HAL open science

Extraction de saillances par un réseau de neurones impulsionnels

Sylvain Chevallier, Philippe Tarroux

To cite this version:

Sylvain Chevallier, Philippe Tarroux. Extraction de saillances par un réseau de neurones impulsionnels. NeuroComp, Oct 2006, Pont-à-Mousson, France. pp.154-157. hal-00529479

HAL Id: hal-00529479

https://hal.science/hal-00529479v1

Submitted on 25 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EXTRACTION DE SAILLANCES PAR UN RÉSEAU DE NEURONES IMPULSIONNELS

Sylvain Chevallier
LIMSI-CNRS UPR 3251
Bât. 508, BP 133 91403 Orsay
email: sylvain .chevallier@limsi.fr

Philippe Tarroux
École Normale Supérieure, Paris
& LIMSI-CNRS UPR 3251, Orsay
email: philippe.tarroux@limsi.fr

ABSTRACT
Nous proposons une architecture de la vision pré-attentionnelle qui permet d'extraire des saillances dans des scènes visuelles complexes. Cette architecture est composée de neurones impulsionnels qui rendent possible l'obtention des saillances directement dans l'ordre de leur importance. Nous décrivons un moyen rapide d'intégrer les contributions d'une somme de PSP au potentiel membranaire. Nous présentons un exemple de résultat obtenu à partir d'images acquises par la caméra d'un robot.
KEY WORDS
Saillance - Neurones Impulsionnels - Attention Visuelle

1 Introduction

L'utilisation de la connaissance des systèmes vivants intéresse depuis longtemps les concepteurs de systèmes artificiels, que ce soit en robotique [3, 7] ou en vision artificielle [25]. Doit-on recréer extensivement toutes les propriétés observées ou peut-on espérer trouver un niveau d'abstraction fonctionnel qui offrirait des solutions intéressantes ? Nous proposons d'explorer une approche bio-inspirée de l'attention en exposant une architecture neuronale, détaillée dans la partie 2 et dont les résultats sont exposés dans la partie 3.

La vision artificielle recherche toujours des solutions pour réduire la complexité des images et traiter ainsi seulement les informations porteuses de sens [13]. Ces informations sont regroupées sous le terme de saillance bien qu'il n'existe pas encore de définition largement approuvée. La saillance visuelle est étroitement liée aux mécanismes pré-attentionnels décrits par les psychologues [22, 24], qui sont complètement automatiques et indépendants du contexte. De plus, des travaux en neurobiologie semblent confirmer l'existence d'une carte de saillance, qui consisterait en une représentation topographique des stimulus pertinents, même si elle est peut-être distribuée sur plusieurs aires corticales [23]. C'est sur ces saillances que les mécanismes attentionnels, issus des fonctions cognitives plus élaborées, vont intervenir. Il existe plusieurs modélisations de l'attention ; Itti [11] propose un état de l'art des modèles biologiquement inspirés et Heinke et Humphreys [10] un état de l'art des modèles issus de la psychologie. Nous proposons une approche différente dans le sens où nous nous sommes intéréssés à l'utilisation de ces modèlisation dans le cadre de la robotique. Nous proposons une implémentation des mécanismes pré-attentifs qui repose sur l'utilisation de réseaux de neurones impulsionnels (ou SNN pour *spiking neural networks*), plus réalistes biologiquement et qui possèdent des propriétés intéressantes [16, 9]. Les neurones impulsionnels permettent de manipuler finement l'information temporelle, ce qui semble être une propriété essentielle des neurones biologiques [1]. Ils sont de plus en plus largement utilisés [2, 18, 14, 8], même si leur capacités théoriques sont encore discutées [15, 20]. Notre travail diffère de celui de Thorpe [21], qui s'intèresse principalement à la catégorisation visuelle ultra-rapide. L'architecture pré-attentive que nous décrivons est une étape dans la mise en place d'un système d'attention artificiel.

