

# Propriétés d'un Pré-processeur de Vision Fovéale

Emmanuel MARILLY<sup>1</sup>, Alain MERCIER<sup>1</sup>, Christophe COROYER<sup>2</sup>, Alain FAURE<sup>1</sup>, Olga CACHARD<sup>1</sup>,

<sup>1</sup>Laboratoire LACOS, Université du Havre  
Place Robert Schuman, 76610, Le Havre, France

<sup>2</sup>Laboratoire SYSCOM, Université de Marne la vallée  
5 Boulevard Descartes, 77454, Marne la Vallée, France

marilly@iut.univ-lehavre.fr

**Résumé** – Nous avons développé un modèle simplifié de rétine artificielle : le modèle Retina. Il est fondé sur un réseau de neurones artificiels organisé de manière à simuler la variation radiale de l'acuité visuelle (vision fovéale). Il s'apparente à la classe des pré-processeurs de vision fovéale. La mise en œuvre de neurones analogiques et impulsionsnels (Hodgkin Huxley) nous a permis d'encoder les informations. Les paramètres pertinents qui encodent le mouvement et les contours sont extraits des signaux de sorties de la Retina à l'aide d'une analyse temps / fréquence (Wigner Ville). L'intérêt principal de ce modèle est sa réponse aux stimuli stationnaires ou en mouvements, différenciée suivant les formes et les vitesses. Ce modèle est doté des propriétés d'adaptativité aux différentes conditions d'illuminations, de robustesse au bruit et sa caractéristique de multi-résolution permet la détection et l'estimation d'une large échelle de vitesses. L'objectif final est l'implémentation de ce capteur sur un système de robotique mobile.

**Abstract** – In the context of mobile robot, we have evolved a first simplified model of foveal vision : the Retina model. It is based on an artificial neural network organised to simulate the radial variation of the visual acuity. The information is encoded through the implementation of analogue and impulse neurons. The main interest of this model is its response to stationary or moving which can be distinguished both according to their shapes and velocities. This model is adaptive and its multi-resolution characteristics allow the detection of a wide range of velocities (low velocities (fovea), high velocities (peripheries zones)). From impulse output signals of Retina, we extract the pertinent parameters that encode the motion and pattern information thanks to a time frequency analysis. The use of a sensor of biological inspiration : the impulse Retina, for the motion detection, the extraction of the velocity vector, the pattern recognition bring many advantages. Indeed, as in the biological retina, our sensor adapts to the various conditions of illumination and is robust to noise. We validate this smart sensor by using a real time sequence image which enables us to study its behaviour under real conditions.

## 1. Introduction:

Il va de soit que la rétine biologique est plus qu'une simple camera vidéo. Elle ne fait pas que la conversion d'une information optique en un signal électrique mais elle effectue un important traitement du signal avant sa transmission aux niveaux supérieurs. Depuis quelques années, il y a eu un gros regain d'intérêt dans la modélisation et la construction de rétines artificielles. Plusieurs équipes de recherches ont déjà fabriqué des circuits électroniques modélisant certains aspects des fonctions rétinienne [1] [2] [3] et d'autres équipes ont développé des modèles numériques [4][5][6]. Chaque implémentation prend en compte certains aspects spécifiques des fonctions rétinienne tels que la transduction logarithmique ou non linéaire des photorécepteurs, l'adaptation locale des photorécepteurs, l'inhibition latérale ou la vision fovéale. En général ces implémentations de modèle rétinien, qu'elles soient logicielles ou électroniques, ne prennent en compte qu'une caractéristique particulière.

Dans le but de reproduire fidèlement le mécanisme de vision naturelle des vertébrés, nous avons élaboré un premier modèle simplifié de vision fovéale : le modèle Retina. C'est un réseau de neurones artificiels organisé de

manière à simuler la variation radiale de l'acuité visuelle. La mise en œuvre de neurones analogiques et impulsionsnels dans notre modèle nous a permis d'encoder les informations rétinienne. L'intérêt principal de ce modèle est sa réponse aux stimuli stationnaires ou en mouvements, différenciée suivant les formes et les vitesses. La Retina impulsionsnelle est adaptative et ses caractéristiques de multirésolutions permettent d'estimer une large échelle de vitesses.

