

# DE LA RETINE À LA PERCEPTION DE SCENES : TRAITEMENT DU SIGNAL ET CORRELATS PSYCHOPHYSIQUES

Jeanny Hérault\*, Aude Oliva\*, William Beaudot#

\* Laboratoire de Traitement d'Images et de Reconnaissance de Formes, I.N.P.G. 46 Av. Félix Viallet, F-38031 Grenoble Cedex.

# Centre Suisse d'Électronique et Microtechnique, Maladière 71, CH-2007 Neuchâtel.

## RÉSUMÉ

Dans cet article, nous présentons plusieurs aspects du traitement spatio-temporel et de la couleur dans le système visuel des vertébrés. Nous montrons que la rétine est un processeur analogique de filtrage, en temps continu et espace échantillonné où temps et espace ne sont pas des variables séparables. Il en résulte d'intéressantes propriétés pour le traitement des objets en mouvement ou pour la perception "Coarse-to-Fine". Nous montrons ensuite que la couleur est traitée, en parallèle avec la luminance, selon un effet de modulation spatiale d'amplitude, la luminance subissant un filtrage passe-haut alors que l'information de couleur est inchangée. Nous montrons ensuite comment les primitives extraites, au niveau cortical, par des bancs de filtres de bandes orientées peuvent s'utiliser en vue de la perception de scènes. Ces résultats issus des données physiologiques et psychophysiques, suggèrent une nouvelle approche du traitement spatio-temporel des images, particulièrement efficace et de mise en oeuvre simple avec des circuits électriques.

## ABSTRACT

In this paper, we present various aspects of the spatiotemporal filtering of B & W and colour images that occurs in the visual system of vertebrates. We show that the retina is an analogue filter, continuous in time and sampled in space, where time and space are non-separable variables. It results from this interesting properties in the processing of moving objects or in "Coarse-to-Fine" perception. We show also that colour is processed, in parallel with luminance, according to a process of spatial modulation of amplitude, luminance being high-pass filtered whereas colour being unchanged. We show further how the primitives, extracted in the primary visual areas by a bank of oriented band-pass filters can be used in the frame of scene perception. These results derived from biological and psychophysical data suggest a new approach in spatio-temporal image processing, particularly efficient and simple to implement on electrical circuits.

## 1. Le système visuel

Nous nous intéresserons ici à ce que les psychologues appellent la vision de bas niveau, c'est à dire celle qui fait intervenir l'œil, la rétine, le corps genouillé latéral (CGL) et l'aire corticale striée V1 ou cortex visuel primaire (voir figure 1).

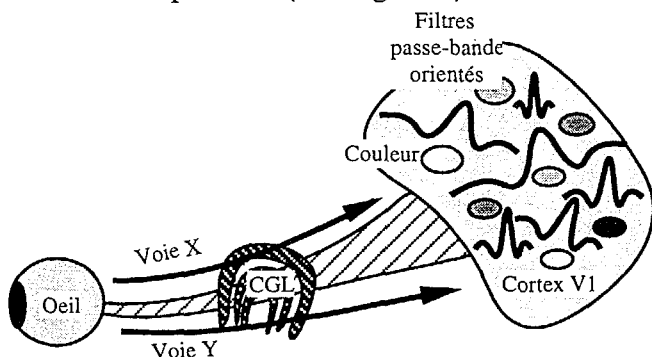


Fig. 1. Schéma du système visuel de l'œil au cortex visuel primaire.

L'œil est le premier élément de traitement de la chaîne, il reçoit la lumière à travers une lentille épaisse, le cristallin, qui focalise les rayons lumineux sur la rétine. L'image produite n'est nette que dans la zone centrale appelée fovéa. A mesure que les rayons

s'inclinent par rapport à l'axe optique du cristallin, l'image formée est de plus en plus floue. Les récepteurs qui tapissent la rétine ont une densité plus grande dans la région centrale qu'en périphérie, semblant respecter ainsi un certain théorème de Shannon.

La rétine [5, 6], nous le verrons plus loin, est composée de plusieurs couches de neurones qui réalisent un premier filtrage spatio-temporel des images, avec une compression de données : alors qu'il y a environ 160 millions de récepteurs, le nerf optique ne compte qu'un million de fibres. Composé des axones des cellules ganglionnaires, le nerf optique comporte deux types de voies : les cellules X qui véhiculent des informations de luminance et de couleur, et les cellules Y, à réponse transitoire qui véhiculent uniquement des informations de luminance [16].

Les voies X et Y traversent le corps genouillé latéral, dont la fonction encore mal définie ne sera pas abordée ici, pour se distribuer sur le cortex visuel primaire selon une topologie respectant localement celle de la rétine. Là, une seconde série de filtres produit des cellules sensibles à diverses bandes de fréquences spatiales et temporelles [6], à diverses

