatial de leur fonction s'apparente bien au profil spatial des champs récepteurs dans le cortex [JP87] où un changement dans la polarité du contraste est perçu comme un mouvement dans la direction opposée [Ans70, AR75]. Donc, la direction est la première variable directement mise en relation avec le mouvement qu'on peut mesurer ([GHA94], cfr. [HW62]).

Les sous-sections suivantes montrent les deux étapes de mon filtrage spatio-temporel causal : le filtrage spatial et le traitement temporel causal.

2.1.1 Le filtrage spatial

Une combinaison linéaire d'échantillons consécutifs de l'image peut produire une différence dans la modulation de la réponse pour deux directions différentes de mouvement [JWF93, JWK+97]. Donc, le spectre d'énergie local associé à un pixel dans une image est défini ici comme le filtrage spatial issu du produit du filtre de Gabor spatial et de l'image.

Notation

Soit $I(x, y, t)$ la distribution d'intensités de l'image $I(x, y)$ dans le temps t.

Soit une séquence d'images en échelle de gris $\mathcal{S} = \{I(x, y, t) \mid I(x, y, t),\ t \in \mathcal{N}^*\}$, de laquelle je prendrai une sous-séquence de taille τ, $\mathcal{S}_\tau = \{I(x, y, \tilde{t}) \mid \tilde{t} = t - \breve{t},\ 0 \leq \breve{t} < \tau,\ \tau \leq t$ et $\tilde{t},\ \breve{t},\ \tau \in \mathcal{N}^*\}$.

Soit $I(x, y, \tilde{t}) \simeq I(x - \upsilon\tilde{t}, y - \nu\tilde{t})$ où $\vec{v} = (\upsilon, \nu)$ est le vecteur de mouvement dans une petite région de l'image.

Soit $\Omega_G = \{(\xi_g, \eta_g) \mid \mid \xi_g \mid \leq \xi_G,\ \mid \eta_g \mid \leq \eta_G,\ \xi_g,\ \eta_g \in \mathcal{Z}$ et $\xi_G,\ \eta_G \in \mathcal{N}^*\}$ la fenêtre de convolution de base du filtrage spatial exprimée en pixels de l'image centrée en $(0, 0)$, de taille $(2\xi_G + 1) \times (2\eta_G + 1)$ et des pas $\xi_G^x \times \eta_G^y$ appliqués sur l'image $I(x, y)$. Nous pouvons voir cette description comme le champ récepteur du neurone localisé en (x, y) (voir la figure II.2.1).

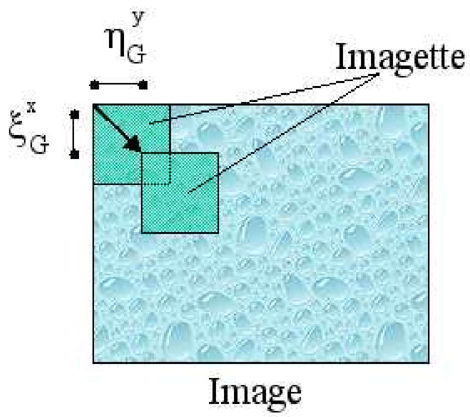


FIG. 2.1 – Détail de l'interprétation de ξ_G^x et η_G^y, les deux pas de déplacement de la fenêtre de convolution (imagette) Ω_G pour parcourir toute l'image.

La modélisation des cellules simples

Par analogie à la fonction des cellules simples dans V1, le filtre de Gabor utilisé dans l'équation II.1.3 sera [PP97, Dau85] :

$$g_\theta(\xi_g, \eta_g) = \frac{1}{2\pi\sigma_{\xi_g}\sigma_{\eta_g}} exp\left(-\frac{\hat{\xi}_g^2}{2\sigma_{\xi_g}^2} - \frac{\gamma^2\hat{\eta}_g^2}{2\sigma_{\eta_g}^2}\right) exp\left(2\pi_i\frac{\hat{\xi}_g}{\lambda} + \phi\right) \tag{2.1}$$

$\hat{\xi_g}$ et $\hat{\eta_g}$ sont définies dans l'équation II.2.2.

Ce filtre est le filtre de Gabor orienté classique sur la région Ω_G où $\sigma_{\xi_g} \times \sigma_{\eta_g}$ est la dimension du champ récepteur. Donc, mes filtres sont implantés comme des noyaux de convolution de l'image où les valeurs des coefficients choisis mettront en valeur les réponses dans les orientations spécifiques, à savoir : $0.23 < \gamma < 0.92$, le quotient d'aspect spatial du champ récepteur (excentricité, [JP87]), λ sa fréquence spatiale préférentielle (longueur d'onde) et ϕ sa symétrie (phase).

L'élection de la bande passante de la fréquence spatiale $0.4 \leq \beta = \sqrt{\sigma_{\xi_g}\sigma_{\eta_g}}/\lambda = 0.56 \leq 2.6$ est motivée par les propriétés des cellules simples dans V1 modélisées par des filtres de Gabor [PP97] [34].

Pour générer mes différents filtres de Gabor orientés classique en 2D je fais la rotation de (ξ_g, η_g) à θ degrés de telle façon que

$$\begin{aligned} \hat{\xi_g} &= \xi_g cos\theta - \eta_g sin\theta \\ \hat{\eta_g} &= \xi_g sin\theta + \eta_g cos\theta \end{aligned} \tag{2.2}$$

Dans la figure II.2.2 je montre deux de mes filtres spatiaux.

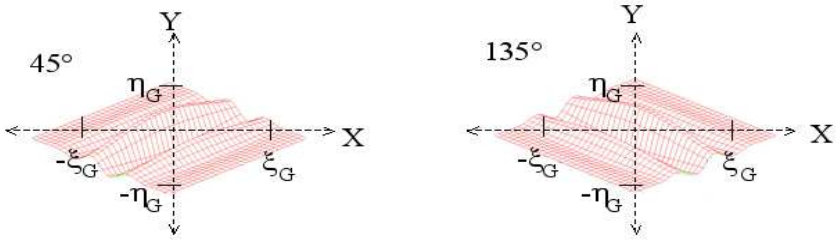


FIG. 2.2 – Deux de mes filtres spatiaux de base.

La version discrétisée de l'équation II.1.3 est alors

$$G_\theta(x, y, \tilde{t}) = \sum_{\xi_g=-\xi_G}^{\xi_G} \sum_{\eta_g=-\eta_G}^{\eta_G} I(x - \xi_g, y - \eta_g, \tilde{t}) g_\theta(\xi_g, \eta_g) \tag{2.3}$$

Afin de rendre robuste mon filtrage aux changements de luminosité, je normalise la réponse des cellules simples par rapport à la luminance moyenne locale de l'image d'entrée. Ainsi l'équation II.1.3 est divisée par le niveau de gris moyen $M(x, y, \tilde{t})$ (voir l'équation II.1.4) dans le champ récepteur et calculée comme une pondération gaussienne de $g_\theta(\xi_g, \eta_g)$ par la moyenne locale des intensités au moyen des coefficients $m(\xi_g, \eta_g)$

$$m(\xi_g, \eta_g) = exp\left(-\frac{\xi_g^2 + \gamma^2\eta_g^2}{2\sigma_{\xi_g}\sigma_{\eta_g}}\right) \tag{2.4}$$

Donc, le quotient dans l'équation II.1.2 est proportionnel au contraste local dans le champ récepteur du neurone modélisé par la fonction $g_\theta(\xi_g, \eta_g)$.

[34] $\beta = 0.56$ est équivalente à 1.0 octave [JP87]

$$1.0 = log_2 \frac{\beta\pi + \sqrt{\frac{ln2}{2}}}{\beta\pi - \sqrt{\frac{ln2}{2}}}$$

Cette mesure est la bande passante de la fréquence spatiale du filtre. Celle-ci varie entre 0.4 et 2.6 octaves chez l'être humain

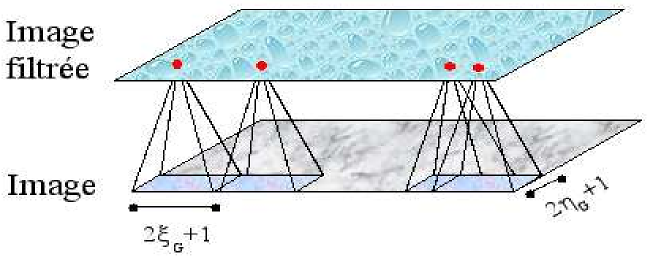


FIG. 2.3 – Détail de la convolution de la fonction d'intensités de l'image avec mes filtres de Gabor orientés. Ma fenêtre de convolution (voir la figure II.2.1) parcours toute l'image avec les pas de déplacement en x et en y comme ont été définis dans §II.2.1.1

La version discrétisée de l'équation II.1.4 est alors

$$M(x, y, \tilde{t}) = \sum_{\xi_g=-\xi_G}^{\xi_G} \sum_{\eta_g=-\eta_G}^{\eta_G} I(x-\xi_g, y-\eta_g, \tilde{t}) m(\xi_g, \eta_g) \tag{2.5}$$

Finalement l'équation II.1.1 génère donc une réponse au contraste similaire à celle mesurée par les cellules simples réelles.

2.1.2 Le traitement temporel causal

Comme nous avons vu dans §I.1.3.2, les neurones de MT répondent fortement aux stimuli visuels dans une direction et une vitesse préférentielle. Cette vitesse est essentiellement construite par l'addition de réponses d'un ensemble en orientation et en temps de neurones de V1.

Dans mon modèle, j'ai plutôt choisi de m'inspirer des réponses des cellules simples par l'utilisation de mes filtres spatio-temporels causaux qui récupèrent la réponse des cellules de MT au travers de sa causalité temporelle (voir §II.2.3.1).

Ainsi, une fois fait le filtrage spatial de toutes les images de la séquence $\mathcal{S}_\tau$ (voir la sous-section précédente), je peux estimer les différentes vitesses supposées, $\Lambda = \{-\mathcal{V}, -\mathcal{V}+1, ..., \mathcal{V}-1, \mathcal{V}\}$.

Mon traitement temporel implique le calcul de la moyenne temporelle sur toute la séquence $\mathcal{S}_\tau$ du filtrage spatial précédemment calculé. C'est-à-dire, pour chacune des $\mathcal{D}$ orientations différentes du filtrage spatial j'aurai un ensemble de zones de recherche en accord à la vitesse supposée de chaque neurone résultant du filtrage spatial.

La figure II.2.4 montre trois cartes qui représentent les réponses du filtrage spatial pour une séquence de taille $\tau = 3$ et pour une orientation donnée. Le traitement temporel que je fais implique la recherche des différentes réponses dans la position et la direction définies par $(x - \hat{\upsilon}_i, y - \hat{\nu}_i)$ dans les différentes images de la séquence S_τ où

$$\hat{\upsilon}_i = \tfrac{t-\check{t}}{\tau-1} v_i cos\theta, \quad \hat{\nu}_i = \tfrac{t-\check{t}}{\tau-1} v_i sin\theta \tag{2.6}$$

pour chaque vitesse v_i supposée dans Λ.

D'autre part, j'ai mis en valeur dans §II.1.2.1 le besoin d'un échantillonnage suffisamment rapide afin de faire les détections des vitesses supposées au niveau local. Ces vitesses seront donc supposées instantanées, linéaires, locales et constantes pour une direction et une vitesse de mouvement données.

La version discrétisée de l'équation II.1.5 est alors

$$F(x, y, \tilde{t}, \theta, v) = \frac{1}{\tau} \sum_{\check{t}=0}^{\tau} S_\theta(x - \hat{\upsilon}, y - \hat{\nu}, t - \check{t}) \tag{2.7}$$

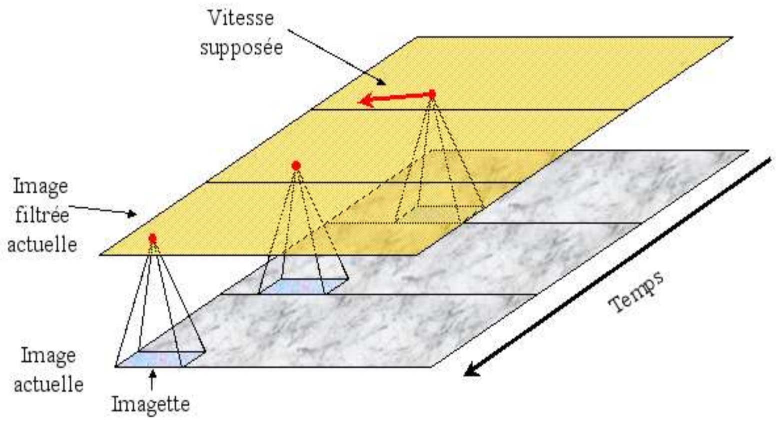


FIG. 2.4 – Représentation graphique d'un point qui bouge dans une direction et une vitesse supposées. Les réponses du filtrage spatial dans les positions indiquées par le point sur les pyramides sont utilisées pour le calcul de la moyenne.

Jusqu'ici, j'ai décrit mon filtrage FSTC. Mais, dans les réponses des filtres de Gabor on en trouve des positives et des négatives. Or, les neurones du cerveau dans l'être humain ne travaillent qu'avec des réponses positives. C'est pourquoi je fais une rectification de demi-onde après le traitement temporel causal pour rendre toutes les réponses positives, i.e. je vais transformer toutes les valeurs négatives en zéros.

Il faut noter que, même si j'ai fait une rectification de demi-onde, les réponses de mon filtrage restent encore avec beaucoup d'information de contraste des objets immobiles dans la scène. Pour la dissiper, je l'élimine par une simple différence temporelle ponctuelle entre deux réponses consécutives dans le temps issues du filtrage FSTC pour le même neurone.

Finalement, comme je manipule des vitesses positives et négatives mais je travaille toujours avec des valeurs positives, j'ai besoin de ré-interpréter les vitesses négatives pour l'affichage de la réponse finale après avoir appliqué mon processus d'intégration des vitesses décrit dans §II.2.3.1, i.e. une vitesse négative dans une direction est égale à une vitesse positive dans la direction antagoniste (direction inverse), mais sur les contrastes inverses.

Par la suite, nous allons discuter la faisabilité de l'implantation matérielle de mon premier module, FSTC. Il s'agit d'un travail réalisé avec l'aide de César TORRES HUITZIL docteur de l'INAOEP [35], post-doctorant dans l'équipe CORTEX.

2.2 Implantation matérielle du FSTC

Les modèles connexionnistes bio-inspirés sont difficiles à implanter parce qu'ils ont un grand nombre d'éléments non-linéaires travaillant de façon concurrente avec un dense schéma d'interconnexions.

[35] Institut National d'Astrophysique, Optique et Électronique de Puebla au Mexique

Dans le monde réel, les tâches à résoudre pour la navigation d'un robot mobile autonome ont besoin d'un support matériel dédié. La technologie des circuits numériques FPGA permet la réalisation d'applications optimisées avec une grande flexibilité par leur reprogrammabilité. De plus, elle fournit une structure hautement parallèle de calcul ainsi qu'un développement assez rapide par rapport aux circuits ASIC's [36] [Ham03, DH00]. Dans [THGCS05b] nous justifions le choix des FPGA comme support d'implantation et dans [Gir00] l'utilisation des FPGA pour l'implantation des réseaux de neurones artificiels est plus largement discutée.

Par la suite, je vais faire un rappel des notions sur les FPGA, puis décrire les deux étapes de mon premier module pour finir avec quelques détails sur leur faisabilité d'implantation en temps réel.

2.2.1 Notion de FPGA

Nous pouvons voir un FPGA comme un réseau prédiffusé programmable par l'utilisateur, i.e. ce réseau est destiné à être programmé sur place, avant d'être utilisé pour une fonction particulière.

Le but principal à l'égard de la réalisation d'un circuit logique à l'aide d'un FPGA est de faire un circuit donné avec le moins de ressources possibles tout en satisfaisant les contraintes imposées de vitesse, ou d'avoir le circuit le plus rapide possible et en utilisant le maximum possible de ressources offertes par la puce donnée.

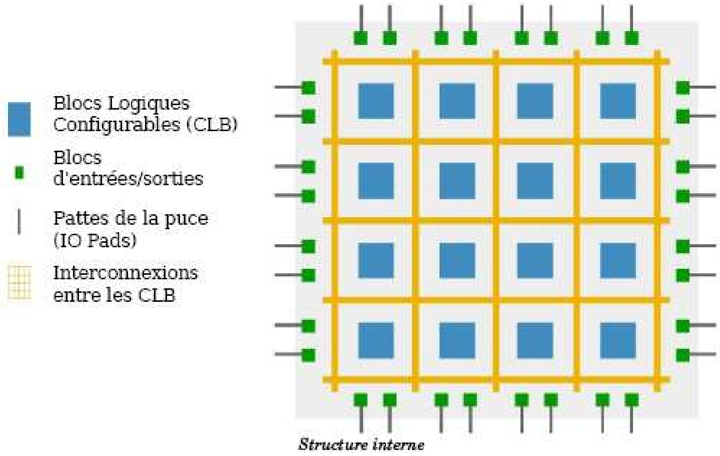


FIG. 2.5 – Architecture générale d'un circuit FPGA.

D'un point de vue pratique, un FPGA est composé de portes logiques à quatre entrées (ET, OU, etc.) organisées en tableaux. L'ensemble des fonctions logiques réalisables par chaque porte logique est représenté par une mémoire (table de conversion/look-up table, LUT) [37] et une bascule bistable (flip-flop) [38] leur permettant de partager localement leur information. L'ensemble de LUT, de flip-flops et de multiplexeurs forme un bloc de configuration logique (CLB) ou un élément de traitement. Tous les CLB sont liés les uns les autres par un réseau d'interconnexions. La communication avec le monde extérieur passe par les blocs d'entrée et de sortie (voir la figure II.2.5).

[36] Application Specific Integrated Circuits.
[37] Tableau permettant d'établir une relation entre deux séries de valeurs.
[38] Dispositif à déclenchements contenu dans un circuit électronique à deux états stables possibles et qui permet de basculer d'un état à l'autre en fonction d'un signal d'entrée et d'un signal de déclenchement.

On programme un FPGA à l'aide du langage de description de matériel pour des circuits intégrées à très haut vitesse (VHDL), puis la synthèse et le routage permettent d'avoir un système cablé de bonne performance.

Trois étapes pour la conception sont nécessaires :

1. Définition et description du circuit à l'aide de VHDL.
2. Compilation des différentes descriptions et création d'un ensemble de fichiers définissant le circuit au niveau hiérarchique le plus bas (niveau de CLB).
3. Placement, routage et création du fichier de configuration.

Les logiciels du langage VHDL intègrent également la possibilité de vérification à plusieurs niveaux : logique, temporelle, etc.

Le placement et le routage sont un problème d'optimisation complexe à cause de l'utilisation d'algorithmes très évolués mais les résultats ne sont pas toujours satisfaisants. En conséquence, on a un compromis entre les procédés entièrement automatiques et ceux assistés par l'opérateur.

2.2.2 Aspect spatial

Afin de mimer la préférence à l'orientation des cellules du cortex visuel, nous avons mis au point un dispositif matériel contenant des modules qui travaillent de façon concurrente sur la même image d'entrée mais avec différentes orientations préférées.

Ce dispositif est semblable à l'organisation en colonnes d'orientation des neurones dans V1 (voir §I.1.3.1) par l'utilisation d'un ensemble de tableaux systoliques [39] 2D. Cet ensemble traite les données de l'image en accord avec le noyau de convolution de type Gabor et met en valeur les réponses des neurones simples dans leur orientation préférentielle.

La figure II.2.6 montre un schéma simplifié en blocs de l'architecture générale de l'implantation du traitement spatial inspiré de [THAE04]. À l'intérieur de chaque tableau il y a un ensemble d'unités de traitement avec la même orientation préférentielle mais avec des champs récepteurs différents.

Chaque tableau systolique 2D reçoit l'image d'entrée et les coefficients de Gabor stockés depuis des mémoires externes, $P = I(x, y)$ et $W_\theta = g_\theta(\xi_g, \eta_g)$, respectivement dans la figure II.2.6. Chaque bloc GWP (fenêtre de traitement de Gabor, "Gabor Window Processor") représente un neurone sélectif à une orientation, le bloc D est le contrôleur de synchronisation et le bloc LDC le collecteur local de données (Local Data Collector).

LDC récupère les résultats des GWP localisés dans une même colonne d'orientation et les envoie vers le collecteur global de données. D'un point de vue fonctionnel, le flot de données passe de haut en bas et de gauche à droite dans le tableau 2D. Par conséquent, les GWP travaillent progressivement dans un pipeline [40] systolique : le pixel courant lu est diffusé à tous les GWP et les coefficients de Gabor sont partagés entre les GWP adjacents.

Finalement, chaque GWP est constitué par un multiplicateur et un accumulateur travaillant en parallèle avec une arithmétique de point fixe.

Chaque GWP fait la convolution du filtre de Gabor orienté assigné avec les intensités de l'image $I(x, y)$ dans l'imagette Ω_G centrée en (x, y).

À chaque cycle d'horloge, chaque GWP exécute trois opérations en parallèle, à savoir :

– La convolution entre l'intensité lumineuse du pixel actuel et le coefficient de Gabor correspondant.

[39] Architecture parallèle spécialisée composée d'un grand nombre de processeurs identiques évoluant de manière synchrone.

[40] Architecture utilisée où nous faisons un enchaînement des traitements des données consécutives.

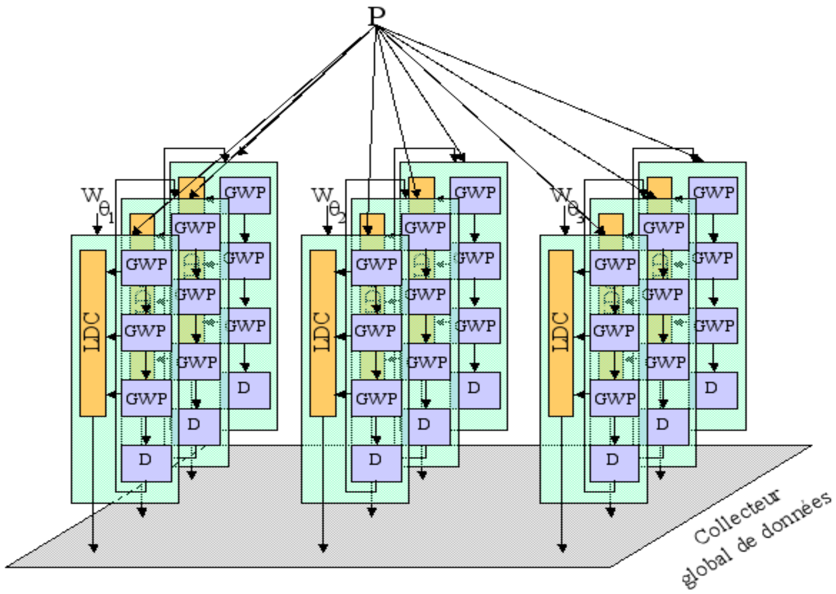


FIG. 2.6 – Diagramme de blocs de l'organisation spatiale de notre tableau systolique. Chaque tableau systolique calcule un filtre orienté de Gabor et l'ensemble des tableaux calcule en parallèle les principales orientations.

- L'accumulation de ce résultat avec les valeurs précédemment calculées, la lecture d'un nouveau coefficient et son stockage dans le registre interne.
- La transmission du coefficient précédent au GWP suivant.

Le processus d'initialisation, le détail du flot de données interne et les pas de calcul pour un tableau systolique simplifié de six GWP organisés dans deux colonnes sont montrés dans la figure II.2.7. Les GWP de la même colonne commencent à faire progressivement le traitement du même pixel d'entrée, mais avec des coefficients de Gabor différents. Pour cet exemple simplifié, nous considérons une imagette de Gabor de 7×7 et nous montrons l'imagette étendue qui contient à la fois les imagettes correspondant aux six GWP de cet exemple.

Après qu'une colonne de coefficients de Gabor a été lue (sept valeurs pour cet exemple) les coefficients doivent être retardés de deux cycles d'horloge (car il faut attendre que les dernières données soient traitées par le dernier GWP de cette colonne) et transmis ensuite à la colonne suivante de GWP comme il est indiqué par le flot de données sur le côté gauche de la figure II.2.7. En général nous aurons $n-1$ cycles d'horloge d'attente pour l'initialisation dans chaque ensemble de n lignes où n est la quantité de GWP physiquement implantés dans une colonne.

2.2.3 Aspect temporel

Le module matériel pour le traitement temporel causal est fondé sur le traitement d'un élément simple qui calcule la moyenne des résultats précédents en les utilisant dans des positions différentes dans la séquence d'images filtrées spatialement. Le calcul de la moyenne est déterminé par la vitesse supposée courante et l'orientation à une position donnée comme nous l'avons vu dans §II.1.2.2 et §II.2.1.2.

Un aspect clef pour une implantation matérielle efficace de la deuxième étape de mon filtrage FSTC est le schéma de segmentation et d'adressage de la mémoire. Dans l'architecture du module proposé, il est supposé que les images filtrées spatialement de la séquence $\mathcal{S}_\tau$ ont besoin d'un

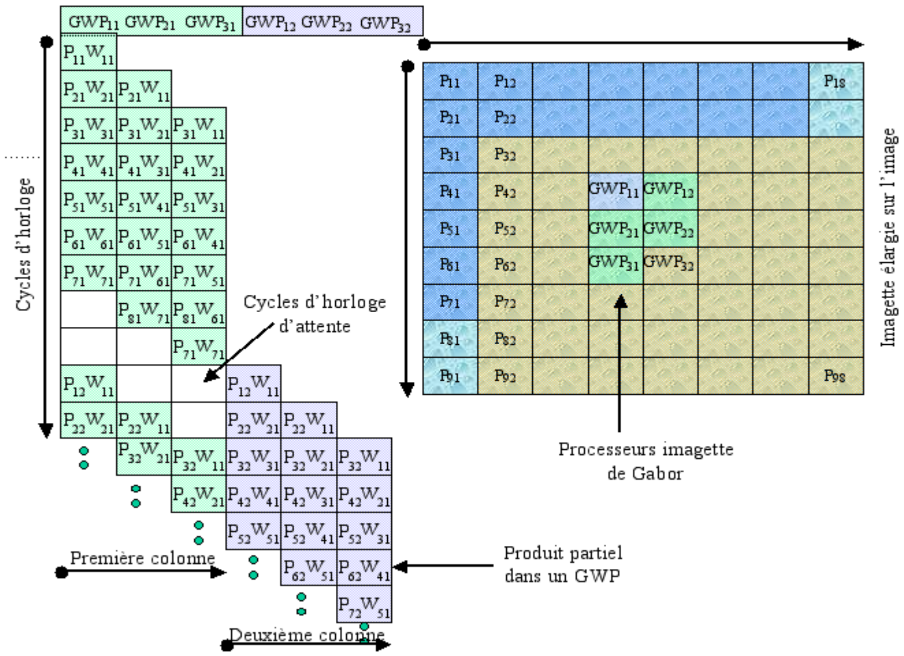


FIG. 2.7 – Le flot de données et les pas de calcul pour un tableau du traitement spatial proposé (GWP). Les calculs internes dans le temps sont montrés à gauche. Les GWP sont étroitement liés dans l'espace et travaillent avec des données communes.

traitement temporel pour une vitesse et une orientation spécifiques. Il faut donc pouvoir les récupérer dans des banques de mémoire indépendantes.

Le module matériel de traitement temporel causal est essentiellement composé d'une unité d'adressage parallèle programmable et d'un module de calcul de la moyenne. La figure II.2.8a montre un diagramme de bloc simplifié de la structure interne du module matériel pour le traitement temporel et son interface.

L'unité d'adressage produit en parallèle les différentes adresses pour avoir accès aux emplacements des différentes valeurs des images filtrées spatialement de la séquence $\mathcal{S}_\tau$ selon les conditions établies par une orientation et une vitesse données. Dans la figure II.2.8b nous montrons trois itérations pour le cas d'un déplacement à $\theta = 0^\circ$, $v = 1$ et $\tau = 3$.

Les valeurs obtenues par les GWP sont envoyées au module qui calculera leur moyenne en utilisant une arithmétique parallèle. Le résultat produit par le module d'adressage est envoyé à un emplacement de mémoire de sortie.

2.2.4 Étude de la satisfaction des contraintes d'embarquement

Les modules matériels proposés pour l'estimation du mouvement dans mon approche bio-inspirée ont été modélisés en utilisant le langage de description matérielle VHDL et en accord avec la description structurelle du premier module de mon modèle : le filtrage FSTC.

Le modèle VHDL pour chaque module a été simulé et validé en utilisant le logiciel Foundation ISE 5.1 de Xilinx. Les modules de traitement spatial et temporel ont été optimisés dans une certaine mesure. Les deux modules matériels ont été synthétisés indépendamment pour un dispositif Virtex FPGA de la Technologie de Synthèse Xilinx (Xilinx Synthesis Technology, XST), avec de bons résultats en termes d'utilisation des ressources matérielles et d'exécution. Les résultats de la synthèse et les chronométrages pour ces modules lorsqu'ils ont été synthétisés indépendamment

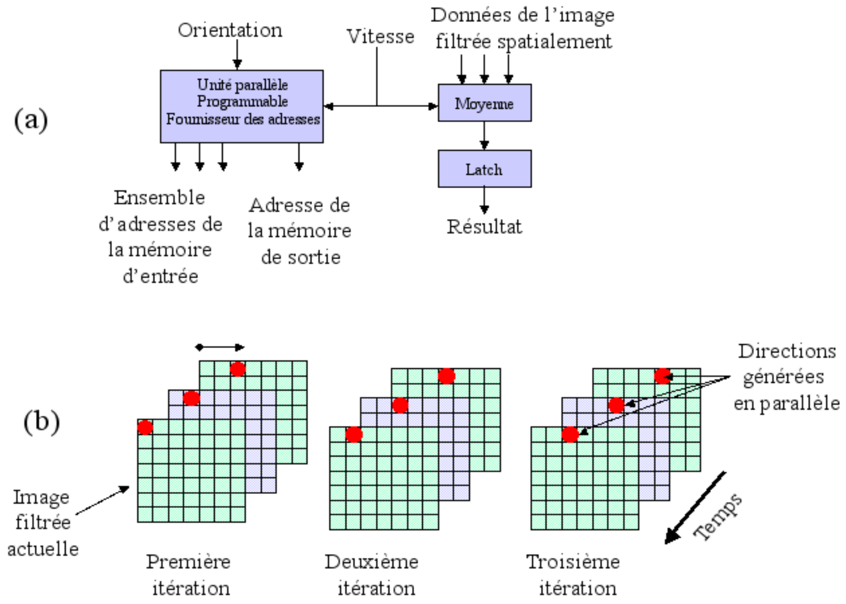


FIG. 2.8 – a) Diagramme simplifié en blocs du module de traitement temporel causal qui calcule la moyenne d'un ensemble des résultats du filtrage spatial dans différentes positions de la séquence $\mathcal{S}_\tau$. b) Exemple de la manipulation des mémoires pour trois valeurs de l'image filtrée spatialement et pour une vitesse supposée.

pour le dispositif XCV1000-4BG560 Virtex sont récapitulés dans le tableau II.2.1.

Le module de traitement spatial a été synthétisé pour une configuration de huit tableaux systoliques 2D, chacun avec 7×7 éléments de traitement. Ainsi, nous avons considéré huit orientations différentes et calculé en parallèle $392 = 8 \times 7 \times 7$ neurones sélectifs à l'orientation.

Ce module utilise l'arithmétique de point fixe avec une longueur de mot de 8 bits pour coder les pixels de l'image d'entrée et pour les différents coefficients de Gabor. Intérieurement, le module utilise un mot de 16 bits pour les opérations d'accumulation et de multiplication. La représentation interne de 16 bits est alors normalisée à une représentation de 8 bits, négligeant les bits les moins significatifs. Selon l'analyse d'exécution reportée par Torres et Arias [THAE04] et la fréquence d'horloge maximale annoncée dans notre synthèse, le temps exigé pour le filtrage spatial d'une image de 128×128 pixels dans les huit différentes orientations est autour de 0.56 millisecondes et le temps exigé pour le traitement temporel pour $V = 5$ différentes vitesses supposées, $D = 8$ différentes orientations et une séquence de taille $\tau = 3$ est autour de 6 millisecondes.

La mise au point du logiciel pour le calcul du filtrage spatial travaillant en série sur un ordinateur mono-processeur Pentium IV et 512 Mo de mémoire principale exige autour de 100 millisecondes pour la même résolution d'image et de tailles d'imagettes. Il faut noter aussi la puissance de consommation étant donnée que le ratio vitesse/consommation sur FPGA est plus grand que dans l'ordinateur Pentium IV.

De cette façon, nous obtenons une amélioration de la performance autour de 100 fois plus élevé que celle de l'ordinateur mono-processeur du type décrit ci-dessus. Ces résultats nous ouvrent la possibilité de la mise en œuvre en temps réel du module FSTC tout entier. Les fortes exigences informatiques et les besoins en largeur de bande de la mémoire de mon modèle bio-inspiré posent de fortes contraintes pour sa mise en œuvre et de même pour être résolues par des processeurs à signal numérique (DSP). À cause de leur structure interne, les FPGA sont convenables pour éviter les bouchons de mémoire vu qu'ils utilisent des registres internes et une mémoire intégrée à plusieurs ports.

Après avoir vu une description détaillée et l'implantation sur des circuits FPGA du premier module (FSTC) de mon modèle connexionniste neuromimétique pour la perception visuelle du mouvement, je vais montrer par la suite quelques exemples d'application de ce premier module.

Synthèse FPGA sur le dispositif Virtex XCV1000-4BG560		
Ressources matérielles	*Filtrage spatial*	*Traitement temporel*
Nombre de tranches	12286/12288	178/12288
Nombre de LUT à 4 entrées	18545/24576	323/24576
Nombre de bascule bistable	8020/24576	96/24576
Pourcentage d'utilisation du FPGA	99%	1%
Fréquence maximale d'horloge	46 Mhz	84.94 Mhz
Temps de traitement	0.56 ms	6 ms

TAB. 2.1 – Résume des synthèses FPGA réalisées.

2.3 Applications

J'ai validé mon premier module avec des séquences d'images synthétiques et réelles qui sont décrites dans l'annexe C. Les conventions appliquées pour l'interprétation des images de sorties et les paramètres libres utilisés sont reportés dans l'annexe D.1.

Par la suite, je vais d'abord définir le processus d'intégration de vitesses que j'appliquerai dans le module MFIA (voir le chapitre suivant). Puis, je montrerai quelques exemples d'application et mes premières conclusions.

2.3.1 Intégration des vitesses

Les cellules de MT répondent à une vitesse préférée [FK84] mais elles sont inhibées par le mouvement opposé à leur vitesse préférée [MNW83].

Selon Sekuler et ses collaborateurs [SWB02], l'être humain est capable de détecter le mouvement d'un objet simple, lorsqu'il a bougé au moins d'une minute d'arc (cette estimation dépend des caractéristiques de l'objet, la vitesse, l'illumination et la taille de la région rétinienne simulée), de plus il est capable de détecter un mouvement cohérent avec un *seuil de mouvement* [41] de 5% [Hei95]. Ce pourcentage varie par rapport à la taille de la scène, la densité de points et la durée d'exposition mais pas avec la direction du mouvement. Un mouvement fortement cohérent dans une direction augmente temporairement la cohérence des seuils pour les directions qui sont dans un azimut de 40° [Hei95]. Ces seuils peuvent varier en dépendant de l'attention dans une direction particulière du mouvement.

Les différentes illusions du mouvement fondées sur le contraste sont dues à l'utilisation des modèles codant les neurones avec une vitesse préférée [MNW86].

Dans mon filtrage FSTC je vais simplement intégrer les différentes orientations et les vitesses par un processus de type WTA. Les résultats du filtrage FSTC comme ceux du mécanisme MFIA

[41] Selon Tayama [Tay00], un seuil de mouvement est le pourcentage minimum de points nécessaires intégrant un signal pour la détection d'un mouvement cohérent.