2 Réseau de neurones impulsionnels

L'architecture proposée permet d'extraire des saillances d'une scène visuelle complexe et a été conçue comme une étape vers la mise en place d'un contrôleur neuronal complet. Elle est composée d'un ensemble de cartes neuronales (des couches 2D de SNN) connectées de façon *feedforward*. L'implémentation réalisée peut inclure différents types de modèles neuronaux en même temps et fonctionne de façon distribuée [4]. Cette architecture se divise en deux voies différentes, pour traiter les hautes et basses fréquences spatiales (resp. en bas et en haut de la figure 1). Ce traitement multi-échelles peut être rapproché de celui des voies parvo- et magno-cellulaires du cortex visuel des primates, d'autant que les traitements en hautes fréquences arrivent après ceux des basses fréquences, avec un délai fixe. Les traitements réalisés dans les basses fréquences regroupent les points saillants sur une carte de pré-activation (PreAc sur la figure 1). L'extraction de saillances est effectuée sur la carte de saillance quand les activations sont concordantes entre la carte de pré-activation et les cartes des hautes fréquences. Les neurones impulsionnels sont de bons détecteurs de synchronie, ce qui permet de les utiliser efficacement sur la carte de saillances pour faire émerger les saillances.

La prise en compte des couleurs dans l'extraction des saillances est réalisée sur des cartes d'opposition de

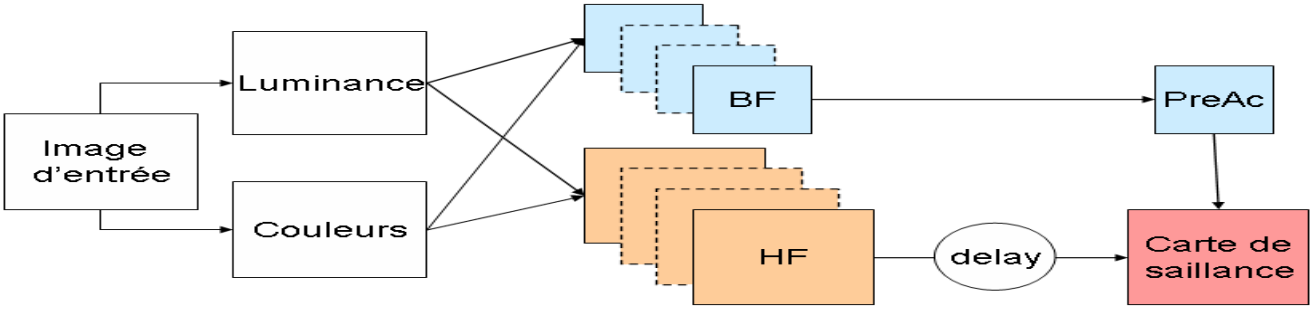

Figure 1. Chaque carré représente une carte 2D de neurones. Les masques de connexions, expliqués dans 2.2, sont figurés par les flèches. Les basses fréquences sont traitées par les cartes BF et les hautes fréquences par les cartes HF."Delay" signifie que les synapses possèdent un délai fixe non nul.

couleurs, rouge-vert et bleu-jaune (Simple Opposition sur la figure 1). Ces cartes sont composées de neurones semblables aux cellules de simple opposition que l'on trouve dans le corps genouillé latéral des primates [17, 19]. Il existe d'autres types de neurones qui traitent les couleurs, comme les cellules de double opposition couleur-luminance par exemple, mais elles restent encore peu étudiées [19]. Cette modélisation est cohérente vis-à-vis des connaissances physiologiques actuelles, même si toutes les propriétés chromatiques et spatiales des différentes cellules du système visuel n'ont pas encore été étudiées extensivement [5].

2.1 Carte neuronale

L'unité de base pour la construction de cette architecture est la carte neuronale. Elle est organisée en couches 2D de NxM neurones ($\frac{N}{2}$x$\frac{M}{2}$ pour les cartes basses fréquences), avec NxM la taille de l'image d'entrée en pixels. L'organisation de ces cartes est "rétinotopique" dans le sens où les propriétés topologiques de l'image d'entrée sont conservées. Les cartes d'entrées convertissent la luminance (ou les couleurs rouge, vert et bleu) des pixels en un courant d'entrée sur leurs neurones correspondants. Ceci revient à convertir la luminance (ou les couleurs) en taux de décharge.