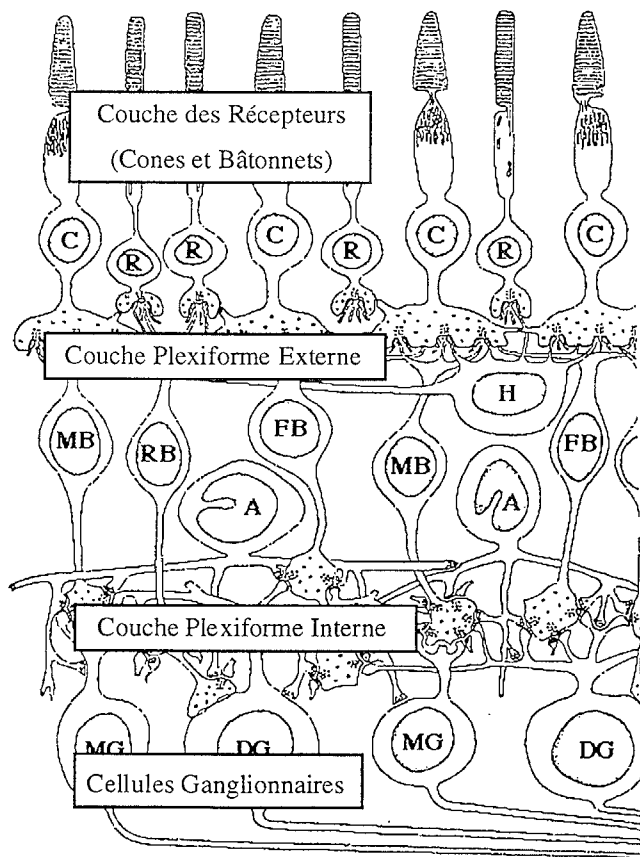


orientations des stimuli et à différents canaux de couleur. On pourrait ainsi dire que l'image rétinienne est ici décomposée en un certain nombre de primitives, ou encore qu'à chaque "pixel" de la rétine y est associé un vecteur dont les composantes sont ces primitives.

Dans la dernière partie de cet exposé, nous utiliserons ces primitives pour proposer un principe de l'analyse globale de scènes [14]. Il permet de rendre compte du fait qu'en une centaine de millisecondes, donc sans pouvoir mettre en œuvre aucune séquence d'analyse, nous pouvons reconnaître un visage ou un paysage familiers.

## 2. La rétine des vertébrés

Les récepteurs rétiniens sont de deux types [7]. Les bâtonnets, pour la vision mésopique et scotopique (faibles luminances), sont à réponse achromatique.



**Fig. 2.** La rétine, avec ses récepteurs, la couche plexiforme externe comprenant les interconnexions entre les récepteurs et les cellules horizontales et bipolaires, la couche plexiforme interne qui contient les interconnexions entre les cellule bipolaires, amacrines et ganglionnaires.

Pratiquement inexistant dans la zone fovéale, ils se rencontrent dans les zones périfovéale et périphériques de la rétine. Les cônes, utilisés en vision photopique (luminance élevée), sont présents principalement en région fovéale et parafovéale, ils présentent trois gammes de sensibilité à la couleur : le rouge, le vert et

le bleu. Dans la fovéa, on ne trouve que des cônes sensibles au rouge et au vert. Les cônes de même couleur sont reliés entre eux par des synapses dites jonctions "gap". Le signal des récepteurs est analogique, il est, dans une certaine mesure, proportionnel au logarithme de l'intensité lumineuse reçue. Nous ne considérerons par la suite que la rétine en zone fovéale et parafovéale.

Dans la couche plexiforme externe, les récepteurs délivrent leur signal aux cellules bipolaires qui assurent la transmission longitudinale vers les cellules ganglionnaires, et aux cellules horizontales qui sont responsables d'un couplage transversal entre les signaux puisqu'elles sont connectées entre elles par des jonctions "gap", comme les cônes. Elles agissent sur les bipolaires avec une action opposée à celle des cônes.

Dans la couche plexiforme interne, les cellules bipolaires excitent les cellules ganglionnaires, sous la médiation des cellules amacrines de différents types qui jouent des rôles multiples comme la diffusion transversale ou encore l'accentuation des transitoires. Des cellules dites "interplexiformes" agissent en retour sur les horizontales et les récepteurs en modulant leurs caractéristiques [2, 20].

## 3. Le modèle des récepteurs de couleur

Les récepteurs étant de trois types (en région parafovéale), et ayant chacun sa place sur la rétine, force est de constater que l'échantillonnage chromatique (relatif à chaque couleur) est le tiers de l'échantillonnage total de la rétine. Comment allons nous modéliser cela? Bien que le maillage hexagonal centré soit le mieux à même de rendre compte du problème [8], il sera plus simple d'en développer ici le modèle pour un espace unidimensionnel.

Soit donc une rétine fictive unidimensionnelle, échantillonnée régulièrement (avec un pas  $k$ ) à trois récepteurs par pixel, fournissant les signaux discrets : rouge  $r(k)$ , vert  $g(k)$  et bleu  $b(k)$ . Nous allons représenter la rétine réelle comme l'entrelacement de trois sous-échantillonnages de ces signaux. Le signal  $r'(k)$  des récepteurs réels rouges est obtenu par le produit du signal  $r(k)$  par une fonction d'échantillonnage de période  $1/3$ , les autres signaux sont obtenus par les produits avec la même fonction décalée de  $\pm 1/3$  selon le cas :

$$r(k) = r_0(k) \cdot \frac{1}{3} \left[ 1 + 2 \cos(2\pi \frac{k}{3}) \right] \quad (1)$$

qui vaut 1 pour  $k=3p$  entier, 0 sinon

$$g(k) = g_0(k) \cdot \frac{1}{3} \left[ 1 + 2 \cos(2\pi \frac{k-1}{3}) \right] \quad (1')$$

qui vaut 1 pour  $k=3p+1$  entier, 0 sinon

$$b(k) = b_0(k) \cdot \frac{1}{3} \left[ 1 + 2 \cos\left(2\pi \frac{k+1}{3}\right) \right] \quad (1'')$$

qui vaut 1 pour  $k=3p-1$  entier, 0 sinon

Le signal composite produit par ce système s'écrit alors simplement  $s_c(k) = r(k) + g(k) + b(k)$ , soit encore :

$$s_c(k) = \frac{r_0(k) + g_0(k) + b_0(k)}{3} \quad (2)$$

$$+ \frac{2}{3} \left[ r_0(k) \cos\left(2\pi \frac{k}{3}\right) + g_0(k) \cos\left(2\pi \frac{k-1}{3}\right) + b_0(k) \cos\left(2\pi \frac{k+1}{3}\right) \right]$$

Il représente donc la somme d'un signal de luminance

$$l(k) = 1/3 \{r_0(k) + g_0(k) + b_0(k)\}$$

échantillonné avec le pas initial  $k$  et d'un signal de chrominance

$$chr(k) = \frac{2}{3} \left[ r_0(k) \cos\left(2\pi \frac{k}{3}\right) + g_0(k) \cos\left(2\pi \frac{k-1}{3}\right) + b_0(k) \cos\left(2\pi \frac{k+1}{3}\right) \right]$$

composé des signaux fictifs de couleur modulés par des fonctions sinusoïdales décalées.