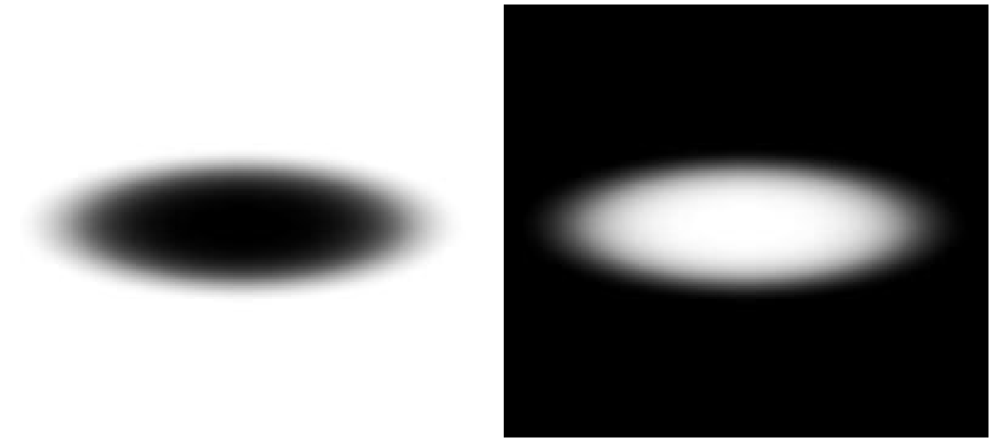
sont intégrés par le processus WTA défini par

$$\hat{H}(x,y,t) = max_{v \in V} \left(\sum_{\theta} H(x,y,t,\theta,v,T) \cdot \vartheta_\theta \right) \tag{2.8}$$

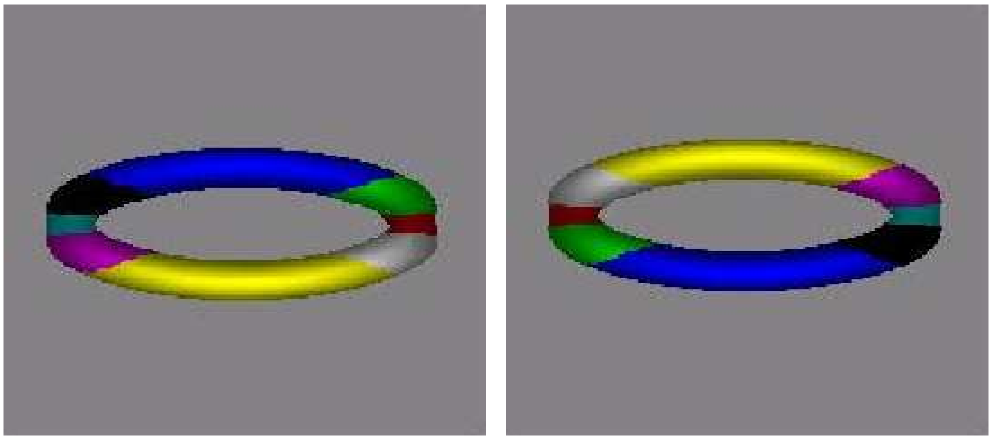
où $\hat{H}(x,y,t)$ est le neurone gagnant et ϑ_θ est le vecteur de direction unitaire θ. Donc, je cherche la somme vectorielle maximale sur tout l'ensemble d'orientations.

2.3.2 Exemples d'application

Mes expériences débutent par l'application de mon filtrage spatial à deux boules statiques montrées dans la figure II.2.9 : une noire sur fond blanc et une autre blanche sur fond noir.



(a) Boule noire (b) Boule blanche



(c) Filtrage spatial boule noire (d) Filtrage spatial boule blanche

FIG. 2.9 – Images pour tester la détection du contraste avec la première étape de mon filtrage FSTC.

Les figures II.2.9c et II.2.9d montrent la sortie de mon filtrage spatial appliqué à la boule noire et à la boule blanche, respectivement. Il faut noter que la sensibilité au contraste est fortement influencée lorsque j'applique mon filtrage FSTC aux images que j'appelle en positif (comme la boule noire) [42] et en négatif (comme la boule blanche) [43]. Notons que les réponses pour la même

[42] Les contrastes dans une image en positif sont référencés à un fond clair.
[43] Les contrastes dans une image en négatif sont référencés à un fond noir.

orientation du contraste, par exemple celle en couleur jaune qui représente un contraste vers le sud (direction à 270 °) dans la figure II.2.9c, sont mises dans la partie inférieure de la boule noire mais dans la partie supérieure de la boule blanche dans la figure II.2.9d. De même pour les sept autres orientations.

Cela nous dit que le traitement sur des images en négatif nous donnera des réponses opposées à celles obtenues dans des images en positif. Par la suite, les interprétations des directions de mouvement seront faites en supposant par défaut que toutes les images des séquences choisies pour mes expériences sont en positif.

Dans les figures II.2.10 et II.2.13 je montre l'application de mon filtrage FSTC au début des séquences du Taxi d'Hambourg et de Yosemite Fly-Through, respectivement. Les résultats du filtrage spatial montrent tous les différents contrastes trouvés dans l'image actuelle alors que les résultats du traitement temporel causal ne montrent que les objets ou personnes en mouvement.

En particulier, dans la figure II.2.10b nous pouvons y voir trois voitures en mouvement : le taxi au centre, la voiture noire en bas et à gauche et le camion en bas et à droite. Les deux derniers ne sont pas si détectables dans la figure II.2.10a. Le piéton en haut et à gauche est presque imperceptible dans mes expériences.

Dans la figure II.2.12, le passage du filtrage spatial au traitement temporel causal me fait perdre l'horizon mais nous montre les orientations de la partie inférieure centrale plus cohérentes. De plus, ce qu'on ne voyait pas à gauche dans le simple filtrage spatial est mis en valeur dans le traitement temporel causal et de façon plus cohérente.

Dans les images II.2.11 et II.2.13 j'ai mis trois images qui correspondent à l'application de mon filtrage FSTC qui est sensible aux changements d'illumination, i.e. je ne l'ai pas pondéré par sa moyenne locale (voir l'équation II.1.2 dans §II.1.2.1). Dans les deux dernières j'ai fait la pondération par la moyenne locale pour le rendre non sensible aux changements d'illumination, dont la deuxième est la séquence normale et la troisième en sens inverse.

L'application de mon filtrage FSTC sensible aux changements d'illumination atténue la perception des objets en faible contraste. Par contre, lorsque j'applique le traitement de l'intensité pour rendre le filtrage non sensible aux changements d'illumination (en référence à l'équation II.1.1), les objets en faible contraste et en mouvement sont mis en valeur pour être détectés. En particulier, dans la figure II.2.13 les résultats par rapport au mouvement des nuages dans la séquence sont plus complexes à interpréter. Avec le filtrage sensible aux changements d'illumination on perçoit "plus de détails" que avec le filtrage non sensible à ces changements. Cela est dû à ce que le filtrage non sensible aux changements d'illumination met en valeur les faibles contrastes mais diminue aussi les contrastes forts.

Enfin, la dernière image montre les résultats de l'application de mon filtrage FSTC à la même séquence mais en sens inverse, i.e. en la parcourant de la dernière image à la première. Dans la séquence inverse, les contrastes spatiaux sont conservés, mais les vitesses détectées sont opposées, et de par l'intégration des vitesses négatives, le résultat montre bien le changement de direction.

2.4 Bilan

Nous approcher du comportement des neurones simples dans V1 est une tâche lourde. J'ai simplement suivi les propositions issues de ma recherche en neurobiologie (voir §I.1) et des premiers modèles proposés (voir §I.2.3).

J'ai donc abouti à un module de filtrage spatio-temporel causal en combinant les idées du filtrage spatial de type Gabor avec une forte hypothèse d'un grand échantillonnage pour supposer des vitesses locales constantes.

Comme le filtrage passe-bande temporel filtre les contrastes statiques de telle façon que seuls les changements en contraste peuvent entrer dans les détecteurs [SLD82], ce type de filtrage temporel peut être implémenté dans un système biologique par des neurones ayant une réponse phasique [CI03]. Dans mon cas, je me suis inspiré de ces idées pour proposer mon filtrage temporel causal avec une rectification temporelle.

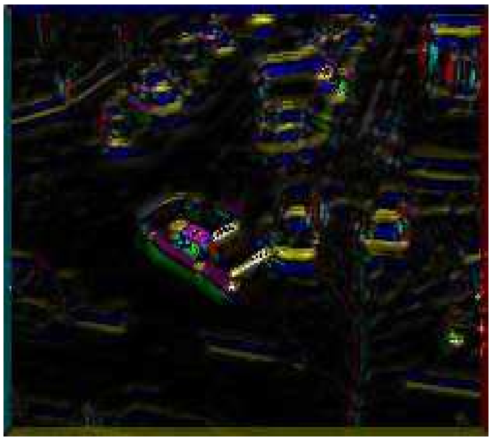
D'autre part, d'après Ibbotson et Clifford [IC01a], les détecteurs élémentaires peuvent préserver la polarité du signal d'entrée de telle façon que les caractéristiques du filtrage dépendent du stimulus de contraste en accordant un filtrage passe-bas temporel pour les images avec un faible contraste et un autre passe-bande pour les images avec un haut contraste.

Une telle stratégie de codification pourrait réduire la transmission de la redondance de l'information à haut contraste alors qu'elle maximise le signal des faibles contrastes [SLD82]. Mais j'ai plutôt choisi de diviser la valeur de sortie par la moyenne de son voisinage (moyenne locale) avec une pondération dépendant de la saturation maximale permise (voir §II.1.2.1).

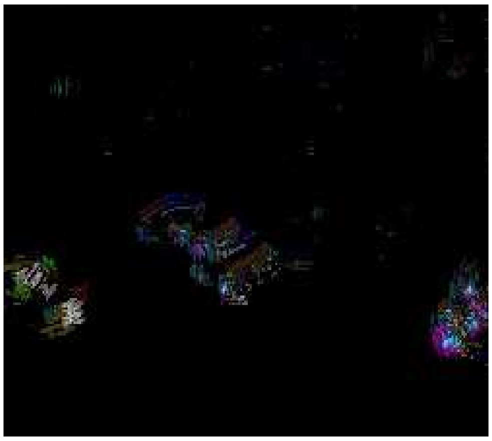
D'autre part, l'architecture proposée pour l'implantation matérielle du premier module de mon modèle connexionniste met en valeur une régularité et une modularité qui la rendent facilement extensible et je peux l'adapter à différentes tailles d'imagettes pour appliquer les filtres de Gabor.

De plus, le traitement parallèle de données et leur réutilisation sont fondés sur la manipulation d'une mémoire de type colonne et un ensemble d'éléments qui travaillent en parallèle. Donc, la dense connectivité et le travail de ces éléments prennent en compte le parallélisme naturel des neurones [Gir00]. Je peux dire que nous avons partiellement exploité le parallélisme neuronal [ibidem].

Jusqu'ici mon modèle reste entièrement local et distribué montrant des zones de mouvement assez diversifiées mais sans une réponse globale. Dans le chapitre suivant, je vais les rendre plus cohérentes par rapport au mouvement grâce à mon mécanisme MFIA et, finalement, dans le dernier chapitre je ferai le couplage complet de mon modèle pour extraire une réponse globale fondée sur la focalisation de l'attention.

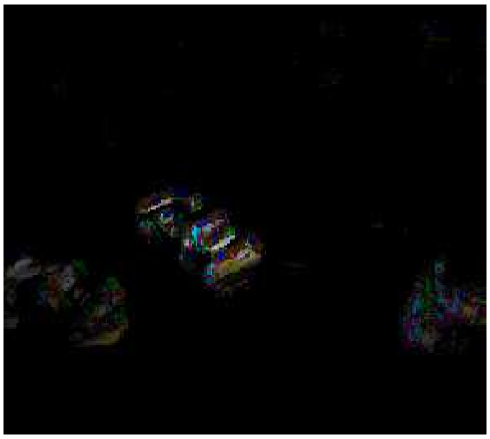


(a) Filtrage spatial

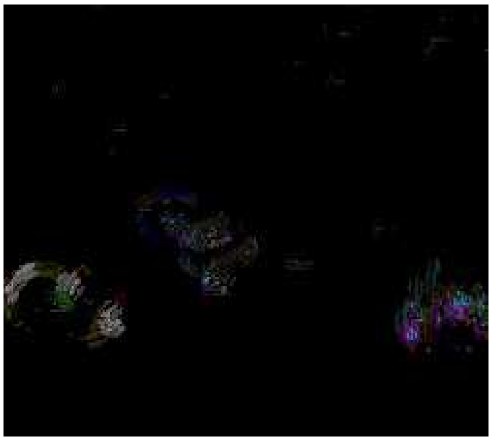


(b) Traitement temporel causal

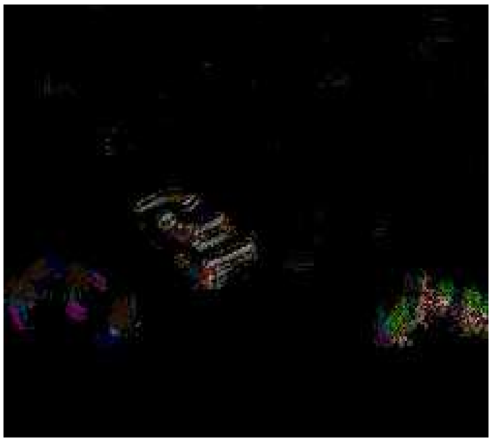
FIG. 2.10 – Application du filtrage FSTC au début de la séquence du Taxi d'Hambourg.



(a) FSTC sensible aux changements d'illumination

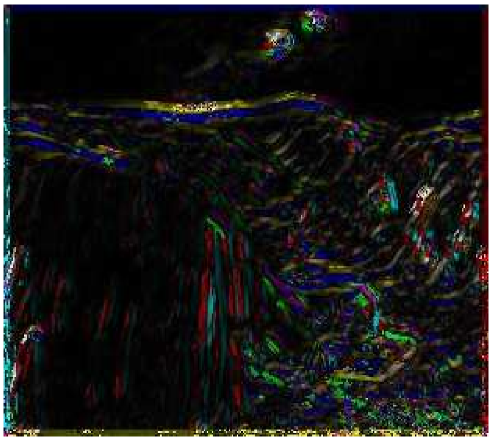


(b) FSTC non sensible aux changements d'illumination

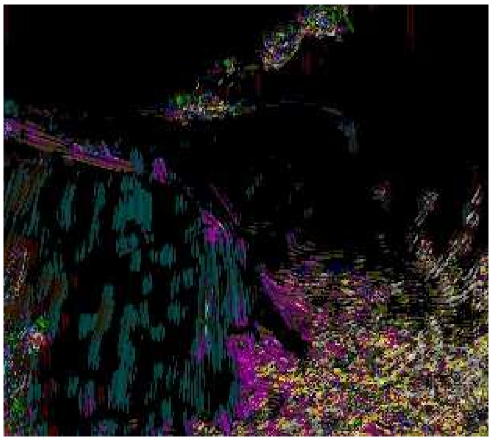


(c) Séquence inverse avec le FSTC non sensible aux changements d'illumination

FIG. 2.11 – Résultats obtenus de l'application du filtrage FSTC à la séquence du Taxi d'Hambourg sensible aux changements d'illumination, non sensible et pour la séquence en sens inverse.

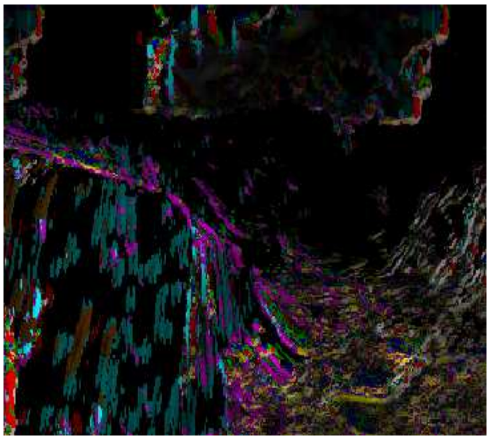


(a) Filtrage spatial

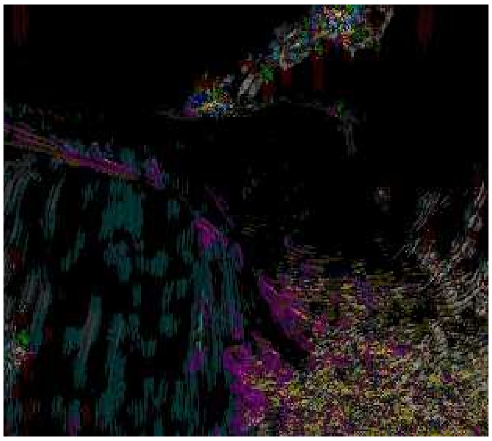


(b) Traitement temporel causal

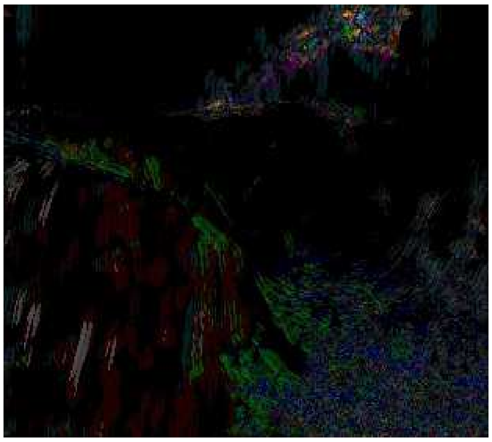
FIG. 2.12 – Application du filtrage FSTC au début de la séquence Yosemite Fly-Through.



(a) FSTC sensible aux changements d'illumination



(b) FSTC non sensible aux changements d'illumination



(c) Séquence inverse avec le FSTC non sensible aux changements d'illumination

FIG. 2.13 – Résultats obtenus de l'application du filtrage FSTC à la séquence de Yosemite Fly-Through sensible aux changements d'illumination, non sensible et pour la séquence en sens inverse.

Chapitre 3

Mécanisme de fortes interactions antagonistes

Sommaire

Le noyau de mon modèle s'inspire des fortes interactions latérales qu'expérimentent les neurones de V1 : une intense interaction locale et distribuée, intra-colonne de type excitatrice/inhibitrice de même qu'inter-colonne de type inhibitrice (voir §II.1.3).

Dans ce chapitre je vais préciser le deuxième module de mon modèle : le mécanisme de fortes interactions antagonistes (MFIA).

Dans un premier temps je vais spécifier les interactions latérales excitatrices et inhibitrices inspirées des interactions intra-colonnes et inter-colonnes d'orientation dans V1, respectivement (§II.3.1).

Ensuite j'analyserai l'équation II.1.7 et ses effets dans mon modèle (§II.3.2) et, ultérieurement, nous ferons une étude de la faisabilité de son implémentation matérielle sur FPGA (§II.3.3).

Finalement, je montrerai quelques exemples d'application (§II.3.4).

3.1 Présentation détaillée du mécanisme MFIA

Grossberg [Gro73, GL75, Gro78a, Gro80, Gro88, Gro82] a suggéré que le système visuel a tendance à chercher des objets du monde extérieur et à les comparer à d'anciens schémas de mémoire connus. Pour expliquer ce processus il a créé un modèle hiérarchique coopératif/compétitif de fonctions qui relient les perceptions sensorielles et le cerveau.

De plus, dans une compétition interdépendante, les gagnants émergent comme le résultat des interactions directes ou indirectes parmi les participants qui ont affecté leur performance [vO03]. En neurobiologie, ce type de compétition correspond aux effets inhibiteurs qu'un neurone reçoit d'autres neurones [Rib92, RB94].

La forte inhibition latérale donne un comportement de "gagnant prend tout" (winner-take-all, WTA) qui empêche que plusieurs neurones détecteurs de mouvement local dans un voisinage deviennent actifs simultanément [Mar03]. Par conséquent, tous les neurones détecteurs de mouvement local tendent à apprendre à répondre aux différentes séquences de mouvement. Mais cette forte inhibition latérale dans l'étape de détection du mouvement local ne peut représenter le mouvement que d'une seule caractéristique visuelle.

Lorsque j'applique le principe décrit ci-dessus à une image, l'inhibition-excitation est utilisée comme un mécanisme pour contrôler l'activité du réseau et produire des réponses émergentes logiques fondées sur des mécanismes semblables à des réseaux de neurones de type WTA. Dans de tels réseaux, un neurone reçoit tous les signaux d'inhibition et d'excitation dans un voisinage ou une sphère d'influence qui règle son activité selon un mécanisme d'interaction prédéterminé. D'habitude, les interactions entre les neurones dans la sphère d'influence ont une force modulée selon leur réponse et la distance de l'un à l'autre.

Donc, pour le module MFIA, je décrirai tout d'abord les différentes sphères d'influence nécessaires pour le mécanisme ; ensuite, sa règle d'actualisation ; plus tard, les poids synaptiques issus de cette règle d'actualisation et, finalement, le calcul de la contribution des différentes vitesses du neurone.

3.1.1 Les sphères d'influence

Dans le cortex visuel primaire, les colonnes d'orientation corticale ont de denses interactions excitatrices internes et de denses interactions inhibitrices aussi auteur d'eux (voir §I.1.3.1).

Dans mon modèle, chaque neurone dans le deuxième module reçoit : un ensemble d'entrées excitatrices $\Omega_{(x,y)}^{\Theta_E}$ ayant quasiment la même direction d'excitation principale qui correspondrait au comportement des neurones d'une colonne corticale dans V1, et un ensemble d'entrées inhibitrices $\Omega_{(x,y)}^{\Theta_I}$ correspondant aux autres directions.

Ensemble de directions

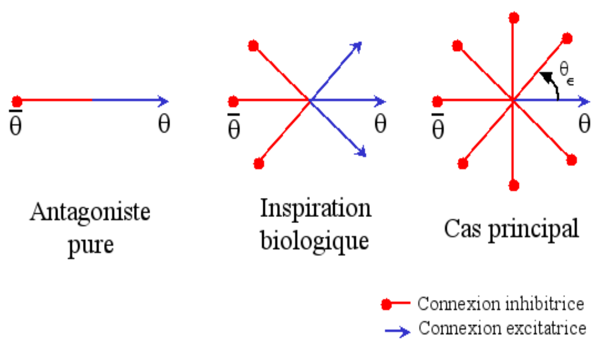


FIG. 3.1 – Différents schémas d'interactions orientées antagonistes testés dans mon mécanisme MFIA.

Les ensembles de directions excitatrices et inhibitrices sont donnés par

$$\begin{aligned} \Theta_E &= \{\theta_e \mid \theta_e = \theta \pm m \cdot \theta_\epsilon, \mid \theta_e - \theta \mid \leq \vartheta_e, \text{ et } m \in \mathcal{N}^*\} \\ \Theta_I &= \{\theta_i \mid \theta_i = \bar{\theta} \pm m \cdot \theta_\epsilon, \mid \theta_i - \bar{\theta} \mid \leq \vartheta_i, \text{ et } m \in \mathcal{N}^*\} \end{aligned} \tag{3.1}$$

où θ est la direction centrale des interactions excitatrices, $\bar{\theta} = \theta + \pi$ est la direction centrale antagoniste, c'est-à-dire, la direction complètement opposée à la direction θ, θ_ϵ l'écart entre deux directions consécutives comme je l'ai défini dans §II.1.2.1. ϑ_e est l'angle maximal d'ouverture depuis la direction centrale excitatrice θ et ϑ_i est l'angle maximal d'ouverture depuis la direction centrale inhibitrice $\bar{\theta}$. Pour simplifier certaines équations communes à l'excitation et à l'inhibition, j'utilise la notation θ_z pour représenter à la fois θ_e et θ_i (le voisinage d'excitation et d'inhibition, respectivement).

La figure II.3.1 montre trois schémas différents d'interaction directionnelle dans mon mécanisme antagoniste. De gauche à droite, le premier schéma montre l'antagonisme pur : une direction avec sa direction antagoniste. Le deuxième schéma suggère l'interaction entre les colonnes d'orientation proches de l'orientation θ et de son antagoniste $\bar{\theta}$. Le troisième est le schéma le plus souvent utilisé dans mon mécanisme (on les a tous testés néanmoins).

Définition des sphères d'influence

Les sphères d'influence excitatrices sont les voisinages d'excitation définis par

$$\Omega^{\Theta_E}_{(x,y)} = \{(\xi_{\theta_e}, \eta_{\theta_e}) \mid\mid x - \xi_{\theta_e} \mid \leq \xi_E, \mid y - \eta_{\theta_e} \mid \leq \eta_E, x, y \in \mathcal{N}^*, \xi_{\theta_e}, \eta_{\theta_e} \in \mathcal{Z}, \theta_e \in \Theta_E\} \tag{3.2}$$

et centrés dans le neurone localisé en (x, y) dans la direction θ_e. ξ_E et η_E sont les deux bornes supérieures de son rayon d'influence excitatrice.

Les sphères d'influence inhibitrices sont les voisinages d'inhibition définis par

$$\Omega^{\Theta_I}_{(x,y)} = \{(\xi_{\theta_i}, \eta_{\theta_i}) \mid\mid x - \xi_{\theta_i} \mid \leq \xi_I, \mid y - \eta_{\theta_i} \mid \leq \eta_I, x, y \in \mathcal{N}^*, \xi_{\theta_i}, \eta_{\theta_i} \in \mathcal{Z}, \theta_i \in \Theta_I\} \tag{3.3}$$

et centrés dans le neurone localisé en (x, y) dans la direction θ_i. ξ_I et η_I sont les deux bornes supérieures de son rayon d'influence inhibitrice.

La figure II.3.2 montre quatre sphères d'influence différentes centrées dans le neurone localisé en (x, y).

Rayon d'influence

Le rayon d'influence du neurone localisé en (x, y) est calculé par

$$RI_\Gamma(H(x, y, \tilde{t}, \theta_z, v, T), s(\tilde{t})) = \xi_{\theta_z} = \eta_{\theta_z} = \left\lfloor \frac{\Gamma \cdot H(x, y, \tilde{t}, \theta_z, v, T)}{s(\tilde{t})} + 0.5 \right\rfloor \tag{3.4}$$

Γ est le rayon d'influence maximal proposé. $\lfloor(\cdot)\rfloor$ est la partie entière de $(\cdot)$ et $s(\tilde{t}) = 2 \cdot max_{x,y,\tilde{t},\theta,v}[F(x, y, \tilde{t}, \theta, v)]$ est la saturation maximale des neurones. Elle est obtenue en cherchant le maximum parmi toutes les directions et les vitesses de tous les neurones dans la séquence des images filtrées par mon filtrage FSTC, $\mathcal{S}^{FSTC}_\tau$.

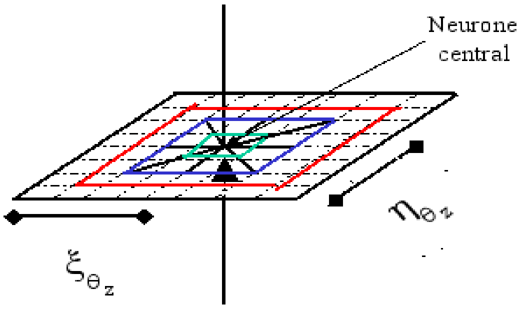


FIG. 3.2 – Quatre sphères d'influence différentes pour le neurone localisé dans la position (x, y).

3.1.2 Règle d'actualisation dans le mécanisme MFIA

L'actualisation des neurones dans le mécanisme MFIA est le résultat de la forte interaction coopérative/compétitive avec leur propre entourage défini par leur sphère d'influence.

Je définis donc l'initialisation du MFIA pour le neurone localisé en (x, y) à l'itération $T = 0$ par

$$H(x, y, \tilde{t}, \theta, v, T = 0) = F(x, y, \tilde{t}, \theta, v) \tag{3.5}$$

i.e. comme la sortie du filtrage FSTC dans la séquence $\mathcal{S}_{\tau}^{FSTC}$ ce qui constitue l'initialisation du mécanisme MFIA.

En discrétisant l'équation II.1.7 et en l'adaptant à mon mécanisme, je propose

$$\begin{aligned} H(x, y, \tilde{t}, \theta, v, T + 1) &= H(x, y, \tilde{t}, \theta, v, T) \\ &+\eta \cdot [-\mathcal{A} \cdot H(x, y, \tilde{t}, \theta, v, T) \\ &+(\mathcal{B} - H(x, y, \tilde{t}, \theta, v, T)) \cdot Exc(x, y, \tilde{t}, \theta, v, T) \\ &-(\mathcal{C} + H(x, y, \tilde{t}, \theta, v, T)) \cdot Inh(x, y, \tilde{t}, \theta, v, T)] \end{aligned} \tag{3.6}$$

pour définir l'actualisation de la réponse du neurone localisé en (x, y) à l'itération $T + 1$.

Le premier terme est l'état antérieur du neurone à actualiser et le deuxième, mon interprétation de son actualisation où, à l'intérieur de ce deuxième terme, le premier terme est la fuite, le deuxième la contribution excitatrice/inhibitrice de son entourage et le troisième la contribution inhibitrice de son entourage. $\mathcal{A}$, $\mathcal{B}$ et $\mathcal{C}$ sont des constantes réelles et η le coefficient d'apprentissage.

De l'équation II.1.7 et d'après Grossberg [Gro73, Gro88], $H(x, y, T) \in [-\mathcal{C}, \mathcal{B}]$ et $H(x, y, T)$ décroît jusqu'à un niveau de repos 0 en l'absence de toutes les entrées.

La contribution excitatrice des neurones dans les sphères d'influence excitatrices $\Omega_{(x,y)}^{\Omega_E}$ est définie par

$$Exc(x, y, \tilde{t}, \theta, v, T) = F(x, y, \tilde{t}, \theta, v) + \sum_{\Omega_{(x,y)}^{\Theta_E}} W_{\Theta_E}(x, y, \xi_{\theta_e}, \eta_{\theta_e}, \theta_e, \theta) \tilde{H}(\xi_{\theta_e}, \eta_{\theta_e}, \tilde{t}, \theta_e, v, T) \tag{3.7}$$

Ainsi, un neurone dans une orientation donnée a un effet excitateur sur lui-même et sur les neurones dans sa sphère d'influence excitatrice. $\tilde{H}(x, y, \tilde{t}, \theta, v, T)$ sera bientôt définie dans §II.3.1.4 et fait intervenir les autres vitesses. Les poids synaptiques W_{Θ_Z} sont décrits dans §II.3.1.3.

D'autre part, un neurone dans l'orientation θ sera inhibé par les neurones antagonistes, c'est-à-dire, par les neurones dans les sphères d'influence inhibitrice $\Omega_{(x,y)}^{\Theta_I}$. Les contributions inhibitrices des neurones dans les sphères d'influence inhibitrices sont définies comme

$$Inh(x, y, \tilde{t}, \theta, v, T) = \sum_{\Omega^{\Theta_I}_{(x,y)}} W_{\Theta_I}(x, y, \xi_{\theta_i}, \eta_{\theta_i}, \theta_i, \theta)\tilde{H}(\xi_{\theta_e}, \eta_{\theta_e}, \tilde{t}, \theta_i, v, T) \tag{3.8}$$

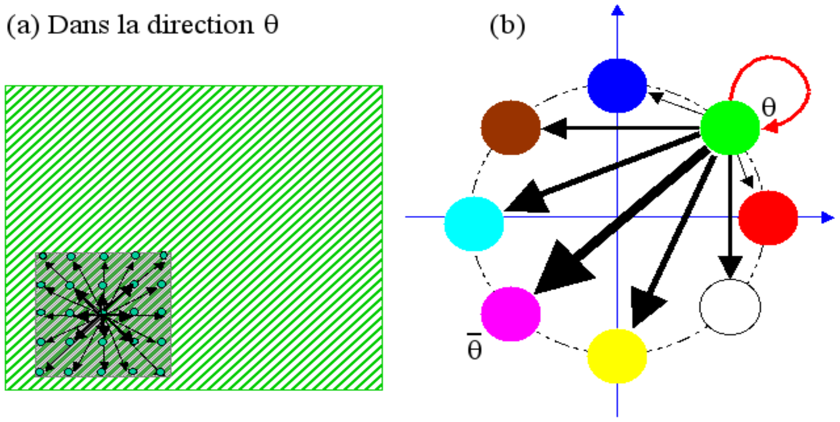


FIG. 3.3 – Les interactions excitatrices (a) et inhibitrices (b) dans le mécanisme MFIA pour le cas (c) de la figure II.3.1.

Dans la figure II.3.3a je montre les interactions excitatrices dans l'orientation principale du neurone localisé en (x, y) et ses interactions inhibitrices correspondantes avec les autres orientations (II.3.3b).

La figure II.3.4 montre la contribution antagoniste des différentes vitesses dans les sphères d'influence respectives sur le neurone localisé dans l'orientation θ_z. Les différentes épaisseurs des connexions dans la carte MFIA montrent les différentes pondérations des interactions inhibitrices antagonistes sur le neurone à actualiser.

Par la suite je vais définir les différents poids synaptiques excitateurs et inhibiteurs.

3.1.3 Les poids synaptiques dans les sphères d'influence

Dans la plupart des modèles neuronaux excitateurs-inhibiteurs, les poids de connexion d'un neurone vers d'autres neurones sont modulés par la distance d'un neurone à l'autre. C'est également le cas dans mon modèle, mais j'y prends aussi en compte d'autres critères que je vais détailler par la suite.

Ainsi, $W_{\Theta_E}(x, y, \xi_{\theta_e}, \eta_{\theta_e}, \theta_e, \theta)$ et $W_{\Theta_I}(x, y, \xi_{\theta_i}, \eta_{\theta_i}, \theta_i, \theta)$, qui sont les poids synaptiques pour les connexions excitatrices et inhibitrices, respectivement, dépendent de :

– La distance euclidienne entre le neurone central (x, y) et son voisinage.
– L'influence du neurone localisé en (x, y).
– La distance angulaire entre l'angle θ_z et θ.

Ainsi, les poids synaptiques excitateurs sont définis comme

$$W_{\Theta_E}(x, y, \xi_{\theta_e}, \eta_{\theta_e}, \theta_e, \theta) = \begin{cases} \frac{|\theta-\theta_e|}{\pi} K(x, y, \xi_{\theta_e}, \eta_{\theta_e}, \theta_e), & si \quad \theta_e \neq \theta \\ K(x, y, \xi_{\theta_e}, \eta_{\theta_e}, \theta_e), & sinon \end{cases} \tag{3.9}$$

et les poids synaptiques inhibiteurs comme

$$W_{\Theta_I}(x, y, \xi_{\theta_i}, \eta_{\theta_i}, \theta_i, \theta) = 1 - \frac{|\ \theta_i - \theta\ |}{\pi} K(x, y, \xi_{\theta_i}, \eta_{\theta_i}, \theta_i) \tag{3.10}$$

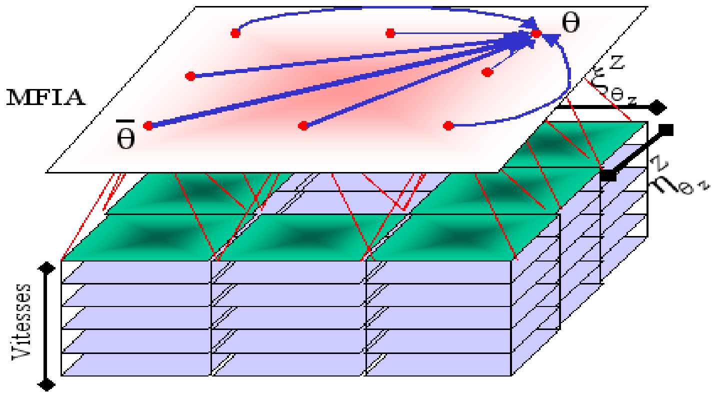


FIG. 3.4 – Règle d'actualisation des neurones dans le mécanisme MFIA. Le neurone à actualiser localisé en (x, y) dans l'orientation θ va exciter à tous ses voisins ayant la même vitesse que lui et étant localisés dans son sphère d'influence excitatrice (montré ici par $\xi_{\theta_z}^Z \times \eta_{\theta_z}^Z$). Les neurones ayant la même orientation θ mais avec des vitesses différentes à celle du neurone à actualiser vont l'inhiber. Ce principe se répète pour chacune des autres orientations dans leur neurone localisé dans la position (x, y) de chaque orientation. La somme pondérée des résultats de chaque ensemble ayant une orientation différente à celle du neurone à actualiser est l'inhibition totale du système.

J'ai mis en valeur dans les équations II.3.9 et II.3.10 la dépendance à la distance angulaire entre θ_z et θ.

La dépendance à la distance euclidienne et le rayon d'influence sont définis par $K(x, y, \xi_e, \eta_e, \theta_z)$ à savoir

$$K(x, y, \xi_{\theta_z}, \eta_{\theta_z}, \theta_z) = \begin{cases} g(d(\cdot), RI_\Gamma(H(\cdot), s(\tilde{t})), \mu, \sigma), & si \quad d(\cdot) \leq RI_\Gamma(H(\cdot), s(\tilde{t})) \\ 0, & sinon \end{cases} \tag{3.11}$$

pour $\xi_{\theta_z} \in \Omega_{(x,y)}^{\Theta z}$ et $\eta_{\Theta_z} \in \Omega_{(x,y)}^{\theta z}$ dans les régions soit d'excitation soit d'inhibition. $d(\cdot) = d(x, y, \xi_{\theta_z}, \eta_{\theta_z})$ est la distance euclidienne de la localisation du neurone (x, y) à celle du neurone $(\xi_{\theta_z}, \eta_{\theta_z})$. $H(\cdot) = H(x, y, \tilde{t}, \theta_z, v, T)$ la réponse du neurone localisé en (x, y) dans l'itération T du mécanisme MFIA et, finalement, $g(d(\cdot), RI_\Gamma(H(\cdot), s(\tilde{t})), \mu, \sigma)$ est la gaussienne centrée sur le neurone en (x, y) avec une moyenne μ et un écart type σ.

3.1.4 Le traitement des vitesses

Comme nous avons vu dans §I.1.3.2, les éléments proches avec des vitesses similaires font une association significative des champs des vitesses avec des liens préférentiels où la vitesse est construite par la somme des sorties pondérées de V1 accordées aux bandes de fréquences spatio-temporelles liées à un plan commun. Chez moi ce plan commun sera chaque colonne d'orientation préférentielle.

L'interaction entre les différentes vitesses est définie par

$$\tilde{H}(x, y, \tilde{t}, \theta_e, v, T) = \left\| \sum_{j=-\mathcal{V}, v \neq v_j}^{\mathcal{V}} \left(\vec{H}(x, y, \tilde{t}, \theta_e, v, T) - \vec{H}(x, y, \tilde{t}, \theta_e, v_j, T) \right) \right\| \tag{3.12}$$

i.e. la somme des différences vectorielles de la réponse du neurone localisé en (x, y) à la vitesse v dans les directions d'excitation θ_e avec toutes les autres vitesses avec les mêmes caractéristiques.

