



HAL
open science

Implémentation d'un mécanisme de "covert attention" avec un réseau de neurones impulsionnels

Sylvain Chevallier, Philippe Tarroux

► **To cite this version:**

Sylvain Chevallier, Philippe Tarroux. Implémentation d'un mécanisme de "covert attention" avec un réseau de neurones impulsionnels. Deuxième conférence française de Neurosciences Computationnelles, "Neurocomp08", Oct 2008, Marseille, France. hal-00331563

HAL Id: hal-00331563

<https://hal.science/hal-00331563v1>

Submitted on 17 Oct 2008

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.

IMPLÉMENTATION D’UN MÉCANISME DE “COVERT ATTENTION” AVEC UN RÉSEAU DE NEURONES IMPULSIONNELS

Sylvain Chevallier
Université Paris-Sud 11, Orsay
& LIMSI-CNRS UPR 3251, 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 implémentation possible du mécanisme de *covert attention* avec des neurones impulsionnels. Nous nous sommes intéressés aux intérêts qu’offre l’utilisation d’un code temporel, tel que celui produit par des neurones impulsionnels, pour les processus attentionnels et de vision bas niveau en vision artificielle. Cette contribution détaille le fonctionnement d’un réseau de neurones capable d’extraire les saillances d’une image et de se focaliser sur un stimulus en mouvement. Les résultats expérimentaux obtenus sur des images naturelles montrent que le réseau décrit est robuste et rapide.

KEY WORDS

Attention - Vision - Neurones impulsionnels - Saillance

1 Introduction

Les récentes avancées en neurosciences computationnelles ont permis de mieux cerner les mécanismes intervenant dans les premières étapes des traitements perceptifs. Les approches bio-inspirées en vision artificielle ont tiré parti de ces avancées pour proposer différents modèles d’attention visuelle [1, 2]. Ces approches bio-inspirées visent un compromis entre réalisme biologique et efficacité computationnelle.

Nous nous sommes intéressés dans ce travail aux mécanismes de “covert attention”, qui désignent la situation dans laquelle le focus attentionnel se déplace sans induire de mouvement oculaire. L’attention sélectionne une partie de la scène visuelle, le focus attentionnel, et est un processus descendant, impliquant des connaissances préalables et des processus de plus haut niveau. La préattention est un processus similaire mais qui est purement ascendant et ne prend en compte que les informations directement disponibles dans les entrées sensorielles. Il existe plusieurs théories explicatives de ce processus préattentionnel [3, 4] et qui s’appuient sur la notion de modalités visuelles, comme par exemple le contraste, l’orientation, le mouvement, les oppositions de couleurs [5, 6]. Les régions d’une scène visuelle sélectionnées par le processus préattentionnel sont appelées saillances. L’hypothèse la plus couramment admise est que les saillances sont regroupées sur une carte de saillance mais dont l’existence n’a pas été

clairement démontrée [7].

Nous proposons d’implémenter un mécanisme préattentif de “covert attention” avec un réseau de neurones impulsionnels. Ce type de modèle neuronal capture un aspect fondamental du fonctionnement des neurones biologiques : ils peuvent coder l’information sous forme d’événements discrets – les trains de spikes. De plus, même avec des modèles impulsionnels très simples, le choix des constantes de temps permet de modifier le comportement d’un neurone : il peut alors se comporter comme un intégrateur temporel ou comme un détecteur de synchronie [8]. Nous avons utilisé ces deux comportements et le codage impulsionnel pour proposer un réseau de neurones qui extrait les saillances d’une image et se focalise sur la saillance la plus importante (Sect. 2). Pour l’évaluation, nous avons enregistré une séquence vidéo sur laquelle on voit un robot se déplacer et les résultats expérimentaux montrent que le réseau est capable de se focaliser sur le robot et de le suivre correctement tout au long de son déplacement (Sect. 3).

2 Description du modèle

Nous avons utilisé des neurones du type intègre-et-décharge à fuite (*Leaky Integrate-and-Fire*, LIF) [9], sans modélisation de la conductance synaptique. Le réseau proposé utilise un simulateur *clock-based* que nous avons développé et qui ne traite que les neurones actifs, c’est-à-dire les neurones qui intègrent des potentiels post-synaptiques (PSP).

Le réseau, schématisé sur la Fig. 1, est un ensemble de cartes neuronales – des couches de neurones 2D – dans lequel on peut distinguer deux voies différentes pour traiter les hautes (HFS) et basses fréquences spatiales (BFS). Les images présentées en entrée du réseau sont décomposées selon leurs différentes modalités visuelles par des filtres neuronaux, expliqués dans la Sect. 2.1, et ces différentes modalités visuelles sont ensuite combinées sur la carte de Saillance (Sect. 2.2). Cette combinaison s’appuie sur le traitement temporel de l’information, les basses fréquences spatiales étant combinées sur la carte de PreAc, qui précède la carte de Saillance. La carte de Focus choisit ensuite la région la plus saillante comme focus d’attention (Sect. 2.3).