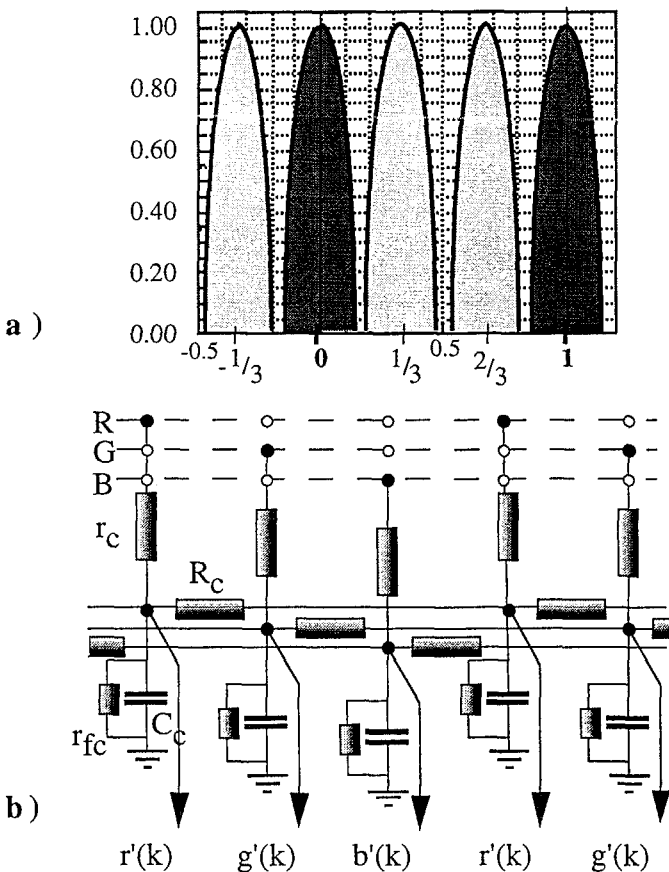


Fig. 3. Modèle de l'échantillonnage couleur (a) et du circuit équivalent du couplage des récepteurs de même couleur (b).

Il en résulte les spectres de fréquence spatiale suivants :

$$\begin{aligned} R(f_x) &= \frac{R_0(f_x)}{3} \left[ \delta(f_x) + \delta\left(f_x - \frac{1}{3}\right) + \delta\left(f_x + \frac{1}{3}\right) \right] \\ G(f_x) &= \frac{G_0(f_x)}{3} \left[ \delta(f_x) + \left\{ \delta\left(f_x - \frac{1}{3}\right) + \delta\left(f_x + \frac{1}{3}\right) \right\} e^{-j2\pi f_x} \right] \\ B(f_x) &= \frac{B_0(f_x)}{3} \left[ \delta(f_x) + \left\{ \delta\left(f_x - \frac{1}{3}\right) + \delta\left(f_x + \frac{1}{3}\right) \right\} e^{+j2\pi f_x} \right] \end{aligned} \quad (3)$$

qui montrent que le spectre :

$$L(f_x) = 1/3 \{R_0(f_x) + G_0(f_x) + B_0(f_x)\}$$

de la luminance est de type passe bas et que le spectre de la chrominance  $Chr(f_x)$  est de type modulation d'un cosinus par les signaux passe bas  $R_0(f_x)$ ,  $G_0(f_x)$  et  $B_0(f_x)$ , comme le montre la figure 3a.

Nous allons maintenant analyser le filtrage réalisé par le couplage entre les cônes. Nous avons dit que les signaux à ce niveau étaient de nature analogique. Appelons  $r_c$  la résistance interne du générateur de tension qui traduit le signal lumineux reçu par le cône en signal électrique,  $r_{fc}$  et  $C_c$  la résistance et la capacité de la membrane cellulaire, et enfin  $R_c$  la résistance de couplage de la jonction "gap" entre deux cônes. On obtient alors un filtre spatio-temporel (figure 3b) continu en temps et échantillonné en espace avec un pas de 3.

Nous avons déjà présenté le calcul d'un tel filtre [3, 9], il se réalise à partir des lois de Kirchoff, en calculant les tensions aux nœuds  $k$  en fonction des tensions aux nœuds  $k-1$  et  $k+1$ , puis en prenant les transformées de Fourier par rapport à la variable temporelle  $t$  continue et par rapport à la variable spatiale  $k$  discrète. On obtient ainsi la fonction de transfert  $G_c$  des cônes, pour chaque couleur en fonction des fréquences spatiale  $f_x$  et temporelle  $f_t$  :

$$G_c(f_x, f_t) = \frac{1}{1 + \beta_c + 2\alpha_c \left( 1 - \cos\left(2\pi \frac{f_x}{0.33}\right) \right) + j 2\pi \tau_c f_t} \quad (4)$$

avec  $\beta_c = r_c/r_{fc}$ ,  $\alpha_c = r_c/R_c$  et  $\tau_c = r_c C_c$ .

On remarquera les propriétés suivantes de cette fonction de transfert. Elle est :

- de type passe-bas spatial et temporel
- causale en temporel et symétrique en spatial
- périodique en spatial (période  $1/3$ ) mais pas en temporel
- à variables temporelle et spatiale non séparables.

Toutes ces remarques nous seront utiles par la suite. Nous obtenons donc en sortie des cônes, trois signaux de couleur  $r'(k)$ ,  $g'(k)$  et  $b'(k)$  correspondant aux filtrages passe-bas de chacun des signaux  $r(k)$ ,  $g(k)$  et  $b(k)$  indépendamment les uns des autres. Nous allons voir maintenant comment ils sont traités par la couche plexiforme externe.