## 2. Le modèle Retina impulsionsnel [7] :

Dans la rétine biologique, l'acuité visuelle décroît du centre de la rétine (fovéa) vers la périphérie. Cette propriété est attribuée à une variation de la densité des photorécepteurs qui décroît du centre à la périphérie. Pour simuler cette distribution non uniforme des photorécepteurs sur la rétine nous utilisons le principe de multirésolution pyramidale de l'image de Burt. Nous reproduisons la rétine en superposant des régions circulaires concentriques ayant chacune un niveau de résolution différente (fig. 1). Pour représenter l'image  $I=\{X_{ij}\}$  où  $X_{ij}$  est le niveau de gris du pixel au niveau de

résolution  $h$ , nous effectuons la convolution de Gauss en chaque point. Nous avons (1) :

$$X_{ij}^h = \sum_{p,q} W(p,q) \cdot X_{(i-2^{h-1}.p, j-2^{h-1}.q)}^{(h-1)} \quad (1)$$

Où  $W$  est le noyau gaussien avec  $p, q \in (-2, -1, 0, 1, 2)$ . Les valeurs ainsi obtenues servent d'entrée au réseau de neurones. Ceci permettra d'organiser les champs réceptifs des neurones.

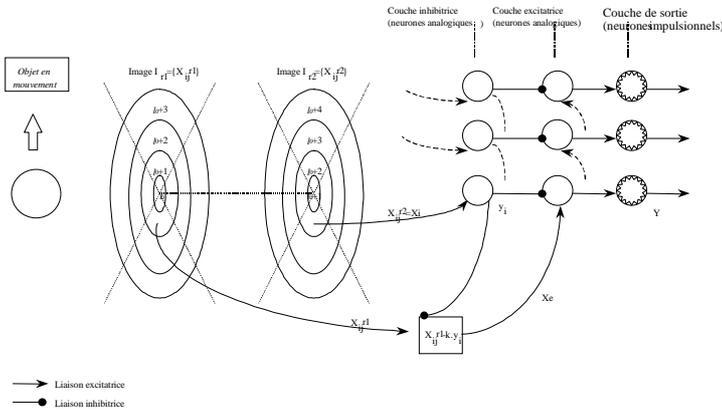


Fig. 1 : Schéma fonctionnel de la Retina

Le réseau comporte deux couches d'entrées ( $X_e$  neurone excitateur et  $X_i$  neurone inhibiteur) et une couche de sortie  $Y$  (neurone impulsif).

Le signal de sortie des neurones appartenant à la couche excitatrice et inhibitrice est la solution d'une équation différentielle du premier ordre en accord avec le modèle de Grossberg [8].

Le signal des neurones impulsifs (fig.2) suit un modèle déduit de celui de Hodgkin Huxley pour décrire la dynamique du potentiel de membrane de ces neurones [9].

Le schéma fonctionnel présenté figure 1 résume le modèle.

### 3. Simulations et commentaires :

Nous simulons la Retina suivant un modèle numérique. Les simulations sont effectuées à partir de stimuli possédant différents mouvements et formes. Les signaux ainsi obtenus ont une structure complexe liée aux caractéristiques des stimuli. L'analyse de ces signaux nous a permis d'extraire les éléments pertinents attachés aux caractéristiques spatio-temporelles. Pour effectuer cette analyse nous avons retenu la transformée de Wigner Ville.

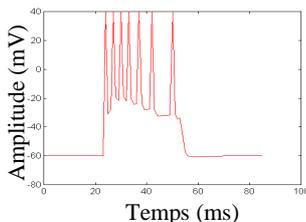


Fig. 2 : Réponse du neurone centre fovéa pour une barre horizontale (2\*3)

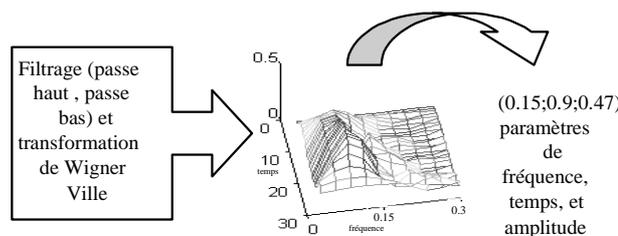


Fig. 3 : Représentation WV passe bas

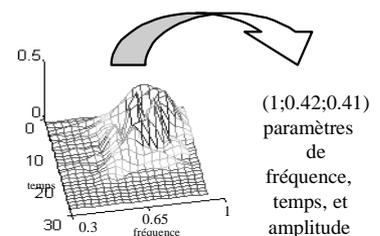


Fig. 4 : Représentation WV passe haut

### 3.1 Analyse temps fréquence :

Le signal impulsionnel délivré par la Retina (fig. 2) est constitué de deux bandes de fréquences (basse et haute fréquence). Nous séparons ces deux bandes de fréquences avant d'effectuer la transformée de Wigner Ville de façon à supprimer le phénomène d'interférence. Nous obtenons donc pour chaque train d'impulsions deux représentations temps fréquence (fig. 3 et 4). De chaque représentation nous extrayons le vecteur d'amplitude maximale composé des paramètres de temps, de fréquence et d'énergie. Nous projetons chacune de ces composantes dans des espaces distincts. Une représentation adaptée des ces paramètres nous permet de visualiser les propriétés et les particularités de chacune de ces composantes.