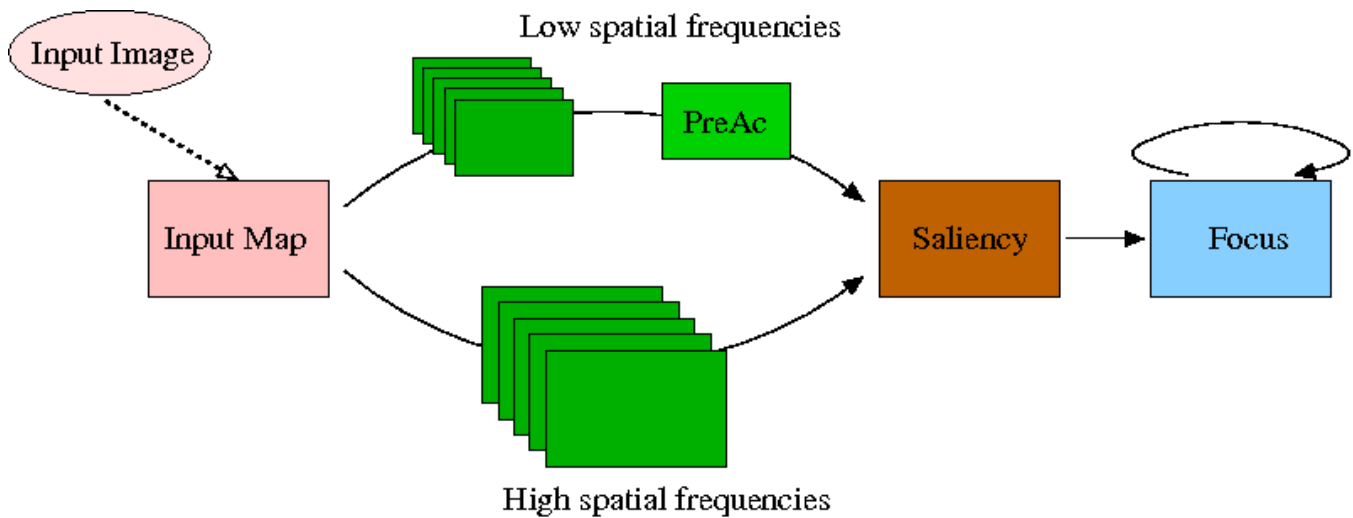


Figure 1. Structure du réseau *feedforward*, composé de cartes de neurones impulsionnels, qui extrait les saillances d'une image d'entrée et se focalise sur la région la plus saillante. L'image d'entrée est traitée pour les hautes (HFS) et basses (BFS) fréquences spatiales.

2.1 Filtres neuronaux

Les images d'entrée sont transformées en trains de spikes, en fonction de la luminance et des couleurs. Chaque neurone de la carte d'Entrée (Fig. 1) est associé au pixel qui lui correspond et la luminance (ou la couleur) détermine le courant qui sera appliqué sur le neurone. Pour une image de taille $L \times H$, il y a huit cartes $L \times H$ et six cartes $\frac{L}{2} \times \frac{H}{2}$.

Les neurones de la carte d'Entrée sont connectés aux deux voies, HFS et BFS, au travers de masques de connexion. Ces masques sont des matrices de poids statiques incluant des délais et qui définissent une projection générique liant deux cartes neuronales. Les valeurs de ces matrices de poids sont similaires aux noyaux de convolution utilisés en traitement d'images. Nous avons utilisé comme noyaux une différence de Gaussiennes (DoG) et quatre Gabor orientés pour les masques de connexion entre la carte d'Entrée et les cartes des voies HFS et BFS.

Ce réseau implémente un filtrage d'image similaire à celui réalisé par des noyaux de convolution classiques. Cependant, la "convolution" réalisée par la propagation des PSP au travers des masques de connexion est appliquée dans un ordre qui dépend de la valeur d'entrée, c'est-à-dire que les coefficients les plus importants sont traités les premiers. De plus, les valeurs d'entrées les plus faibles ne sont pas traitées, car seules les entrées permettant aux neurones de décharger sont propagées. Ce filtrage fonctionnel et le fait que l'implémentation réalisée ne traite que les neurones recevant des PSP permet une exécution rapide de ces traitements (voir la Sect. 3).