#### 4. Cellules horizontales et bipolaires

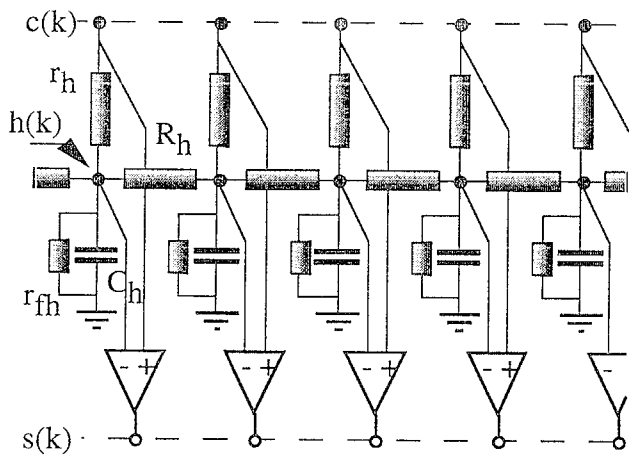
Nous décrivons ici le modèle proposé pour la





première fois par Carver Mead [13] auquel nous avons apporté une modification [3]. Le signal  $c(k) = r'(k) + g'(k) + b'(k)$  est appliqué aux cellules horizontales via des synapse chimiques (figure 4).

Il est considéré comme fourni en chaque point  $k$  par un générateur de tension  $c(k)$  de résistance interne  $r_h$ . Les cellules horizontales ont pour résistance et capacité membranaire  $r_{fh}$  et  $C_h$  respectivement, elles sont reliées entre elles par des jonctions gap de résistance  $R_h$ . Il en résulte, au niveau des cellules horizontales, une fonction de transfert spatio-temporelle  $G_h(f_x, f_t)$  de même type que celle des cônes, avec des paramètres différents, mais cette fois avec un pas d'échantillonnage de 1 et non plus de 3.



**Fig. 4.** Modèle de la couche plexiforme externe. Les signaux des cônes  $c(k)$  excitent les cellules bipolaires (triangles), et les cellules horizontales qui sont couplées par des jonctions gap ( $R_h$ ) qui inhibent les bipolaires. Les sorties sont  $s(k)$ .

Les cellules bipolaires (triangles sur la figure 4) effectuent la différence entre le signal  $c(k)$  des cônes et le signal  $h(k)$  des horizontales. L'ensemble cellules horizontales-bipolaires, qui réalise la différence entre le signal  $c(k)$  et sa version filtrée passe-bas  $h(k)$ , produit donc un signal  $s(k)$  qui est le filtrage passe-haut de  $c(k)$  par un filtre de fonction de transfert  $G_{hb}(f_x, f_t) = 1 - G_h(f_x, f_t)$ . Cette fonction de transfert a pour expression :

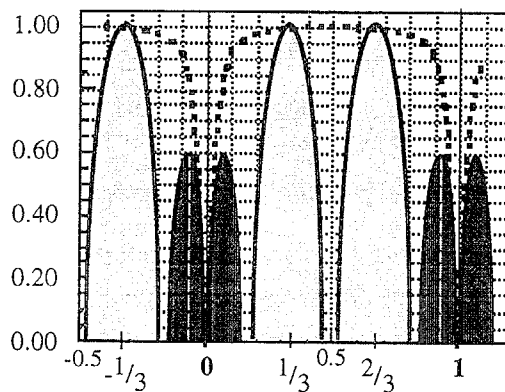
$$G_{hb}(f_x, f_t) = \frac{\beta_h + 2\alpha_h (1 - \cos(2\pi f_x)) + j 2\pi \tau_h f_t}{1 + \beta_h + 2\alpha_h (1 - \cos(2\pi f_x)) + j 2\pi \tau_h f_t} \quad (5)$$

avec  $\beta_h = r_h/r_{fh}$ ,  $\alpha_h = r_h/R_h$  et  $\tau_h = r_h C_h$ .

Son allure est tracée en pointillés sur le diagramme de la figure 5. On remarque donc sur cette figure que le spectre du signal  $s(k)$  est celui des signaux luminance plus chrominance filtré par  $G_c(f_x, f_t)$  puis modulé par  $G_{hb}(f_x, f_t)$ .

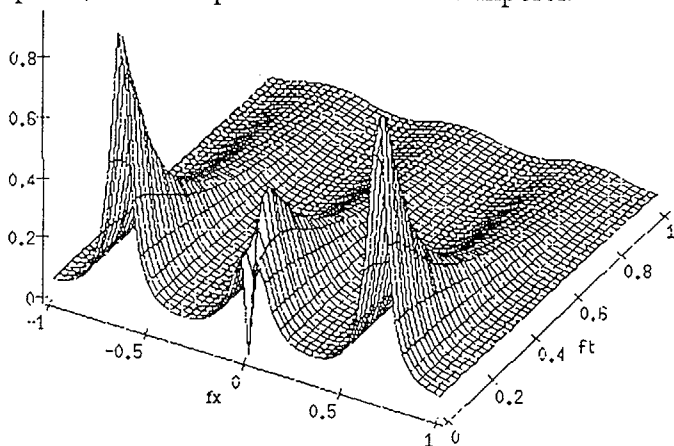
Compte tenu des bandes fréquentielles respectives de la luminance et de la chrominance, leur traitement est totalement différent : le signal de

luminance se trouve filtré passe-haut, donc subit une augmentation de contraste spatio-temporel, alors que le signal de chrominance est pratiquement inchangé.



**Fig. 5.** Effet du filtrage de l'ensemble des cellules bipolaires et horizontales sur les signaux de luminance et de chrominance.

La fonction de transfert spatio-temporelle complète de la couche plexiforme externe est représentée à la figure 6. Elle montre que la différence entre les traitements de luminance et de couleur, bien que due initialement à un aspect d'échantillonnage spatial, est aussi présente au niveau temporel.