Les neurones qui composent ces cartes sont du type intègre-et-décharge linéaire à fuite (*linear leaky integrate-and-fire*), tels que décrits par Gerstner [9]. Leur potentiel membranaire V est régi par l'équation suivante :

$$\begin{cases} \tau\dot{V} = g_L(V - E_L) + PSP(t) \text{ , if } V \geq \vartheta \\ \text{spike sinon} \end{cases} \quad (1)$$

Avec τ, g_L et E_L, les constantes du modèle et $PSP(t)$, l'influence des potentiels post-synaptiques. Nous utilisons cependant un modèle particulier pour ces PSP. Au lieu d'utiliser une fonction α, par exemple $\frac{t}{\tau}e^{\frac{-t}{\tau}}$, nous avons préféré un modèle de différence de gaussiennes (voir équation 2) pour décrire l'intégration membranaire d'un PSP :

$$\begin{cases} \dot{q}_0(t) = -k_0 q_0(t) \\ \dot{q}_1(t) = k_0 q_0(t) - k_1 q_1(t) \end{cases} \quad (2)$$

avec comme solution résultante :

$$\begin{cases} q_0(t) = \overline{q}_0 e^{-k_0 t} \\ q_1(t) = \frac{k_0}{k_1 - k_0}\overline{q}_0 \left[e^{-k_0 t} - e^{-k_1 t}\right] \end{cases} \quad (3)$$

Où $q_1(t)$ est l'influence du PSP sur la membrane du neurone et k_0, k_1 et $\overline{q}_0$ sont des contantes caractéristiques de la synapse. Un PSP en différence de gaussiennes a un décours temporel très similaire à celui d'une fonction α puisqu'une fonction α est un cas particulier d'une différence de gaussiennes [9]. Ci-après, nous appelerons $q_j(t, \hat{t}_j)$ la contribution à t du spike j émis à $\hat{t}_j$ sur le neurone post-synaptique. Le poid synaptique est inclus dans la constante $K_j = \frac{k_0}{k_1 - k_0}\overline{q}_{0,j}$. Dans le cas où les contributions des PSP sont additives, on peut rapidement calculer les contributions d'une somme de PSP sur le potentiel de membrane. En effet, l'équation décrivant l'évolution du potentiel de membrane après qu'un train de PSP a atteint le neurone est donnée par :

$$PSP(t) = \sum_{j=1}^{n} q_j(t, \hat{t}_j) \quad (4)$$

$$= \sum_{j=1}^{n} K_j \left[e^{-k_0(t-\hat{t}_j)} - e^{-k_1(t-\hat{t}_j)}\right] \quad (5)$$

$$PSP(t) = e^{-k_0 t}\sum_{j=1}^{n} K_j e^{k_0 \hat{t}_j} - e^{-k_1 t}\sum_{j=1}^{n} K_j e^{k_1 \hat{t}_j} \quad (6)$$

Si on pose :

$$\overline{K}_p(n) = \sum_{j=1}^{n} K_j e^{k_p \hat{t}_j} \quad (7)$$

on peut écrire :

$$PSP(t) = \overline{K}_0 e^{-k_0 t} - \overline{K}_1 e^{-k_1 t} \quad (8)$$

Il est possible de calculer, pour chaque pas de temps, la contribution du train de PSP en une seule étape, en utilisant

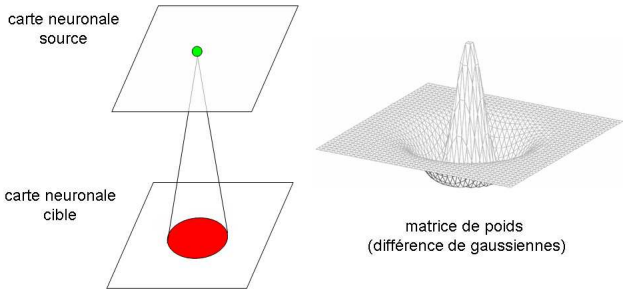

Figure 2. Un neurone (en vert ou gris clair) de la carte neuronale source émet un spike et envoie des PSP à tous les neurones en rouge (ou gris sombre) sur la carte neuronale cible. Le poids de chaque synapse est fixé dans une matrice de poids, à droite. La matrice de poids représentée est une différence de gaussiennes, qui agit comme un détecteur de contrastes.

le résultat de l'équation 8 dans l'équation 1 du potentiel de membrane. Si un nouveau PSP arrive sur le neurone, il suffit de mettre à jour les valeurs de $\overline{K}_0(n)$ et $\overline{K}_1(n)$ avec l'équation 9 :

$$\overline{K}_p(n) = \overline{K}_p(n-1) + K_n e^{k_p \hat{t}_n} \tag{9}$$

où $\overline{K}_n$ est le poids synaptique du n-ième PSP.