Pour les orientations inhibitrices je fait de même, i.e. j'applique l'équation II.3.12 à l'orientation θ_i. $\|\vec{\cdot}\|$ est la norme du vecteur $\vec{\cdot}$.

D'autre part,

$$\vec{H}(x, y, \tilde{t}, \theta, v, T) = (H(x, y, \tilde{t}, \theta, v, T)cos\theta, H(x, y, \tilde{t}, \theta, v, T)sin\theta) \tag{3.13}$$

est la projection cartésienne de la réponse $H(x, y, \tilde{t}, \theta, v, T)$ du neurone localisé en (x, y) dans l'itération T du mécanisme MFIA.

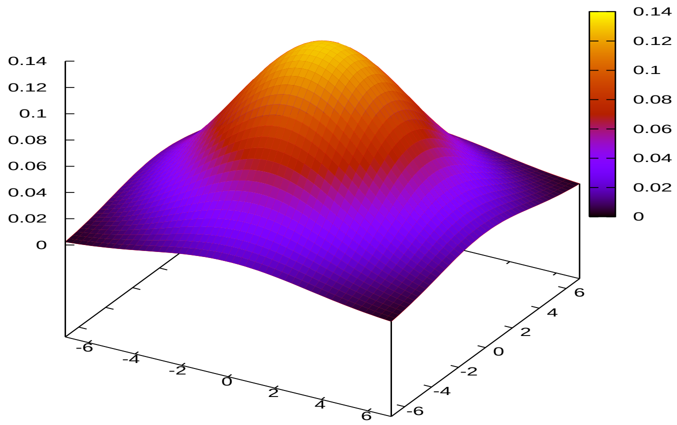


FIG. 3.5 – Pondérations par rapport à la distance euclidienne entre le neurone centré dans sa sphère d'influence (neurone pivot) et les autres neurones dans sa sphère d'influence.

La figure II.3.5 montre les différentes pondérations appliquées aux distances euclidiennes par la définition de la fonction gaussienne normale centrée dans le neurone localisé en (x, y) de moyenne μ et écart type σ.

3.2 Analyse statistique de l'équation d'actualisation

Le module du mécanisme MFIA de mon modèle connexionniste fait émerger des zones de mouvement plus ou moins cohérent (voir §II.3.4 ci-après), mais la maîtrise de ses paramètres au niveau expérimental reste encore problématique.

Ce paramétrage reste difficile maintenant, mais à un moment donné, dans l'état d'avancement des travaux, les résultats étaient tels que des doutes sont apparus quant au comportement de l'équation elle-même, ce qui a motivé l'étude théorique confiée à un DEA de profil mathématique-vision pour l'étude d'éventuels problèmes sur la convergence locale de l'équation II.1.7, d'une part, et pour caractériser les tâches perceptives qui peuvent émerger du mécanisme MFIA, d'autre part. Nous n'en rapportons ici que les lignes essentielles, pour plus de détails voir [Jac04].

Nous ne cherchons pas précisément la solution à l'équation II.1.7 (les techniques de résolution utilisées dans [Gro73, Gro88] ne sont par ailleurs pas applicables dans notre cas multidimensionnel), nous cherchons plutôt à étudier sa dynamique dans la version discrétisée.

Les contraintes portent sur le degré de connectivité et de récurrence, la topologie des traitements, la dynamique des cellules nerveuses et leur relations excitatrices et inhibitrices.

Dans les expérimentations, surtout celles précédant cette étude théorique, deux types de problèmes émergeaient principalement pour réaliser l'apprentissage de la détection de mouvement dans une direction et une vitesse données : l'effet de flou dans les mouvements détectés et une détection de mouvement moins performante dans les zones de l'image faiblement contrastées.

Dans la figure II.3.6, nous montrons la sortie du mécanisme MFIA pour la séquence Taxi d'Hambourg d'après nos expériences. On peut bien y voir, le taxi (au centre) avec deux zones plus au moins homogènes dans une direction. La voiture à gauche et le camion à droite sont faiblement repérés, mais ce type de sorties ne sont plus valables dans la version actuelle de mon modèle connexionniste.

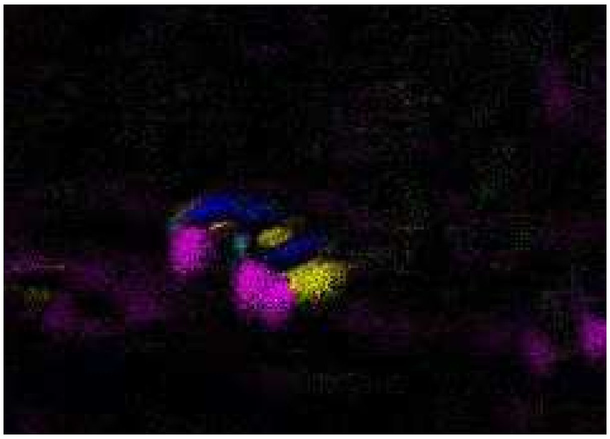


FIG. 3.6 – Sortie du mécanisme MFIA pour la séquence du Taxi d'Hambourg avant l'analyse statistique de l'équation d'actualisation II.1.7.

D'après Grossberg [Gro73, Gro88] le modèle hiérarchique coopératif/compétitif de fonctions qui relient les perceptions sensorielles passe par l'équation d'activation connue comme modèle additif où la différence entre ses entrées excitatrices et ses entrées inhibitrices actualise l'activité du neurone en considération.

La généralisation du modèle additif est le modèle "shunting inhibition" (voir §I.1.3.1). Dans un tel réseau les entrées excitatrices stimulent le neurone pour qu'il atteigne un maximum fini, alors que les entrées inhibitrices inhibent le neurone pour qu'il atteigne un minimum fini ($\mathcal{B}$ et $-\mathcal{C}$ dans l'équation II.1.7, respectivement).

Ce type de réseau est capable, d'après Grossberg [Gro73, Gro88], de traiter les rapports, conserver et normaliser l'activation totale (limitation de la capacité), avoir un certain niveau d'adaptation, supprimer le bruit, perfectionner le contraste, avoir une mémoire à court terme, effectuer une amplification énergique des entrées correctes et une suppression des entrées mal adaptées. Mais l'équation II.1.7 est vectorielle multi-dimensionnelle et l'application directe des résultats de Grossberg n'est pas évidente. Ne disposant pas d'outil analytique permettant d'étudier la qualité de la convergence locale du mécanisme employé, nous nous sommes tournés vers une approche statistique estimant cette convergence locale à partir de multiples contours et trajectoires générés aléatoirement et bruités. Nous décrivons les grandes lignes de cette étude statistique du comportement de l'équation II.1.7 dans la sous-section suivante.

3.2.1 Application de la méthode du point fixe

L'analyse de l'équation II.1.7 nous ramène à une analyse vectorielle multi-dimensionnelle. Vue la complexité de l'analyse, nous avons plutôt choisi d'étudier le problème de la convergence vers une réponse cohérente stable. Pour cela, notre première approche a été d'étudier si la fonction de mise à jour de l'ensemble des neurones était une fonction localement contractante, la notion de contractance (globale néanmoins) impliquant la convergence voulue. Cette sous-section débute par quelques rappels sur la notion de contractance.

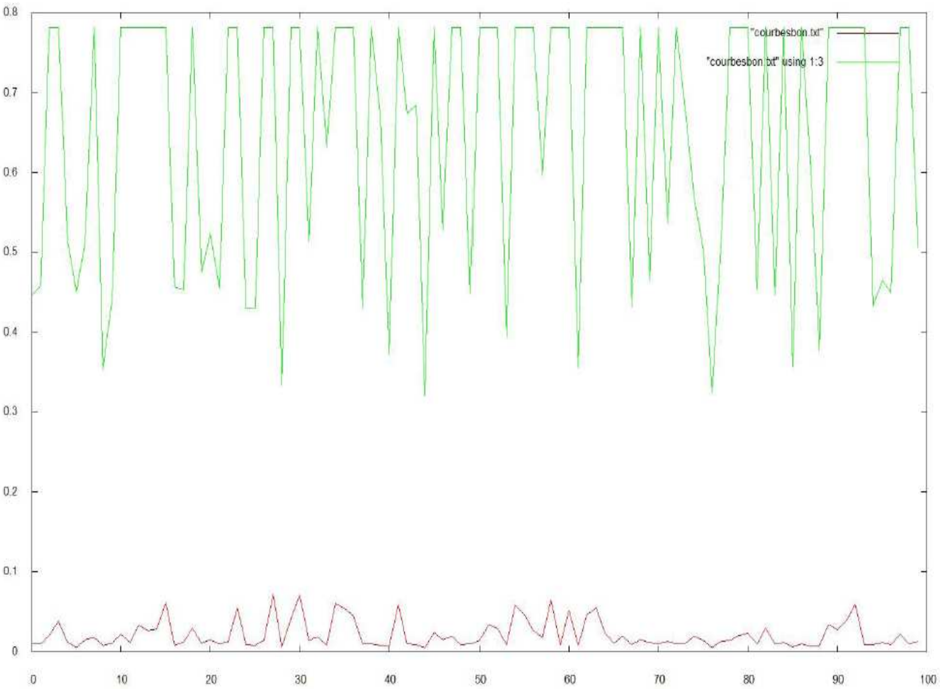


FIG. 3.7 – Valeurs de contractance moyenne et maximale obtenues pour 100 séquences d'images.

Soit

$$h : f_n \longrightarrow f_n + \epsilon\sigma(f_n) \tag{3.14}$$

où h représente la fonction de mise à jour discrète, f_n est le vecteur des sorties des neurones dans le mécanisme MFIA et $\sigma(x)$ correspond à l'ensemble d'équations du mécanisme MFIA. Donc, dans notre problème, l'évolution de l'ensemble des neurones correspond à la suite récurrente

$$f_{n+1} = h(f_n) \tag{3.15}$$

Si

$$\|h'\| < 1 \tag{3.16}$$

alors h est contractante, c'est à dire quel que soit x et y, $\|h(y) - h(x)\| < \|y - x\|$. Et h étant contractante, la suite f_n converge. Dans notre cas, nous allons devoir nous limiter à une étude locale de la contractance, le long de la trajectoire suivie par f_n dans son espace multidimensionnel. Nous cherchons donc à savoir si $\|h'\|$ est localement strictement inférieur à 1.

Pour que notre fonction $h(\vec{u})$ soit localement contractante, il suffit que la norme de sa différentielle locale soit inférieure à 1, c'est à dire :

$$\forall \|\partial(\vec{u})\| \to 0, \quad max \tfrac{\|h(\vec{u}+\delta(\vec{u}))-h(\vec{u})\|}{\|\delta(\vec{u})\|} < 1 \tag{3.17}$$

L'objectif est d'observer la variation du mécanisme MFIA lorsque l'on modifie très légèrement un neurone donné et son voisinage. Pour cela nous faisons une étude sur un grand nombre de séquences d'images réelles et d'images synthétiques dans lesquelles nous choisissons $\partial(\vec{u})$ de façon aléatoire.

3.2.2 Résultats

Le phénomène que nous désirions observer était la présence de contractance dans mon mécanisme MFIA. Par la suite et en fonction des résultats, nous souhaitions savoir si elle dépendait des paramètres de l'équation II.1.7, comment elle évoluait en fonction du contraste ou du bruit présent dans une image.

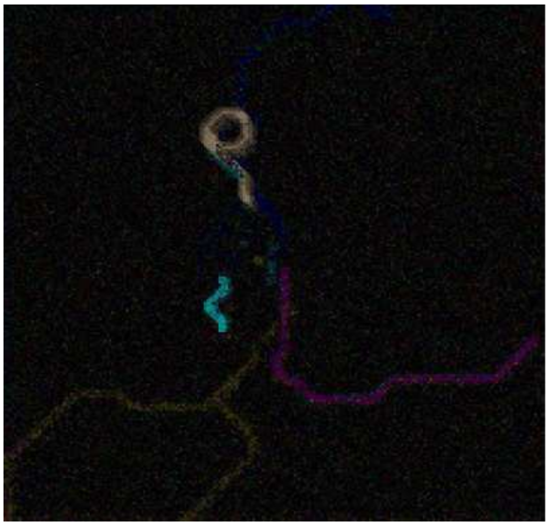


FIG. 3.8 – Les différentes couleurs du contraste indique sa direction du mouvement dans une image de synthèse (voir les codes de couleurs pour les différentes directions dans la figure D.1).

D'après les expériences sur des centaines d'images, si nous considérons l'ensemble de l'image, l'indice de contractance moyenne ($\|h'\|$) est relativement proche de 0, donc une contractance forte. La figure II.3.7 montre les courbes de contractance moyenne en rouge et de contractance maximale en vert pour cent différentes séquences d'images. Ce résultat paraît raisonnable étant donné les grandes zones uniformes de l'image. Les valeurs locales maximales oscillent entre 0.4 et 0.8 et peuvent varier en fonction des paramètres de la fonction d'actualisation de mon mécanisme MFIA.

Lorsque l'on fait varier les paramètres $\mathcal{A}$, $\mathcal{B}$ et $\mathcal{C}$ pour une image donnée, nous obtenons des résultats assez similaires.

Si $\mathcal{B}$ et $\mathcal{C}$ sont supérieurs à 1, le paramètre $\mathcal{A}$ ne semble pas intervenir pour la contractance. Cela signifie que le terme d'amortissement de la fonction joue un rôle moindre comparé aux phénomènes d'excitation et d'inhibition des cellules.

Si $\mathcal{B}$ et $\mathcal{C}$ sont inférieurs à 1, le paramètre $\mathcal{A}$ joue un rôle dans la contractance.

On observe dans l'ensemble des cas que la courbe est similaire, la modification des paramètres revient à effectuer une translation sur les valeurs représentatives de la courbe (pour plus de détails voir [Jac04]).

Un autre aspect qui a attiré notre attention est la variation de la contractance en fonction du contraste des contours de l'image. Plus les contours sont contrastés moins l'image est contractante. Cette conclusion est raisonnable car lorsque les contours sont en mouvement, les interactions excitatrices et inhibitrices sont plus importantes et par conséquent la variation également. En revanche, lorsqu'on se trouve dans une zone très peu contrastée, le contour se confond avec le fond et l'image apparaît quasiment uniforme.

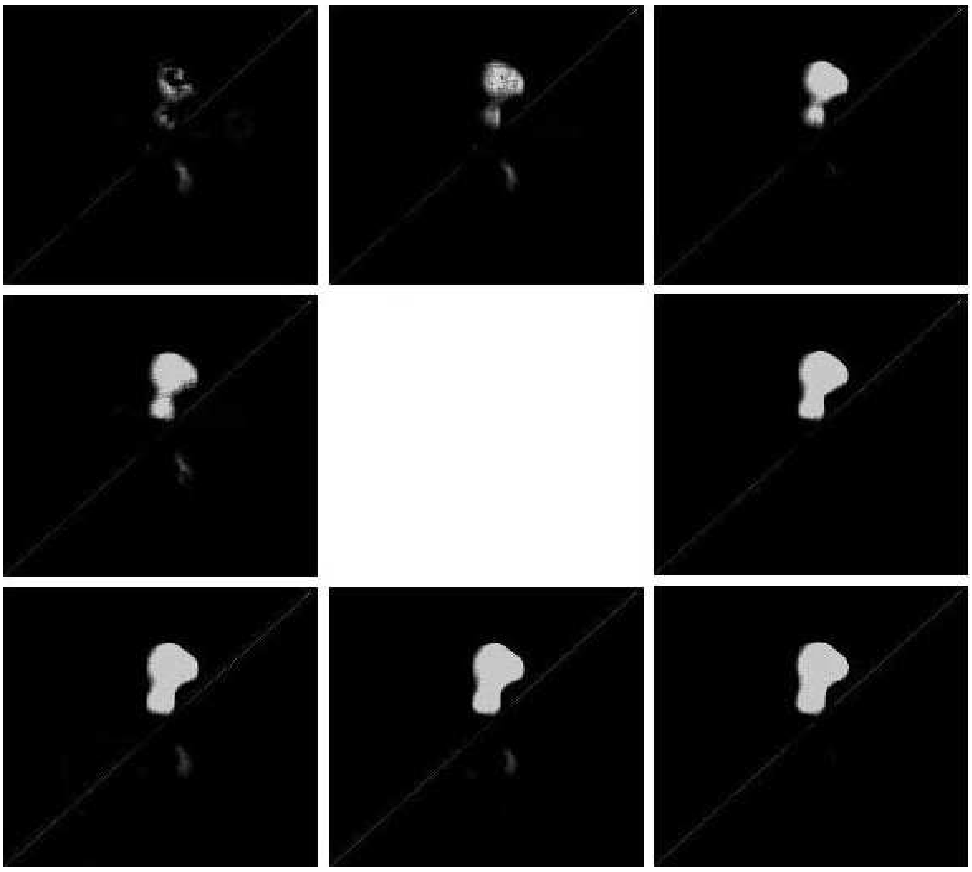


FIG. 3.9 – Contractance obtenue pour l'image II.3.8. Les huit orientations montrées sont (de gauche à droit et de haut en bas) : 135 °, 90 °, 45 °, 180 °, 0 °, 225 °, 270 ° et 315 °.

Lors du filtrage FSTC, l'image obtenue est relativement bruitée. Nous avons observé la contractance de l'image en fonction de l'intensité du bruit présent. Globalement, nous pouvons dire que moins l'image est bruitée, plus elle est contractante. Cependant une image très bruitée sera aussi contractante et le mouvement détecté sera beaucoup moins précis.

Finalement, nous avons cherché à représenter les variations de contractance locale pour les différents pixels d'une image (voir la figure II.3.9). On a représenté dans cette figure en noir les zones fortement contractantes et en blanc les zones plus faiblement contractantes. On peut remarquer que les courbes faiblement contrastées apparaissent comme fortement contractantes. Seul le contour fortement contrasté est mis en valeur sur ces images.

En conclusion, la dynamique de l'équation II.1.7 semble stable ce qui n'est pourtant pas si évident dans la pratique (contractance locale insuffisante et images synthétiques peut-être insuffisamment représentatives). En ce qui concerne l'étalonnage des paramètres de l'équation, les résultats ne sont pas ceux escomptés, en effet, il aurait été appréciable d'avoir des différences significatives de contractance d'une image en fonction des paramètres, afin de les borner ou de fixer mieux une valeur précise pour chacun (pour plus de détails voir [Jac04]). Cependant, notre intuition d'une convergence locale plus ou moins stable a été validée.

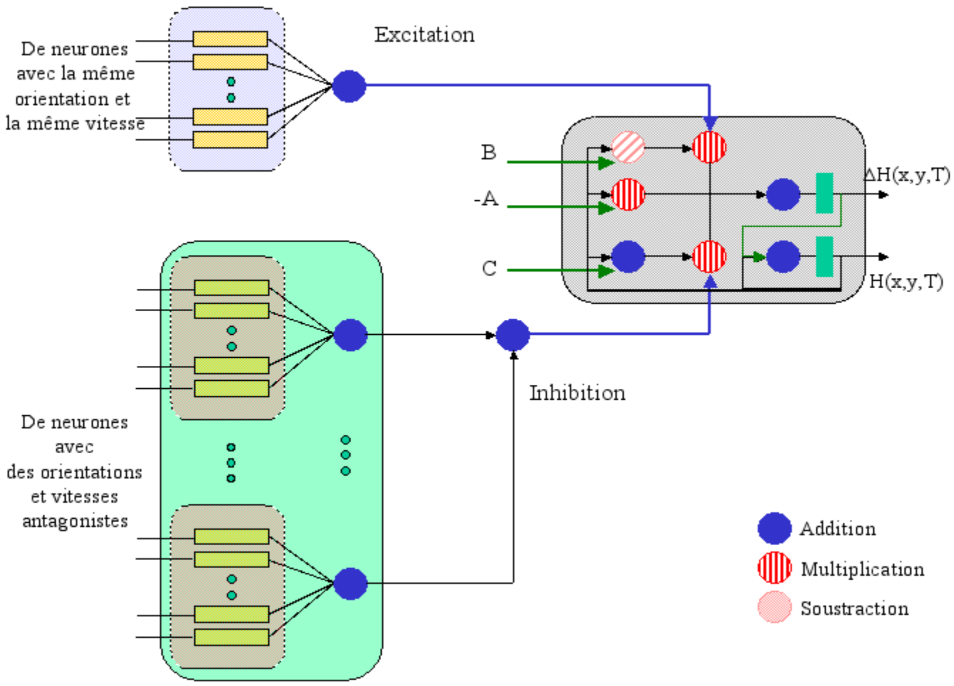


FIG. 3.10 – Diagramme de blocs simplifié du modèle matériel d'un neurone issu du mécanisme MFIA.

3.3 Faisabilité de l'implantation matérielle du MFIA

Comme nous avons vu dans §II.3.1, le module MFIA est fondé sur un réseau excitateur-inhibiteur où un neurone interagit réciproquement avec d'autres à l'intérieur d'un voisinage ou sphère d'influence. L'influence manifestée d'un neurone sur d'autres dépend de trois facteurs : l'état interne du neurone, la distance de l'un à l'autre et leur niveau d'antagonisme comme je l'ai expliqué dans §II.1.3 et §II.3.1.

3.3.1 Vers un parallélisme neuronal

La figure II.3.10 montre un diagramme de bloc simplifié du modèle matériel du mécanisme MFIA utilisé dans l'approche bio-inspirée proposée.

Le mécanisme de fortes interactions excitatrices-inhibitrices décrit dans §I.1.3.1 a pour but de faire émerger des réponses cohérentes. Ce phénomène a été modélisé dans mon modèle dans le module MFIA et pris en compte pour l'étude de faisabilité d'implantation sur FPGA.

Les neurones de ce mécanisme sont essentiellement composés d'un soma (le rectangle à droite dans la figure II.3.10) et un ensemble de synapses (les sous-modules dans MFIA symbolisés par de petits rectangles à gauche de la figure II.3.10) pour calculer l'inhibition-excitation.

Le changement du niveau d'activité d'un neurone dans mon mécanisme MFIA est donc calculé par les éléments à l'intérieur du rectangle à droite dans la figure II.3.10, comme exprimé dans l'équation II.3.6.

D'après la figure II.3.10 on voit un seul neurone qui récupère l'information nécessaire capturée par ses synapses. Cette information est traitée puis le neurone émet un potentiel d'action

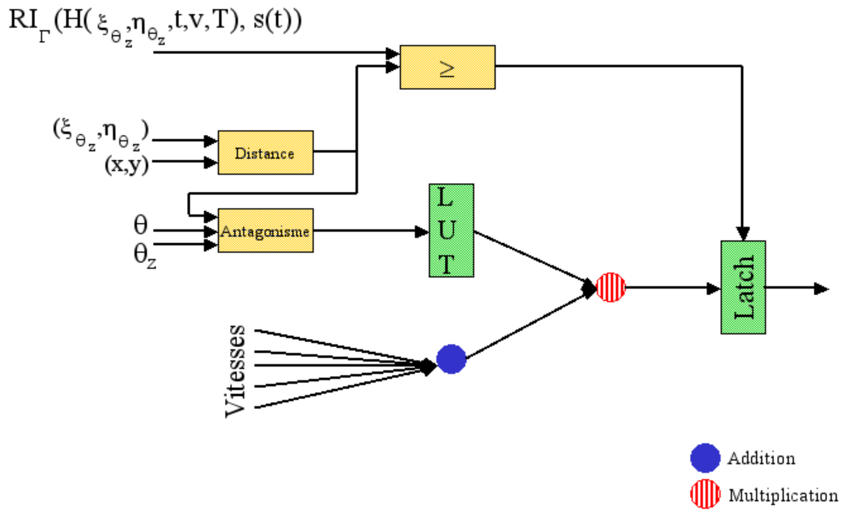


FIG. 3.11 – Diagramme de blocs simplifié du modèle matériel d'une synapse issue du mécanisme MFIA.

dépendant de son état interne. Notre but ici est d'activer un ensemble de neurones qui émettent leur potentiel d'action en parallèle.

3.3.2 Vers un parallélisme synaptique

Dans le diagramme de blocs de la figure II.3.11 nous montrons la structure matérielle d'une synapse dans le mécanisme MFIA. Elle prend les données des neurones localisés dans la sphère d'influence centrée sur le neurone indiqué. Les données sont ajoutées et pondérées par un coefficient qui dépend de la distance entre deux neurones et un poids d'antagonisme proportionnel calculé par le module antagoniste. Une table de recherche, LUT, stocke les coefficients pour ne pas les recalculer à chaque fois.

Le composant de calcul de la distance utilise les coordonnées spatiales pour calculer la distance. Si la sphère d'influence du neurone associé à la synapse ($RI_\Gamma(x, y)$ dans la figure II.3.11) est plus grande que la distance calculée, alors ce neurone contribuera pour changer le niveau d'activité du neurone central, autrement le neurone central ne sera pas affecté par ce neurone lié.

L'influence complète sur un neurone est calculée en ajoutant toutes les contributions des neurones dans sa sphère d'influence. Dans mon approche connexionniste bio-inspirée, la sphère maximale d'influence est de 15×15 (soit un rayon de taille 7) et elle est centrée autour du neurone localisé en (x, y). Alors, les neurones alentour contribuent potentiellement à l'actualisation de l'état interne du neurone central ou pivot.

3.3.3 Limites de l'implantation matérielle

Les besoins matériels pour une pleine implantation parallèle rendent sa mise en œuvre peu pratique avec des dispositifs logiques programmables. Par contre, un module matériel d'un neurone a été conçu avec un nombre réduit de synapses et calcule la contribution totale des neurones.

Le module inclut huit ensembles de synapses, chacun des ensembles calcule et accumule la contribution des neurones aux alentours locaux qui partagent la même orientation, mais des vitesses différentes. L'utilisation de synapses envoyées par un canal multiplexeur dans le temps fournit quelques opportunités d'éviter les bouchons de mémoire par les tableaux systoliques de neurones en pipeline. Puisque les calculs neuronaux du mécanisme MFIA sont locaux, ils peuvent être reformulés comme des opérations à base d'imagettes, comme celles du filtrage FSTC, mais avec des opérations beaucoup plus complexes.

Sous ces considérations, un tableau systolique linéaire de quinze neurones a été utilisé pour mon mécanisme MFIA et appliqué à tous les neurones issus de mon filtrage FSTC.

L'opération fonctionnelle du tableau est assez semblable à celle décrite dans §II.2.2.2 pour une colonne de GWP en raison de la nature fondée sur des fenêtres d'opérations préservant le traitement systolique et en pipeline.

Synthèse FPGA sur le dispositif Virtex XCV100-4BG560	
Ressources matérielles du FPGA	*Module MFIA*
Nombre de tranches	9876/12288
Nombre de LUT à 4 entrées	19286/24576
Nombre de bascule bistable	2684/24576
Pourcentage d'utilisation du FPGA	80%
Fréquence maximale d'horloge	40 Mhz
Temps de traitement	$\sim 10s$

TAB. 3.1 – Résume de la synthèse sur FPGA réalisée pour le module MFIA.

Le module connexionniste a été synthétisé pour une configuration de quinze neurones, chacun avec huit ensembles de synapses [44]. Un temps restant d'environ 20 ms [45] est disponible pour le mécanisme MFIA, ce qui n'est pas assez pour la configuration actuelle, puisque plusieurs itérations sont nécessaires et exige environ dix secondes pour achever le traitement. Il faut noter que le temps nécessaire dans l'ordinateur mono-processeur Pentium IV avec 512 Mo de mémoire est d'environ deux heures.

3.4 Applications

Les paramètres libres de l'équation II.1.7 ont été fixés d'après les suggestions de Moga [Mog00]. Les paramètres libres finalement utilisés sont reportés dans l'annexe D.2.

De la figure II.3.12 à la figure II.3.17 je montre l'entrée, la sortie et un détail du comportement du mécanisme MFIA. Les séquences utilisées sont décrites dans l'annexe C.

Les valeurs montrées dans les graphiques correspondent aux données d'avant et d'après l'actualisation d'un ensemble de neurones choisis aléatoirement. Je compte le nombre de neurones actifs (faiblement ou fortement) [46] avant leur actualisation ($ActifAvant$) et après leur actualisation ($ActifApres$). De plus, je compte le nombre de neurones qui deviennent inactifs à cause des interactions antagonistes ($NonActifApres$).

D'après chacun des graphiques des figures énoncées et dans le cas de caméras statiques, c'est-à-dire pour des séquences d'images sans mouvement propre, on peut bien voir un motif caractéristique qui n'est pas le même pour la séquence Yosemite Fly-Through (voir la figure II.3.13e) où on a un mouvement propre.

[44] Ces ensembles correspondent aux huit orientations différentes prises par mon modèle.

[45] En considérant que l'être humain a besoin de 25 à 30 images par seconde pour avoir la sensation de mouvement, dans mon modèle nous avons besoin donc d'achever tous les 30 ms environ tout le traitement pour une seule image. 10 ms ont été déjà utilisés pour le module de filtrage FSTC.

[46] Un neurone est actif si son rayon d'influence est supérieur à 0.5 (voir l'équation II.3.4).

Je vais analyser la séquence Taxi d'Hambourg, mais l'interprétation est presque la même pour les autres séquences. D'après la figure II.3.12e, le critère d'arrêt pour mon mécanisme MFIA est le changement dans la pente d'$NonActifApres$ et d'$ActifApres$, avant même leur croisement. À ce niveau, j'ai un recouvrement des zones en mouvement. Même si ce recouvrement n'est pas cohérent par rapport à l'orientation, il est cohérent par rapport au mouvement. Au-delà du point de changement de la pente des lignes $NonActifApres$ et $ActifApres$ (à l'itération 13 du mécanisme MFIA), le mécanisme disperse du bruit en commençant par les zones de mouvement vers toute l'image. Le point critique de ce type de bruit commence à partir du point de croisement de ces deux lignes (à l'itération 22 du mécanisme dans ce cas là). Le critère heuristique retenu est donc d'arrêter le mécanisme MFIA lorsque ce point critique est atteint.

Le tableaux II.3.2 montre les points de croisement et les points du changement de la pente des lignes $NonActifApres$ et $ActifApres$ à l'instant $t = 4$ pour les séquences d'images décrites dans l'annexe C.

Séquence	Changement Pente		Croisement
	$NonActifApres$	$ActifApres$	
Taxi d'Hambourg	13	13	22
Yosemite Fly-Though	20	20	–
Dispute "A"	13	14	45
Dispute "B"	13	13	44
Parcourir	13	13	35
Couloir	8	8	29

TAB. 3.2 – Points d'arrêt du mécanisme MFIA d'après le comportement des variables $NonActifApres$ et $ActifApres$.

3.5 Bilan

Les fortes interactions dans V1 sous un mécanisme de compétition/coopération pourraient aboutir à la discrimination de différentes types de mouvement dans une séquence d'images. D'après les séquences d'images décrites dans l'annexe C, toutes montrent un motif caractéristique pour celles correspondant à une caméra statique (voir le graphique dans la figures II.3.12 et des figures II.3.14 à II.3.17) où on perçoit le croisement des lignes $NonActifApres$ et $ActifApres$. Ce n'est pas le cas pour les séquences qui montrent un mouvement propre comme la séquence Yosemite Fly-Though.

Le critère d'arrêt du mécanisme MFIA a été trouvé de façon empirique et après plusieurs essais.

Il reste encore à tester la possible généralisation du comportement des lignes $NonActifApres$ et $ActifApres$ pour une séquence donnée. D'après les résultats obtenus, il apparaît que leur comportement nous aidera à distinguer le mouvement des objets dans une séquences d'images capturée par une caméra statique, i.e. sans mouvement propre, de celui de la caméra, avec mouvement propre.

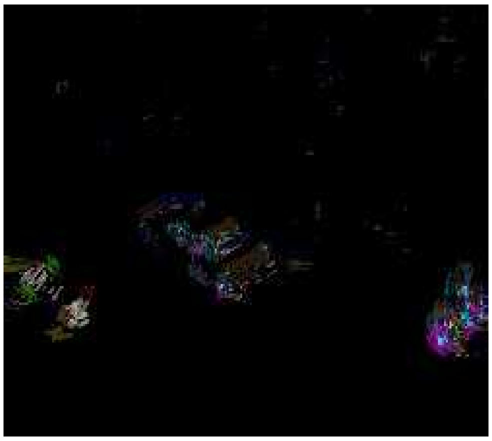
D'autre part, le coût de traitement de mon mécanisme est très lourd. Je suis passé d'un traitement de l'ordre de $\mathcal{D} \times \mathcal{V} \times \tau$ (directions, vitesses supposées et taille de la séquence à analyser, respectivement) fois la taille de l'image dans mon filtrage FSTC à un autre de $\mathcal{D} \times \mathcal{V} \times$

$\tau \times (2 \cdot \Gamma + 1)^2 \times n$ (orientations, vitesses supposées, taille de la séquence, taille maximale du rayon d'influence et le nombre d'interactions du mécanisme MFIA, respectivement).

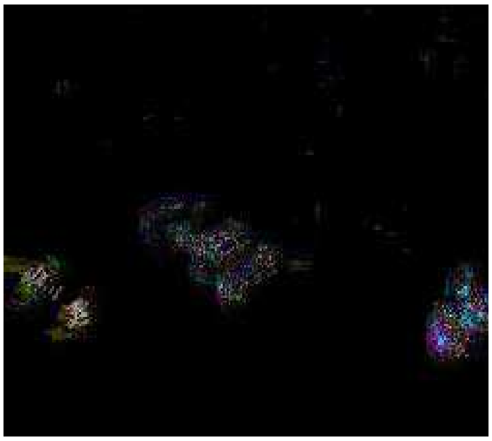
Les résultats issus de l'implantation sur FPGA montrent la nécessité de surmonter le surcoût en temps de système et les difficultés de mémoire du modèle bio-inspiré proposé pour la perception visuelle de mouvement. L'architecture présentée emploie un parallélisme trop partiel au niveau des connexions et au niveau du neurone. Ainsi, un parallélisme complet exige l'optimisation des ressources par une mise au point adéquate des étapes du module MFIA du modèle proposé.

Cependant, en raison de la régularité et de la nature modulaire de l'architecture, il est possible de prolonger le parallélisme à l'aide de tableaux systoliques plus grands en fonction de la disponibilité des ressources matérielles. Mais, les optimisations proposées sont encore loin d'une implantation réelle et pour le moment nous concluons momentanément à la non-faisabilité en temps réel du module MFIA. Par conséquent, je n'implante pas sur FPGA le troisième module.

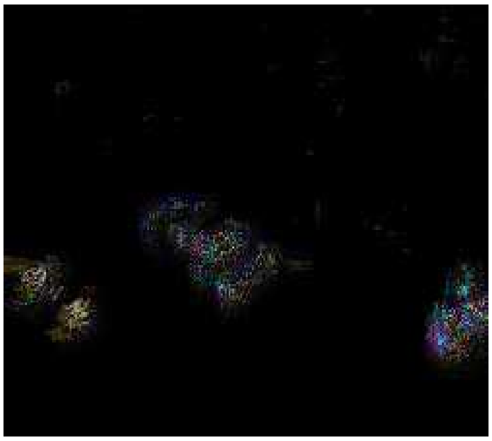
Dans le chapitre suivant je vais finalement décrire le troisième module de mon modèle et faire l'intégration des trois modules. Ce dernier module est ciblé pour la focalisation de l'attention.