Nous choisissons trente neurones dans chacun des anneaux. Ces neurones sont choisis de telle sorte que leurs champs réceptifs recouvrent toute la rétine. Des ces neurones nous extrayons les vecteurs composés de la localisation en temps, fréquence et énergie.

La projection de ces différentes composantes dans des repères distincts nous donne une représentation temporelle, fréquentielle et énergie passe haut et passe bas. De ces représentations plusieurs conclusions peuvent être tirées.

### 3.2 Représentation fréquentielle :

Les représentations fréquentielles définissent la fréquence spatiale des stimuli. Ces signatures dépendent exclusivement de la forme globale des stimuli, ainsi on peut envisager de les utiliser en vue de la reconnaissance de forme. En effet, la localisation fréquentielle permet d'extraire des traits caractéristiques propres à l'étape de reconnaissance automatique (bords et coins). Nous travaillons sur des images dont le niveau de gris (ng) varie entre 0 et 255. Nous avons placé dans la fovéa un carré de taille 5x5 et nous obtenons la figure 5 qui correspond à la représentation fréquentielle haute fréquence. Cette représentation correspond aux réponses des neurones dont les champs récepteurs sont excités. Pour notre exemple nous n'obtenons que la réponse des 25 neurones qui correspondent à la surface du carré. Les fréquences les plus importantes se localisent en bordure du stimulus et fournissent l'information relative aux contours (figure 6).

Pour classifier ces fréquences spatiales nous utilisons un réseau de neurones multicouches à retro-propagation du gradient de l'erreur.

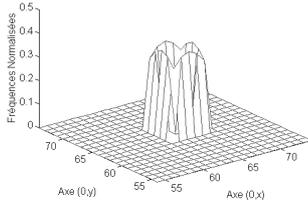


Fig. 5 : Représentation fréquentielle

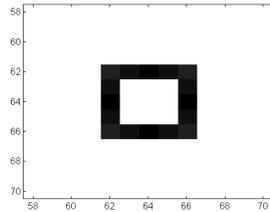


Fig. 6 : Contour du stimulus

Nous avons réparti nos stimuli suivant 3 classes distinctes : les carrés, les barres verticales (largeur inférieure à la hauteur) et les barres horizontales (largeur supérieure à la hauteur). Notre réseau comporte 3 couches : une couche d'entrée de 29 neurones sur laquelle on applique les valeurs des fréquences spatiales extraites de la fovéa, une couche cachée de 60 neurones et une couche de sortie de 3 neurones. Notre rétine associée à ce classifieur nous donne d'excellents résultats car pour plus de 251 stimuli différents nous avons un taux de reconnaissance de 100 % dans des conditions parfaites (sans bruit, illumination constante). De manière à observer les propriétés d'adaptativité de la Retina et la généralisation du réseau de neurones nous allons présenter à notre capteur des stimuli d'intensités lumineuses diverses et nous allons voir comment réagit notre système. La figure 7 représente les variations du taux d'erreur de classification en fonction de l'intensité lumineuse pour tous les carrés, barres verticales et barres horizontales entrant dans la fovéa.

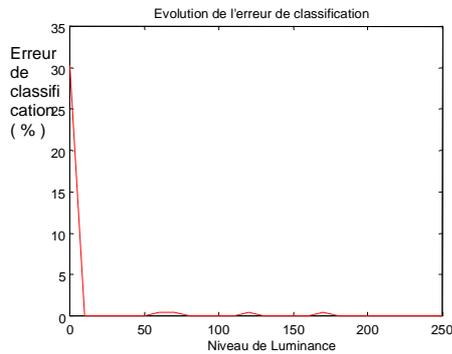


Fig. 7 : Evolution de l'erreur de classification pour différentes luminosités

Excepté pour une très faible luminosité inférieure à 10 (sur 250 niveaux de gris) nous obtenons une reconnaissance des formes de près de 100%. Cet exemple montre l'intérêt d'un tel capteur permettant de s'adapter aux conditions d'éclaircements et ce de manière automatique. Le taux global d'erreur est de 1,7% mais cette valeur est principalement due à la mauvaise classification des stimuli de très basses intensités lumineuses. Si l'on ne prend pas en compte ces très basses valeurs on obtient alors un taux d'erreur inférieur à 0,4 %. On peut considérer ces résultats comme satisfaisants. Cet exemple montre l'intérêt d'un tel capteur permettant de s'adapter aux conditions d'éclaircements et ce de manière automatique. Cette simulation a pour but d'observer le comportement du système quand nous sommes en présence d'images bruitées, chose très commune lorsque nous travaillons avec des images