2.2 Masques de connexion

Quand un neurone sur une carte neurale source émet un spike, il envoie des PSP à NxN neurones sur les cartes neuronales cibles (figure 2). Tous les neurones d'une carte partagent le même pattern de connexions, c'est-à-dire les même projections efférentes. Nous appelons masque de connexion cet ensemble de projections générique, identique pour tous les neurones d'une carte. Les masques sont implémentés sous formes de matrices NxN de poids et de délais, ces poids et délais sont statiques et définis à l'initialisation du programme. Les masques implémentent des types de filtrages qui sont analogues à ceux des *simple cells* [12], observés dans le cortex visuel. Ce sont des traitements similaires aux filtres matriciels de type ondelettes de Gabor, utilisés dans les traitements d'images. Nous utilisons aussi des masques de connexion en différence de gaussiennes, connues aussi sous le nom de chapeau mexicain à cause de leur forme caractéristique. Les masques offrent des avantages pour les calculs distribués, car ils permettent de réduire le nombre et la taille des messages échangés entre les processeurs. Il suffit que la carte source envoie un seul message à chaque carte cible et ce message est traduit en multiples PSP une fois arrivé sur la carte cible.

3 Résultats

Les résultats ont été obtenus à partir des images acquises durant une exploration d'un robot Koala dans les couloirs de notre laboratoire. Les images ont été réduites en 76x56 pixels. La figure 3 montre un exemple de résultat obtenu, en mettant l'accent sur la prise en compte des couleurs. L'image d'entrée est sur la figure 3.c et les zones saillantes sont montrées en 3.a et 3.b, avec les points les plus saillants en blanc. Les saillances sont ordonnées en utilisant les propriétés temporelles des neurones impulsionnels : plus un neurone de la carte de saillance (voir figure 1) émet un spike de façon précoce, plus ce point est considéré comme saillant. Ainsi, les images 3.a et 3.b sont obtenues en affichant l'ordre d'émission des spikes de la carte de saillance. C'est en fait un codage par rang comme celui utilisé par [18, 14], à la différence que le SNN que nous avons utilisé ne se limite pas à un seul spike par neurone. Cette hiérarchisation des saillances est difficile à obtenir autrement, car les algorithmes qui permettent d'ordonner les régions saillantes, les *winner-takes-all* par exemple, sont complexes et coûteux. Cependant, les temps d'exécution moyens de notre algorithme restent relativement élevés comparés aux techniques de traitements d'images, car le nombre de neurones utilisés est très grand (environ 10^5).

En traitement d'image, tous les pixels sont traités dans un ordre arbitraire (ligne/colonne). L'architecture présentée permet de ne traiter que les pixels suffisamment lumineux (ou colorés) et ceci, dans un ordre fonctionnel, car il dépend de leur luminosité (ou de leur couleur). Par exemple, les pixels sont traités dans l'ordre de leur luminance. L'apparition des résultats est donc progressive et dans l'ordre des saillances. Il est ainsi possible d'arrêter les traitements dès que l'on a trouvé une réponse convenable. La dynamique du réseau est comparable à celle d'un algorithme *anytime*, qui améliore en permanence la qualité de ses résultats tout en restant capable de fournir une réponse à n'importe quel moment.

4 Conclusion

Nous avons présenté une architecture de la vision pré-attentive capable d'extraire des saillances de scènes visuelles complexes, en s'inspirant des traitements du système visuel biologique. Comme les temps d'exécution restent élevés, nous avons proposé un moyen de réduire le temps de calcul de l'intégration des PSP sur le potentiel de membrane. L'utilisation de neurones impulsionnels offre la possibilité d'intégrer un apprentissage sur les traitements intermédiaires pour implémenter, en plus des traitements pré-attention, des mécanismes attentionnels descendants.