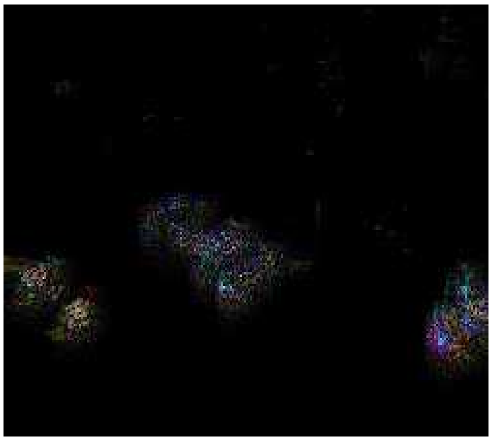
(a) Entrée au mécanisme



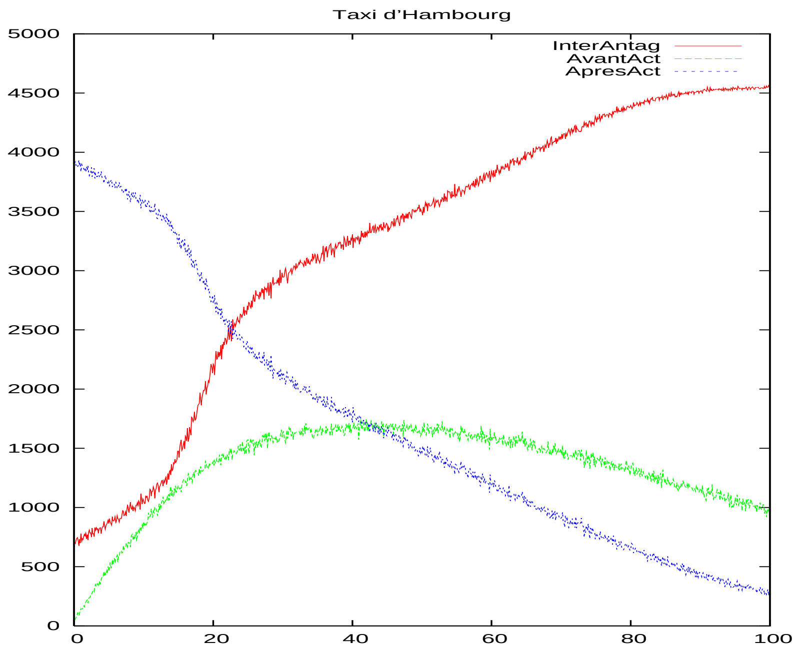
(b) Itération 5



(c) Itération 10

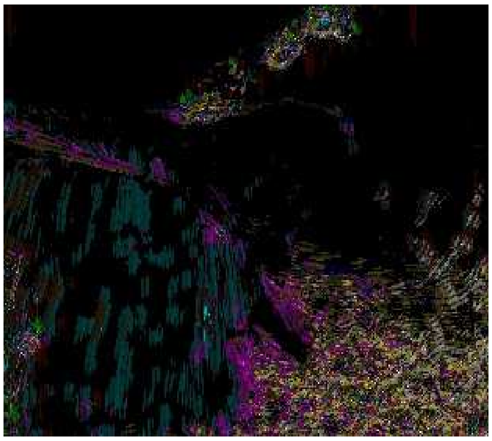


(d) Sortie du mécanisme

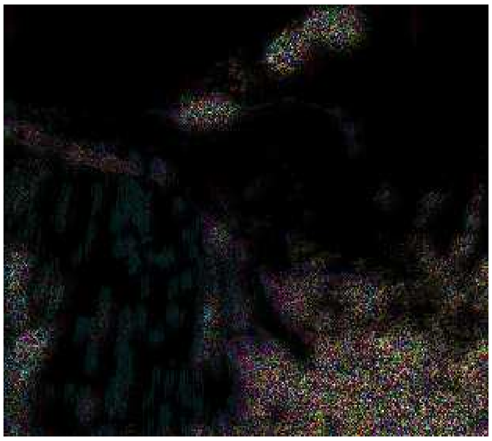


(e) Comportement du mécanisme d'actualisation

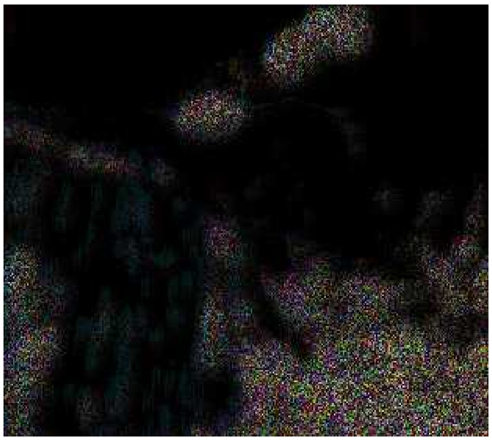
FIG. 3.12 – Résultats obtenus de l'application du mécanisme MFIA à la séquence Taxi d'Hambourg : les premiers pas, la sortie et le comportement du mécanisme d'actualisation.



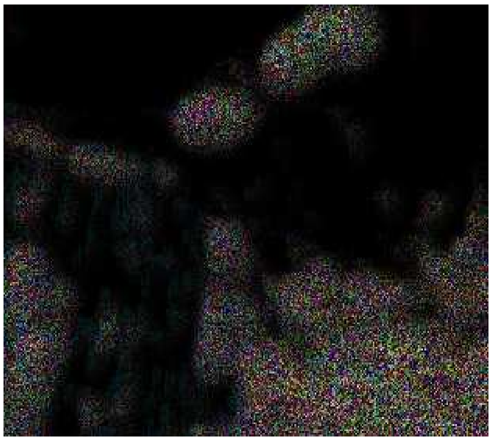
(a) Entrée au mécanisme



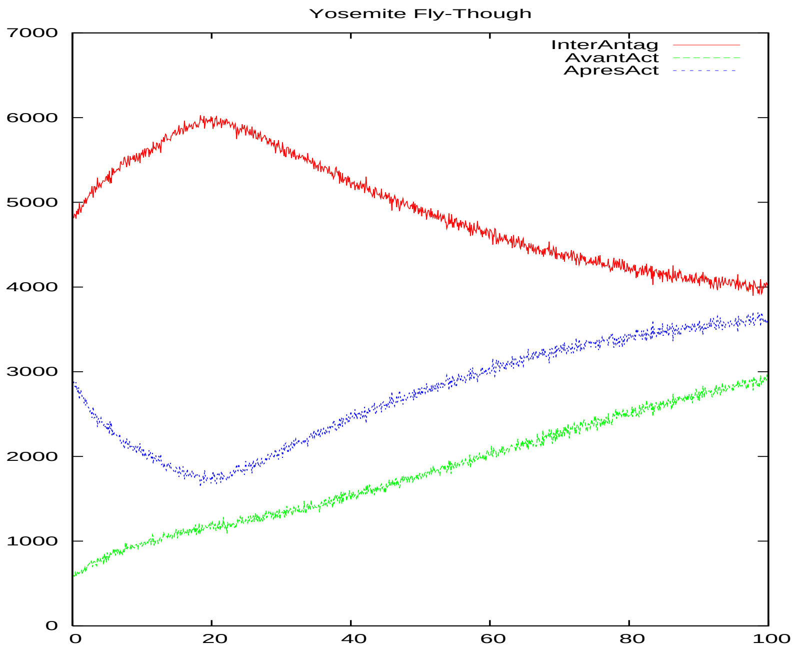
(b) Itération 5



(c) Itération 10

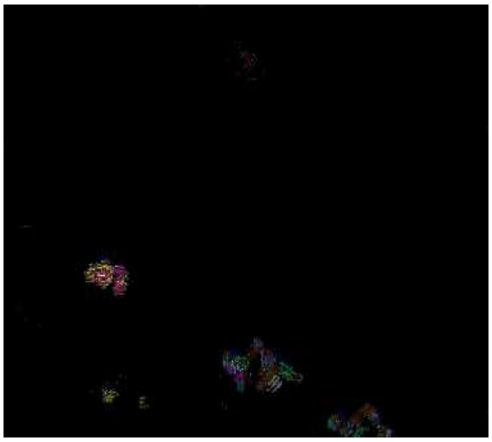


(d) Sortie du mécanisme

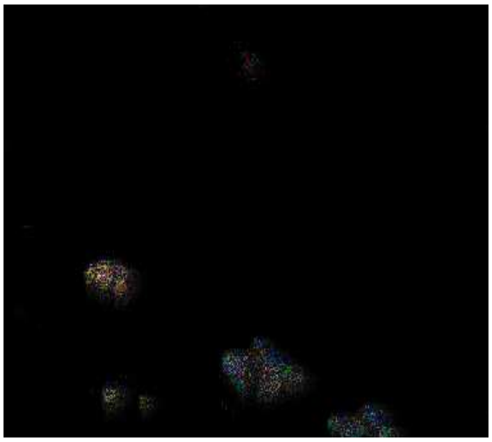


(e) Comportement du mécanisme d'actualisation

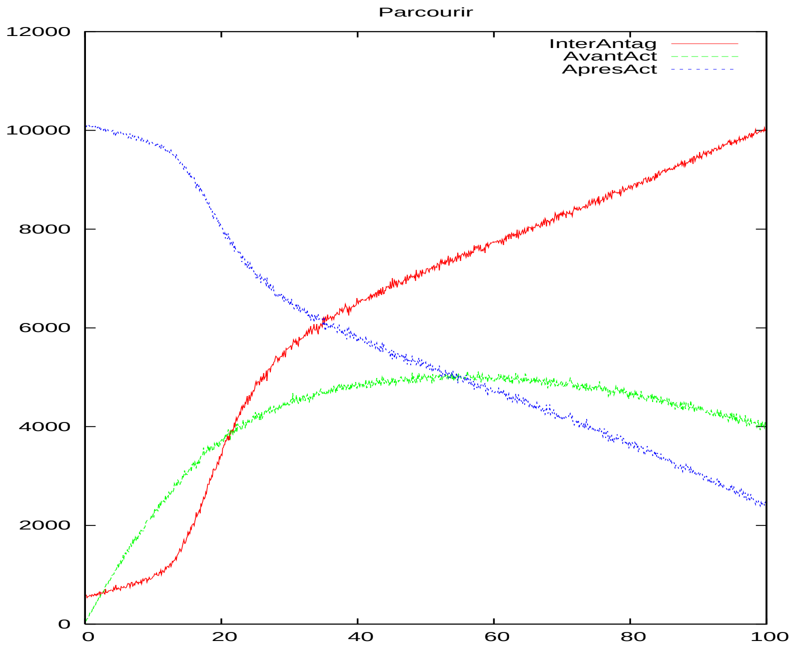
FIG. 3.13 – Résultats obtenus de l'application du mécanisme MFIA à la séquence Yosemite Fly-Through : les premiers pas, la sortie et le comportement du mécanisme d'actualisation.



(a) Entrée au mécanisme

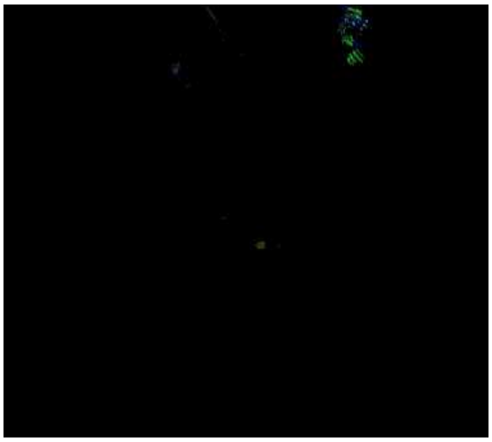


(b) Sortie du mécanisme

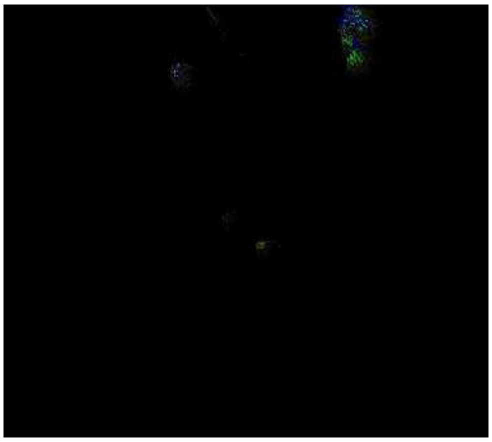


(c) Comportement du mécanisme d'actualisation

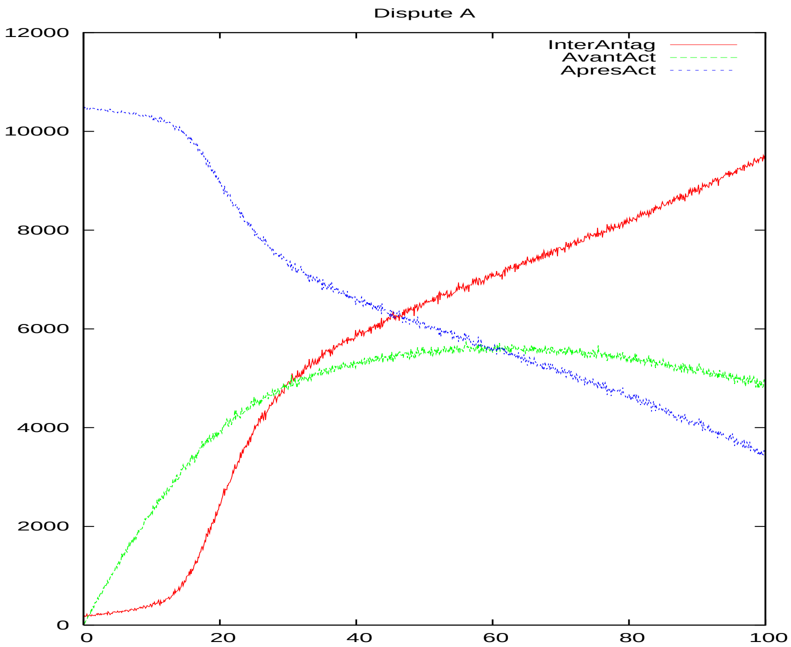
FIG. 3.14 – Résultats obtenus de l'application du mécanisme MFIA à la séquence Parcourir : l'entrée, la sortie et le comportement du mécanisme d'actualisation.



(a) Entrée au mécanisme

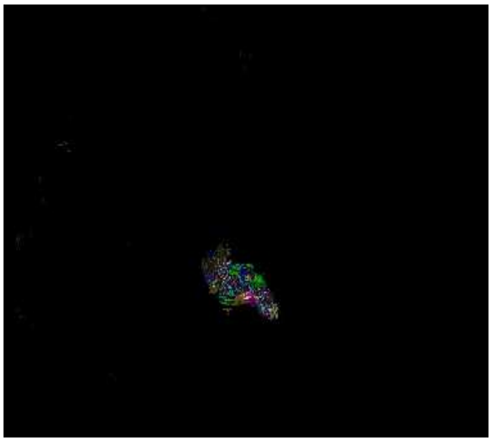


(b) Sortie du mécanisme

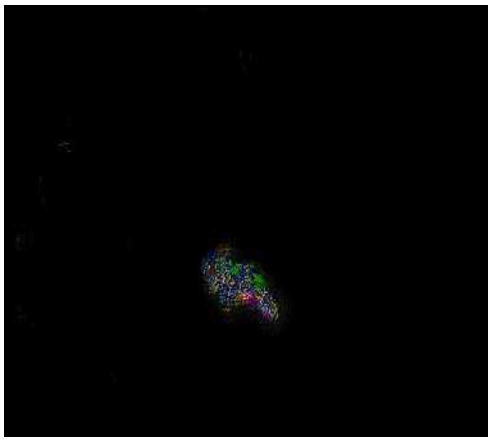


(c) Comportement du mécanisme d'actualisation

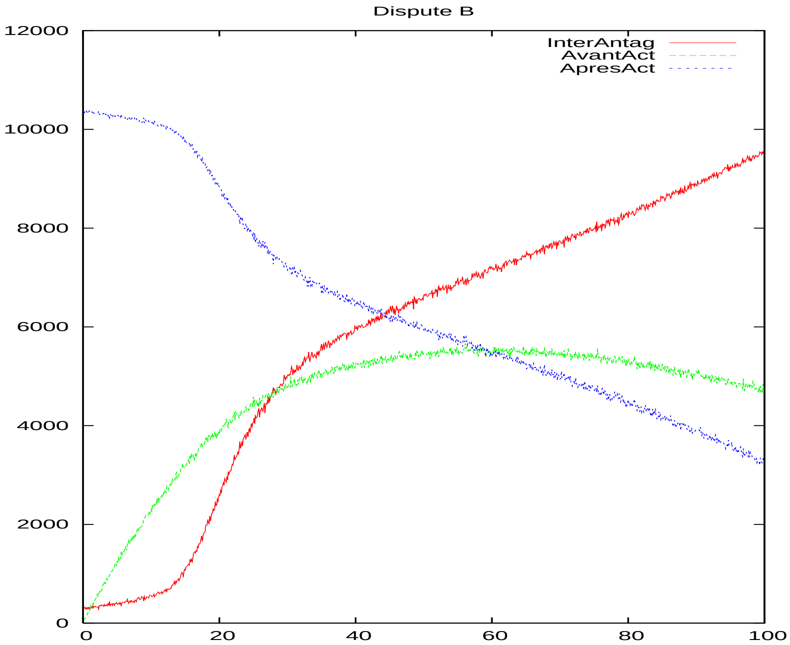
FIG. 3.15 – Résultats obtenus de l'application du mécanisme MFIA à la séquence Dispute "A" : l'entrée, la sortie et le comportement du mécanisme d'actualisation.



(a) Entrée au mécanisme

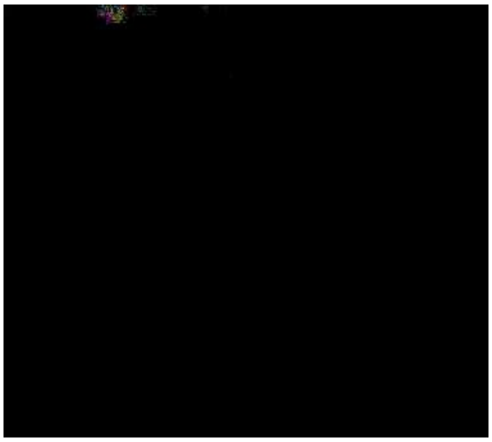


(b) Sortie du mécanisme

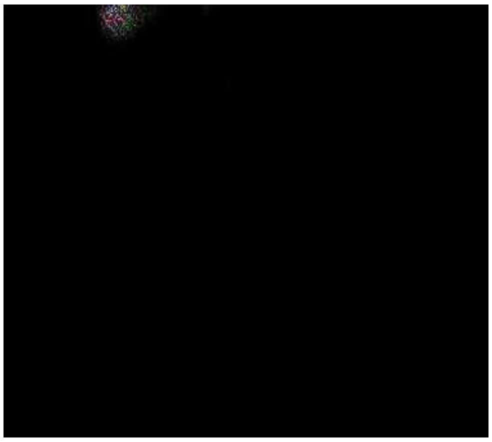


(c) Comportement du mécanisme d'actualisation

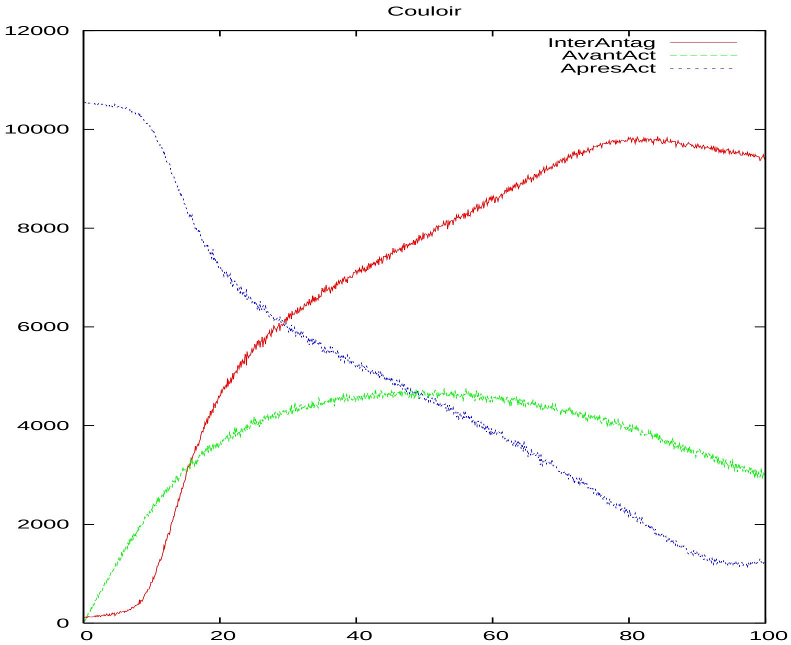
FIG. 3.16 – Résultats obtenus de l'application du mécanisme MFIA à la séquence Dispute "B" : l'entrée, la sortie et le comportement du mécanisme d'actualisation.



(a) Entrée au mécanisme



(b) Sortie du mécanisme



(c) Comportement du mécanisme d'actualisation

FIG. 3.17 – Résultats obtenus de l'application du mécanisme MFIA à la séquence Couloir : l'entrée, la sortie et le comportement du mécanisme d'actualisation.

Chapitre 4

Focalisation de l'attention

Sommaire

Le dernier module de mon modèle connexionniste met en place les interactions denses d'une population neuronale inspirée des interactions des aires supérieures telles que MT, MST et PP (voir §II.1.4).

Dans ce chapitre je décrirai le couplage des modules précédents avec un sous-ensemble du modèle pour la focalisation de l'attention développé au sein de l'équipe CORTEX. Donc, tout d'abord nous allons discuter du besoin d'une interprétation globale des signaux distribués fournis par mon mécanisme MFIA (§II.4.1), après je décrirai de façon générale le modèle de focalisation de l'attention de Rougier et ses collaborateurs [RV05, VRA05] (le modèle que j'appelle FANF et qui est décrit dans §II.4.2), ensuite j'exposerai son adaptation pour créer mon troisième module et les liens créés pour son intégration totale à mon modèle (§II.4.3), puis je montrerai quelques exemples d'application (§II.4.4).

4.1 Vers une interprétation du signal distribué

Dans les trois modules de mon modèle connexionniste neuromimétique nous nous sommes vivement inspirés de la neurobiologie pour avoir un modèle modulaire, compétitif et coopératif à travers de fortes interactions latérales, mais le maintien des ces interactions de façon localisée et distribuée est notamment soutenu par les deux premiers modules, par contre le troisième aboutira à une réponse globale.

Mon mécanisme MFIA nous apporte une grande quantité d'information pour la direction et la vitesse du mouvement d'un objet mais il reste toujours local et distribué sans aboutir pour lui-même à une réponse globale. Si mon filtrage FSTC correspond à un traitement de bas niveau, mon mécanisme MFIA quant à lui ne s'identifie ni comme de bas niveau ni comme de haut niveau, il reste entre tout les deux mais il n'arrive pas à nous donner une réponse globale souhaitée : réponse qu'un module d'analyse et d'interprétation notamment à haut niveau peut nous fournir. De plus, nous avons beaucoup d'information sur chaque neurone issu de mon mécanisme ($D \times V$ différentes directions et vitesses supposées) qui demande d'être analysée et interprétée pour aboutir à une réponse globale.

Dans ma recherche parmi les modèles corticaux pouvant répondre à ce besoin, j'ai choisi le modèle de Rougier et ses collaborateurs [RV05, VRA05] et je l'ai adapté à mes besoins d'une réponse globale pour la perception visuelle du mouvement. Vu que ce modèle s'inspire des aires supérieures du cortex visuel, de même que mon modèle s'inspire des premières (V1 et MT), le couplage des deux se montre comme la suite naturelle dans la conception d'un modèle bio-inspiré, distribué, et globalement basé sur des mécanismes d'excitation/inhibition, pour la perception visuelle du mouvement.

L'interprétation globale permise par ce couplage est la poursuite d'un objet ou d'une personne dans une scène dynamique, ce qui représente un exemple de réponse globale émergeant à partir d'interactions locales. D'autres tâches pourraient être envisagées, comme l'extraction du mouvement propre, la reconstruction de formes 3D, ... chacune d'elles nécessitant de la même manière un travail spécifique de couplage avec un modèle adéquat d'interprétation de haut niveau.

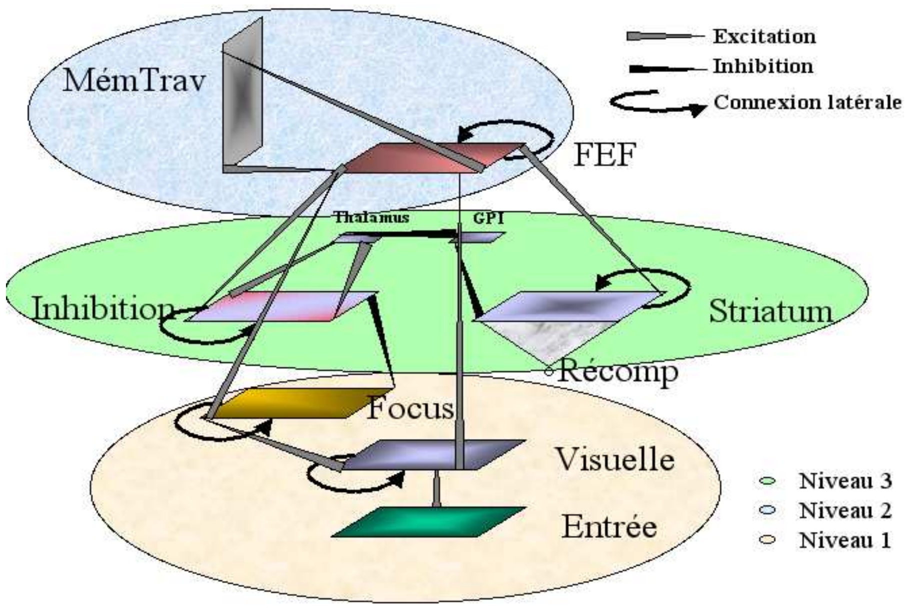


FIG. 4.1 – Architecture générale du modèle de focalisation de l'attention proposé par Rougier et ses collaborateurs [RV05, VRA05] : l'émergence de l'attention (niveau 1), la fixation de l'attention (niveau 2) et la commutation de l'attention (niveau 3).

Les interactions issues de ce couplage vont intégrer les réponses de la mémoire à court terme (mon mécanisme MFIA) combinées à une myriade d'associations dans la mémoire à long terme (la deuxième étape du troisième module, §II.4.2). selon Grossberg [Gro73, GL75, Gro78b, Gro78a, Gro80, Gro82], ce mélange d'interactions brutes de deux mémoires font émerger l'attention comme un de ses produits. Donc, l'attention est vue comme une fonction de filtres harmonisables qui rendent sensibles les réseaux neuronaux par l'ajustement de leurs poids et de leurs seuils dépendant des signaux arrivants.

Par la suite je vais présenter le modèle FANF pour la focalisation de l'attention (§II.4.2) puis les démarches que j'ai fait pour l'adapter à mon modèle et constituer ainsi son troisième module (§II.4.3).

4.2 Présentation du modèle pour la focalisation de l'attention

Le modèle FANF pour la focalisation de l'attention est un modèle dynamique fondé sur la théorie des champs neuronaux continus (CNFT) [Tay99]. Cette théorie explique l'attention comme une propriété émergente d'une population neuronale.

Si la plupart des modèles font émerger une auto-organisation issue de la compétition latérale des interactions dans le cortex [AA82, Ama79, Koh97, SM93, SM97, MBCS97], ils n'exploitent pas l'aspect dynamique de ces neurones. Par l'application de la théorie CNFT la dynamique de formation des motifs pour les champs neuronaux homogènes à inhibition latérale peut être expliqué. Donc, dans certaines conditions, un réseau neuronal d'attracteurs continus est capable de maintenir focalisée une bulle d'activité en relation directe avec l'excitation fournie par la stimulation.

Cette propriété émergente des populations neuronales est identifiée comme l'attention. L'attention est donc vue comme un mécanisme distribué et intégré sur les structures sensorimotrices. Son but est de permettre la focalisation sur des petites régions de l'espace d'entrée à travers d'un traitement hiérarchique autorisant les comportements moteurs appropriés. Donc, l'attention est un processus motivé d'intégration [RV05, VRA05].

La figure II.4.1 décrit les neuf cartes du modèle FANF. On constate à l'intérieur de ce modèle trois étapes de traitement : l'émergence de l'attention, la fixation de l'attention et la commutation de l'attention. Nous retenons les cartes correspondant aux niveaux 1 et 2 (l'émergence de l'attention et la fixation de l'attention) dans cette figure.

Je vais adapter ce modèle (FANF) pour la perception visuelle du mouvement et le coupler avec mes deux premiers modules. Donc, le but de notre troisième module (FA) issu de l'adaptation du modèle de focalisation de l'attention est de faire émerger une seule bulle d'activité issue des entrées afférentes originaires du mécanisme MFIA, puis de fixer l'attention sur cette bulle et de permettre sa poursuite. Pour cela je décrirai le processus d'émergence, puis la fixation de l'attention sur la bulle d'activité qui émerge, ensuite je décrirai succinctement une étape présente dans le modèle de focalisation de l'attention mais non encore utilisée dans mes applications : la commutation de l'attention, et finalement je montrerai la discrétisation des équations utilisées pour l'actualisation de l'état interne de leurs neurones.

4.2.1 L'émergence de l'attention

En prenant en compte l'intégration que font les neurones de MT (puis de MST) sur l'information qui vient de V1 (puis de MT, respectivement, voir §I.1.3.2 et §I. 1.2.5) et les interactions des neurones dans MT et MST, la compétition réalisée par les neurones issus des cartes de la première étape de ce module se focalisera sur la zone de mouvement plus fort.

Cette étape est composée de trois cartes : d'entrée (*Entrée*), visuelle (*Visuelle*) et de focalisation (*Focus*).

La première carte est prise comme une simple carte de reliefs produites par le mécanisme MFIA. Elle sert simplement à fournir l'information nécessaire dans le cadre et contexte du troisième module.

La deuxième carte reçoit les entrées excitatrices de la première avec des connexions de type champs récepteurs, i.e. les entrées dans un voisinage local (imagette) sont pondérées et intégrées dans une seule sortie. Ce processus d'intégration de l'information est similaire à celui réalisé dans la rétine (bâtonnets - cellules magnocellulaires, etc.). Dans cette carte il y a des connexions latérales denses. Ces interactions latérales limitent le nombre de bulles d'activité qui peuvent émerger à tout moment. Ainsi, l'activité de cette deuxième carte ne montre que les stimuli des entrées les plus saillantes.

Finalement, la troisième carte reçoit les entrées excitatrices de la deuxième. La carte *Focus* a des champs récepteurs plus larges que dans la carte *Visuelle* afin de permettre l'émergence d'une seule bulle d'activité. Dans cette carte il y a une connexion latérale complète, i.e. tous ses neurones sont liés à tous les autres par des connexions latérales.

Jusqu'ici, pour l'émergence de l'attention, nous avons fait émerger une seule bulle d'activité parmi d'autres possibles proposées par la deuxième carte.

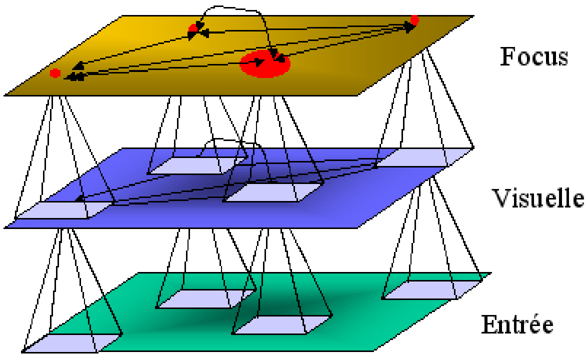


FIG. 4.2 – Détail de l'architecture dans l'étape d'émergence de l'attention (première étape de notre troisième module). Les cartes *Visuelle* et *Focus* appliquent les principes issus de la CNFT.

La figure II.4.2 montre un détail de l'architecture pour la première étape de notre troisième module : l'émergence de l'attention. Nous pouvons y voir de façon très schématique la contribution de la carte d'entrée vers la carte visuelle où une forte interaction latérale a lieu (représentée par les flèches en noir qui relient quatre neurones de la carte visuelle). De même pour la carte focus où, par ailleurs, nous pouvons percevoir l'émergence d'une seule bulle d'activité.

4.2.2 La fixation de l'attention

En appliquant les principes de la théorie des champs neuronaux continus (CNFT) [Ama77, FC75, Tay99] nous pouvons maintenir localisée une bulle d'activité en relation directe avec l'excitation fournie par les stimuli sortants de la première étape.

Avec l'application d'un réseau récurrent, l'activité tend vers une courbe lisse dans une position dépendant des neurones les plus actifs. Cela est possible si la cible a une continuité spatio-temporelle observable, c'est-à-dire que son mouvement est tel qu'on peut le suivre.

Ainsi, pour maintenir fixe l'attention sur la bulle d'activité émergée dans la première étape, la deuxième étape est composée de deux cartes : le champ frontal des yeux (*FEF*) et la mémoire de travail (*MémTrav*) dans le but de mémoriser les stimuli d'un passé récent de la bulle d'activité fournie par la carte Focus.

Les deux cartes de cette étape sont connectées mutuellement de telle façon que la mémoire de travail fait une sorte de dispositif de rebond reflétant l'activité de la carte *FEF*. La seule carte connectée à la carte *MémTrav* est la carte *FEF*. Cette dernière reçoit les entrées excitatrices des cartes *Visuelle* et *Focus*. C'est la jonction des résultats de ces deux cartes qui renforce la focalisation sur une seule bulle d'activité.

Dès qu'une seule bulle d'activité a émergé dans la carte *FEF* son information est transmise à la carte *MémTrav*. Donc, une fois que le stimulus y a été focalisé, cette carte commence à rebondir sur la carte *FEF* et si une autre bulle d'activité émerge dans la carte *Focus*, la carte *MémTrav* aura encore les traces de l'ancienne bulle d'activité avant de rebondir sur la nouvelle bulle. Cette période de latence est utilisée pour maintenir l'attention sur la première bulle si la bulle qui vient d'apparaître disparaît quelques instants. Par contre, si la nouvelle bulle d'activité est maintenue dans la carte *Focus*, la carte *FEF* la prendra et la transmettra à la carte *MémTrav* et cette dernière éliminera au fur et à mesure les traces de l'ancienne en même temps qu'elle récupère l'information de la nouvelle.

Jusqu'ici, en fonction des stimuli d'entrée et de l'émergence des différentes bulles d'activité dans le temps, nous avons une commutation automatique d'une bulle d'activité à l'autre. Dans la troisième étape nous allons choisir quand nous voulons commuter l'attention ou essayer de la maintenir même s'il y a d'autre bulles d'activité qui viennent d'être focalisées.

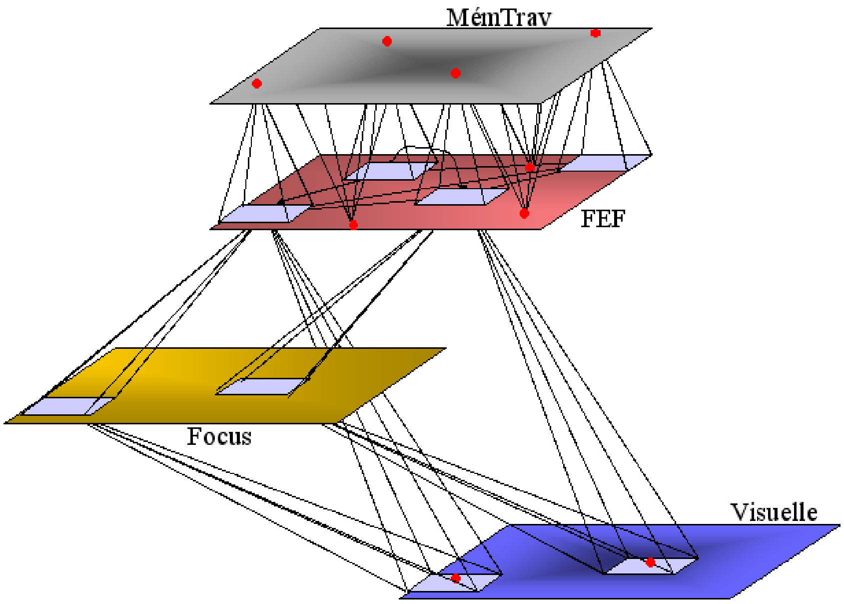


FIG. 4.3 – Détail de l'architecture dans l'étape de focalisation de l'attention (deuxième étape de notre troisième module). De plus à l'application des principes de la CNFT, nous avons une mémoire de travail qui rebondit sur la carte *FEF*.

La figure II.4.3 montre la forte interaction mutuelle entre les cartes *FEF* et *MémTrav*. On y perçoit la contribution excitatrice des cartes *Focus* et *Visuelle* dans la carte *FEF*. Il faut noter la bijection des cartes *FEF* et *MémTrav* où le rebond de *MémTrav* renforce l'activité dans *FEF*.

4.2.3 La commutation de l'attention

Cette étape n'est pas retenue dans mon modèle. Je la présente brièvement car elle correspond à des fonctionnalités additionnelles totalement compatibles avec mon modèle connexionniste.

La troisième étape du modèle FANF est sensée permettre la commutation volontaire de l'attention. Pour cela Rougier et ses collaborateurs[RV05, VRA05] se sont inspirés des interactions faites par les aires sous-corticales telles que le Thalamus et le Striatum qui constituent les cibles majeures des cellules corticales afférentes et qui participent à des boucles internes diverses qui modulent l'activité de la transmission d'information dans la boucle : *cortex* → *striatum* → *globus pallidus* [47] → *thalamus* → *cortex* [ZPMS+04, WGJ97].

Ainsi, la troisième étape est composée de quatre cartes : striatum (*Striatum*), *GPI*, thalamus (*Thalamus*), d'inhibition (*Inhibition*), et d'un neurone de récompense (*Récomp*).

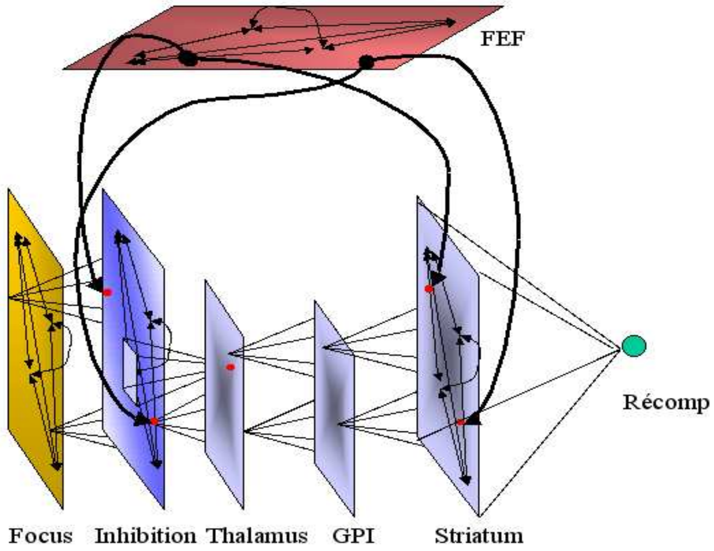


FIG. 4.4 – Détail de l'architecture de l'étape de la commutation de l'attention. On perçoit le neurone *Récomp* responsable du déclenchement du processus inhibitoire nécessaire pour permettre la commutation de l'attention.