réelles. Nous allons donc pouvoir observer les perturbations qu'entraînent le bruit dans la classification. Le bruit ajouté aux stimuli sera centré et nous désignerons par amplitude du bruit la valeur crête - crête du signal, en niveaux de gris. Le bruit ajouté aux stimuli est gaussien et centré. Nous faisons varier l'amplitude du bruit et pour chaque amplitude du bruit nous présentons 100 stimuli d'intensité lumineuse 250 ng au classifieur (réseau de neurones décrit précédemment). L'évolution du nombre de classifications correctes en fonction du niveau de bruit est représentée figure 8.

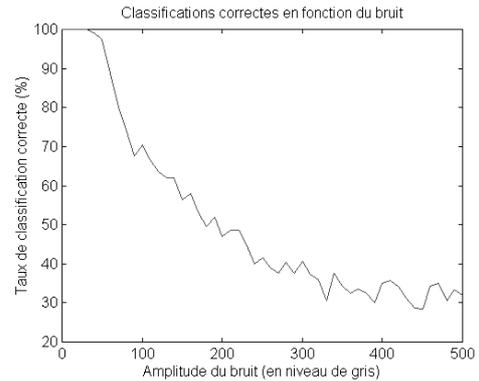


Fig. 8 : Evolution de la classification en fonction du bruit.

Le taux de classification reste satisfaisant jusqu'à un bruit d'amplitude 60 niveaux de gris. Au-delà de cette limite le taux s'altère de manière non négligeable. La sensibilité s'explique par le fait que nous effectuons la classification à partir des données de la fovéa qui est la zone la plus sensible au bruit. Les zones périphériques de la rétine sont moins sensibles au bruit mais ont une résolution plus faible. Il y aura donc un compromis à faire si l'on souhaite une forte résolution dans la reconnaissance des objets avec un système sensible au bruit ou un système moins sensible au bruit mais de plus faible résolution. Malgré cela nous pouvons considérer que notre système de classification reste efficace pour des niveaux de bruits habituels.

### 3.3 Représentation temporelle :

La représentation temporelle est une représentation du mouvement dans l'espace (x,y,t) ou x et y sont les deux dimensions spatiales et t la dimension temporelle. Nous nous plaçons dans le même cas défini par Adelson [10] ou l'estimation du mouvement s'effectue par la mesure d'une pente d'un plan.

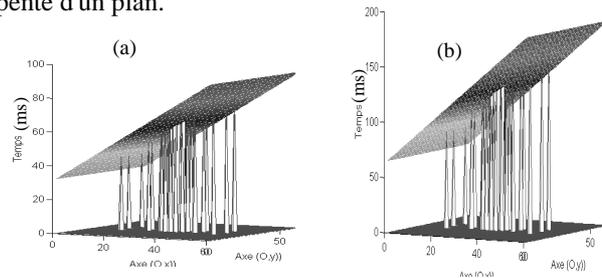


Fig. 9 : Diagrammes spatio-temporels d'un objet de vitesse 1 pixel/image pour (a) et 0.5 pixel/image pour (b)

Les diagrammes spatio-temporels d'un objet possédant un mouvement rectiligne uniforme dans le plan rétinien de vitesse 1 pixel/image et 0.5 pixel/image sont représentés figure 9. La simple estimation de la pente du plan nous permet de définir le vecteur vitesse du stimulus se déplaçant sur la Retina. L'estimation de la vitesse se fait en prenant en compte la réponse des neurones des différents anneaux. Notre capteur possédant des propriétés d'adaptativité, son application sur des images réelles ne pose aucun problème.