**Fig. 6.** Fonction de transfert spatio-temporelle. De la gauche vers le bas : fréquence spatiale, du bas vers la droite : fréquence temporelle.

## 5. Quelques propriétés

### 5.1. Perception de la luminance et de la chrominance

D'après notre modèle, on peut prévoir qu'une image en noir et blanc (représentée uniquement par la partie gris sombre du spectre sur la figure 5) sera traitée avec un effet d'augmentation de contrastes. C'est un phénomène bien connu sous le nom de "bandes de Mach", l'expérience se pratique en regardant des marches de niveaux de gris sur une surface uniformément éclairée : près de chaque limite, le côté sombre paraît plus sombre et le côté clair, plus clair

Au contraire, pour une image à contrastes de

couleurs iso-luminantes, on n'observe aucune accentuation. Ce phénomène est aussi bien connu en psychophysique, quoique moins évident à réaliser dans la vie courante : ainsi, il est très difficile de focaliser son attention sur la limite entre deux couleurs iso-luminantes, et même, si l'on persiste à essayer, cette limite finit par disparaître et il faut recommencer le processus de focalisation!

De plus, souvenons nous que la couche des cellules horizontales-bipolaires reçoit en entrée une succession spatiale de signaux de couleurs :  $r'(k)$ ,  $g'(k)$  et  $b'(k)$ . Qu'en est-il de la succession correspondante des signaux en sortie des bipolaires? Pour le savoir, il nous faut appliquer à  $s(k)$  les trois fonctions de sous-échantillonnage de pas 3 définies par les formules (1). On trouve alors [8] que les cellules bipolaires successives présentent des réponses de type opposition de couleur, dans l'ordre : Rouge-Cyan, Vert-Magenta, Bleu -Jaune.

Ces différentes propriétés ont été vérifiées expérimentalement en biologie [6, 16, 19], soit dans les cellules ganglionnaires X de la rétine, soit dans leurs prolongements (région parvocellulaire du corps genouillé latéral). Ces cellules répondent de deux manières différentes: à une stimulation de luminance pure, par un effet d'accentuation des contrastes spatio-temporels, à une stimulation de chrominance pure, par un effet d'opposition de couleur (Rouge-Vert, ou Bleu -Jaune) mais de type seulement passe-bas.

### 5.2. Aspect temporel du filtre rétinien

Nous avons vu que, dans la fonction de transfert de la couche plexiforme externe, les variables temporelle et spatiale n'étaient pas séparable. Il en résulte que, si l'on présente à la rétine une image à partir d'un instant  $t_0$  et si on la maintient, le filtrage spatial qu'elle subira va évoluer au cours du temps.

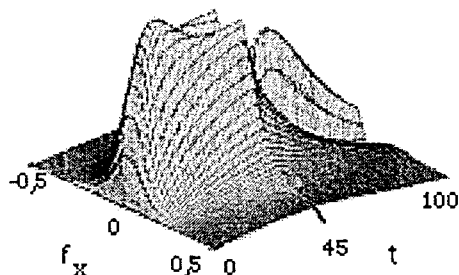


Fig. 7. Évolution temporelle de la fonction de transfert spatiale de la rétine lorsque la stimulation est une image présentée pendant une durée de 45 ms.

Si l'image que l'on présente est un échelon unité temporel :  $i(x, t) = i_0(x) U(t)$ , on constate que la fonction de transfert spatiale est, pour  $t$  voisin de 0, de type passe-bas. Puis, à partir de quelques 20 ms, on voit la partie centrale du spectre se creuser et présenter un aspect passe-haut, c'est à dire d'augmentation de

contrastes. Cela signifie qu'en ouvrant les yeux, on perçoit d'abord les grandes masses de la scène, puis seulement après, les détails. Si nous clignons des yeux assez rapidement, seule l'image des grandes masses sera perçue.

Si maintenant, après avoir regardé une image assez longtemps, nous fermons les yeux, que se passe-t-il? La simulation d'un tel cas est présentée à la figure 7. Lorsque la sortie est stabilisée, la perception est celle des détails. Mais lorsqu'on supprime l'excitation, on voit disparaître la partie haute fréquence du spectre spatial puis apparaître un "rebond" des basses fréquences, qui finit par s'évanouir. Ce rebond, c'est en fait l'image filtrée passe-bas par les cellules horizontales qui reste, en négatif, après l'arrêt de la stimulation des récepteurs. On peut réaliser soi-même cette expérience en regardant une image à forts contrastes pendant un certain temps : lorsque l'on ferme les yeux, la scène apparaît, en négatif et plus floue.

### 5.3. Détection du mouvement

Encore à cause de l'inséparabilité des variables spatiale et temporelle, le comportement de la rétine est particulier pour les parties d'images où le temps et l'espace sont couplés, c'est à dire pour les objets en mouvement.

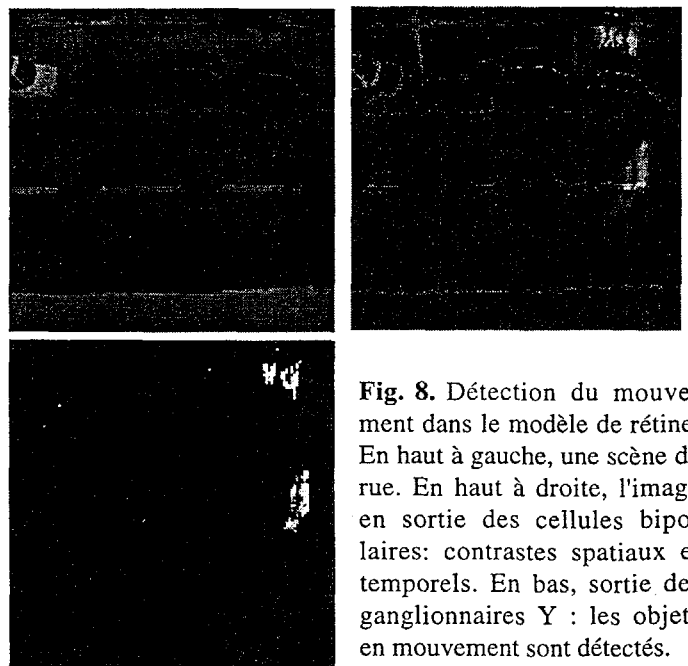


Fig. 8. Détection du mouvement dans le modèle de rétine. En haut à gauche, une scène de rue. En haut à droite, l'image en sortie des cellules bipolaires: contrastes spatiaux et temporels. En bas, sortie des ganglionnaires Y : les objets en mouvement sont détectés.