L'idée est d'inhiber les stimuli de l'ancienne bulle d'activité au niveau de la carte *Focus* en s'aidant de la carte *MémTrav* et des interactions des quatre cartes de cette troisième étape.

Normalement, la carte *Striatum* reçoit une faible excitation de la carte *FEF* et son activité est presque nulle. Lorsqu'on active le neurone *Récomp*, une bulle d'activité dans la carte *Striatum* émergera (la même qu'elle avait avant mais renforcée) en reproduisant celle de la carte *FEF*. C'est à ce moment-là qu'on déclenche le processus inhibiteur.

La carte *Striatum* envoie ses inhibitions vers la carte *GPI*. Les neurones de *GPI* ne répondent pas aux faibles entrées, mais dès qu'ils les reçoivent fortes, ils répondent en envoyant une forte inhibition aux neurones de la carte *Thalamus*. Cette carte a un comportement similaire à la carte *GPI*, ses neurones ne répondent pas aux faibles entrées, mais répondent fortement dès que les entrées sont fortes.

Ensuite, la carte d'*Inhibition* et de *Thalamus* sont mutuellement excitées. Elles sont connectées de la même façon que les cartes *FEF* et *MémTrav*, mais le rebond ici est réglé par l'inhibition tonique de la carte *GPI* sur la carte *Thalamus*.

La carte d'*Inhibition* reçoit les faibles excitations qui viennent de la carte *FEF*. Elles ne sont pas suffisantes pour générer une bulle d'activité, mais dès qu'elle est excitée par le *Thalamus*, la carte d'*Inhibition* reproduit la bulle d'activité issue de la carte *FEF*.

Finalement, la forte activité déclenchée dans la carte *Inhibition* par le processus inhibiteur suscité par le neurone *Récomp* sur la carte *Striatum* fait que l'ancienne bulle d'activité émergée commence à disparaître dans la carte *Focus*. Ce processus permet la focalisation vers une autre

[47] Ensemble formé par les parties interne et moyenne du noyau lenticulaire. Cet ensemble constituerait la partie la plus ancienne du striatum.

bulle d'activité. Donc, nous recommençons à nouveau dès la deuxième étape.

Quand le neurone *Récomp* ne déclenche aucune activité, la faible inhibition envoyée par la carte *Inhibition* vers la carte *Focus* ne suffit pas pour éliminer la bulle d'activité qui émerge. Une fois que le neurone *Récomp* déclenche une activité, le processus d'inhibition augmente de telle façon qu'il est capable d'inhiber suffisamment la carte *Focus* pour qu'elle change de cible.

La figure II.4.4 montre un détail des interactions entre les cartes qui favorisent la commutation de l'attention. Une fois que le neurone *Récomp* a déclenché son activité, les cartes *Striatum*, *GPI* et *Thalamus* s'activent pour que la dernière rebondisse sur la carte *Inhibition* de la même façon que *MémTrav* rebondit sur *FEF*. L'ensemble des cartes *GPI* et *Thalamus* permet d'avoir une mémoire interne d'inhibition avec des champs récepteurs élargis. Finalement la carte *Inhibition* envoie ses inhibitions vers la carte *Focus* pour éliminer l'ancienne bulle d'activité et permettre qu'une autre bulle émerge si elle existe.

Nous pouvons aussi observer l'excitation envoyée par la carte *FEF* vers les cartes *Inhibition* et *Striatum*. Dans le cas où le neurone *Récomp* est inactif, ces deux dernières cartes n'influencent pas les autres car leur activité est presque nulle.

4.2.4 Discrétisation des équations de la CNFT

Nous avons brièvement parcouru le modèle FANF. Pour mon modèle, nous allons dans la suite montrer la discrétisation de la règle d'actualisation de l'état interne des neurones représentée par l'équation II.1.8 proposée par Rougier et ses collaborateurs [RV05, VRA05] de la façon suivante :

$$J_{\mathcal{M}}(x,y,t+1) = \varpi \left[J_{\mathcal{M}}(x,y,t) + \frac{1}{\kappa} \left(-(J_{\mathcal{M}}(x,y,t) - baseline) + \frac{1}{\alpha} N(x,y,t) \right) \right] \tag{4.1}$$

où $J_{\mathcal{M}}(x,y,t)$ est le neurone à actualiser positionné en (x,y) au temps t dans la carte $\mathcal{M}$. κ est le temps constant, α le facteur de pondération pour les influences externes, *baseline* l'énergie minimale détectable, et

$$N(x,y,t) = \sum_{\not{\mathcal{M}}} w_{\not{\mathcal{M}}}((x,y)-(\hat{x},\hat{y}))J_{\not{\mathcal{M}}}(x,y,t) + \sum_{\mathcal{M}} w_{\mathcal{M}}((x,y)-(\bar{x},\bar{y}))J_{\mathcal{M}}(x,y,t) \tag{4.2}$$

la contribution inter-cartes par la voie des champs récepteurs de la carte $\not{\mathcal{M}}$ sur la carte $\mathcal{M}$ et la contribution intra-carte par les interactions latérales dans la carte $\mathcal{M}$, premier et deuxième terme respectivement.

Enfin,

$$\varpi(x) = \begin{cases} 0 & si \quad x < 0 \\ 1 & si \quad x > 1 \\ x & sinon \end{cases} \tag{4.3}$$

permet de borner la valeur d'actualisation pour $J_{\mathcal{M}}(x,y,t)$.

Les pondérations $w_{\not{\mathcal{M}}}(x,y)$ et $w_{\mathcal{M}}(x,y)$ sont définies dans les équations II.1.9 et II.1.10, respectivement.

4.3 Adaptation du modèle FANF à la perception du mouvement

Le modèle FANF est ciblé pour la focalisation de l'attention avec la possibilité de la commuter (§II.4.2). Dans cette section je vais montrer son adaptation à mes besoins pour la perception visuelle du mouvement et son intégration à mon modèle connexionniste.

Du modèle FANF je ne retiens que les deux premières étapes, à savoir : l'émergence de l'attention et la fixation de l'attention. Ces étapes nous fournissent la position dans l'espace de la bulle d'activité focalisé, i.e. la position de cette bulle dans la carte *Focus*. Cette position nous aide pour repérer l'objet en mouvement et pour décrire sa trajectoire.

Pour mieux m'en servir, j'ai conçu un module intermédiaire d'intégration entre le mécanisme MFIA et le module FA (§II.4.3.1) pour faciliter le couplage et l'intégration des trois modules (§II.4.3.2).

4.3.1 Intégration des directions et des vitesses

Comme le module FA récupère les caractéristiques les plus saillantes dans une image de taille fixe 40×40, j'ai besoin de réduire fortement la dimension de la carte d'intégration des vitesses proposée dans §II.2.3.1.

Dans l'architecture montrée dans la figure II.4.5, les cartes d'intégration de vitesse et de direction sont affectées soit par un fort processus de "winner-take-all" (WTA) soit par une simple réduction de la taille en appliquant une gaussienne au résultat obtenu par l'application de l'équation II.2.8 à la carte *Intégration des vitesses*. Dans la figure II.4.5 je montre ce dernier traitement qui a été le plus couramment utilisé, néanmoins j'ai aussi testé le premier.

La première option est définie par

$$\overline{H}(x,y,t) = \frac{1}{\xi_{CN} \cdot \eta_{CN}} \sum_{\xi_{CN} \times \eta_{CN}} max_{\theta,v} H(x,y,\tilde{t},\theta,v,T) \tag{4.4}$$

où $\xi_{CN} \times \eta_{CN}$ est la taille du champ récepteur nécessaire pour réduire les sorties de mon mécanisme MFIA et les adapter pour le module FA.

Et la deuxième option est définie par

$$\overline{H}(x,y,t) = \frac{1}{\xi_{CN} \cdot \eta_{CN}} \sum_{\xi_{CN} \times \eta_{CN}} g_{\xi_{CN} \times \eta_{CN}}(x,y,t) H(x,y,\tilde{t},\theta,v,T) \tag{4.5}$$

avec $g_{\xi_{CN} \times \eta_{CN}}(x,y,t)$ la gaussienne appliquée au voisinage du neurone de sortie du mécanisme MFIA centré en (x,y,t) et $H(x,y,\tilde{t},\theta,v,T)$ la réponse du mécanisme MFIA à l'itération T.

4.3.2 Intégration des trois modules

Une fois parcouru chacun des modules de mon modèle connexionniste neuromimétique et en faisant les adéquations nécessaires avec le module intermédiaire, je me dispose à les intégrer.

La figure II.4.5 montre l'intégration de mon modèle connexionniste avec ses trois modules FSTC, MFIA et FA. On y voit aussi le module intermédiaire (accolade *Intégration*) pour l'adaptation du flot d'information que j'ai traité pour faire l'intégration du mécanisme MFIA vers le module FA. Dans ce module intermédiaire l'intégration de la vitesse est l'entrée au module FA.

Tout d'abord, les huit différentes cartes 3D fournies par le filtrage FSTC sont envoyée sans altérations vers le mécanisme MFIA. L'intégration ultérieure du deuxième module vers le troisième a suivi le traitement fourni par le module intermédiaire d'intégration (§II.4.3.1).

Dans la section suivante, je vais montrer quelques exemples d'application de la détection visuelle du mouvement pour la poursuite d'objets en mouvement dans des séquences d'images réelles.

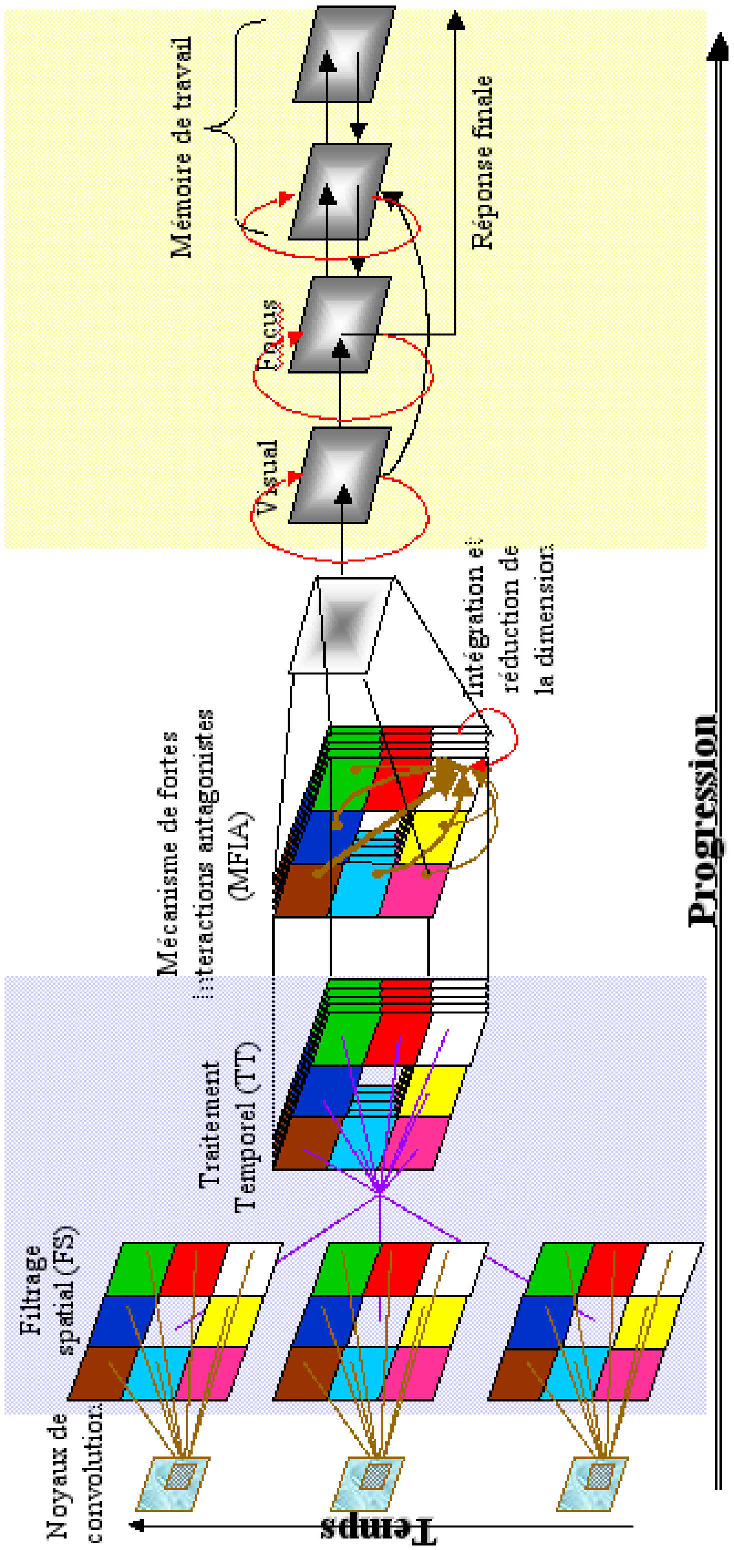


FIG. 4.5 – Architecture globale de mon modèle avec ses trois modules intégrés. D'en bas en haut on perçoit : l'image d'entrée puis le filtrage FSTC, le premier module faisant un filtrage spatio-temporel causal. Ensuite, le mécanisme d'interactions antagonistes, MFIA. Après, une étape intermédiaire pour faire le couplage de mon modèle avec les deux premières étapes du modèle FANF. Et, finalement, notre modèle FA pour la focalisation de l'attention. Il faut noter que je fais une forte réduction de la dimension de la carte *Vitesse* à la carte *Entrée*.

4.4 Applications

Lors du traitement d'une séquence avec chaque image nous cherchons à repérer la bulle d'activité obtenue dans la carte *Focus*.

La description des séquences d'images réelles utilisées est détaillée dans annexe C et les valeurs utilisées pour les paramètres libres du troisième module sont rapportées dans l'annexe D.3.

Dans tous les cas mon modèle se focalise toujours sur la première personne en mouvement qui apparaît dans la scène. Dans aucun cas nous n'avons forcé le changement de l'attention malgré les possibilités (voir les travaux de Rougier et ses collaborateurs [RV05, VRA05] sur cet aspect).

La figure II.4.6 montre la trajectoire résultante pour chacune de cinq parties de la séquence Dispute "A".

La première partie montre deux trajectoires correspondant à deux personnes qui bougent : une qui monte en haut et à droite de la scène et l'autre qui descend depuis le milieu. La troisième personne n'apparaît pas car elle est très petite par rapport aux suppositions de détection d'une bulle d'activité suffisante.

Dans la deuxième partie on voit clairement la trajectoire de la personne qui descend. On perçoit aussi en haut quelques traces de la troisième personne qui n'a pas été détectée dans la première partie.

Dans la troisième partie, on continue a voir faiblement les traces de la personne en haut de la scène.

Dans la quatrième, la plus complexe, on voit clairement la trajectoire suivie par la personne qui vient du côté droit et en haut. On perçoit aussi quelques traces de la personne en haut. Toutes les deux vont se rencontrer, se disputer puis sortir en courant dans des trajectoires divergentes. Il faut noter que le moment de la dispute peut être repéré par le pic dans la trajectoire dans la figure II.4.6d.

Finalement, la cinquième partie ne montre pas les personnes qui bougent en bas de la scène. On ne voit que quelques points en bas à gauche et en haut.

La figure II.4.7 montre la trajectoire résultante pour chacune de trois parties de la séquence Dispute "B".

La première partie montre la trajectoire suivie par une seule personne.

La deuxième partie montre clairement la trajectoire suivie par une des deux personnes qui vont se disputer. À gauche on perçoit les traces de la deuxième personne. Dans le côté gauche et au milieu, on voit des points concentrés. Ils indiquent le mouvement perçu par la dispute de ces deux personnes.

La troisième partie montre les traces de la trajectoire suivie par la personne ayant apparu à gauche dans la deuxième partie de la séquence. On peut y voir une faible séparation vers le haut, c'est l'autre personne.

En faisant un filtrage asymétrique (j'ai mis $\phi = \frac{\pi}{2}$) j'obtiens une trajectoire plus claire que celle montrée en utilisant un filtrage symétrique ($\phi = 0$).

La figure II.4.8 montre la trajectoire résultante pour chacune de trois parties de la séquence Parcourir.

La première partie montre deux trajectoires différentes correspondant successivement à deux des quatre personnes qui bougent dans cette partie de la séquence. La première trajectoire correspond à la personne qui parcourt toute la scène (trajectoire à gauche) et l'autre à la personne qui arrive au centre de la scène, l'atteint puis revient sur ses pas (trajectoire à droite).

J'ai fait la même manipulation que dans la séquence Dispute "B" (voir la figure II.4.8b) pour cette première partie de la séquence Parcourir et on perçoit que la trajectoire obtenue par un filtrage asymétrique est plus forte mais avec plus de petits pics.

Dans la deuxième partie, on voit deux petits points qui apparaissent de deux autres personnes en mouvement : en bas à gauche et en haut.

Finalement, la troisième partie montre la trajectoire faite par une personne qui parcourt toute la scène, reste, va-et-vient et retourne sur ses pas. On perçoit aussi que, pendant que cette personne reste sur la scène avec un faible mouvement, on peut voir d'autres personnes qui bougent aussi (voir en bas et à gauche).

La figure II.4.9 montre la trajectoire résultante pour chacune des trois parties de la séquence Couloir.

Dans la première partie on ne perçoit presque rien en mouvement.

La deuxième partie montre la personne qui sort du magasin et commence à marcher vers le fond du couloir. La trajectoire décrite n'est pas très lisse.

Dans la troisième partie on voit clairement le passage au milieu de la scène et de gauche à droite d'une femme. On perçoit une interruption dans sa trajectoire. Il s'agit d'une colonne qui a caché la femme pour quelques instants. Mais on la revoit à nouveau un peu avant de sortir de la scène.

Finalement, dans la quatrième partie, on ne perçoit pas le mouvement des personnes au fond du couloir. Elles sont très petites par rapport aux conditions d'émergence supposées pour une bulle d'activité dans la carte *Focus*.

4.5 Bilan

La contribution des différentes interactions entre les aires du cortex visuel supérieur, celles du cortex visuel primaire et des aires sous-corticales nous permet d'inférer des mécanismes d'association et d'inhibition pour la commutation de l'attention.

Le module présenté fait interagir différentes cartes neuronales qui s'excitent les unes les autres, quelques unes font des inhibitions et la plupart ont de denses interactions latérales. La population neuronale ainsi décrite fait tout d'abord émerger une seule bulle d'activité parmi les autres possibles dans la carte *Visuelle*, puis elle garde les traces de la bulle d'activité focalisée (obtenue dans la carte *Focus*) par un rebond de la mémoire de travail (*MémTrav*) sur la carte *FEF*. Ces cartes permettent de maintenir l'attention sur la bulle stockée temporairement dans cette mémoire.

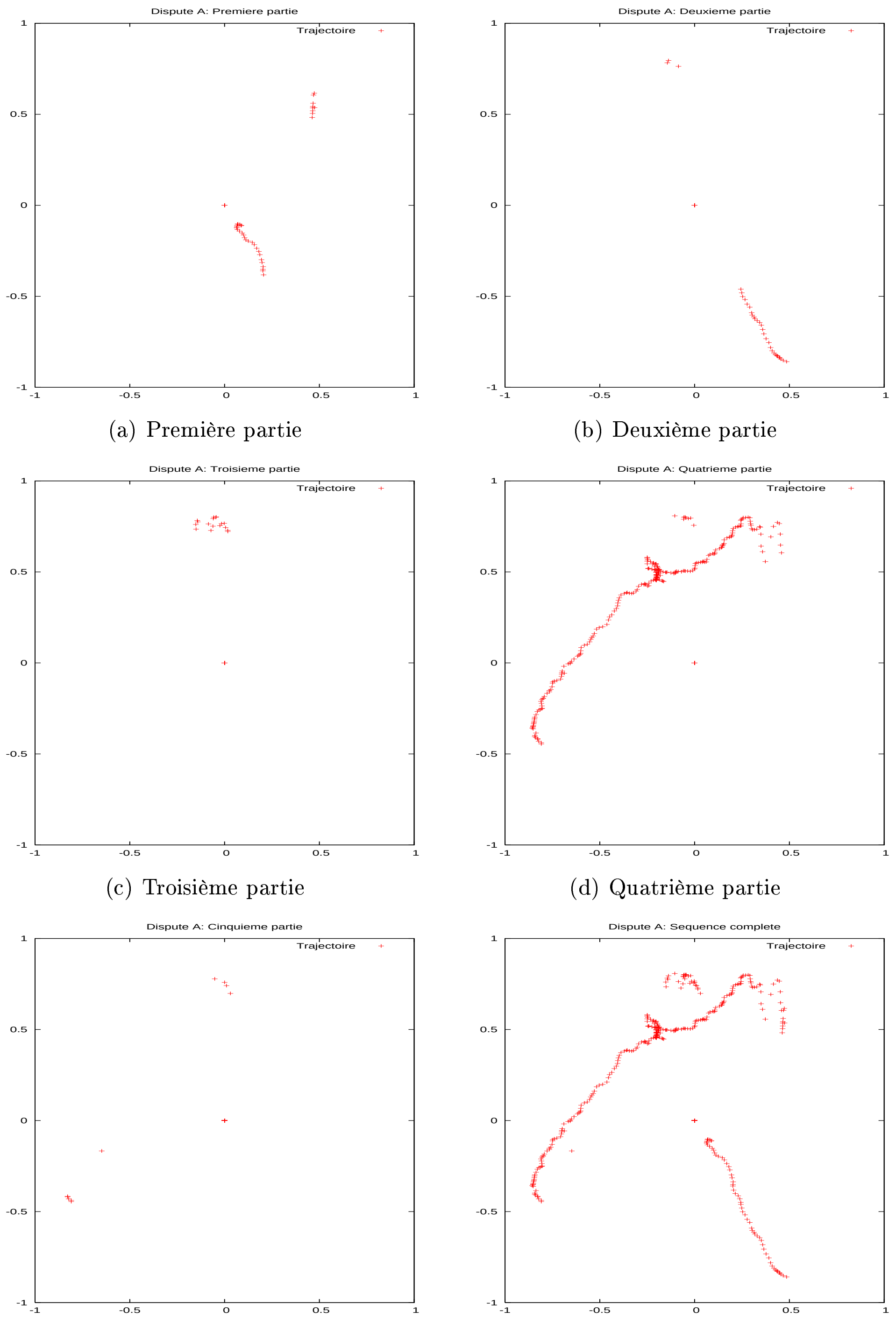
D'après les expériences faites dans §II.4.4, mon modèle connexionniste neuromimétique pour la perception visuelle du mouvement est capable de détecter les personnes ou objets en mouvement et de les poursuivre en tenant compte du fait qu'ils sont dans un environnement où d'autres personnes ou objets bougent aussi. Mon système apparaît suffisamment robuste pour ne pas perdre la focalisation de l'attention sur la cible originale. En outre, il est aussi capable de poursuivre des objets dans des scènes complexes dans lesquelles on n'a aucune information a priori. Néanmoins, dans différentes scènes complexes, notre système commute automatiquement l'attention vers d'autres cibles plus saillantes. Avoir des différentes commutations dans une période du temps assez court est dû à la faible détection par mon filtrage FSTC du mouvement de l'ancienne cible.

D'autres modèles ont été proposés pour le suivi d'un objet en temps réel [PGM00, GMP99, ZS98] en s'inspirant aussi des interactions entre les aires MT, MST et pariétal postérieur (PP).

La mémoire utilisée dans [ZS98] a une fonction de codificateur et la nôtre plutôt celle d'un miroir. La différence des architectures de ces modèles rend difficile une comparaison directe avec la nôtre. Ses performances ne sont pas rapportés dans la littérature accessible.

D'autre part, l'intégration des directions et des vitesses dans mon modèle nous rend des trajectoires plus claires et continues par rapport à leur utilisation de façon indépendante (voir les figures II.4.7c et II.4.7d par exemple), i.e. il est mieux d'utiliser l'intégration proposée par l'équation II.4.5 où je fais une intégration tout d'abord parmi les directions puis parmi les vitesses que de le faire simplement comme un WTA sur le tout (cfr à l'équation II.4.4). Tout cela est aussi un mécanisme qu'on trouve dans les passages d'information de V1 vers MT et de MT vers MST.

Tous les modules de mon modèle connexionniste, bien que très différents sont intensément basés sur des excitations et des inhibitions locales et distribuées. En partant d'un filtrage spatio-temporel causal, nous arrivons à des populations de neurones qui partagent fortement leur information avec leurs neurones voisins. Les uns avec une orientation préférentielle, les autres en les inhibant par leur orientation antagoniste, le tout en faisant émerger une bulle d'activité retenue et nécessaire pour la poursuite.



(a) Première partie

(b) Deuxième partie

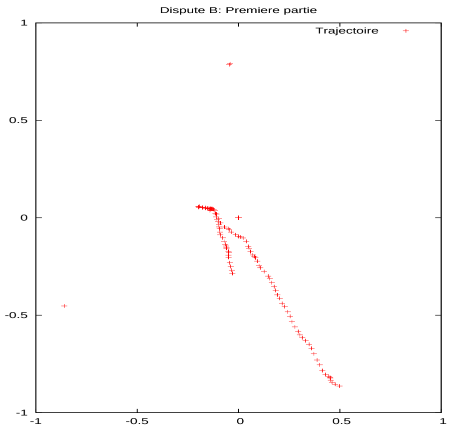
(c) Troisième partie

(d) Quatrième partie

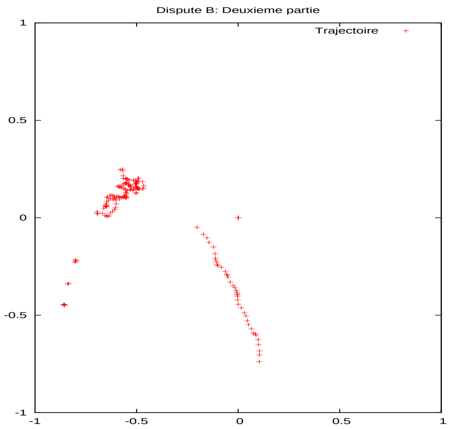
(e) Cinquième partie

(f) Trajectoire complète

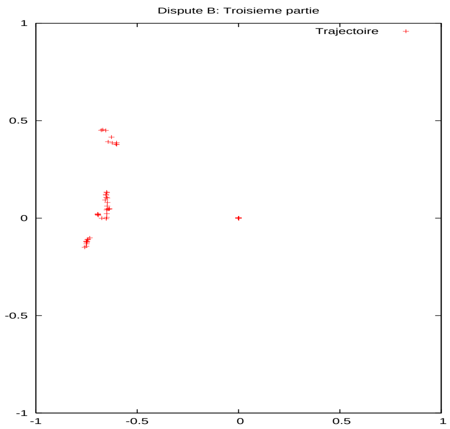
FIG. 4.6 – Trajectoires obtenues de l'application de mon modèle à la séquence Dispute “A”. La trajectoire pour chaque partie de la séquence et la trajectoire complète.



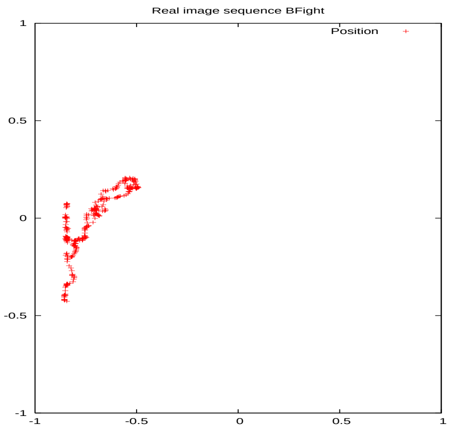
(a) Première partie



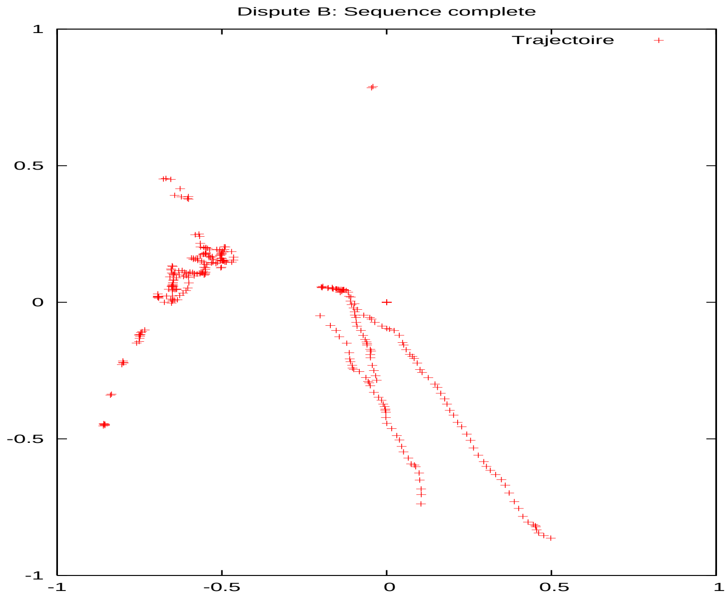
(b) Deuxième partie



(c) Troisième partie

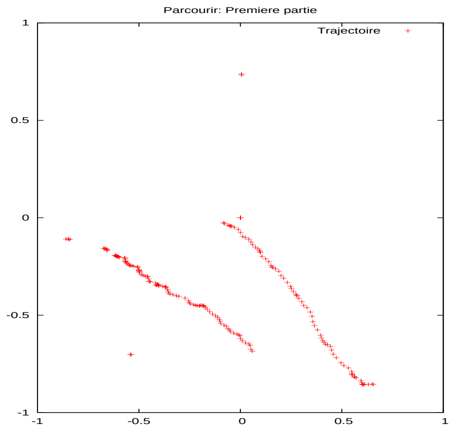


(d) Variation pour la deuxième et troisième partie

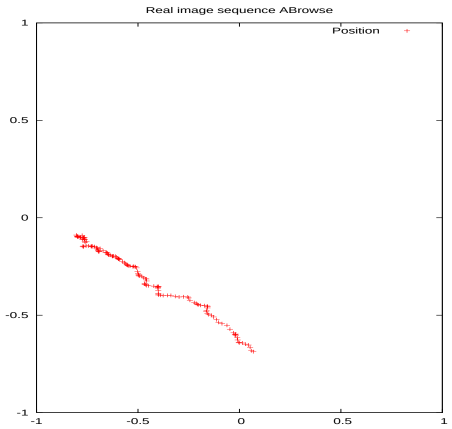


(e) Trajectoire complète

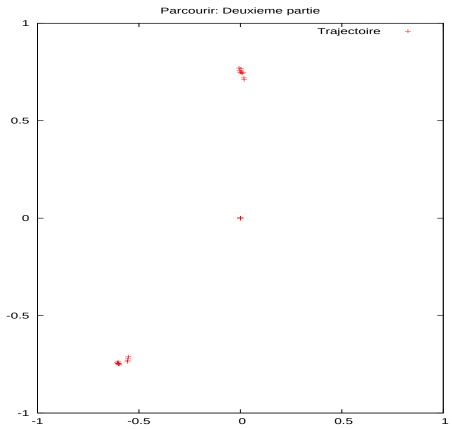
FIG. 4.7 – Trajectoires obtenues de l'application de mon modèle à la séquence Dispute "B". La trajectoire pour chaque partie de la séquence et la trajectoire complète. La figure (d) est une variation des deux dernières partie obtenues à partir du filtrage FSTC avec $\phi = \frac{\pi}{2}$



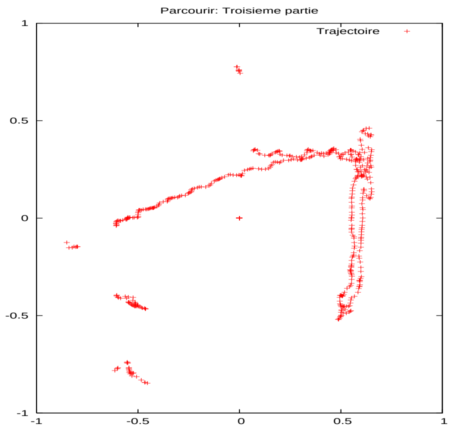
(a) Première partie



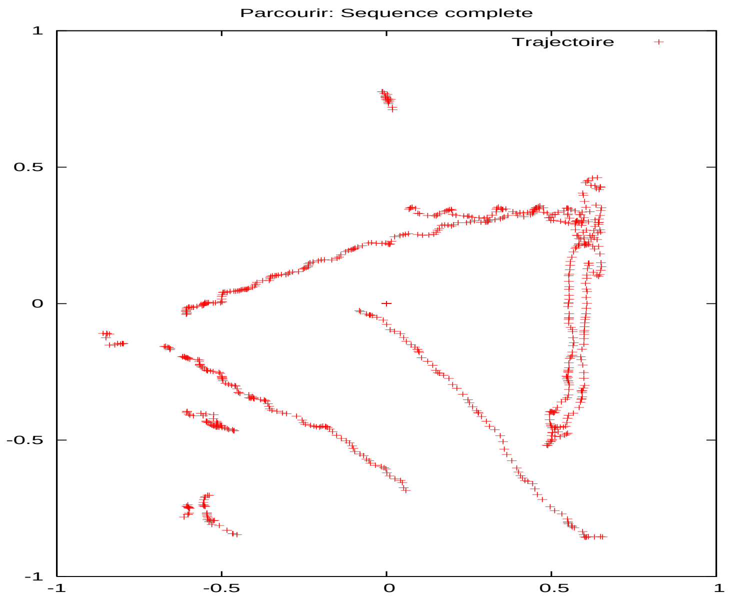
(b) Une variation de la première partie



(c) Deuxième partie

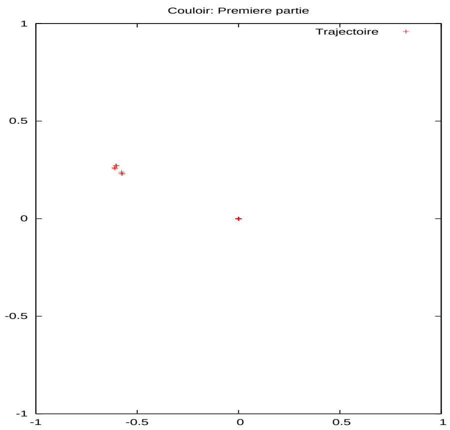


(d) Troisième partie

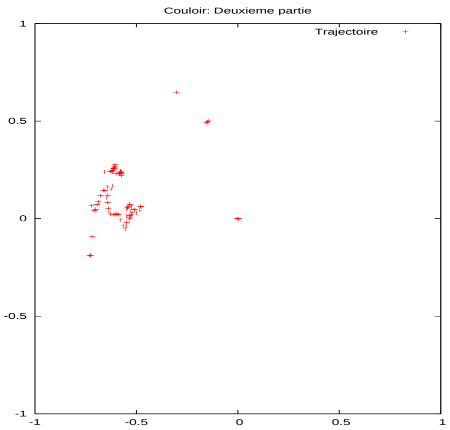


(e) Trajectoire complète

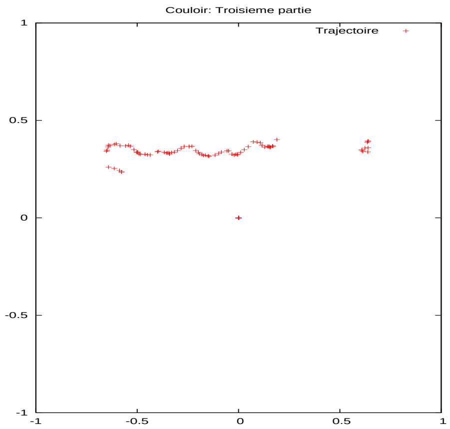
FIG. 4.8 – Trajectoires obtenues de l'application de mon modèle à la séquence Parcourir. La trajectoire pour chaque partie de la séquence et la trajectoire complète. La figure (b) est une variation depuis le filtrage FSTC avec $\phi = \frac{\pi}{2}$.