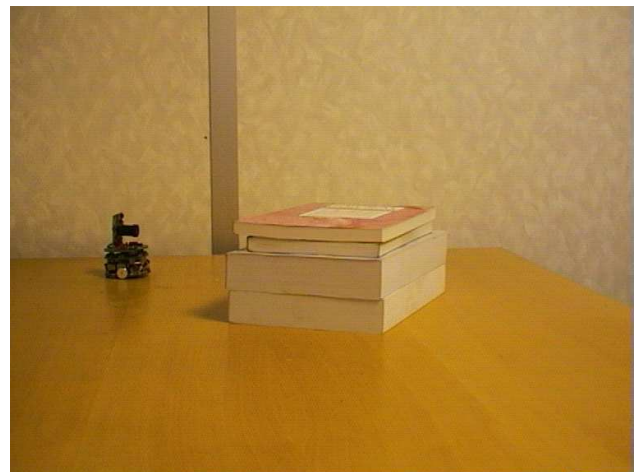


Figure 2. Image extraite de la séquence vidéo. Un robot se déplace dans la scène visuelle et constitue le stimulus à détecter.

2.2 Carte de saillance

La carte de Saillance (Fig. 1) regroupe toutes les informations issues des différentes modalités visuelles de chaque fréquence spatiale. Les neurones de la carte de Saillance ont été réglés pour se comporter comme des détecteurs de synchronie. Ainsi, les saillances émergent de cette carte seulement si une localisation spatiale donnée reçoit des spikes de différentes cartes neuronales codant pour différentes modalités visuelles.

Un point saillant est représenté par un spike émis par un neurone de la carte de Saillance. De cette manière les saillances sont encodées temporellement et surviennent de

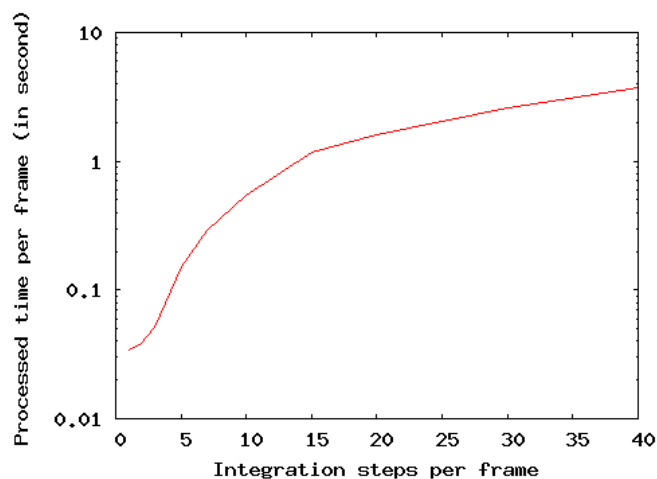
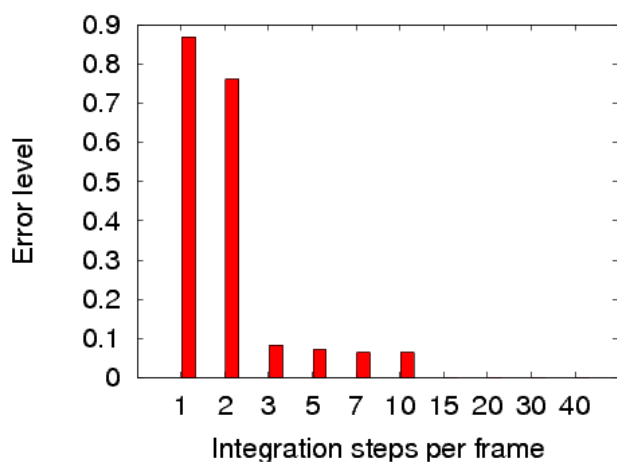


Figure 3. *Gauche* : Niveau d'erreur exprimé pour différents pas d'intégration par trame ($N = 1, 2, 3, 4, 5, 7, 10, 15, 20, 30, 40$). *Droite* : temps moyen écoulé (en seconde, échelle logarithmique) pour le traitement d'une trame en fonction du nombre de pas d'intégration par trame.

façon ordonnée. En effet, le neurone qui code pour la saillance la plus forte sera le premier à émettre un spike, et ainsi de suite. Comparativement, les autres approches bio-inspirées [2] utilisent un algorithme de *Winner-Takes-All* (WTA) pour trouver la saillance la plus forte. La détermination de la seconde saillance nécessite d'inhiber la saillance précédente et de relancer l'algorithme de WTA. Le réseau présenté utilise une carte neuronale récurrente pour implémenter un processus de sélection et les résultats expérimentaux montrent que ce type de réseau implémente un WTA implicite et rapide.