À la figure 8, nous présentons au modèle de rétine une scène de rue où des piétons se déplacent, l'œil étant supposé fixe. Après le transitoire d'établissement, l'image en sortie des bipolaires montre une sélection des contours des objets, une sensibilité au bruit spatial et temporel des capteurs, et une réponse particulière aux objets en mouvement.

Si l'on rajoute à notre modèle la simulation des cellules ganglionnaires Y "ON et OFF" qui, à partir du



signal des bipolaires et des cellules amacrines transitoires, effectuent un filtrage temporel passe-haut et une sélection entre les objets noirs sur fond blanc et les objets blancs sur fond noir, on obtient une sélection des objets en mouvement. Les bruits spatiaux et temporels n'étant pas corrélés, ils sont éliminés.

L'architecture de traitement d'une rétine est intéressante à deux points de vue pour l'ingénieur. D'abord, elle est facilement réalisable en circuits intégrés VLSI analogiques, voir [4, 13] et figure 9. Ensuite, du fait que le temps est continu, il n'y a pas d'ambiguïté pour l'estimation des vitesses d'objets mobiles comme c'est le cas dans les caméras où temps et espace sont échantillonnés [9, 10]. On peut imaginer que la détection des objets mobiles puisse se faire même si l'échantillonnage spatial ne respecte pas le théorème de Shannon, puisqu'en présence de mouvement, la périodicité spatiale des spectres ne présente plus le phénomène de recouvrement. Ceci ne serait pas toujours vrai avec un échantillonnage temporel.

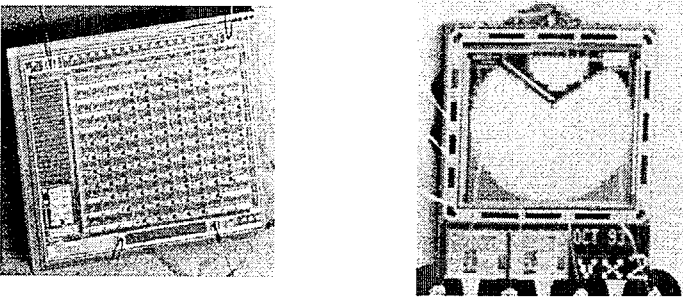


Fig. 9. Deux rétines en silicium réalisées au CSEM.

## 6. Corrélats psychophysiques

Dans ce qui précède, nous avons étudié les premières couches du système visuel du point de vue du traitement du signal et nous avons donné quelques pistes sur leur relations avec des données expérimentales en biologie et en psychophysique. Par une démarche inverse, nous allons étudier maintenant quelques problèmes de psychophysique et en chercher les interprétations possibles en matière d'architecture neuronale.

### 6.1. Perception "Coarse-to-Fine"

On présente à un sujet une image pendant une durée relativement brève (20 à 30 ms), de telle sorte qu'il lui soit impossible de dire ce qu'il a "vu". On cherche ensuite si les informations qu'il a reçues pendant ce flash lui permettent de reconnaître mieux, ou moins bien une seconde scène qui lui est présentée pendant 100 ms. Il n'a donc pas le temps d'explorer visuellement cette seconde scène et la mesure de son temps de réponse est un indice. On s'aperçoit alors que son temps de réponse dépend du spectre spatial de

la première image qui lui a été présentée. Pour obtenir une meilleure acuité des tests psychophysiques, la première image est une chimère [14, 15, 17] composée des basses fréquences spatiales d'une scène et des hautes fréquences spatiales d'une seconde scène (figure 10). Par exemple, on présente l'image basse-fréquence d'un salon additionnée à l'image des hautes fréquences d'une route.

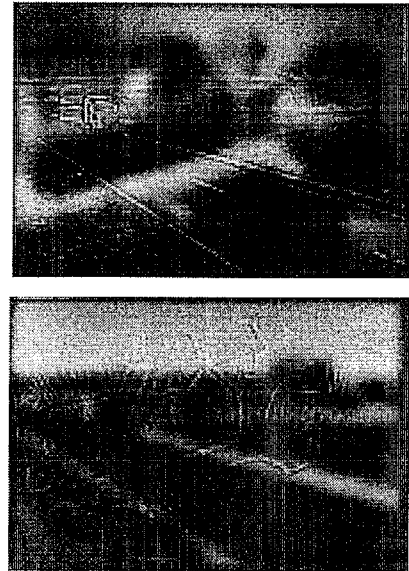


Fig. 10. Chimères utilisées pour mettre en évidence la perception "Coarse-to-Fine".

Pour un sujet "vierge", c'est à dire sans préoccupation attentionnelle précise, si la seconde image présentée est l'image complète d'un salon, elle sera bien plus rapidement nommée que si c'est une route. Autrement dit, ce sont seulement les basses fréquences qui sont perçues durant les 20 ms de la première image et qui aident à la reconnaissance de la seconde. La perception d'une scène, en l'absence de tâche attentionnelle, s'effectue dans le sens BF=>HF, c'est à dire "Coarse-to-Fine". Le modèle de rétine est compatible avec ce principe, puisqu'il montre (figure 7) que lors de la présentation d'une image, son filtrage spatial est d'abord passe-bas, puis ensuite passe haut.