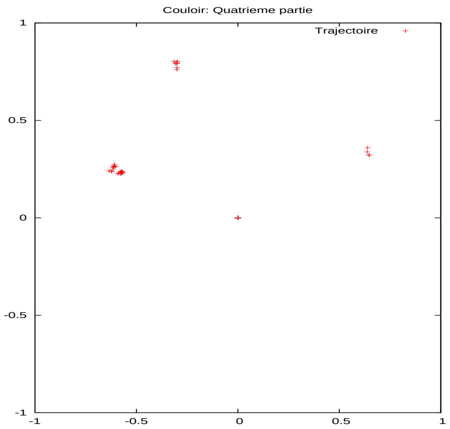
(a) Première partie



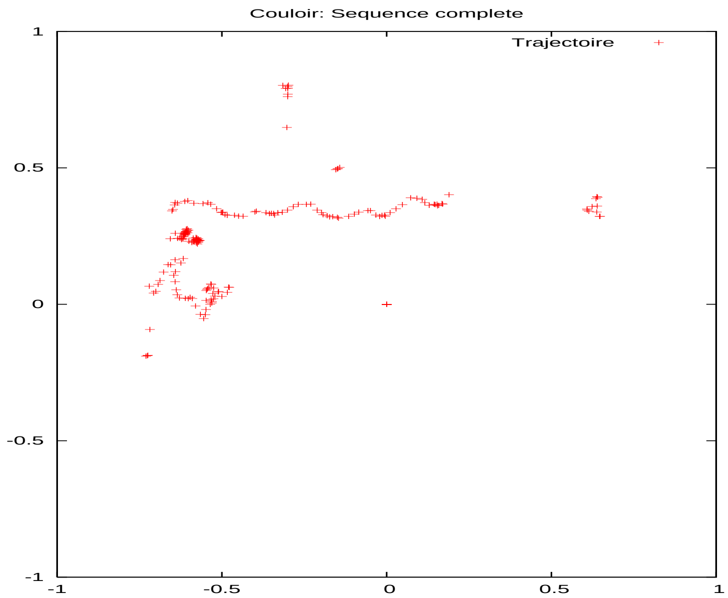
(b) Deuxième partie



(c) Troisième partie



(d) Quatrième partie



(e) Trajectoire complète

FIG. 4.9 – Trajectoires obtenues de l'application de mon modèle à la séquence Couloir. La trajectoire pour chaque partie de la séquence et la trajectoire complète.

Conclusion et perspectives

La recherche en connexionnisme fait aujourd'hui face à des fortes contraintes pour doter un robot d'une totale autonomie. Cette autonomie implique différentes tâches à mettre en place et à faire coopérer de façon constructive pour que le robot soit capable de répondre et de s'adapter aux fréquents changements de son environnement.

Parmi les tâches en robotique autonome les plus souvent demandées et le plus coûteuses en calcul sont celles issues de la perception visuelle : pour la reconnaissance des formes, d'une part, et pour la localisation spatiale, d'autre part.

La communauté scientifique s'est tout d'abord impliquée pour concevoir des systèmes performants pour la reconnaissance de formes. Parmi ces systèmes, ceux issus de l'application des réseaux de neurones artificiels ont bien montré leur capacité de généralisation pour la classification de motifs, d'adaptation lorsqu'on modifie ou ajoute des nouveaux motifs à la base de données appris et de reconstruction des motifs partiellement corrompus.

D'autre part, on a commencé à s'orienter un peu plus du côté de la localisation spatiale. Si le traitement "statique" des images a fait grandement évoluer les démarches de la reconnaissance des motifs, le traitement "dynamique" d'une séquence d'images nous apporte l'information temporelle montrant l'évolution de ces motifs dans une scène et complétant leur information spatiale.

Dans ce cadre, la recherche en perception du mouvement, dont les grandes notions sont apparues au fur et à mesure, pousse la recherche en neurophysiologie et motive une forte interaction inter-disciplinaire avec les domaines scientifiques relativement nouveaux que sont la robotique et l'informatique.

Partout différents chercheurs essaient de modéliser l'être vivant parce qu'on le voit comme une source immense d'inspiration pour mettre en place des systèmes artificiels qui émulent certains de ses comportements et, dans mon cas, pour les tâches liées à la perception visuelle du mouvement.

Dans un contexte évolutif, fortement modulaire et intégrateur, les modèles bio-inspirés fournissent une approche qui essaie de mettre en place les différents aspects du traitement cortical pour aller au-delà des contraintes, parfois prohibitives, de la modélisation classique, à savoir : la haute complexité des systèmes d'équations différentielles à résoudre, la relaxation inhérente à leur application, la forte sensibilité aux changements de l'environnement, l'application en temps réel et les contraintes de généralisation.

D'une part, la neurobiologie nous dit que dans notre cerveau il existe un grand réseau de coopération-compétition distingué en modules différents qui interagissent en pro-action, en rétro-action et par auto-organisation. D'autre part, les essais pour mettre en place ces idées demandent des ressources matérielles assez importantes.

Les modèles classiques nous fournissent déjà les premières approches, mais une intégration de très haut niveau, puis des couplages avec d'autres systèmes font appel à une modélisation plus robuste et ouverte.

Donc, la bio-inspiration fournit des idées de modularité fonctionnelle et de leurs interrelations qu'un modèle informatique peut mettre en place pour nous aider à faire émerger des comportements autonomes. Voilà un des principaux axes dans la recherche en robotique et informatique liée par les neurosciences.

En outre, la perception du mouvement visuel ne dépend pas a priori de l'interprétation ou de la reconnaissance de la forme, mais elle est fondée sur la conjonction d'un processus d'intégration tendant à combiner les signaux de vitesses locales de façon à produire des zones de mouvement cohérent, et d'un processus de segmentation tendant, au contraire, à différencier les signaux locaux de manière à individualiser les objets les uns des autres.

La quantité d'information fournie par le nerf optique (autour de 10^8 à 10^9 bits par seconde) excède largement la capacité du cerveau pour la traiter complètement et l'assimiler comme une expérience consciente. Naturellement, notre système nerveux en choisit certaines parties depuis l'entrée pour être traitées de façon préférentielle en déplaçant le processus de l'attention d'un endroit à l'autre de façon sérielle : c'est l'attention visuelle sélective. Mais l'estimation du mouvement est aussi une façon de réduire la redondance de l'information visuelle. La perception du mouvement visuel ne dépend pas a priori de l'interprétation ou de la reconnaissance de la forme. Elle est fondée sur la conjonction d'un processus d'intégration et d'un processus de segmentation. Le premier tend à combiner l'ensemble des signaux locaux des vitesses de façon à produire des zones de mouvement cohérent, et le deuxième tend, au contraire, à différencier les signaux locaux de manière à individualiser les objets les uns des autres. Ainsi, le mouvement peut être construit comme un évènement qui se développe dans l'espace et dans le temps.

Enfin, mettre en relation la bio-inspiration et les modèles connexionnistes a fait émerger un vaste domaine de recherche où la modularité, la rétro-action et les fortes interactions latérales émergent comme une solution consistante pour pallier les durs chemins classiques de la perception du mouvement.

Dans ce cadre, j'ai proposé un modèle connexionniste neuromimétique permettant de traiter un ensemble de tâches de perception visuelle dynamique (robot/objets environnants en mouvement). Cela implique le compromis de maintenir la satisfaction des contraintes (bio-inspiration, traitements locaux massivement distribués) en vue d'une implantation ultérieure temps-réel embarquée sur circuit FPGA (Field Programmable Gate Array) au cœur de la définition du modèle proposé, malgré la complexité des tâches en jeu.

Bref, je peux résumer mes apports de la façon suivante :

- Mon modèle connexionniste neuromimétique pour la perception visuelle du mouvement est bio-inspiré, densément interconnecté, modulaire, fortement localisé et distribué. Il fait la détection de mouvement dans différentes directions et pour différentes vitesses supposées en réduisant l'effet de flou dans les zones de mouvement.
- Les trois modules de mon modèle connexionniste neuromimétique font tout d'abord une analyse de bas niveau pour la détection du mouvement local (FSTC), puis la détection des aires de mouvement cohérent (MFIA), pour finir avec une analyse de haut niveau pour l'émergence de la focalisation de l'attention. Ces trois modules sont assemblés dans mon modèle bio-inspiré densément interconnecté par le biais des connexions pondérées excitatrices et inhibitrices. La plupart des interactions excitatrices correspondent aux champs récepteurs en pro-action alors que celles inhibitrices correspondent aux interactions latérales.
- Même si d'autres modèles bio-inspirés pour la poursuite [PGM00, SH98, ZS98] ont été proposés, mon modèle traite l'information des signaux lumineux depuis son acquisition et a été construit sous de fortes contraintes de localité, hautement distribué et avec de denses connexions excitatrices et inhibitrices interconnectées en vue de l'implantation matérielle embarquée ultérieure.
- Par rapport au filtrage FSTC, j'ai volontairement choisi de traiter l'aspect temporel de

façon différente des modèles classiques. Pour moi le traitement temporel est complètement différent du traitement spatial. J'ai tout d'abord fait un traitement spatial par la simple application de filtres de Gabor 2D et inspiré des réponses des cellules simples de V1 puis une interprétation causale de la vitesse qui implique l'aspect temporel. Ceci est mon filtrage spatio-temporel causal (FSTC).
- Mon mécanisme MFIA est fortement inspiré des interactions inter- et intra-colonnes d'orientation dans V1 avec un rayon d'influence dynamique pour chaque neurone. Dû à ce rayon, les contraintes de localité sont respectées et, par diffusion et donc de façon distribuée, mon système aboutit à homogénéiser les zones de mouvement présentes avant de commencer à bruiter l'image depuis chaque zone du mouvement repéré.
- Le module FA met en œuvre les denses interactions inter- et intra-aire des aires V1, MT et MST. Ces interactions sont issues d'une population des neurones complètement interconnectée selon le principe de la théorie des champs neuronaux continus (CNFT). Les interactions coopératives-compétitives des populations neuronales des cinq différentes cartes utilisées font émerger une seule bulle d'activité qui est stockée et repérée par rebond des deux cartes supérieures pour maintenir la focalisation de l'attention sur elle et la poursuivre.
- L'intégration des trois modules de mon modèle connexionniste neuromimétique a été indépendante du filtrage FSTC vers le mécanisme MFIA, d'une part, et l'intégration des orientations et des vitesses se fait avec une forte réduction de la dimension des sorties du mécanisme MFIA vers FA, d'autre part.
- Nous avons implanté le première module (FSTC) sur des circuits FPGA. Le gain a été 1 à 100. L'étude de la faisabilité d'implantation du deuxième module (MFIA) nous a permis d'explorer les capacités d'implantation du parallélisme au niveau neuronal et au niveau des connexions synaptiques.

J'ai choisi quelques résultats de la neurobiologie pour m'inspirer et pouvoir concevoir mon modèle connexionniste neuromimétique pour la perception visuelle du mouvement. Il faut mettre en valeur que je n'ai pas modélisé les interactions des colonnes corticales dans V1 ni les interactions de MT et MST, je me suis simplement inspiré du traitement cortical de l'information visuelle pour la perception du mouvement. Donc, mon modèle a plusieurs aspects à améliorer parmi lesquels nous pouvons énoncer les suivants :
- J'ai fait la détection du mouvement dans huit directions différentes parmi les 36 possibles biologiquement et j'ai supposé au moins cinq vitesses différentes en considérant un haut échantillonnage. Tout cela est très gourmand en capacité de calcul, en espace de mémoire et en temps (autour de deux heures pour 100 époques [48] dans le mécanisme MFIA pour une image de 128×128).
- Le paramétrage pour chaque module est devenu ma difficulté principale. Pour le premier module (FSTC) j'ai pris les valeurs des paramètres suggérés dans la bibliographie étudiée. Pour le deuxième (MFIA) et le troisième (FA) nous les avons fixé par rapport à nos expériences.
- La sortie de mes détecteurs est modulée par le mouvement mais mes détecteurs peuvent aussi répondre fortement au mouvement propre dans l'image [CI03]. Je n'ai pas encore séparé le mouvement propre du mouvement des objets.
- J'ai choisi une seul bande de fréquence spatiale pour mes expériences de filtrage. D'après la neurobiologie, il s'agit plutôt d'un ensemble des bandes de fréquences qui interagissent

[48]Nombre d'itérations nécessaires pour que tous les neurones soient actualisés.

pour aboutir à une réponse dépendant de la fréquence spatiale et temporelle.
- D'après Schiller et ses collaborateurs [SSM86], l'utilisation d'un mécanisme d'excitation et d'inhibition dans un canal simple comme je le fais est moins efficace par rapport à la codification de l'augmentation et de la diminution de l'information par des processus excitateurs opposés.
- La dynamique d'activation des neurones dans le cortex moteur primaire (M1) [GLP+89] a montré que le vecteur de population, c'est-à-dire, la somme vectorielle des directions préférentielles des neurones pondérées par leur activité, tourne de la direction initiale indiquée par le stimulus vers la direction finale correspondant à celle du mouvement à effectuer. La direction de ce vecteur de population est une meilleure prédiction neuronale de la direction du mouvement effectué que la direction préférentielle des neurones dont la fréquence de la réponse est maximale [GLP+89, GSK86, GTL93], mécanisme plus simple que j'ai utilisé pour le moment.

D'après la recherche continue en neurobiologie et des aspects que j'ai volontairement omis dans la conception de mon modèle connexionniste neuromimétique j'envisage dans un futur proche de compléter mon modèle par :
- L'introduction des connexions en rétro-action du module FA vers le module MFIA. Ces connexions nous donneront une architecture plus proche des données neurobiologiques et nous aideraient à renforcer l'émergence pour une poursuite plus robuste d'un objet en mouvement dans des environnements dynamiques.
- Comme conséquence de la rétro-action énoncée ci-dessus, le traitement du mouvement propre nous permettra de généraliser mon modèle vu qu'au niveau du mécanisme MFIA nous avons déjà des suggestions de discrimination du mouvement propre de mouvement des objets.
- Pour le suivi simultanée de différentes bulles d'activité et en même temps sélectif à des vitesses préférées, nous pouvons ajouter des cartes *Focus* supplémentaires s'inhibant mutuellement mais qui garderaient chacune une seule bulle d'activité.
- Pour atteindre l'auto-organisation dans mon mécanisme MFIA je devrai permettre aux connexions pondérées des inhibitions latérales et aux poids excitateurs ascendants de varier en accord avec une règle d'apprentissage [HM94, MAH94]. Une règle d'apprentissage inhibitrice anti-hebbienne peut générer des connexions inhibitrices latérales qui deviennent sélectives dans une voie qui permet de représenter le mouvement transparent [Mar03]. Cette règle renforce les connexions inhibitrices parmi les neurones détecteurs de mouvement local qui représentent des mouvements similaires et affaiblit les connexions inhibitrices parmi les neurones détecteurs du mouvement local qui représentent des mouvements non similaires.

En conclusion, la recherche pluridisciplinaire pour doter un robot d'une totale autonomie peut se suivre dans le contexte des approches bio-inspirées où la modularité, l'intégration, la coopération-compétition et le caractère évolutif des interactions inter- et intra-modules nous offrent les outils nécessaires pour palier les contraintes parfois prohibitives de la modélisation classique.

Ces interactions pour une tâche particulière (dans mon cas pour la perception visuelle du mouvement) couplées aux interactions issues d'autre tâches (comme la reconnaissance de formes, la détection des odeurs, la planification de trajectoires, etc.) peuvent faire émerger des comportements tout d'abord non mis en relation puis par apprentissage soit supervisé soit non supervisé soit par renforcement, les associer pour aboutir à une tâche précise plus complexe à résoudre. Cela impliquerait un module capable d'intégrer les éventuels propriétés surgies des interactions de ces différentes tâches mises ensemble.

Dans ce contexte, la mise en série de différentes tâches de perception puis de la mise en étroite relation de leur résultats nous donnerait une autre façon de voir le comportement des interactions des neurones de notre cerveau qui fonctionnent en parallèle et de façon distribuée. Donc, on commencerait plutôt pour la modélisation d'une tâche précise, avant de considérer le couplage avec d'autres tâches déjà modélisées en gérant les propriétés issues de ce couplage pour enfin permettre la pluralité de ces associations en vue de la création d'une intelligence neuromimétique pouvant interagir avec les systèmes issus de l'intelligence artificiel classique.

Annexe A

Parcours historique de la recherche dans la perception du mouvement

Dans la recherche en perception du mouvement, les grandes notions sont apparues au fur et à mesure en neurophysiologie et grâce à une forte interaction inter-disciplinaire avec des nouveaux domaines scientifiques comme la robotique et l'informatique.

A.1 La recherche en neurophysiologie

Depuis que Santiago Ramón y Cajal [RyC94] a montré que le système nerveux est constitué par l'assemblage d'unités distinctes appelés neurones, la recherche en neurophysiologie a déclenché différentes lignes d'exploration, notamment dans l'organisation et l'inter-coopération de ces ensembles de neurones.

A.1.1 Le champ récepteur

La notion de champ récepteur est née des travaux d'Adrian en 1928 [Adr28] et les interactions inhibitrices latérales par ceux de Hartline en 1938 (dans les espèces inférieures) et de Kuffler en 1958 (chez les mammifères) [Har38, Kuf53].

En 1953, la notion de champ récepteur est devenue une pièce importante pour la détection du mouvement [Bar53, Rei61] et pour la recherche sensorielle et l'inhibition spatio-temporelle [Rei61].

A.1.2 La sélectivité à la direction

La sélectivité à la direction dans la rétine chez les mammifères proposée par Aubert en 1886 [Aub86] obéit à une direction préférentielle et à une nulle après un processus d'inhibition-excitation [BH63].

D'autre part, Hubel et Weisel ont montré que les neurones du cortex visuel sont sélectifs à la direction et leur sélectivité au mouvement serait une propriété sensorielle fondamentale fondée sur la localité de leur champs récepteurs [HW59, HW62, HW68].

En 1968, Hubel et Weisel [HW68] ont montré qu'un tiers des cellules de V1 chez le singe sont sélectives à la direction et localisées préférentiellement dans la couche IV et VI de V1.

A.2 Les modèles bio-inspirés

Partout les chercheurs essaient de modéliser l'être vivant parce qu'on le voit comme une source immense d'inspiration pour mettre en place des systèmes artificiels qui émulent certains de ses comportements. Dans notre cas, pour les tâches liées à la perception visuelle du mouvement.

A.2.1 Les pionniers

Exner a été le premier à proposer un simple détecteur de mouvement [Exn88b, Exn88a, Exn94]. Il a présenté une figure d'un réseau de neurones faisant des comparaisons entre deux stimulations successives dans différents points de la rétine (l'espace) et dans différents points dans le temps. Ce réseau serait le premier modèle neuronal pour la détection du mouvement.

Quarante ans après, des travaux de Brown [Bro31] émergeait un modèle pour l'estimation de la vitesse.

A.2.2 Le modèle de base

Hassenstein et Reichardt sont à l'origine du premier modèle informatique pour la détection du mouvement [HR56, Rei61] en considérant que les cellules du cerveau répondent de façon individuelle à une direction donnée. Il s'agit d'un dispositif logique simple capable de détecter le mouvement dans une direction et de s'inhiber dans la direction contraire par des pré-filtrages spatiaux et temporels : un simple détecteur de corrélation multiplicatif à base de deux sous-unités de détecteurs de mouvements opposés.

Ce modèle a grandement inspiré la recherche connexionniste jusqu'à nos jours [vSS84, vSS85, IC01b, IC01a] en mettant en valeur les propriétés de ce pré-filtrage qui renforce les caractéristiques des détecteurs qu'il nourrit.

Annexe B

Modèles énergétiques et approches multirésolution

B.1 Un parcours parmi les modèles énergétiques

Dans leur forme mathématiquement parfaite, les modèles d'énergie et de Reichardt produisent des sorties identiques [vSS85, EBA92], mais dans les systèmes biologiques ces deux principes ne sont pas distingués dans toutes les étapes.

Adelson et Bergen [AB85] ont proposé des filtres d'énergie du mouvement pour modéliser la sensibilité des cellules simples et complexes du cortex visuel primaire où chaque réponse spatiale est multipliée par chaque réponse temporelle. Les filtres résultants sont alors sommés et soustraits pour produire quatre filtres sélectifs directionnellement : une paire sélective à gauche et l'autre paire sélective à droite, avec les membres de chaque paire en quadrature. Les réponses de chaque paire sont calculées au carré puis sommées pour obtenir l'énergie associée à une direction particulière.

Leur modèle d'énergie spatio-temporel 2D est fondé sur des couples de filtres linéaires transversaux orientés dans l'espace-temps. En extrayant l'énergie de mouvement à gauche et à droite, Adelson et Bergen [AB85] ont produit une énergie de réponse opposée équivalente à un détecteur élaboré de Reichardt pour un ensemble de filtres donnés.

Watson et Ahumada [WAJ85] ont défini un mécanisme sélectif à l'orientation par similitude aux mesures psychophysiques de la sensibilité au mouvement chez l'être humain. Ils ont utilisé un ensemble de fonctions spatiales 2D de Gabor et une fonction temporelle de Gabor dont les premières sont accordées à différentes orientations pour l'estimation de la vitesse locale de l'image. L'intégration des réponses de ces filtres discriminent les mesures locales à chaque filtre avec un groupe directionnel qui fournit une composante linéairement indépendante du vecteur de vitesse.

Le modèle Adelson et Movshon [AM82] utilise les composantes de premier ordre plus la règle d'intersection des contraintes. La première étape extrait les composantes de la vitesse et la direction de façon codifiées. Dans la deuxième, deux ou plus composantes sont recombinées pour codifier les motifs de direction et vitesse.

Heeger a présenté un modèle informatique pour l'estimation de la vitesse de l'image en utilisant les paires de quadrature des filtres de Gabor spatio-temporels [Hee88]. Il a défini une famille de filtres de Gabor accordés à la même bande de fréquence spatiale, mais avec une différentes orientations spatiales. L'amplitude de la fréquence spatiale accordée est donc invariante et les autres ensembles de filtres peuvent être destinés à différents canaux de fréquences. En outre, un simple ensemble de tels filtres peut être mis en cascade à travers une pyramide gaussienne (voir l'annexe B.2.1) en vue de couvrir différents canaux.

Son modèle fait des convolutions par des filtres de Gabor qui exploitent leurs propriétés de séparabilité où chaque convolution est une cascade de lissages gaussiens passe-bas et de filtres de Gabor passe-bande. Il minimise l'erreur de prédiction par la technique de Newton.

Le modèle de Lu et Sperling [SL95] est un modèle énergétique plus une rectification d'onde complète dans les deux dimensions de l'image pour convertir l'information de second ordre en premier ordre.

Le modèle Bowns [Bow01] est un modèle énergétique incorporant un mécanisme parallèle de poursuite des caractéristiques. Une amélioration que lui-même a faite est de poursuivre les caractéristiques au niveau des composantes. Dans ce cas là, la direction est perpendiculaire à l'orientation de la composante.

B.1.1 Estimation par une banc de filtres spatio-temporels de Gabor

La représentation pyramidale ou *représentation multi-résolution* est le résultat d'une succession de filtrages passe-bas et de sous-échantillonnages. Cette approche est utilisée couramment dans l'estimation du mouvement dans le but d'estimer les grands déplacements [HB93].

Les grandes vitesses dans l'image initiale à haute résolution apparaissent comme des petites vitesses dans une image à basse résolution d'un niveau supérieur de la pyramide. La méthode de Heeger utilise la multi-résolution en proposant les domaines de vitesse suivants correspondant à chaque niveau de la pyramide. Avant le traitement, on fait un pré-filtrage de chaque image pour se situer plus près de l'hypothèse de spectre blanc. En utilisant un filtrage spatial de différences gaussiennes (DOG, *Difference of Gaussiens*, un filtre passe-bande) on blanchit l'allure du spectre de chaque image en éliminant les très basses fréquences (la composante continue et l'énergie qui l'entoure) et en nettoyant les hautes fréquences qui sont généralement dues au bruit.

Spinei [Spi98] a utilisé 36 filtres de Gabor disposés en 3 gammes de vitesses (ou niveaux de la pyramide). Chaque gamme de vitesse a 12 filtres disposés à différentes fréquences spatiales et temporelles. Donc cette architecture assure un recouvrement assez compact de l'espace fréquentiel.

Par ailleurs, il est parti des méthodes de filtrage contrôlé [49], en s'inspirant d'une technique d'estimation par un banc de filtres de Gabor calculant directement la vitesse, vers l'estimation du mouvement par triades de Gabor et sa généralisation dans le cas du mouvement transparent [50].

Pour le calcul des énergies des réponses des filtres et en utilisant le théorème de Parseval on a :

$$\int\int\int_{R^3} |F(f_x, f_y, f_t)|^2 df_x d_f y df_t = \int\int\int_{R^3} |I(x,y,t)|^2 dxdydt \tag{B.1}$$

$$\sim \sum_{R^3} |I(x,y,t)|^2 \tag{B.2}$$

On peut donc calculer directement l'énergie en utilisant les images issues des filtres d'énergie de Gabor ($|I(x,y,t)|^2$) car elle est invariante à la phase des textures. Ensuite, il est nécessaire d'intégrer (dans le domaine discret on fait la somme) cette expression dans le domaine spatio-temporel pour remplir les conditions du théorème de Parseval et retrouver l'énergie de la réponse

[49] Il a traité différemment les régions proches des frontières de mouvement et les régions éloignées, afin de diminuer les effets de lissage intrinsèques à la méthode initiale.

[50] Deux vitesses dans un même point.

du filtre de Gabor. En outre, comme on demande une mesure locale dans chaque pixel, on effectue l'intégration sur un support spatio-temporel limité. Pour cela, on effectue une convolution avec la gaussienne à la base du filtrage de Gabor 3D [Hee88].

B.1.2 Quelques équivalences et variantes

Adelson et Bergen ainsi que Simoncelli [AB86, Sim93] ont montré que l'approche fondée sur le gradient de Lucas et Kanade [LK81] est équivalente aux méthodes énergétiques. D'autre part, Poggio et Reichardt [PR76] ont montré que le modèle proposé par Reichardt [Rei69] est équivalent aux méthodes énergétiques. Ces modèles sont présentés comme des modèles biologiques pour la détection du mouvement chez l'être humain [WAJ85, GY90].

Grzywacz et Yuille [GY90] ont proposé des filtres de Gabor 3D sélectifs à l'orientation pour mesurer l'énergie de mouvement dans l'espace de la fréquence. L'estimation de la vitesse est obtenue par intégration des réponses des différentes populations de filtres sur une base locale.

B.2 Approche Multirésolution en vision par ordinateur

Une des approches les plus utilisées dans la vision par ordinateur est l'approche multirésolution : un traitement hiérarchique à plusieurs niveaux qui réduit la dimension de l'entrée au fur et à mesure (généralement dans un facteur de deux).

Soit

$$\hat{f}_0 = g * f_0 \tag{B.3}$$

avec f_0 le signal initial, $\hat{f}_0$ le signal filtré, g le filtre passe-bas et $*$ l'opérateur de convolution. Dans le domaine de Fourier

$$\hat{F}_0 = G \cdot F_0 \tag{B.4}$$

Ce qui donne la formule générale

$$f_1 = \hat{F}_0 \downarrow 2 = (g * f_0) \downarrow 2 \tag{B.5}$$

où le symbole $f_0 \downarrow k$ indique le sous-échantillonnage de pas k. Et par itération

$$f_{k+1} = (g * f_k) \downarrow 2 \tag{B.6}$$

Parmi les approches multirésolution les plus utilisées nous trouvons : la pyramide gaussienne, d'erreurs, de différences gaussiennes et la laplacienne. Je vais par la suite les décrire.

B.2.1 Pyramide gaussienne

Soit I une image et g_σ le filtre gaussien 2D

$$g_\sigma(x, y) = \frac{1}{2\pi\sigma} exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{B.7}$$

On définit la pyramide gaussienne par

$$G_0 = IG_k = (g_\sigma * G_{k-1}) \downarrow 2 = (g_\sigma * g_\sigma * ...g_\sigma * G_0) = (g_{k,\sigma} * G_0) \downarrow 2^k \tag{B.8}$$

B.2.2 Pyramide des erreurs

Si on fait un filtrage avec un échantillonnage, on perdra, en général, de l'information.
On quantifie la perte d'information, E_k, par l'écart entre l'image G_k et la prédiction que l'on peut faire à partir de l'image G_{k+1}

$$E_k = G_k - (g_\sigma * G_{k+1} \uparrow 2) \tag{B.9}$$

Une propriété intéressante est que la suite $Gn, E_n, ..., E_0$ constitue un codage exact de l'image I

$$G_k = (g_\sigma * G_{k+1} \uparrow 2) + E_k \tag{B.10}$$

B.2.3 Pyramide des différences de gaussiennes (DOG)

Si on assemble les équations B.9 et B.8 dans sa première partie, on obtient

$$E_k = G_k - ((g_\sigma * G_k) \downarrow 2) \uparrow 2 = G_k - g_{2,\sigma} * G_k \tag{B.11}$$

et, d'après l'équation B.8 dans sa dernière partie,

$$E_k = (g_{k,\sigma} - g_{k+2,\sigma}) * G_0 = DOG_k(G_0) \tag{B.12}$$

B.2.4 Pyramide laplacienne

Il y a analogie de forme entre une différence de gaussiennes et la dérivée seconde (laplacienne en 2D) d'une gaussienne, à savoir :

$$\nabla^2 g_\sigma(x, y) = -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2 + y^2}{2\sigma^2}\right) exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{B.13}$$

donc,

$$DOG_k(G_0) \approx G_0 * \nabla^2 g_{k\sigma} = \nabla^2 G_0 * g_{k\sigma} = \nabla^2(G_0 * g_{k\sigma}) \tag{B.14}$$

où $G_0 = I$ et $g_{k,\sigma} = (g_{k-1} * g_\sigma) \downarrow 2$. Le niveau k de la laplacienne contient une information de type contour.

Annexe C

Description des séquences d'images utilisées

Dans cette annexe je décris les séquences d'images utilisées. Toutes les séquences qui vont être décrites sont des séquences d'images réelles et seule la séquence Yosemite est une séquence d'images synthétiques.

Les séquences d'images hormis celle du Taxi d'Hambourg et de Yosemite sont issues d'une caméra de surveillance. Ces séquences sont disponibles au public dans le site Internet du projet EC Funded CAVIAR : projet/IST 2001 37540 (*http ://homepages.inf.ed.ac.uk/rbf/CAVIAR/*). Elles ont été générées le 11 juillet 2003 et le 20 janvier 2004.

La séquence Couloir montre les faits dans un des couloirs d'un centre commercial à Lisbonne au Portugal et les autres dans le hall du laboratoire INRIA à Grenoble en France.

La résolution de ces séquences a été de demi-résolution PAL standard (384×288 pixels, 25 images par seconde) et compressées en utilisant MPEG2.

C.1 Taxi d'Hambourg

La séquence d'images de la fameuse scène du taxi d'Hambourg montre trois véhicules en mouvement : au centre, le taxi (voiture blanche) qui tourne à sa droite ; en bas et à gauche, une voiture noire qui se déplace vers la "droite", et en bas et à droite un camion qui se déplace vers la "gauche". De plus, en haut et à gauche il y a une personne qui marche vers la gauche.

Il faut noter que dans cette séquence il y a un léger mouvement de la caméra (mouvement propre) de gauche à droite et de haut en bas. Ce mouvement bruite l'estimation du flot optique.

Dans la figure C.1 je montre un détail de la séquence du taxi d'Hambourg.

FIG. C.1 – Séquence de la scène du taxi d'Hambourg.

Pour mes expériences, la taille de cette séquence a été de 21 images. Ces images sont en format JPEG, en échelle de gris et de 256×190 pixels chacune.

C.2 Séquence Yosemite

La fameuse séquence d'images synthétiques "Yosemite Fly-Through" a été produite par Lynn Quam au sein de SRI International.

Cette séquence montre le survol d'un engin parmi les collines. Donc, on a un mouvement propre qui génère un flot optique divergent où le focus d'expansion est situé autour du centre de l'image. La vitesse d'en bas et à gauche est de 5 pixels par image et les nuages se déplacent vers la droite d'environ 2 pixels par image.

Donc, le mouvement du terrain est divergent et celui des nuages dans l'horizon plus au moins constant.

La séquence que j'utilise est composée de 15 images de 316×252 pixels et est en format JPEG et en échelle de gris. La figure C.2 montre un détail de cette séquence.

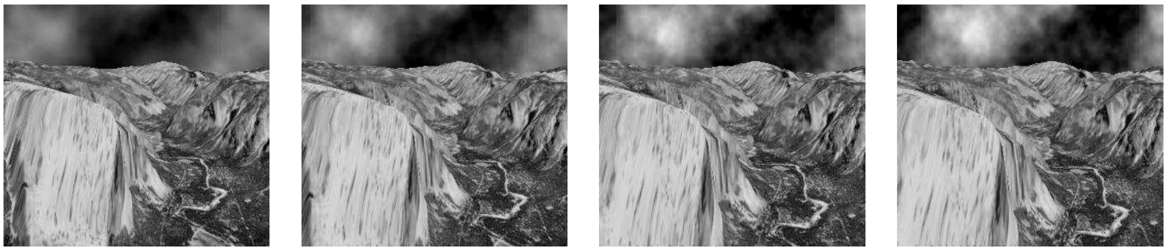


FIG. C.2 – Séquence de la scène du "Yosemite Fly-Through".

C.3 Séquence Dispute "A"

La séquence d'images réelles Dispute "A" issue de la caméra de surveillance dans le hall du laboratoire de l'INRIA à Grenoble peut être organisée en cinq parties : dans la première partie, on voit trois personnes qui marchent dans des directions différentes ; dans la deuxième, il n'y en a que deux qui marchent aussi dans des directions différentes ; dans la troisième, une seule personne au centre et en haut qui marche ; dans la quatrième, deux personnes marchent, s'approchent et commencent à se disputer, puis ils s'en vont en courant dans deux chemins différents. Finalement, dans la cinquième partie, on arrive à voir trois personnes qui bougent en bas de l'image.

La quatrième partie est très complexe puisque les deux personnes qui se rencontrent se disputent entre elles, se cachent entre elles et changent d'orientation à un moment donné.

Dans la figure C.3 je montre dans chaque ligne de haut en bas les cinq différentes parties de cette séquence.

Cette séquence a 551 images RGB en format JPEG.

C.4 Séquence Dispute "B"

La séquence d'images réelles Dispute "B" issue de la caméra de surveillance dans le hall du laboratoire de l'INRIA à Grenoble peut être organisée en trois parties : dans la première partie, on voit une seule personne qui marche, reste au centre de la scène peu de temps, puis il sort de la scène par la partie en bas et à droite ; dans la deuxième, une personne qui marche apparaît au fond et à gauche de la scène puis une deuxième personne qui marche aussi apparaît en bas et

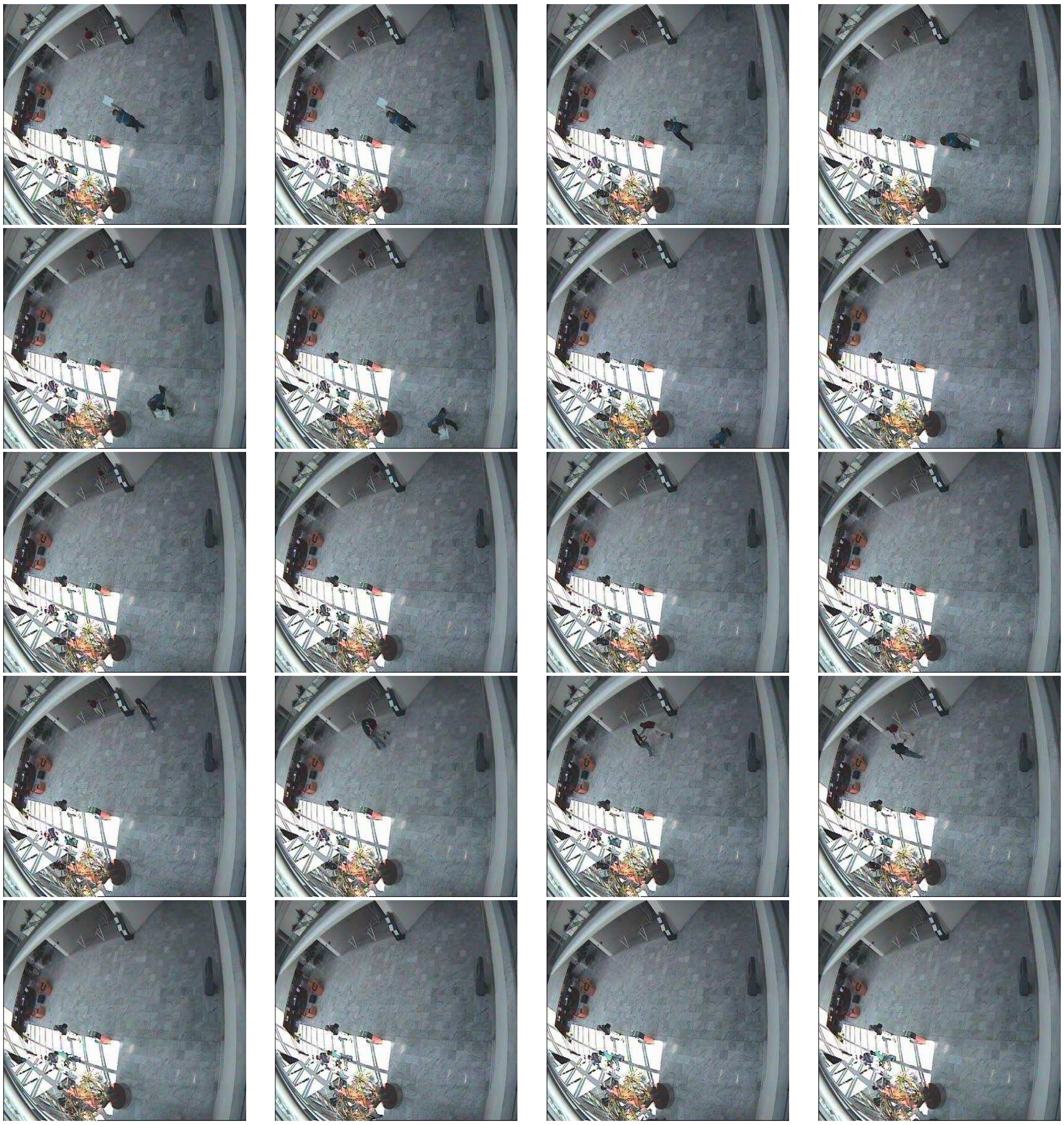


FIG. C.3 – Séquence d'images réelles, Dispute "A", issue d'une caméra de surveillance.