Tableau 1: Comparaison



Vitesse réelle sur axe (O,x)	Vitesse mesurée sur axe (O,x)
1.45	1.49
2.9	3.08
5.81	5.85
11.6	11.5
23	25

Fig. 10 : Séquence d'images

Nous avons donc à partir d'une séquence d'images réelles (Fig. 10) estimé le vecteur vitesse et comparé avec la vitesse théorique du stimulus. Les résultats sont présentés tableau 1. Nos résultats sont très proches des résultats théoriques malgré les différences d'éclaircissements entre les images successives de la séquence et le bruit présent dans l'image. Le fait de prendre en considération la réponse de l'ensemble des neurones des différents anneaux, permet d'obtenir une estimation de la vitesse moyennée et donc de minimiser l'erreur. Mais notre capteur fovéale dispose de zones de résolutions différentes et chacune de ces zones estime la vitesse de manière différente. La répartition des vitesses mesurées en fonction de l'excentricité est montrée fig. 11. A chaque zone correspond une bande d'estimation de vitesse optimum, petite vitesse pour la fovéa et grande vitesse en périphérie. Ce capteur détermine donc au mieux la vitesse d'un objet quelle que soit sa vitesse.

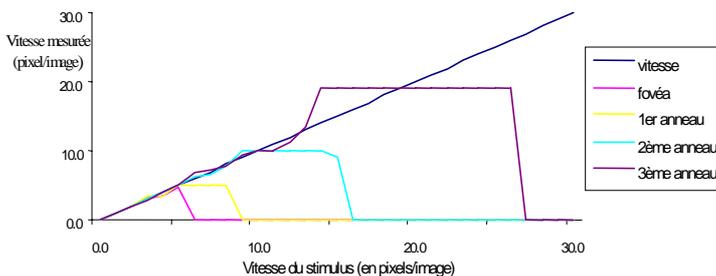


Fig. 11 : Vitesses mesurées par les différentes zones de la rétine

Les vitesses maximales mesurées sont respectivement de 5, 8, 15, et 26 pixels/image pour la fovéa et les trois anneaux de la Retina. La fovéa est sensible aux basses vitesses avec une bonne précision tandis que les anneaux mesurent des plages de vitesses de plus en plus grandes mais avec une précision de plus en plus faible. Ces plages de sensibilité sont dues à la

positions des neurones sur la rétine, à la taille des champs récepteurs et à la taille du stimulus.

## 4. Conclusion :

La Retina impulsionnelle, comme le système visuel humain, est soumise à l'interaction spatio-temporelle, c'est à dire que la perception simultanée spatiale et temporelle n'est pas séparable [11]. Mais les particularités de notre modèle : multi-résolution (fovéa pour forme et mouvement lent / anneau pour mouvement rapide), traitement par canaux (canal temporel et canal fréquentiel proche de la décomposition effectuée par le cerveau) et le taux d'échantillonnage variable de nos neurones suivant leurs spécificités nous permettent d'extraire à la fois les informations relatives au mouvement et à la forme des stimuli. Le fait de pouvoir à la fois estimer le mouvement et la forme des stimuli nous a permis d'appliquer notre pré-processeur de vision fovéale à la poursuite sélective des stimuli et nous offre des perspectives intéressantes.

## Références

- [1] Mahowald. "Silicon retina with adaptive photoreceptors". In SPIE/SPSE Symposium on Electronic Science and Technology: From Neurons to Chips, pages 52-58, Orlando, florida April 1991
- [2] Sandini. "A foveated retina-like sensor using CCD technology". In C. Mead and Ismail editors, Analog VLSI and Neural Network Implementations, pages 1-23. Dekluwer Publ., Boston 1989.
- [3] Torralba, Herault. "Circuits neuromorphiques pour l'estimation du mouvement". Grets'i'97, Grenoble 1997.
- [4] Chang et Yeh. "A retina-like image acquisition system with wide range light adaptation". SPIE Visual Communications and image processing 91: Image Processing 1606:456-469 1991
- [5] Y. Zeevi, R. Ginosar, "Neural Computer for Foveating Vision Systems" in Advanced Neural Computers. Eckmiller (editor). Elsevier Science Publishers 1990, p 323-330.
- [6] Faure, "Cybernétique des réseaux neuronaux" p172-179, éditions Hermès 1998.
- [7] Marilly, Coroyer, Cachard, Faure, "The impulse Retina : a smart sensor" Proceedings of EUSIPCO-98, vol III, p1561-1564. 8-11 sept 1998, Rhodes, Greece.
- [8] Grossberg : "Non linear neural networks : principles, mechanisms, and architecture" 1988 vol1 p17-61 neural networks.
- [9] Hodgkins et Huxley, 1952, "A quantitative description of membrane current and its application to conduction and excitation in nerve", Journal of physiology 117, pp500-544.
- [10] Adelson et Bergen, "spatiotemporal energy models for the perception of motion" J. Optical Society of America, vol A2, no 2 pp 284-299, 1985.
- [11] Watson. "Handbook of perception and Human performance", Vol 1, Sensory Processes and perception, Chapter 6, Temporal sensitivity. John Wiley, 1986.