References

[1] Sander Bohte. The evidence for neural information processing with precise spike-times: A survey. *Natural Computing*, 3:195–206,

Figure 3. (c) image d'entrée. (a) saillances extraites de l'image d'entrée, les pixels les plus blancs sont les plus saillants et les plus sombres les moins saillants. Ces saillances ont été obtenu en travaillant seulement sur la luminosité. (b) idem, mais en travaillant cette fois sur la luminosité et les couleurs. (d) Les différences d'activation entre (a) et (b) sont représenté par les zones entourées en pointilliés.

2004.

[2] C. Brody and J.J. Hopfield. Simple networks for spike-timing-based computation, with application to olfactory processing. *Neuron*, 37:843–852, 2003.

[3] R. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

[4] S. Chevallier, P. Tarroux, and H. Paugam-Moisy. Saliency extraction with a distributed spiking neural network. In *Proc. of European Symposium on Artificial Neural Networks*, pages 209–214, 2006.

[5] D. Dacey and O. Packer. Colour coding in the primate retina: diverse cell types and cones-specific circuitry. *Current Opinion in Neurobiology*, 13:421–427, 2003.

[6] A .Delorme, J. Gautrais, R. van Rullen, and S. Thorpe. Spikenet: a simulator for modeling large networks of integrate-and-fire neurons. *Neurocomputing*, 26-27:989–96, 1999.

[7] D. Floreano, J. Godjevac, A. Martinoli, F.Mondada, and J-D. Nicoud. *Advances in Intelligent Autonomous Agents*, chapter Design, Control, and Applications of Autonomous Mobile Robots. Kluwer, 1998.

[8] D. George and F. Sommer. Computing with inter-spike interval codes in networks of integrate and fire neurons. *Neurocomputing*, 65-66:415–420, 2005.

[9] W. Gerstner and W. Kistler. *Spiking Neuron Models*. Cambridge University Press, 2002.

[10] D. Heinke and G. Humphreys. *Connectionist Models in Psychology*, chapter Computational Models of Visual Selective Attention: A Review. Houghton, G., in press.

[11] L. Itti, G. Rees, and J. Tsotsos, editors. *Neurobiology of Attention*, chapter Models of Bottom-Up Attention and Saliency. Elsevier, 2005.

[12] JP. Jones and LA. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

[13] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[14] P. Kornprobst, T. Viéville, and I.K. Dimov. Could early visual processes be sufficient to label motions? In *Int. Joint Conf. on Neural Networks*, IJCNN'05, pages 1687–1692. IEEE-INNS, 2005.

[15] R. Legenstein, C. Neger, and W. Maas. What can a neuron learn with spike-time-dependent plasticity? *Neural Computation*, 17(11):2337–2382, 2005.

[16] W. Maass and C. Bishop, editors. *Pulsed Neural Networks*. MIT-Press, 1999.

[17] R. Masland. The fundamental plan of the retina. *Nature neuroscience*, 4(9):877–886, 2001.

[18] L. Perrinet. Efficient source detection using integrate-and-fire neurons. In W. Duch et al., editor, *Int. Conf. on Artificial Neural Networks*, volume 3696 of *Lecture Notes in Computer Science*, pages 167–172. Springer-Verlag, 2005.

[19] R. Shapley and M. Hawken. Neural mechanisms for color perception in the primary visual cortex. *Current Opinion in Neurobiology*, 12:426–432, 2002.

[20] J. Sima and J. Sgall. On the nonlearnability of a single spiking neuron. *Neural Computation*, 17(12):2635–2647, 2005.

[21] S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

[22] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 1980.

[23] S. Treue. Visual attention: the where, what, how and why of saliency. *Current Opinion in Neurobiology*, 13:428–432, 2003.

[24] J. Wolfe. *Seeing*, chapter Visual attention, pages 335–386. Academic Press, 2me edition, 2000.

[25] Florentin Wörgötter, Norbert Krüger, Nicolas Pugeault, Dirk Calow, Marcus Lappe, Karl Pauwels, Marc Van Hulle, Sovira Tan, and Alan Johnston. Early cognitive vision: Using gestalt-laws for task-dependent, active image-processing. *Natural Computing*, 3:293–321, 2004.