Fig. C.4 – Séquence d'images réelles, Dispute "B", issue d'une caméra de surveillance.

à droite de la scène. Tous les deux se rencontrent au centre gauche de la scène et se disputent. Finalement, dans la troisième partie de cette séquence les deux personnes qui se sont rencontrées se dispersent et se rejoignent à nouveau en restant toujours dans le côté centre gauche de la scène.

La deuxième partie est aussi complexe car, dans la dispute, il y a un moment où la deuxième personne cache presque complètement la première.

Dans la figure C.4 je montre dans chaque ligne de haut en bas les trois différentes parties de cette séquence.

Cette séquence a la même taille et les mêmes caractéristiques que la séquence Dispute "A".

C.5 Séquence Parcourir

La séquence d'images réelles Parcourir peut être organisée en trois parties : dans la première partie on voit quatre personnes qui marchent : deux on en voit clairement deux autour du centre de la scène. De ces deux personnes une restera au centre puis reviendra sur ses pas et l'autre continuera à marcher jusqu'au centre gauche de la scène. Dans la deuxième partie, il n'y a pas de mouvement et dans la troisième, une autre personne se balade dans la scène : elle arrive à la machine au centre droit de la scène, reste, marche en bas, revient à la machine et marche à gauche jusqu'au moment où la séquence est arrêtée.

Dans la figure C.5 je montre la première et la troisième partie de cette séquence.

Cette séquence a 1043 image et a les mêmes caractéristiques que la séquence Dispute "A".

FIG. C.5 – Séquence d'images réelles, Parcourir, issue d'une caméra de surveillance.

C.6 Séquence Couloir

La séquence d'images réelles Couloir est issue de la caméra de surveillance d'un magasin. Cette séquence peut être organisée en quatre parties : dans la première partie un groupe de personnes discute et bouge faiblement dans le même endroit au centre haut de la scène. Dans la deuxième partie, la personne "X" sort du magasin (au centre gauche de la scène), marche vers le fond du couloir (au centre haut de la scène) et le groupe de personnes qui discutait marche aussi au fond du couloir. Au fur et à mesure, le groupe de personnes et la personne "X" se confondent, forment trois puis deux groupes. Dans la troisième partie, on voit une femme sortir du magasin et traverser la scène de gauche à droite pendant que le groupe de personnes continue à marcher vers le fond du couloir. Cette femme passe parmi les colonnes qui la cachent, on la revoit puis elle disparaît de la scène.

Finalement, dans la quatrième partie le groupe de personnes continue à marcher, une personne du groupe reste, les autres continuent, puis le groupe tourne à droite et disparaît et un peu plus loin la personne "X" tourne aussi à droite et disparaît de la scène.

On peut voir dans cette séquence une réflexion des silhouettes des personnes sur le sol et pour la femme aussi sur la colonne.

Dans la figure C.6 je montre dans chaque ligne les quatre différentes partie de cette séquence.

Cette séquence a 1119 image et a les mêmes caractéristiques que la séquence Dispute "A".

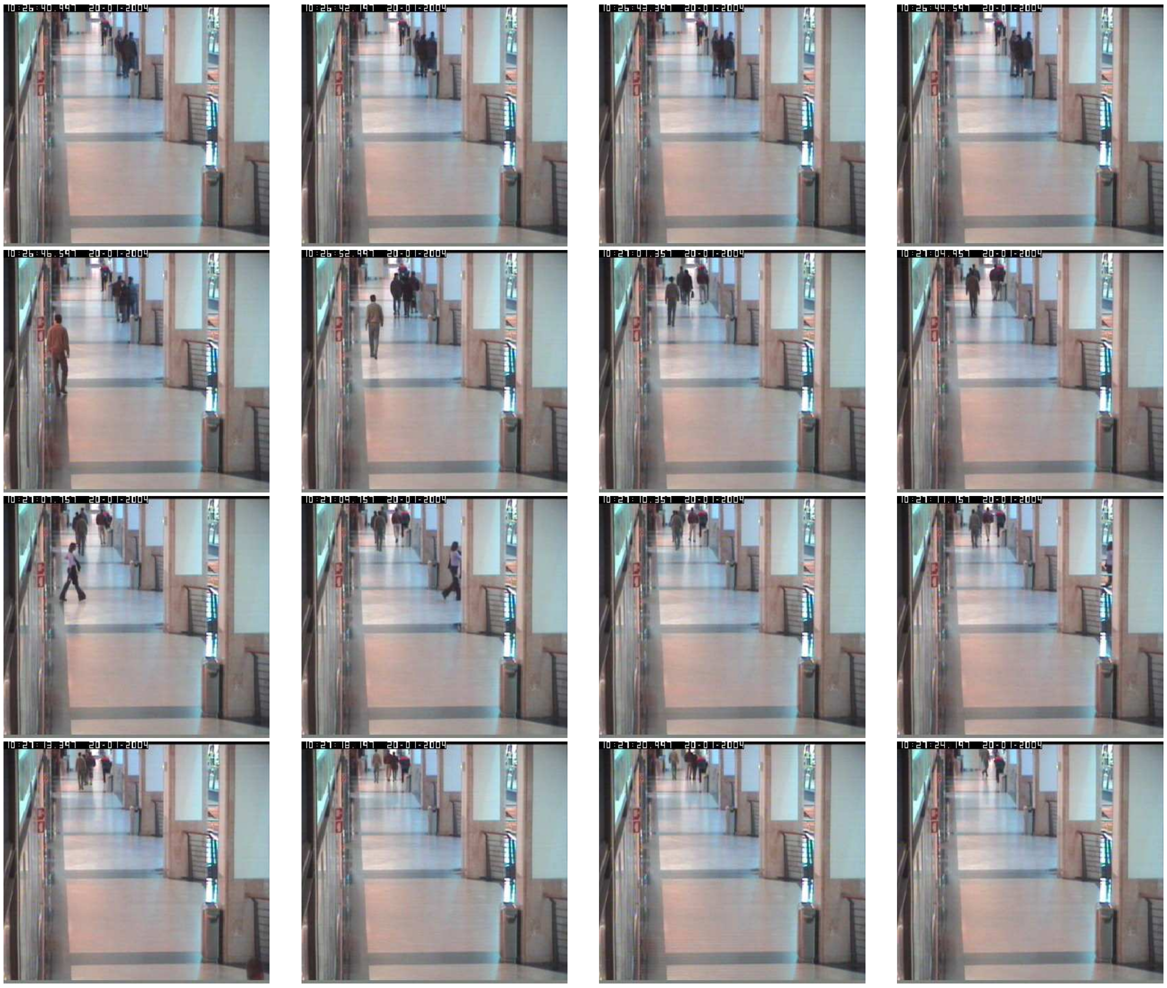


FIG. C.6 – Séquence d'images réelles, Couloir, issue de la caméra de surveillance d'un magasin.

Annexe D

Conventions et paramétrage

Je montre ici les conventions prises en charge pour mon modèle connexionniste et les valeurs de ses différents paramètres pour chaque module.

D.1 Pour le filtrage FSTC

En sachant que les colonnes d'orientation chez l'être humain ont un écart d'environ 10° entre deux colonnes consécutives (voir la section 1.3.1 de la première partie), j'ai choisi de m'approcher grossièrement à 45° en raison du lourd calcul qu'implique mon filtrage FSTC.

Ainsi, la détection de mouvement dans mon modèle est effectuée dans chacune des huit directions possibles. À chaque direction j'associe une couleur. Dans la figure D.1 nous pouvons voir le code de couleurs associé à chacune des huit directions de mouvement que nous pouvons percevoir. Ce codage a été utilisé dans tous les trois modules de mon modèle.

Un mouvement dans une direction donnée est représenté par une couleur particulière. L'intensité de la couleur dépend du contraste présent dans l'image initiale.

Les paramètres libres pour mon premier module, FSTC, ont été fixés en partie en accord avec les suggestions de [PP97, KP00].

Donc la taille des séquences d'images testées a été de $\tau = 3$, le nombre de vitesses de $\mathcal{V} = 5$ (positives et négatives comprises), le nombre de directions détectables $\mathcal{D} = 8$, et les paramètres que j'ai fixés pour l'équation 2.1 de la deuxième partie ont été : la fréquence spatiale de $\lambda = 0.03125 * Width$ pour un champ récepteur de $\xi_g = \eta_g = 0.56 * \lambda * Width$ avec $\gamma = 0.5$ comme excentricité et une phase de $\phi = 0^\circ$ où $Width$ est la largeur de l'image.

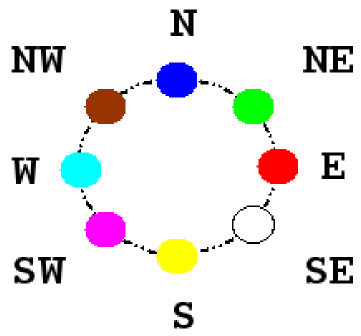


FIG. D.1 – Définition de couleurs pour chacune des huit différentes directions utilisées.

La taille de l'imagette de convolution a été fixée à $\xi_G = \eta_G = 3$ avec des pas $\xi_G^x, \eta_G^y \in \{1, 2, 3\}$.

En ce qui concerne le traitement pour rendre les résultats du filtrage indépendants de l'illumination j'ai suivi les suggestions de [PP97] et j'ai fixé $\mathcal{R} = 5$ et $\mathcal{P} = 3$.

Le choix de ces paramètres a été grandement influencé par les limitations d'implantation matérielles.

D.2 Pour le mécanisme MFIA

Dans V1, les fortes interactions inter- et intra-colonnes d'orientation ne sont pas complètement denses au sens strict. Les interactions sont plutôt situés dans un petit voisinage aussi bien intra-colonnes qu'inter-colonnes (voir §I.1.3.1).

Dans mon modèle j'ai choisi un rayon d'influence maximal de $\xi_E = \eta_E = \xi_I = \eta_I = 7$ qui définit les voisinages d'excitation et d'inhibition dans les sphères d'influence Θ_E et Θ_I avec une ouverture maximale de $\vartheta_e = \vartheta_i = 45^\circ$.

Les paramètres libres pour résoudre l'équation 1.7 de la deuxième partie ont été fixés en accord avec les suggestions de [Gro73, Gro88, Mog00], à savoir : $\mathcal{A} = \frac{2*max(\mathcal{B},\mathcal{C})}{4}$ où $\mathcal{B}$ et $\mathcal{C}$ sont un pourcentage de la valeur de saturation maximale dans la séquence $\mathcal{S}_\mathcal{T}$ issue du filtrage FSTC. Dans mon cas je les ai fixés avec des pourcentages de 95% et 95%, 75% et 50%, 50% et 75% pour $\mathcal{B}$ et $\mathcal{C}$, respectivement.

Le coefficient d'apprentissage η a pris des valeurs entre 0.0001 et 0.1.

D.3 Pour la focalisation de l'attention

Pour le troisième module (FA) j'ai fixé les paramètres libres tels qu'ils ont été rapportés par [RV05, VRA05], à savoir : $\kappa = 1$ et $\alpha = 13$ pour toutes les cartes.

La taille et l'énergie minimale détectable (baseline) pour les différentes cartes sont montrées dans le tableau D.1.

Carte	*Taille*	*Baseline*
Visuelle	40×40	0.00
Focus	40×40	-0.05
FEF	40×40	-0.20
MémTrav	40×40	0.00

TAB. D.1 – Tailles et énergies minimales détectables (Baseline) pour les différentes cartes du troisième module.

Les différentes pondérations de connexions inter- et intra-cartes sont montrées dans le tableau D.2 (ChR signifie champ récepteur).

Carte source	*Carte cible*	*Type*	*D*	*d*	*E*	*e*
Entrée	Visuelle	ChR	2.00	2.00	—	—
Visuelle	Visuelle	Latéral	2.50	2.00	1.00	4.00
Visuelle	Focus	ChR	0.25	2.00	—	—
Focus	Focus	Latéral	1.70	4.00	0.65	17.00
Visuelle	FEF	ChR	0.25	2.00	—	—
Focus	FEF	ChR	0.20	2.00	—	—
FEF	FEF	Latéral	2.50	2.00	1.00	4.00
FEF	MémTrav	ChR	2.35	1.50	—	—
MémTrav	FEF	ChR	2.40	1.50	—	—

TAB. D.2 – Table des différentes pondérations d'interaction inter- et intra-cartes issues du module de focalisation de l'attention.

Annexe E

Publications

Cette thèse a donné lieu aux publications et communications suivantes :

Revues internationales

- *Avec comité de lecture*
  - [THGCS05b] : César Torres Huitzil, Bernard Girau, and Claudio Castellanos Sánchez. *On-chip visual perception of motion : A bio-inspired connectionist model on FPGA*. Neural Networks 18(5-6) :557-565, 2005.
- *Sur sélection/invitation*
  - [CSGA04a] : Claudio Castellanos Sánchez, Bernard Girau, and Frédéric Alexandre. *Approche connexionniste de la perception du mouvement.* In Ph. Dumas and Henri Dou, editors, International Journal for Information Sciences for Decision Making (ISDM), Special Issue Colloque MajecSTIC'03, volume 13, pages 72-81, Toulon/Marseille - France, January 2004.
  - [CSAG03] : Claudio Castellanos Sánchez, Frédéric Alexandre, and Bernard Girau. *Towards bio-inspired neural networks for visual perception of motion.* European Research Consortium for Informatics and Mathematics (ERCIM News), 55 :10-11, October 2003.

Conférences internationales

- [CSG05] : Claudio Castellanos Sánchez and Bernard Girau. *Dynamic pursuit with a bio-inspired neural model.* In J. Blanc Talon, D. Popescu, W. Philips, and P. Scheunders, editors, Advanced Concepts for Intelligent Vision Systems, ACIVS 2005, LNCS 3708 :284-291, September 2005.
- [THGCS05a] : César Torres Huitzil, Bernard Girau, and Claudio Castellanos Sánchez. *Digital implementation of a bio-inspired neural model for motion estimation.* In International Joint Conference on Neural Networks 2005, August 2005.
- [CSGA04b] : Claudio Castellanos Sánchez, Bernard Girau, and Frédéric Alexandre. *A connectionist approach for visual perception of motion.* In Leslie Smith, Amir Hussain, and Igor Aleksander, editors, Brain Inspired Cognitive Systems (BICS 2004), pages BIS3-1 :1-7, September 2004.

Conférences nationales

- [CSGA03] : Claudio Castellanos Sánchez, Bernard Girau, and Frédéric Alexandre. *Approche connexionniste de la perception du mouvement.* In Philippe Bouche, Aguinaldo Garrido de Ceita, Gwenaëlle Grosset, Olivier Labarthe, and Julien Serment, editors, Manifestation des

Jeunes Chercheurs des Sciences et Technologies de l'Information et de la Communication, MajecSTIC'2003, pages 236-242, Domaine Universitaire de st. Jérôme, Avenue Escadrille Normandie Niémen, 13397 Marseille CEDEX 20, France, October 2003. Association des doctorants et docteurs de LIS et le Laboratoire de Sciences de l'Information et des Systèmes (LIS).

Bibliographie

[AA82] S. Amari and M. A. Arbib, editors. *Competition and cooperation in neural networks*. Springer-Verlag, 1982.

[AB85] E. H. Adelson and J. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(7) :284–299, 1985.

[AB86] E. H. Adelson and J. Bergen. The extraction of spatiotemporal energy in human and machine vision. In *Proc. IEEE Workshop on Visual Motion*, pages 151–156, Charleston, 1986.

[ADG84] T. D. Albright, R. Desimone, and C. G. Gross. Columnar organization of directionally selective cells in visual area mt of the macaque. *Journal of Neurophysiology*, 51 :16, 1984.

[Adr28] E. D. Adrian, editor. *The Basis of Sensation*. Christophers, London, 1928.

[AG00] D. Ascher and N. M. Grzywacz. A bayesien model for the measurement of visual velocity. *Vision Research*, 40 :3427–3434, 2000.

[Alb84] T. D. Albright. Direction and orientation selectivity of neurons in visual area mt of the macaque. *Journal of Neurophysiology*, 52 :1106–1130, 1984.

[Ale90] Frédéric Alexandre. *Une modélisation fonctionnelle du cortex : la Colonne Corticale. Aspects visuels et moteurs.* PhD thesis, Henri Poincaré, Nancy, France, 1990.

[AM82] E. Adelson and J. Movshon. Phenomenal coherence of moving visual patterns. *Nature*, 300(9) :523–525, 1982.

[Ama77] S.-I. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27 :77–87, 1977.

[Ama79] S.-I. Amari. Formation of topographic maps and columnar microstructures. *Biological Cybernetics*, 35 :63–72, 1979.

[Ana89] P Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal Computer Vision*, 2 :283–310, 1989.

[And97] R. A. Andersen. Neural mechanisms of visual motion perception in primates. *Neuron*, 18 :865–872, 1997.

[Ans70] S. Anstins. Phi mouvement as a substraction process. *Vision Research*, 20 :1411–1430, 1970.

[AR75] S. M. Anstis and B. J. Rogers. Illusory reservals of visual depth and movement during changes in contrast. *Vision Research*, 15 :957–961, 1975.

[AS99] K. C. Anderson and R. M. Siegel. Optic flow selectivity in the anterior superior temporal polysensory areas, stpa, of the behaving monkey. *The Journal of Neurosciences*, 19(7) :2681–2692, 1999.

[Aub86] H. Aubert. Die bewegungsempfindung. *Pflugers Arch. Ges. Physiol.*, 39 :347–370, 1886.

[BA98] Randolph Blake and Thomas Satoru Aiba. Detection and discrimination of optical flow components. *Japanese Psychological Research*, 40(1) :19–30, 1998.

[Bar53] H. B. Barlow. Summation and inhibition in the frog's retina. *Journal Physiology*, 119 :69–88, 1953.

[Bar81] H. B. Barlow. Critical factors limiting the design of the eye and visual cortex : The ferrier lecture. *Proceedings Royal society London B*, 212 :1–34, 1981.

[BB95a] S. S. Beauchemin and J. L. Barron. The computation of optical flow. Technical report, Department of Computer Science, June 1995.

[BB95b] S. S. Beauchemin and J. L. Barron. The structure of occlusion in fourier space. *Vision Inteface*, pages 112–119, 1995.

[BBL02] A. Bodnarova, M. Bennamoun, and S. Latham. Optimal gabor filters for textile flaw detection. *Pattern Recognition*, 35 :2973–2991, 2002.

[BFB94] J. J. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. In *Proceedings Conference on Computer Vision and Pattern Recognition*, pages 236–242, Champaigne, 1994.

[BH63] H. B. Barlow and R. M. Hill. Selective sensitivity to direction of motion in the ganglion cells of the rabbit's retina. *Science*, 139 :412–414, 1963.

[BH92] G. Beckers and V. Homberg. Cerebral visual motion blindness : Transitory akinetopsia induced by transcanial magnetic stimulation of human area v5. In *Proc. R. Soc. Lond.*, volume 249, pages 173–178, 1992.

[BHKG91] H. Barman, L. Haglund, H. Knutsson, and G. Granlund. Estimation of velocity, acceleration, and disparity in time sequences. In *Proceedings IEEE Workshop on Visual Motion*, pages 223–242, Princeton, 1991.

[Bla92] M. J. Black. *Robust incremental optical flow.* PhD thesis, Yale University, Yale, 1992.

[BMA+96] D. Bradley, M. Maxwell, R. A. Andersen, M. S. Banks, and K. Shenoy. Mechanisms of heading perception in primate visual cortex. *Science*, 273 :1544–1547, 1996.

[BMV98] D. C. Burr, M. C. Morrone, and L. M. Vaina. Large receptive fields for optic flow detection in humans. *Vision Research*, 38 :1731–1743, 1998.

[Bow01] L. Bowns. Ioc, vector sum, and squaring : three different motion effects or one ? *Vision Research*, 41(7) :965–972, 2001.

[BQ01] O. Braddick and N. Quian. The organization of global motion and transparency. In J. M. Zanker and J. Zeil, editors, *Motion Vision : Computational, Neural, and Ecological Constraints*, pages 85–112, Berlin, Heidelberg, Germany, 2001.

[BQA95] D. Bradley, N. Quian, and R. Andersen. Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature*, 373 :609–611, 1995.

[Bra65] R Bracewell. *The Fourier Transform and Its Applications.* New York : McGraw-Hill, 1965.

[Bra93] O. Braddick. Segmentation vs integration in visual motion processing. *Trends in Neuroscience*, 16 :263–268, 1993.

[Bra97] O. Braddick. Local and global representations of velocity : transparency, opponency and global direction perception. *Perception*, 26 :995–1010, 1997.

[Bro31] J. F. Brown. The visual perception of velocity. *Psychol. Forsh*, 14 :199–232, 1931.

[BS91] V. Braintenberg and A Schüz, editors. *Anatomy in the cortex.* Springer-Verlag, 1991.

[BSD01] P. J. Bex, A. J. Simmers, and S. C. Dakin. Snakes and ladders : teh role of temporal modulation in visual contour integration. *Vision Research*, 41 :3775–3782, 2001.

[BSNM93] K. H. Britten, M. N. Shadlen, W. T. Newsome, and W. T. Movshon. Responses of neurons in macaque mt to stochastic motion signals. *Visual Neuroscience*, 10 :1157–1169, 1993.

[BT80] S. T. Barnard and W. B. Thompson. Disparity analysis of images. In *IEEE Transaction PAMI*, pages 333–340, 1980.

[BT92] R. Born and R. Tootell. Segregation of global and local motion processing in macaque middle temporal cortex. *Nature*, 357 :497–499, 1992.

[BV98] Scott A. Beardsley and Lucia M. Vaina. Computational modeling of optical flow selectivity in mstd neurons. *Network Computation in Neural Systems*, 9 :467–493, 1998.

[BWV03] Scott A. Beardsley, Robert L. Ward, and Lucia M. Vaina. A neural network model of spiral-planar motion tuning in mstd. *Vision Research*, 43 :577–595, 2003.

[Cam97] T. Camus. Real-time quantized optical flow. *The Journal of Real-Time Imaging (special issue on Real-Time Motion Analysis)*, 3 :71–86, 1997.

[Can86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6) :679–698, 1986.

[CBC+98] J. C. Culhman, S. A. Brandt, P. Cavanagh, N. G. Kanwisher, A. M. Dale, and R. B. Tootell. Cortical fmri activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, 80 :2657–2670, 1998.

[CI03] C. W. G. Clifford and M. R. Ibbotson. Fundamental mechanisms of visual motion detection : models, cells and functions. *Progress in Neurobiology*, 68 :409–437, 2003.

[CM89] P. Cavanagh and G. Mather. Motion : the long and short of it. *Spatial Vision*, 4 :103–129, 1989.

[CR76] C. Cafforio and F. Rocca. Methods for measuring small displacements of television images. *IEEE Transactions on Information Theory*, IT-22 :573–579, 1976.

[CSAG03] Claudio Castellanos Sánchez, Frédéric Alexandre, and Bernard Girau. Towards bio-inspired neural networks for visual perception of motion. *European Research Consortium for Informatics and Mathematics (ERCIM News)*, 55 :10–11, October 2003.

[CSG05] Claudio Castellanos Sánchez and Bernard Girau. Dynamic pursuit with a bio-inspired neural model. In J. Blanc Talon, W. Philips, D. Popescu, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems, ACIVS 2005*, pages LNCS 3708 :284–291. Antwerp University, September 2005.

[CSGA03] Claudio Castellanos Sánchez, Bernard Girau, and Frédéric Alexandre. Approche connexionniste de la perception du mouvement. In Philippe Bouche, Aguinaldo Garrido de Ceita, Gwenaëlle Grosset, Olivier Labarthe, and Julien Serment, editors, *Manifestation des Jeunes Chercheurs des Sciences et Technologies de l'Information et de la Communication - MajecSTIC'2003*, pages 236–242, Domaine Universitaire de st. Jérôme, Avenue Escadrille Normandie Niémen, 13397 Marseille CEDEX 20, France, October 2003. Association des doctorants et docteurs de LIS et le Laboratoire de Sciences de l'Information et des Systèmes (LIS).

[CSGA04a] Claudio Castellanos Sánchez, Bernard Girau, and Frédéric Alexandre. Approche connexionniste de la perception du mouvement. In Ph. Dumas and Henri Dou, editors, *International Journal for Information Sciences for Decision Making (ISDM), Special Issue Colloque MajecSTIC'03*, volume 13, pages 72–81, Toulon/Marseille - France, January 2004.

[CSGA04b] Claudio Castellanos Sánchez, Bernard Girau, and Frédéric Alexandre. A connectionist approach for visual perception of motion. In Leslie Smith, Amir Hussain, and Igor Aleksander, editors, *Brain Inspired Cognitive Systems (BICS 2004)*, pages BIS3–1 :1–7, September 2004.

[DAN98] G. C. De Angelis and William T. Newsome. Cortical area mt and the perception of stereoscopic depth. *Nature*, 394 :677–680, 1998.

[Dau80] J. G. Daugman. Two-dimensional spectral analysis of cortical receptive fields profiles. *Vision Research*, 20 :847–856, 1980.

[Dau85] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2 :1160–1169, 1985.

[Dau88] John G. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7) :1169–1179, 1988.

[Dav03] Jules Davidoff. Color perception. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 210–215. Bradford Book MIT Press, 2003.

[DBH92] A. M. Derrington, D. R. Badcock, and S. A. Holroyd. Analysis of the motion of 2-dimensional patterns : evidence for a second-order process. *Vision Research*, 32(4) :699–707, 1992.

[DD03] John Daugman and Cathryn Downing. Gabor wavelets for statistical pattern recognition. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 414–420. Bradford Book MIT Press, 2003.

[DH93] A. M. Derrington and G. B. Henning. Detecting and discriminating the direction of motion of luminance and colour grattings. *Vision Research*, 33 :799–811, 1993.

[DH00] André De Hon. The density advantage of configurable computing. *IEEE Computer*, 33(4) :41–49, 2000.

[DKM+95] R. J. Douglas, C. Koch, M. Mahowald, K. A. C. Martin, and H. H. Suarez. Recurrent excitation in neocortical circuits. *Science*, 269 :981–985, 1995.

[DMM96] Rodney Douglas, Misha Mahowald, and Kevan Martin. Neuroinformatics as explanatory neuroscience. *Neuroimage*, 4 :S25–S28, 1996.

[DMW89] R. J. Douglas, K. A. C. Martin, and D. Whitterigde. A canonical microcircuit for neocortex. *Neural Computation*, 1 :480–488, 1989.

[DPW99] R. F. Dougherty, W. A. Press, and B. A. Wandell. Perceived speed of colored stimuli. *Neuron*, 24 :893–899, 1999.

[DVCM+00] R. L. De Valois, N. P. Cottaris, L. E. Mahon, S. D. Elfar, and J. A. Wilson. Spatial and temporal receptive fields in geniculate and cortical cells and directional selectivity. *Vision Research*, 40 :3685–3702, 2000.

[DVK91] Karen K. De Valois and Frank Kooi. Fonctional classification of parallel visual pathways. In Arne Valberg and Barry B. Lee, editors, *From Pigments to Perception : Advances in Understanding Visual Processes*, pages 165–171, 233 Spring Street, New York, N. Y., 10013, 1991. Plenum Press and NATO ASI Series.

[DW88] M. R. Dürsteler and R. H. Wurtz. Pursuit and optokinetic deficits following chemical lesions of cortical areas mt and mst. *Journal of Neurophysiology*, 60 :940–965, 1988.

[DW91a] C. Duffy and R. Wurtz. Sensitivity of mst neurons to optic flow. i. a continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6) :1329–1345, 1991.

[DW91b] C. Duffy and R. Wurtz. Sensitivity of mst neurons to optic flow. ii. mehcanisms of response selectivity revealed by small field stimuli. *Journal of Neurophysiology*, 65(6) :1346–1359, 1991.

[DW95] C. Duffy and R. Wurtz. Response of monkey mst neurons to optic flow stimuli with shifted centers of motion. *The Journal of Neurosciences*, 15 :5192–5208, 1995.

[DW97a] C. Duffy and R. Wurtz. Medial superior temporal area neurons respond to speed patterns in optical flow. *The Journal of Neurosciences*, 17(8) :2839–2851, 1997.

[DW97b] C. Duffy and R. Wurtz. Planar directional contributions to optical flow responses in mst neurons. *Journal of Neurophysiology*, 77 :782–796, 1997.

[EA02] Emad N. Eskandar and John A. Assad. Distinct nature of directional signals among parietal cortical areas during visual guidance. *Journal of Neurophysiology*, 88 :1777–1790, 2002.

[EBA92] R. C. Emerson, J. R. Bergen, and E. H. Adelson. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, 32 :203–218, 1992.

[ECVK86] R. Emerson, M. Citron, W. Vaughn, and S. Klein. Substructure in directionally selective complexe receptive fields of cat. *Investigative Ophtalmology and Visual Science Supplement*, 27 :16, 1986.

[ET91] R. G. Erickon and P. Thier. A neuronal correlate of spatial stability during periods of self-induced visual motion. *Experimental Brain Research*, 86 :608–616, 1991.

[Exn88a] S. Exner. Einige beobachtungen uber bewegungsnachbilder. *Centr. Physiol.*, 1 :135–140, 1888.

[Exn88b] S. Exner. Über optische bewegungsempfindungen. *Biol. Wentralbl.*, 8 :437, 1888.

[Exn94] S. Exner. Entwuft zu einer phsysiologischen erklärung der psychischen erscheinungen. *I. Teil. Deuticke, Leipzig*, pages 37–140, 1894.

[FB99] Hervé Frezza-Buet. *Un modèle du cortex pour le comportement motivé d'un agent neuromimétique autonome.* PhD thesis, Henri Poincaré, Nancy, France, 1999.

[FC75] J. Feldman and J. D. Cowan. Large-scale activity in neural nets i. theory with application to motoneuron pool responses. *Biological Cybernetics*, 17 :29–38, 1975.

[FCMMDFG03] Antonio Fernández Caballero, José Mira Mira, Ana E. Delgado, and Miguel A. Fernández Graciani. Lateral interaction in accumulative computation : a model for motion detection. *Neurocomputing*, 50 :341–364, 2003.

[FCMMFGD03] Antonio Fernández Caballero, José Mira Mira, Miguel Ángel Fernández Graciani, and Ana E. Delgado. On motion detection through a multi-layer neural network architecture. *Neural Networks*, 16 :205–222, 2003.

[FHH93] David J. Field, Anthony Hayes, and Robert F. Hess. Contour integration by the human visual system : evidence for a local association field. *Vision Research*, 33(2) :173–194, 1993.

[FJ90] D. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal Computer Vision*, 5 :77–104, 1990.

[FK84] D. J. Fellemann and J. H. Kaas. Receptive-fields properties of neurons in middle temporal visual area (mt) of the owl monkeys. *Journal of Neurophysiology*, 52 :487–513, 1984.

[FL95] D. J. Fleet and K. Langley. Computational analysis of non-fourier motion. *Vision Research*, 34(22) :3057–3079, 1995.

[FT79] C. I. Fennema and W. B. Thompson. Velocity determination via scenes containing several moving images. *Computer Graphics and Image Processing*, 9 :301–315, 1979.

[FT02] Winfried A. Fellez and John G. Taylor. Establishing retinotopy by lateral-inhibition type homogeneous neural fields. *Neurocomputing*, 48 :313–322, 2002.

[FWQ02] Julián Martín Fernández, Brendon Watson, and Ning Qian. Computing relief structure form motion with a distributed velocity and disparity representation. *Vision Research*, 42 :883–898, 2002.

[Gab46] D. Gabor. Theory of communications. *J. Inst. of Elect. Eng.*, 93 :429–457, 1946.

[GAS94] M. S. A. Graziano, R. A. Andersen, and R. J. Snowden. Tuning of mst neurons to spiral motions. *The Journal of Neurosciences*, 14 :54–67, 1994.

[GG84] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *PAMI*, 6(6) :721–741, 1984.

[GHA94] Norberto M. Grzywacz, Julie M. Harris, and Franklin R. Amthor. Computational and neural constraints for the measurement of local visual motion. In Andrew T. Smith and Robert J. Snowden, editors, *Visual Detection of Motion*, pages 19–50. Academic Press Limited, 1994.

[Gib50] J. J. Gibson, editor. *The Perception of the Visual World.* Houghton Mifflin, Boston, MA, 1950.

[Gil85] C. Gilbert. Horizontal integration in the neocortex. *Trends in Neuroscience*, 8 :160–165, 1985.

[Gir00] B. Girau. Fpna : Interaction between fpga and neural computation. *International Journal of Neural Systems*, 10(3) :243, 2000.

[GKL92] D. Goldreich, R. J. Krauzlis, and S. G. Lisberg. Effect of changing feedback delay on spantaneous oscillations in smooth pursuit eye mouvement of monkeys. *The Journal of Neurosciences*, 67 :625–638, 1992.

[GL75] S. Grossberg and D. S. Levine. Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. *Journal of Theoretical Biology*, 53 :341–380, 1975.

[GLP+89] A. P. Georgopoulos, J. T. Lurito, M. Petrides, A. B. Schwarttz, and J. T. Massey. Mental rotation of the neuronal population vector. *Science*, 243 :234, 1989.

[GM03] Norberto M. Grzywacz and David K. Merwine. Directional selectivity. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 353–358. Cambridge, Massachusetts 02142, USA, 2003.

[GMN+99] B. Galvin, B. McCane, K. Novins, D. Mason, and S. Mills. Recovering motion fields : An evaluation of eight optical flow algoritms. Technical report, Computer Science Department, 1999.

[GMP99] S. Grossberg, E. Mingolla, and C. Pack. A neural model of motion processing and visual navigation by cortical area mst. *Cerebral Cortex*, 9 :878–895, 1999.

[GN03] Norberto M. Grzywacz and Anthony M. Norcia. Directional selectivity in the cortex. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 309–311. Bradford Book MIT Press, 2003.

[Gro73] Stephen Grossberg. Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52 :217–257, 1973.

[Gro76] S. Grossberg. On the development of feature detectors in the visual cortex with application to learning and reaction-diffusion systems. *Biological Cybernetics*, 21 :145–158, 1976.

[Gro78a] Stephen Grossberg. Competition, decision, and concensus. *Journal of Mathematical Analysis and Applications*, 66 :470–493, 1978.

[Gro78b] Stephen Grossberg. A theory of human memory : Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell, editors, *Progress in theoretical biology*, pages 233–374, New York, 1978. Academic Press.

[Gro80] Stephen Grossberg. Biological competition : Decision rules, pattern formation, and oscillations. *Proceedings of the National Academiy of Sciences*, 77 :2338–2343, 1980.

[Gro82] Stephen Grossberg, editor. *Studies of mind and brain : Neural principles of learning, perception, development, cognition, and motor control.* Reidel Press, 1982.

[Gro88] Stephen Grossberg. Nonlinear neural networks : Principles, mechanisms, and architectures. *Neural Networks*, 1 :17–61, 1988.

[GSA03] Norberto M. Grzywacz, Evelyne Sernagor, and Franklin R. Amthor. Directional selectivity in the retina. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 816–840. Bradford Book MIT Press, 2003.

[GSK86] A. P. Georgopoulos, A. B. Schwarttz, and R. E. Kettner. Neural population coding of movement direction. *Science*, 233 :1416–1419, 1986.

[GTL93] A. P. Georgopoulos, M. Taira, and A. Lukashin. Cognitive neurophysiology of the motor cortex. *Science*, 260 :47–52, 1993.

[GWM95] N. M. Grzywacz, S. N. J. Watamaniuk, and S. P. McKee. Temporal coherence theory for the detection and measurement of visual motion. *Vision Research*, 35 :3183–3203, 1995.

[GY90] N. M. Grzywacz and A. L. Yuille. A model for the estimate of local image velocity by cells in the visual cortex. In *Proceedings Royal society London B*, volume 239, pages 129–161, London, 1990.

[H0́1] Jeanny Hérault. De la rétine biologique aux circuits neuromorphiques. In Jean-Michel Jolion, editor, *Les systèmes de vision*, pages 55–96, Paris, 2001. Hermes Science Europe Ltd.

[Ham03] Dan Hammerstron. Digital vlsi neural networks. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 349–352. Cambridge, Massachusetts 02142, USA, 2003.

[Har38] H. K. Hartline. The response of single optic nerve fibres of the vertebrate eye to illumination of retina. *Am. J. Physiol.*, 121 :400–415, 1938.

[HB93] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12) :1217–1232, 1993.

[HBD74] G. H. Henry, P. O. Bishop, and B. Dreher. Orientation, axis and direction as stimulus parameters for striate cells. *Vision Research*, 14 :767–777, 1974.

[HBD+99] D. J. Heeger, G. M. Boynton, J. B. Demb, E. Seidemann, and W. T. Newsome. Motion opponency in visual cortex. *The Journal of Neurosciences*, 19 :7162–7174, 1999.

[Hee87] David J. Heeger. Model for the extraction of the image flow. *Journal of the Optical Society of America A*, 4(8) :1455–1471, 1987.