### 6.2. Perception de scènes

On a l'habitude de dire, en vision par ordinateur, qu'une scène est composée d'objets et que ces objets sont composés de primitives. On en déduit alors que pour analyser une scène, il faut d'abord la segmenter en primitives, puis en objets pour enfin pouvoir la reconnaître.

Or notre cerveau nous permet de reconnaître un visage ou une scène en quelques 100 ms, c'est à dire le temps de quelques impulsions nerveuses. Il n'a donc pas matériellement le temps de procéder à une analyse séquentielle comme la précédente. Comment peut-il donc s'y prendre?



Dans sa thèse [14], à partir de considérations psychologiques sur la perception coarse-to-fine, A. Oliva a postulé qu'une scène pouvait être caractérisée par l'ensemble des énergies en sortie de filtres de bandes orientés, dans la gamme des basses et très basses fréquences. Elle a donc pris une série d'une quarantaine d'images variées appartenant à diverses catégories, scènes de villes, de routes, d'autoroutes, ou de salons.

Puis, sur le modèle des filtres orientés du cortex

cérébral, elle les a filtré par un ensemble de filtres de GABOR pour quatre orientations allant de 0 à 135 degrés, et pour trois bandes de fréquences basses et très basses de  $1/64$  à  $1/16$  de la fréquence maximale. Chaque image était donc représentée par un vecteur à douze composantes, les énergies des douze filtres de bande orientés. Il s'agissait alors de classer ces vecteurs sur la base de leur proximité dans un espace à 12 dimensions, et ceci de manière automatique, sans l'intervention d'un superviseur.

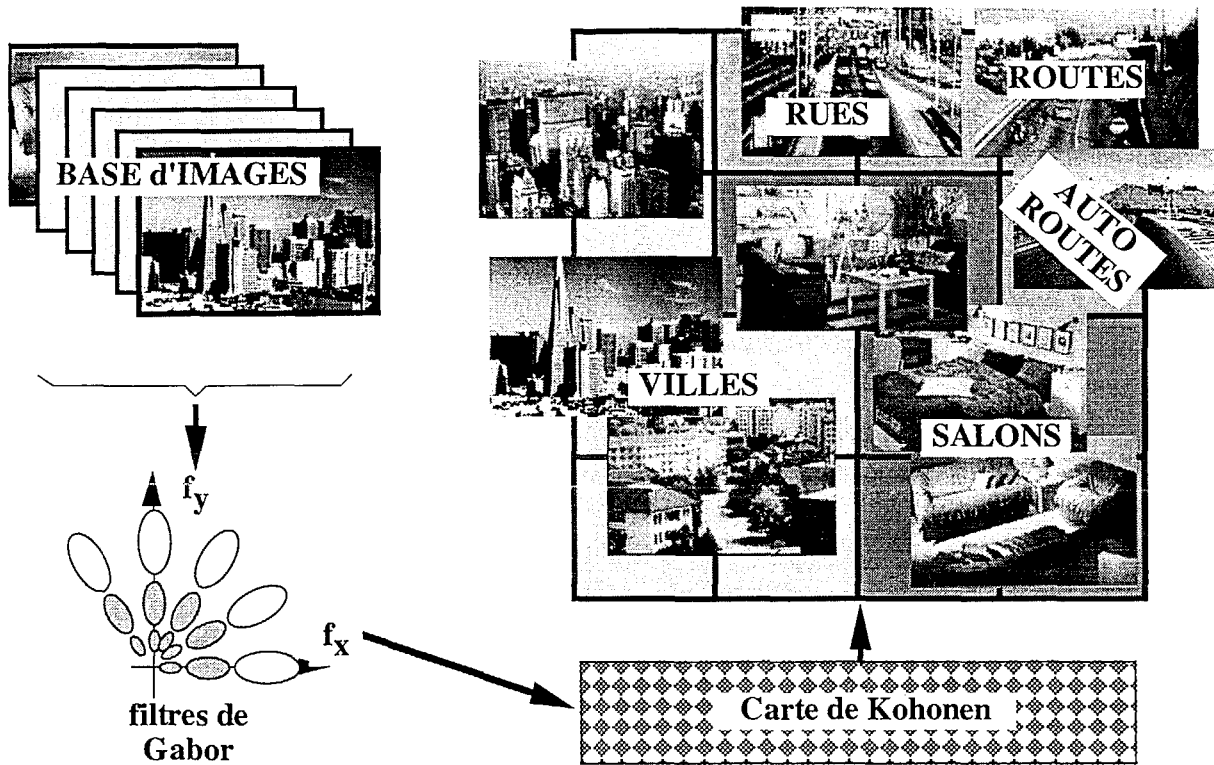


Fig. 11. Filtrage d'images par des filtres de bande orientés, et leur classification non supervisée par une carte auto-organisatrice de Kohonen : les différentes scènes sont classées par similarité, selon une logique voisine de celle d'un opérateur humain.

Ce classement a été effectué de manière non supervisée grâce à une carte auto-organisatrice de Kohonen [11]. Dans ce type de réseau neuromimétique, les unités ou "neurones" disposent leurs vecteurs d'entrée de manière à effectuer une quantification vectorielle de l'espace des données sous la contrainte de leur voisinage. Ainsi, deux neurones voisins dans la carte auront des vecteurs voisins pointant dans l'espace d'entrée. L'apprentissage se fait sans superviseur : lors de la présentation d'une donnée au réseau, le neurone dont le vecteur est le plus proche de la donnée est autorisé s'adapter vers la donnée, il entraîne dans ce processus ses voisins dans un certain rayon. Après convergence de l'apprentissage, l'espace d'entrée est pavé par les vecteurs des neurones (prototypes), la topologie locale du pavage représentant celle des neurones.