[Hee88] D. J. Heeger. Optical flow using spatiotemporal filters. In *International Journal Computer Vision*, pages 279–302, 1988.

[Hee92a] D. J. Heeger. Half-squaring in responses of cat simple cells. *Visual Neuroscience*, 9 :427–443, 1992.

[Hee92b] D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9 :181–198, 1992.

[Hee93] D. J. Heeger. Modeling simple-cell direction selectivity with normalized, half-square, linear operators. *Journal of Neurophysiology*, 70 :1885–1898, 1993.

[Hei95] S. M. Z. Heindenreich. Evidence that luminant and equiluminant motion signals are integrated by directionally selective mechanims. *Perception*, 24 :879–890, 1995.

[HJ92] D. H. Heeger and A. D. Jepson. Subspace methods for recovering rigid motion i : Algorithm and implementation. *International Journal Computer Vision*, 7 :95–117, 1992.

[HL03] Robert F. Hess and Timothy Ledgeway. The dectection of direction-defined and speed-defined spatial contours : one mechanism or two ? *Vision Research*, 43 :597–606, 2003.

[HM94] R. S. Hubbart and J. A. Marshall. Self-organizing neural network model of the visual inertia phenomenon in motion perception. Technical Report TR-94-001, University of North Carolina, 1994.

[HN93] Makoto Hirahara and Takashi Nagano. A neural network model for visual motion detection that can explain psychophysical and neurophysiological phenomena. *Biological Cybernetics*, 68(3) :247–252, 1993.

[Hod64] A. L. Hodgkin, editor. *The conduction of the nervous impulse*. Liverpool University Press, Liverpool, 1964.

[HR56] B. Hassenstein and W. Reichardt. Systemtheoretische analyse der zeitreihenfolgen und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers chlorophanus. *Zeitschrift für Naturforschung*, 11b :513–524, 1956.

[HR81] P. Hammond and J. Reck. Influence of velocity on directional tuning of complex cells in cat striate cortex for texture motion. *Neuroscience Letters*, 19 :309–314, 1981.

[HR03] Ellen C. Hildreth and Constance S. Royden. Motion perception. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 585–588. Bradford Book MIT Press, 2003.

[HRH01] Alexander C. Huk, David Ress, and David J. Heeger. Neural basis of the motion aftereffect reconsidered. *Neuron*, 32 :161–172, 2001.

[HS80] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. Technical Report 572, Artificial Intelligence Laboratory, April 1980.

[HS93] Berthold K. P. Horn and Brian G. Schunck. “determining optical flow" : a retrospective. *Artificial Intelligence*, 59 :81–87, 1993.

[HW59] D. H. Hubel and T. N. Weisel. Receptive fields of single neurons in the cat's striate cortex. *Journal Physiology*, 148 :574–591, 1959.

[HW62] D. H. Hubel and T. N. Weisel. Receptive fields, binocular interaction and functional architecture in the cats visual cortex. *Journal Physiology*, 160 :106–154, 1962.

[HW68] D. H. Hubel and T. N. Weisel. Receptive fields and functional architecture of monkey striate cortex. *Journal Physiology*, 195 :215–243, 1968.

[HW72] D. H. Hubel and T. N. Weisel. Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. *The Journal of Comparative Neurology*, 146 :421–450, 1972.

[IA02] Mahamadou Idrissa and Marc Archeroy. Texture classification using gabor filters. *Pattern Recognition Letters*, 23 :1095–1102, 2002.

[IC01a] M. R. Ibbotson and C. W. G. Clifford. Characterising the temporal delay filters of biological motion detectors. *Vision Research*, 41 :2311–2323, 2001.

[IC01b] M. R. Ibbotson and C. W. G. Clifford. Interaction between on and off signals in direction motion detectors feeding the nucleus of optic track of the wallaby. *Journal of Neurophysiology*, 86 :997–1005, 2001.

[Imb83] Michel Imbert. Neurobiologie de l'image. *La Recherche*, 144 :600–613, 1983.

[IT03] U. J. Ilg and P. Thier. Visual tracking neurons in primate area mst are activated by smooth-poursuit eyes mouvements of an "imaginary" target. *Journal of Neurophysiology*, 90 :1489–1502, 2003.

[Jac04] Florence Jacquey. Parallèlisme connexionniste et perception visuelle du mouvement. LORIA-CORTEX Stage de DEA, Nancy, France, September 2004.

[JC95] A. Johnston and C. W. G. Clifford. A unified account of three apparent motion illusions. *Vision Research*, 35 :1109–1123, 1995.

[JD03] Robert A. Jacobs and Melissa Dominguez. Visual development and the acquisition of motion velocity sensitivities. *Neural Computation*, 15(4) :761–781, 2003.

[JF91] A. K. Jain and F. Farrokhmia. Unsupervised texture segmentation usign gabor filters. *Pattern Recognition*, 24(12) :1167–1186, 1991.

[JP87] J. Jones and L. Palmer. An evaluation on the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6) :1233–1258, 1987.

[JWF93] B. Jagadeesh, H. S. Wheat, and D. Ferster. Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. *Science*, 262 :1901–1904, 1993.

[JWK+97] B. Jagadeesh, H. S. Wheat, L. L. Kontsevich, C. W. Tyler, and D. Ferster. Direction selectivity of synaptic potentials in simple cells of the cat visual cortex. *Journal of Neurophysiology*, 78 :2772–2789, 1997.

[KB91] J. J. Kulikowki and P. O. Bishop. Fourier analysis and spatial representation in the visual cortex. *Experientia*, 37 :160–163, 1991.

[KD88] J. Konrad and E. Dubois. Multigrid bayesian estimation of image motion fields using stochastic relaxation. In *Proceedings of 2nd International Conference on Computer Vision*, pages 354–362, 1988.

[KK00] Z. Kourtzi and N. Kanwisher. Activation in human mt/mst by static images with implied motion. *Journal Cognitive Neuroscience*, 12 :48, 2000.

[KLP+01] Roland Kern, Maik Lutterkas, Christian Petereit, Jens P. Lindemann, and Martin Egelhaaf. Neural processing of behaviourally generated optic flow : experiments and models simulations. *Network Computation in Neural Systems*, 12 :351–369, 2001.

[KMS03] Christof Koch, Chun-Hui Mo, and William Softky. Single-cell models. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 1044–1049. Cambridge, Massachusetts 02142, USA, 2003.

[Koe86] J. J. Koenderink. Optical flow. *Vision Research*, 26 :161–180, 1986.

[Koe96] Étienne Koechlin. *Représentation et processus congitifs dans le cortex cérébral : Décision et dynamique d'activation bayesiennes dans les populations de neurones corticaux*. PhD thesis, École des Hautes Études en Sciences Sociales, November 1996.

[Koh84] T. Kohonen, editor. *Self-Organizing and Associative Memory.* Springer, Berlin, 1984.

[Koh97] T. Kohonen, editor. *Self-Organizing Maps.* Springer, Berlin, second edition, 1997.

[KP00] P. Kruizinga and N. Petkov. Computational model of dot-pattern selective cells. *Biological Cybernetics*, 83 :313–325, 2000.

[KSWY94] K. Kawano, M. Shidara, Y. Watanabe, and S. Yamane. Neural activity in cortical area mst of alert monkey during ocular following responses. *Journal of Neurophysiology*, 71 :2305–2324, 1994.

[KT87] J. K. Kearney and W. B. Thompson. An error analysis of gradient-based methods for optical flow estimation. In *IEEE Transaction PAMI*, pages 229–244, 1987.

[Kuf53] S. W. Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16 :37–68, 1953.

[LH02] Timothy Ledgeway and Robert F. Hess. Rules for combining the outputs of local motion detectors to define simple contours. *Vision Research*, 42(5) :653–660, 2002.

[LK81] B. Lucas and T. Kanade. An iterative image registration technique with an applications to stereo vision. *DARPA IU Workshop*, pages 121–130, 1981.

[LM75] J. Limb and J. Murphy. Estimating the velocity of moving images in television signals. *Computer Graphics and Image Processing*, 4 :311–327, 1975.

[LN03] Jing Liu and William T. Newsome. Funtional organization of speed tuned neurones in visual area mt. *Journal of Neurophysiology*, 89 :246–256, 2003.

[LN04] Peter E. Latham and Sheila Nirenberg. Computing and stability in cortical networks. *Neural Computation*, pages 1385–1412, 2004.

[LPB01] M. S. Livingstone, C. C. Pack, and R. T. Born. Two-dimensional substructure of mt receptive fields. *Neuron*, 30 :781–793, 2001.

[LRO93] L. Lagae, S. Raiguel, and C. A. Orban. Speed and direction selectivity of macaque middle temporal neurons. *Journal of Neurophysiology*, 69 :19–39, 1993.

[MAGN85a] J. Movshon, E. Adelson, M. Gizzi, and W. Newsome. The analysis of moving visual patterns. *Experimental Brain Research*, 11 :117–151, 1985.

[MAGN85b] J. A. Movshon, E. H. Adelson, M. S. Gizzi, and W. T. Newsome. The analysis of moving visual patterns. In C. Chagas, R. Gattas, and C. Gross, editors, *Pattern Recognition mechanisms*, pages 117–151, New-York, 1985. Springer-Verlag.

[MAH94] J. A. Marshall, R. K. Alley, and R. S. Hubbart. Learning to represent visual depth from occlusion events. In Arbib reference, 1994.

[Mar80] S. Marcelja. Mathematical description of the reponses of simple cortical cells. *Journal of the Optical Society of America A*, 70(11) :1297–1300, 1980.

[Mar82] D. Marr, editor. *Vision.* Freeman, San Francisco, CA, 1982.

[Mar03] Jonathan A. Marshall. Motion perception : Self-organization. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 389–391. Bradford Book MIT Press, 2003.

[MBCS97] Risto Miikkulainen, James A. Bednar, Yoonsuck Choe, and Joseph Sirosh. Self-organization, plasticity, and low-level visual phenomena in a laterally connected map model of the primary visual cortex. In R. L. Goldstone, P. G. Schyns, and D. L. Medin, editors, *Psychology of Learning and Motivation*, volume 36, pages 257–208, San Diego, CA, 1997. Academic Press.

[Mer91] William H. Merigan. P and m pathway specialization in the macaque. In Arne Valberg and Barry B. Lee, editors, *From Pigments to Perception : Advances in Understanding Visual Processes*, pages 117–125, 233 Spring Street, New York, N. Y., 10013, 1991. Plenum Press and NATO ASI Series.

[MF90] V. Markandey and B. E. Flinchbaugh. Multispectral constraints for optical flow computation. In *Proc. of ICCV*, pages 38–41, 1990.

[MG94] R. Muller and M. W. Greenlee. Effect of contrast and adaptation on the perception of the direction and speed of drifting gratings. *Vision Research*, 34 :2071–2092, 1994.

[Min03] Ennio Mingolla. Neural models of motion integration and segmentation. *Neural Networks*, 16 :939–945, 2003.

[MN87] J. H. R. Maunsell and W. T. Newsome. Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10 :363–401, 1987.

[MN96] J. A. Movshon and W. T. Newsome. Visual response properties of striate cortical neurons projecting to area mt in macaque monkeys. *The Journal of Neurosciences*, 23 :7733, 1996.

[MNGG01] B. McCane, K. Novins, D. Grannitch, and B. Galvin. On benchmarking optical flow. *Computer Vision and Image Undestanding*, pages 126–143, 2001.

[MNW83] A. Mikami, W. T. Newsome, and R. H. Wurtz. Mechanisms of direction and speed selectivity in the middle temporal visual area (mt) of the macaque monkey. *Investigative Ophtalmology and Visual Science Supplement*, 24 :107, 1983.

[MNW86] A. Mikami, W. T. Newsome, and R. H. Wurtz. Motion selectivity in macaque visual cortex i : Mechanisms of direction and speed selectivity in extrastriate area mt. *Journal of Neurophysiology*, 55 :1038–1327, 1986.

[Mog00] Sorin Moga. *Apprendre par imitation : une nouvelle voie d'apprentissage pour les robots autonomes*. PhD thesis, Université de Cergy-Pontoise, Cergy-Pontoise, France, September 2000.

[Mou79] V. B. Mouncastle. An organizing principle for cerebral fonction : The unit module and the distributed system. In F. O. Schmitt and F. G. Worden, editors, *The neuroscience fourth study program*, pages 1115–1139, Cambridge, 1979. MIT Press.

[MPJP85] J. McLean-Palmer, J. Jones, and L. Palmer. New degrees of freedom in the structure of simple receptive fields. *Investigative Ophtalmology and Visual Science Supplement*, 26 :265, 1985.

[MSN86] S. P. McKee, G. H. Silverman, and K. Nakayama. Precise velocity discimination despite random variation in temporal frequency and contrast. *Vision Research*, 26 :609–619, 1986.

[MSWM93] M. Miyauchi, M. Seki, A. Watanabe, and A. Miyauchi. Interpretation of optical flow through complex neural network. *New Trends in Neural Computation*, pages 645–650, 1993.

[MU81] D. Marr and S. Ullman. Directional selectivity and its use in early visual processing. In *Proc. R. Soc. Lond.*, volume B211, pages 151–180, 1981.

[Muk93] N. Mukawa. Estimation of shape, reflection coefficients and illuminant direction from image sequences. In *ICVV*, pages –, 1993.

[MVE83a] J. H. R. Maunsell and D. C. Van Essen. The connections of the middle temporal visual area (mt) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neurosciences*, 3 :2563–2586, 1983.

[MVE83b] J. H. R. Maunsell and D. C. Van Essen. Functional properties of neurones in middle temporal visual area of the macaque monkey. i. selectivity for stimulus direction, speed and orientation. *Journal of Neurophysiology*, 49 :1127–1147, 1983.

[MVM+02] J. A. Mazer, W. E. Vinje, J. McDermott, P. H. Schiller, and J. L. Gallant. Spatial frequency and orientation tuning dynamics in area v1r. *Proccedings National Academy of Sciences*, 99 :1645–1650, 2002.

[Nag87] H. H. Nagel. On the estimation of optical flow : Relation between different approaches and some new results. *Artificial Inteligence*, 33 :299–324, 1987.

[Nag89] H. H. Nagel. On a contraint equation for the estimation of displacement rates in image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1) :13–30, 1989.

[Nak85] K. Nakayama. Biological image motion processing : a review. *Vision Research*, 25 :625–660, 1985.

[NGM83] W. T. Newsome, M. S. Gizzi, and J. A. Movshon. Spatial and temporal properties of neurons in macaque mt. *Investigative Ophtalmology and Visual Science Supplement*, 24 :106, 1983.

[NHU04] Takashi Nagano, Makoto Hirahara, and Wakako Urushihara. A general model for visual motion detection. *Biological Cybernetics*, 91 :99–103, 2004.

[NJ99] S. Nishida and A. Johnston. Influence of motion signals on the perceived position of spatial pattern. *Nature*, 397 :610–612, 1999.

[NS94a] S. Nishida and T. Sato. Motion detection mechanism in the human visual system. *J. Ins. TV Eng.*, 48 :157–163, 1994.

[NS94b] S. J. Nowlan and T. J. Sejnowski. Filter selection model for motion segmentation and velocity integration. *Journal of the Optical Society of America A*, 11 :3177–3200, 1994.

[NS95] S. J. Nowlan and T. J. Sejnowski. A selection model for motion processing in area mt of primates. *The Journal of Neurosciences*, 15 :1195–1214, 1995.

[NT91] R. Navarro and A. Tabernero. Gaussian wavelet transform : two alternative fast implementations for images. *Multidimensional Systems and Signal Processing*, 2(4) :421–436, 1991.

[NWK88] W. T. Newsome, R. H. Wurtz, and H. Komatsu. Relation of cortical areas mt and mst to pursuit eyes mouvement. ii. differentiation of retinal from extraretinal inputs. *Journal of Neurophysiology*, 60 :604–620, 1988.

[OK03] Bruno A. Olshausen and Christof Koch. Selective visual attention. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 837–840. Bradford Book MIT Press, 2003.

[O'R90] J. K. O'Reagan. Les "vrais" mystères de la vision. *Non publié*, pages 1–12, 1990.

[Orb97] G. A. Orban. Visual processing in macaque area mt/v5 and its satellites (mstd and mstv). In K. S. Rockland, J. H. Kaas, and A. Peters, editors, *Cerebral Cortex*, pages 359–434, New York, 1997. Plenum.

[OS92] M. Ogata and T. Sato. Motion-detection model with two stages : Spatiotemporal filtering and feature matching. *Journal of the Optical Society of America A*, 9(3) :377–387, 1992.

[PA83] J. M. Prager and M. A. Arbib. Computing the optical flow : The match algorithm and prediction. *Computer Vision, Graphics and Image Processing*, 24 :271–304, 1983.

[PB01] C. C. Pack and R. T. Born. Temporal dynamics of a neural solution to the aperture problem in visual area mt of macaque brain. *Nature*, 409 :1040–1042, 2001.

[Per92] J. A. Perrone. A model of self-motion estimation within primate extrastriate visual cortex. *Journal of the Optical Society of America A*, 9 :177–194, 1992.

[PGM00] Christopher Pack, Stephen Grossberg, and Ennio Mingolla. A neural model of smooth pursuit control and motion perception by cortical area mst. Technical Report CAS/CNR-TR-99-023, Department of Cognitive and Neural Systems and Center for Adaptive Systems, 677 Beacon St, Boston, MA 02215, September 2000.

[PGP00] A. Pantle, D. P. Gallogly, and O. C. Piehler. Direction biasing by brief apparent motion stimuli. *Vision Research*, 40 :1979–1991, 2000.

[Pir98] Patrick Pirim. Procedé et dispositif fonctionnant en temps réel pour le repérage et la localisation d'une zone en mouvement relatif dans une scène, ainsi que pour la détermination de la vitesse et de la direction du déplacement. Technical Report WO 98/05002, Carlos Magnus Limited, France, February 1998.

[PLW01] Alvaro Pascual Leone and Vincent Walsh. Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292 :510–512, 2001.

[PM98] C. Pack and E. Mingolla. Global induced motion and visual stability in an optic flow illusion. *Vision Research*, 38 :3083–3093, 1998.

[PN98] A. J. Parker and W. T. Newsome. Sense and the single neuron : Probing the pshysiology of perception. *Annual Review of Neuroscience*, 21 :227–277, 1998.

[PP97] N. Petkov and P.Kruizinga. Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli : bar and grating cells. *Biological Cybernetics*, 76 :83–96, 1997.

[PR76] T. Poggio and W. E. Reichardt. Visual control of orientation behavior in the fly : Part iii : Towards the underlying neural interactions. *Q. Rev. Biophys.*, 9 :377–438, 1976.

[PR83] D. Pollen and S. Ronner. Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on Systems, Man, and Cybernetics*, 13 :906–916, 1983.

[PS94] J. A. Perrone and L. S. Stone. A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research*, 34 :2917–2938, 1994.

[PS00] R. E. Phinney and R. M. Siegel. Speed selectivity for optical flow in area 7a of the behaving macaque. *Cerebral Cortex*, 10 :413–421, 2000.

[PT01] J. Perrone and A. Thiele. Speed skills : measuring the visual speed analyzing properties of primate mt neurones. *Nature Neuroscience*, 4(5) :526–532, 2001.

[PvGPO94] M. Proesmans, L. van Gool, E. Pauwels, and A. Oosterlinck. Determination of optical flow and its discontinuities using non-linear diffusion. In *3er European Conference on Computer Vision, ECCV'94*, volume 2, pages 295–304, 1994.

[QAA94] N. Quian, R. A. Andersen, and E. H. Adelson. Transparent motion perception as detection of unbalanced motion signals. ii. psysiology. *The Journal of Neurosciences*, 14 :7367–7380, 1994.

[RB94] R. R. Ribchester and J. A. Barry. Spatial versus consumptive competition at polyneuronally innervated neuromuscular junctions. *Experimental of Physiology*, 79 :465–494, 1994.

[RDSW92] J. A. Reggia, C. D'Autrechy, C. G. Sutton, and M. A. Weinrich. Competitive distribution theory of neocortical dynamics. *Neural Computation*, 4 :287–317, 1992.

[Ree01] Geraint Rees. Attention, adaptation, and the motion aftereffect. *Neuron*, 32 :6–8, 2001.

[Rei61] W. Reichardt. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith, editor, *Sensory Communication*, pages 303–317, New York, 1961. Wesley.

[Rei69] W. Reichardt. Movement perception in insects. In W. Reichardt, editor, *Processing of optical data by organisms and machines*, pages 465–493, London, New York, 1969. Academy.

[Rib92] R. R. Ribchester. Cartels, competition and activity-dependent synapse elimination. *Trends in Neuroscience*, 15 :389, 1992.

[Rod98] R. W. Rodieck, editor. *The first steps in seeing.* Sinauer, Sunderland, MA, 1998.

[RSD+00] J. A. Reggia, C. G. Sutton, C. D'Autrechy, S. Cho, and S. L. Armentrout. Cortical inhibition as explained by competitive distribution hypothesis. In M. D. Fraser, editor, *Netwok Models for Control and Processing*, pages 31–62, Portland, 2000. IntellectTM.

[RV03] N. Rishikesh and Y. V. Venkatesh. A computational model for the develepment of simple-cell receptive fields spanning the regimes before and after eye-opening. *Neurocomputing*, 50 :125–158, 2003.

[RV05] N. Rougier and J. Vitay. Emergence of attention within a neural population. *Accepted to Neural Networks*, 2005.

[RyC94] Santiago Ramón y Cajal. The croonian lecture : La fine structure des centrs nerveux. *Proceedings Royal society London*, 55 :–, 1894.

[RZZ+03] Hengyi Rao, Tiangang Zhou, Yan Zhuo, Silu Fan, and Lin Chen. Spatiotemporal activation of the two visual pathways in form discrimination and spatial location : A brain mapping study. *Human Brain Mapping*, 18 :79–89, 2003.

[SBG+00] C. Senior, J. Barnes, V. Giampietro, M. J. Brammer, E. T. Bullmore, A. Simmons, and A. S. David. The functional neuroanatomy of implicit motion perception or "representational momentum". *Curr. Biol.*, 10 :16, 2000.

[Sch85] B. G. Schunck. Image flow : Fundamentals and future research. In *IEEE CVPR*, pages 560–571, San Francisco, California, 1985.

[SEC84] R. Shapley and C. Enroth-Cugell. Visual adaptation and retinal gain control. *Progess in Retinal Research*, 3 :263–346, 1984.

[Ser93] Margaret Euphrasia Sereno, editor. *Neural Computation of Pattern Motion : Modeling stages of motion analysis in the primate visual cortex*. Bardford Books, The MIT Press, Cambridge, Massachusetts, 1993.

[SGLF02] Peggy Seriès, Sébastien Georges, Jean Lorenceau, and Yves Frégnac. Orientation dependent modulation of apparent speed : a model based on the dynamics of feed-forward and horizontal connectivity in v1 cortex. *Vision Research*, 42 :2781–2797, 2002.

[SH92] E. P. Simoncelli and David J. Heeger. A computational model for perception of two dimensional pattern velocities. *Investigative Ophtalmology and Visual Science Supplement*, 33 :–, 1992.

[SH98] Eero P. Simoncelli and David J. Heeger. A model of neural responses in visual area mt. *Vision Research*, 38(5) :743–761, 1998.

[Sim93] E. P. Simoncelli. *Distributed analysis and representation of visual motion*. PhD thesis, Massachussetts Intitute of Technologie, Cambrige, M. A., October 1993.

[Sin91] Ajit Singh, editor. *Optic Flow Computation, a unified perspective*. IEEE Computer Society Press, 1991.

[SKS00] Paul R. Schrater, David C. Knill, and Eero P. Simoncelli. Mechanisms of visual motion detection. *Nature Neuroscience*, 3(1) :64–68, 2000.

[SKT+96] H. Sato, N. Katsuyama, H. Tamura, Y. Hata, and T. Tsumoto. Mechanisms underlying orientation selectivity of neurons in the primary visual cortex of the macaque. *Journal Physiology*, 494 :757–771, 1996.

[SL95] G. Sperling and Zhong-Lin Lu. The functional architecture of human visual motion perception. *Vision Research*, 35(19) :2697–2722, 1995.

[SLD82] M. V. Srinivasan, S. B. Laughlin, and A. Dubs. Predictive coding : a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B*, 216 :427–459, 1982.

[SM93] Joseph Sirosh and Risto Miikkulainen. How lateral interaction develops in a self-organizing feature map. In *Proceedings in the IEEE International Conference on Neural Networks*, San Diego, CA, 1993.

[SM97] Joseph Sirosh and Risto Miikkulainen. Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*, 9 :577–594, 1997.

[Smi03] Robert G. Smith. Retina. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 816–840. Bradford Book MIT Press, 2003.

[Sno98] R. Snowden. Shifts in perceived position following adaptation to visual motion. *Curr. Biol.*, 8 :1343–1345, 1998.

[SNS95] D. C. Somers, S. B. Nelson, and M Sur. An emergent model of orientation selectivity in cat visuel cortical simple cells. *The Journal of Neurosciences*, 15 :5448, 1995.

[SPH98] Adrian Spinei, Denis Pellerin, and Jeanny Hérault. Spatiotemporal energy-based method for velocity estimation. *SignalProc*, 65 :347–362, 1998.

[Spi98] Adrian Spinei. *Estimation du mouvement par triades de filtres de Gabor : Application au mouvement d'objets transparents*. PhD thesis, Laboratoire des Images et des Signaux, October 1998.

[SPK99] M. V. Srinivasan, Michael Poteser, and Karl Kral. Motion detection in insect orientation and navigation. *Vision Research*, 39 :2749–2766, 1999.

[SS94] Andrew T. Smith and Robert J. Snowden, editors. *Visual Motion of Detection*. Academic Press Limited, Cambridge, Great Britain, 1994.

[SSM86] P. H. Schiller, J. H. Sandell, and J. R. H. Maunsell. Functions of the on and off channels of the visual system. *Nature*, 322 :824–825, 1986.

[Ste90] P. Sterling. Retine. In M. Shephered, editor, *The Synaptique Organisation of the Brain*, pages 170–213. Oxford University Press, 1990.

[Ste02] Peter Sterling. How neurons compute direction. *Nature*, 420 :375–376, 2002.

[STEA91] R. J. Snowden, S. Treue, R. G. Erickson, and R. A. Andersen. The response of area mt and v1 neurons to transparent motion. *The Journal of Neurosciences*, 11 :2768–2785, 1991.

[Sto01] Alan Alfred Stocker. *Constraint Optimization Networks for Visual Motion Perception - Analysis and Synthesis*. PhD thesis, Swiss Federal Institute of Technology Zürich, Zürich, Switzerland, 2001.

[SUN95] M. Stemmler, M. Usher, and E. Neibur. Lateral interactions in primary visual cortex : A model bridging physiology and psychophysics. *Science*, 269 :1877–1880, 1995.

[SvHMO99] S. Sunaert, P. van Hecke, G. Marchal, and G. A. Orban. Motion-responsive regions of the human brain. *Experimental Brain Research*, 127 :355, 1999.

[SWB02] Robert Sekuler, Scott N. J. Watamaniuk, and Randolph Blake. Motion perception. *Steven's Handbook of Experimental Psychology*, 1 :121–176, 2002.

[SYT$^+$86] H. Saito, M. Yukie, K. Tanaka, K. Hikosaka, Y. Fukada, and E. Iwai. Integration of direction signal of image motion in the superior temporal sulcus of the macaque monkey. *The Journal of Neurosciences*, 6 :145–157, 1986.

[TAAH95] S. Treue, R. A. Andersen, H. Ando, and E. C. Hildreth. Structure-from-motion : Perceptual evidence for surface interpolation. *Vision Research*, 35 :139–148, 1995.

[Tay99] J. G. Taylor. Neural 'bubble' dynamics in two dimensions : foundations. *Biological Cybernetics*, 80(6) :393–410, 1999.

[Tay00] T. Tayama. The minimum temporal thresholds for motion detection of grating patterns. *Perception*, 29 :761–769, 2000.

[TB81] W. B. Thompson and S. T. Barnard. Lower-level estimation and interpretation of visual motion. *IEEE Computer*, 14 :20–28, 1981.

[TD94] A. M. Thomson and J. Deuchars. Temporal and spatial properties of local circuits in neocortex. *Trends in Neuroscience*, 17 :119–126, 1994.

[TFS89] S. Tanaka, Y. Fukada, and H. Saito. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *The Journal of Neurosciences*, 62 :626–641, 1989.

[THAE04] César Torres Huitzil and Miguel Arias Estrada. Real-time image processing wtih a compact fpga-based systolic architecture. *Journal on Real-Time Imaging*, 10 :177–187, 2004.

[THGCS05a] César Torres Huitzil, Bernard Girau, and Claudio Castellanos Sánchez. Digital implementation of a bio-inspired neural model for motion estimation. In *International Joint Conference on Neural Networks 2005*, pages –, August 2005.

[THGCS05b] César Torres Huitzil, Bernard Girau, and Claudio Castellanos Sánchez. On-chip visual perception of motion : A bio-inspired connectionist model on fpga. *Neural Networks*, 18(5-6) :557–565, June 2005.

[Tho82] P. Thompson. Perceived rate of movement depends on contrast. *Vision Research*, 22 :377–380, 1982.

[TK03] Thomas R. Tucker and Lawrance C. Katz. Recruitement of local inhibitory networks by horizontal connections in layers 2/3 of ferret visual cortex. *Journal of Neurophysiology*, 89 :501–512, 2003.

[TMH+97] R. B. H. Tootell, J. D. Mendola, N. K. Hadjikhani, P. J. Ledden, A. K. Liu, J. B. Reppas, M. I. Sereno, and A. M. Dale. Functional analysis of v3a and related areas in human visual cortex. *The Journal of Neurosciences*, 17 :7060–7078, 1997.

[TP99] J. T. Todd and V. J. Perotti. The visual perception of surface orientation from optical motion. *Percept. Psychophys.*, 61 :1577–1589, 1999.

[TPH95] A. Teuner, O. Pichler, and B. J. Hosticka. Unsupervised texture segmentation of images using tunied matched gabor filters. *IEEE Transactions on Image Processing*, 4(6) :863–870, 1995.

[TRK+95] R. B. H. Tootell, J. B. Reppas, K. K. Kwong, R. Malach, R. T. Born, T. J. Brady, B. R. Rosen, and J. W. Belliveau. Functional analysis of human mt and related visual cortical areas using magnetic resonance imaging. *The Journal of Neurosciences*, 15 :3215–3230, 1995.

[UGVT88] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60 :79–97, 1988.

[VH93] V. G. Vaidya and R. M. Haralick. Wigner distribution for 2d motion estimation from noisy images. *Journal Visual Communication and Image Representation*, 4(4) :281–297, 1993.

[vO03] Arjen van Ooyen, editor. *Modelling Neural Development*. A Bradford Book, The MIT Press, USA, 2003.

[VP89] A. Verri and T. Poggio. Motion field and optical flow : Qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5) :490–498, 1989.

[VRA05] J. Vitay, N. P. Rougier, and F. Alexandre. A distributed model of spatial visual attention. In Stefan Wermter and Günther Palm, editors, *Biomimetic Neural Learning for Intelligent Robotics*, volume LNAI 3575, pages 54–72. Springer-Verlag, 2005.

[vSS84] J. P. H. van Santen and G. Sperling. Temporal covariance model of human motion perception. *Journal of the Optical Society of America A*, 1 :451, 1984.

[vSS85] J. P. H. van Santen and G. Sperling. Elaborated riechardt detectors. *Journal of the Optical Society of America A*, 2 :300–321, 1985.

[WAJ85] A. B. Watson and A. J. Ahumada Jr. Model of human visual motion sensing. *Journal of the Optical Society of America A*, 2(2) :322–341, 1985.

[Wel89] L. Welch. The perception of moving plaids reveals two motion-processing stages. *Nature*, 337 :734–736, 1989.

[Wen90] J. Weng. A theory of image matching. In *Proceedings of 3rd International Conference on Computer Vision*, pages 200–209, Osaka, 1990.

[WGJ97] Terence H. Williams, Nedzad Gluhbegovic, and Jean Y. Jew, editors. *The Human Brain : Dissections of the Real Brain*. The University of Iowa, 1997.

[WMG94] S. N. J. Watamaniuk, S. P. McKee, and N. M. Grzywacz. Detecting a trajectory embedded in random direction motion noise. *Vision Research*, 35 :65–77, 1994.

[Woo90] R. J. Woodham. Multiple light source optical flow. In *Proc. of ICCV*, pages 42–46, 1990.

[WR97] G. Wallis and E. T. Rolls. Invariant face and object recognition in the visual system. *Prog. Neurobiol.*, 51 :167–194, 1997.

[WSR96] S. Wuerger, R. Shapley, and N. Rubin. "on the visually perceived direction of motion" by hans wallach : 60 year later. *Perception*, 25 :1317–1367, 1996.

[WT95] A. B. Watson and K. Turano. The optimal motion stimulus. *Vision Research*, 35 :325–336, 1995.

[WWB88] A. M. Waxman, J. Wu, and F. Bergholm. Convected activation profiles and receptive fields for realt time measurement of short range visual motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 717–723, Ann Arbor, MI, 1988.

[WWS90] Allen M. Waxman, Jian Wu, and Michael Seibert. Computing visual motion in the short and the long : From receptive fields to neural networks. In *Visual motion : Proceedings IEEE Computer Society*, pages 156–164, Irvine, CA, USA, 1990.

[YG89] A. L. Yuille and N. M. Grzywacz. A winner take all mechanisms based on presynaptic inhibition. *Neural Computation*, 1 :334–347, 1989.

[ZPMS+04] C. F. Zink, G. Pagnoni, M. E. Martin Skurski, J. C. Chappelow, and G. S. Berns. Human striatal responses to monetary reward depend on saliency. *Neuron*, 42 :509–517, 2004.

[ZS98] Richard S. Zemel and Terrence J. Sejnowski. A model for encoding multiple object motions and self-motion in area mst of primate visual cortex. *The Journal of Neurosciences*, 18(1) :531–547, 1998.

[ZZ01] J. M. Zanker and J. Zeil, editors. *Motion Vision : Computational, Neural, and Ecological Constraints*. Springer-Verlag, Berlin, Heidelberg, Germany, 2001.

Résumé

La recherche en connexionnisme est en face des fortes contraintes d'autonomie en robotique notamment issues des tâches perceptives visuelles qui n'ont pas encore été satisfaites par les modèles neuronaux utilisés jusqu'à aujourd'hui, surtout celles liées à la perception du mouvement. Dans ce cadre, je propose un modèle connexionniste neuromimétique permettant de traiter un ensemble de tâches de perception visuelle dynamique (robot/objets environnants en mouvement). Cela implique le compromis de maintenir la satisfaction des contraintes (bio-inspiration, traitements locaux massivement distribués) en vue d'une implantation ultérieure temps-réel embarquée sur circuit FPGA (Field Programmable Gate Array) au cœur de la définition du modèle proposé, malgré la complexité des tâches en jeu.

Le modèle connexionniste proposé pour la perception visuelle du mouvement est constitué de trois modules : le premier opère un filtrage spatio-temporel causal issu des filtres de Gabor et inspiré des réponses des cellules simples du cortex visuel primaire, V1. Le deuxième met en place un mécanisme distribué de fortes interactions localisées fondé sur un principe antagoniste inspiré de l'organisation en colonnes d'orientation dans V1. Finalement, en m'inspirant des propriétés des champs récepteurs des neurones de MT et MST (aire temporelle médiane et supérieure médiane, respectivement), nous intégrons les réponses du second module et les envoyons au troisième. Ce dernier fait émerger un seul objet en mouvement à travers l'évolution en différentes cartes des interactions latérales, en pro-action et en rétro-action d'une population neuronale densément interconnectée selon le principe de la CNFT (Continuum Neural Field Theory). L'attention sur l'objet émergé nous permet donc de le suivre.

Mots-clés: Réseaux de neurones, perception visuelle du mouvement, filtrage spatio-temporel, interactions antagonistes, filtres de Gabor, champs neuronaux.

Abstract

The connectionist research is in front of the strong constraints of autonomy in robotics issue of visual perceptive tasks that have not yet been satisfied by the neural models defined until now, particularly to the linked to motion perception ones. In this context, we propose a neuromimetic connectionist model allowing to deal with a set of tasks for dynamic visual perception (robot/objects in movement). It implies the compromise to satisfy some major constraints (bio-inspiration, massively distributed local processing aimed at being embedded on FPGA circuits for real-time processing) that must be kept central in the definition of our proposed model, in spite of the complexity of the considered tasks.

Our connectionist model for the visual perception of motion consists of three modules : the first one performs a causal spatio temporal filtering based on Gabor filters and inspired by the responses of the simple cells of the primary visual cortex, V1. The second one carries out a distributed mechanism using strong localized interactions based on an antagonist principle inspired by the organization of V1 as a set of orientation columns. Finally, we integrate the responses of the second module and we send them to the third one, as inspired by the properties of the receptive fields of the MT and MST (middle temporal and middle superior temporal areas, respectively). This last module brings to the foreground a single moving object through the evolution in different maps of the laterals, feed-forward and feed-back interactions of a neural population that is densely interconnected according to the principle of the CNFT (Continuum Neural Field Theory). Globally, a single moving object is focused and followed as a result of these interactions.

Keywords: neural networks, visual perception of motion, spatio-temporal filters, antagonist interactions, Gabor filters, neural fields.