A cause de cette topologie, les neurones se

spécialisent chacun pour un type d'image et lorsqu'on présente en entrée du réseau les vecteurs descriptifs de diverses images, ils se trouvent classés par les neurones dans des régions spécialisées par type de scène.

Le réseau a donc détecté, au sein des composantes des images, des corrélations typiques de scènes, sans qu'on lui ait appris que telle et telle images étaient similaires ou appartenaient à une même classe de scènes.

Il est intéressant de constater qu'il a été possible ici de classer des scènes sans avoir à en analyser le contenu sémantique. Une telle opération, ne nécessitant pas d'analyse séquentielle du contenu, mais basée uniquement sur des corrélations entre composantes peut se réaliser en des temps très courts sur des architectures parallèles. C'est probablement des mécanismes de ce genre qui permettent à notre



cerveau ses performances remarquables en matière de reconnaissance rapide.

## 7. Conclusion

Nous avons montré quelques exemples du traitement biologique de l'information. D'abord le traitement spatio-temporel des images par la rétine, avec le multiplexage en fréquences spatiales de la couleur et de la luminance, la détection de mouvement et la perception "coarse-to-fine". Nous avons vu ensuite l'utilisation d'un modèle du traitement cortical en bandes fréquentielles orientées pour la

caractérisation non supervisée de scènes. Ces différents algorithmes ont un trait commun : leur grande simplicité de mise en œuvre sur des structures de calcul parallèles et analogiques. Il s'agit là des tout premiers pas d'une nouvelle méthode d'approche, encore rudimentaire. Mais nous pensons qu'il est possible maintenant pour l'ingénieur de rechercher quelque inspiration dans le modèle biologique. Bien sûr, les avions qui volent ne le font pas sur le modèle des oiseaux, mais que dire de nos ordinateurs qui n'atteignent même pas les capacités de traitement d'une cervelle d'oiseau?...

## 8. Références

- [1] ATICK J., ZHAOPING LI, REDLICH A., (1990) *Color coding and its interaction with spatiotemporal processing in the retina*. Technical report IASSNS-HEP-90/75, NYU-NN-90/3
- [2] BEAUDOT W. (1994) *Le traitement neuronal de l'information dans la rétine des vertébrés : un creuset d'idées pour la vision artificielle*. Thèse de Doctorat de l'INPG, Grenoble
- [3] BEAUDOT W, PALAGI P, HÉRAULT J (1993) *Realistic Simulation Tool for Early Visual Processing including Space, Time and Colour Data*. International Workshop on Artificial Neural Networks 93, Barcelona, June 1993
- [4] BOUVIER G, MAHNI A, HÉRAULT J. (1995) *A contrast and motion-sensitive silicon retina*, ESSIRC-95
- [5] BUSHBAUM G., GOTTSCHALK A., (1983) *Trichromacy, opponent colours coding and optimum colour information transmission in the retina*. Proc. of the Royal Society of London, B 220, 89-113
- [6] DE VALOIS R., DE VALOIS K. (1990), *Spatial Vision* Oxford Psychology Series N° 14. Oxford University Press
- [7] DOWLING JE (1987) *The retina: An approachable part of the brain*. MA: Harvard University Press, Cambridge
- [8] HÉRAULT J. *Vertebrate retina: sub-sampling and aliasing effects can explain colour-opponent and colour constancy phenomena*. Invited paper in Int. Conference on Artificial Neural Networks (ICANN'94), Sorrento, Italy, 1994
- [9] HÉRAULT J., BEAUDOT W. *The Vertebrate Retina: A Model of Spatiotemporal Image Filtering*. 14<sup>ème</sup> Colloque GRETSI, Juan Les Pins, Septembre 1993
- [10] HÉRAULT J, BEAUDOT W (1993) *Motion Processing in the Retina: About a Velocity Matched Filter*. European Symposium on Artificial Neural Networks 93, Brussels, April 1993
- [11] KOHONEN T. (1984) *Self-Organisation and Associative Memory*. Springer-Verlag, Washington DC, 2nd edition
- [12] KREMERS J., LEE B.B., KAISER P. K., (1992) *Sensitivity of the macaque retinal ganglion cells and human observers to combined luminance and chromatic temporal modulation*. J. Opt. Soc. Am. A, Vol 9, N° 9
- [13] MEAD C, MAHOWALD M (1988) *A silicon model for early visual processing*. Neural Networks, 1:1, 91-97
- [14] OLIVA A. (1995) *Perception de Scènes : Traitement fréquentiel du signal visuel, aspects psychophysiques et neurophysiologiques*. Thèse de Doctorat de l'INPG, spécialité Sciences Cognitives, Grenoble
- [15] OLIVA A., SCHYNS P.G., TIBERGHEN G. *Regularity of spatial organization in scenes categories and coarse-to-fine identification*. Perception, supp. 22, 102, 1993
- [16] RICHTER J, ULLMAN S (1982) *A model for the temporal organization of X- and Y-type receptive fields in the primate retina*. Biological Cybernetics 43, 127-145
- [17] SCHYNS P.G., OLIVA A.. *From blobs to boundary edges: Evidence for time and spatial scale scene recognition*. , Psychological Science, vol 5, 5, 1994
- [18] SRINIVASAN MV, LAUGHLIN SB, DUBS A (1982) *Predictive coding: a fresh view of inhibition in the retina*. Proceedings of The Royal Society of London, B 216, pp 427-459
- [19] WÄSSLE H, BOYCOTT BB (1991) *Functional architecture of the mammalian retina*. Physiological Reviews, Vol. 71, n°2, pp 447-480
- [20] WITKOVSKY P, STONE S, TRANCHINA D (1989) *Photoreceptor to Horizontal Cell Synaptic Transfer in the Xenopus Retina: Modulation by Dopamine Ligands and a Circuit Model for Interactions of Rod and Cone Inputs*. Journal of Neurophysiology, vol. 62, n°4, pp 864-881