2.3 Covert Attention

Quand une saillance est détectée, les spikes de la carte de Saillance sont envoyés à la carte de Focus, qui est auto-connectée (Fig. 1). Le masque d'auto-connexion est un DoG, ce qui excite les voisins proches et inhibe les voisins éloignés. Cette carte neuronale récurrente a besoin des spikes de la carte de Saillance pour maintenir une activité stable et ne peut maintenir une activité auto-entretenue. Si la saillance bouge, l'activité de la carte de Focus la suit aussi longtemps que la saillance reste dans la partie positive de la DoG [10, 11].

3 Résultats expérimentaux

Nous avons utilisé des images acquises avec une caméra Sony EVID31 pour évaluer ce réseau de neurones. Il s'agit d'une séquence de 30 trames où l'on peut voir un robot Khepera se déplacer de gauche à droite et en profondeur. Durant cette séquence le réseau se focalise sur le stimulus (le robot) et le suit. Les trames sont des images de 760x570 pixels et sont réduites à 76x56 pixels pour servir d'entrée au réseau de neurones. Comme un pixel correspond à un

neurone, le réseau est composé d'environ 53 000 neurones. Pour ces expériences, nous avons utilisé uniquement les informations de luminance.

La première trame est présentée au réseau pendant 20 pas d'intégration. Cette étape d'initialisation amorce l'activité du réseau et permet de laisser la carte de Focus émettre ses premiers spikes. Chaque trame est ensuite présentée pendant N pas d'intégration de notre simulateur. Les résultats pour les différentes valeurs de N sont présentés sur la Fig. 3.

Pour évaluer les performances de ce réseau, nous avons utilisé la distance euclidienne entre le barycentre du stimulus et le barycentre de l'activité comme mesure d'erreur. Le barycentre du stimulus correspond au barycentre de l'ensemble des pixels du stimulus pour une trame donnée. Le barycentre de l'activité est défini comme étant le barycentre de tous les spikes émis par la carte de Focus pendant les N pas d'intégration correspondant à la présentation d'une trame donnée.

Le diagramme de gauche de la Fig. 3 montre que le niveau d'erreur décroît quand le réseau dispose de plus de temps (de pas d'intégration) pour traiter chaque trame. Ce niveau d'erreur décroît rapidement dès qu'il y a suffisamment de pas de temps d'intégration pour traiter chaque trame. Il est intéressant de remarquer que le niveau d'erreur est peu élevé pour 3 pas d'intégration par trame, surtout si l'on se réfère au temps de traitements correspondant, présenté sur la partie droite de la Fig. 3. Tous les résultats présentés ont été obtenus avec une machine de bureau (Intel Core2 cadencé à 1,86 GHz). Le temps de traitement d'une trame, en utilisant 3 pas d'intégration par trame, montre que le réseau peut traiter 20 trame par seconde en maintenant un faible taux d'erreur. Ces résultats confirment que les réseaux de neurones impulsionnels peuvent être utilisés dans un contexte temps réel.

4 Conclusion

Nous avons proposé un réseau de neurones impulsionnels capable trouver les saillances d'une image et de se focaliser sur la plus importante. Le comportement de détecteur de synchronie des neurones de la carte de Saillance permet d'extraire les saillances. L'implémentation du processus de "covert attention" est rendu possible par les traitements temporels réalisés par les neurones impulsionnels. L'évaluation de ce réseau a été réalisée avec des images naturelles et a montré que le réseau était capable de se focaliser sur un stimulus en mouvement. Les temps de traitements mesurés indiquent que ce réseau est utilisable dans un environnement temps réel.

References

- [1] Laurent Itti and Christopher Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.
- [2] L. Itti. Models of bottom-up attention and saliency. In L. Itti, G. Rees, and J.K. Tsotsos, editors, *Neurobiology of Attention*, pages 576–582. Elsevier, San Diego, USA, 2005.
- [3] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.
- [4] J. Wolfe. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1(2):202–238, 1994.
- [5] A. Treisman and J. Souther. Search asymmetry: a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology - General*, 114:285–310, 1985.
- [6] J. Wolfe. Visual attention. In *Seeing*, pages 335–386. Academic Press, 2ème édition, 2000.
- [7] Jillian H. Fecteau and Douglas P. Munoz. Saliency, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, 10(8):382–390, 2006.
- [8] P. König, A.K. Engel, and W. Singer. Integrator or coincidence detector? the role of the cortical neuron revisited. *Trends in Neurosciences*, 19(4):130–137, 1996.
- [9] Wulfram Gerstner and Werner Kistler. *Spiking Neuron Models: Single Neurons, Population, Plasticity*, chapter 4.1. Cambridge University Press, New York, NY, USA, 2002.
- [10] S. Chevallier and P. Tarroux. Visual focus with spiking neurons. In M. Verleysen, editor, *European Symposium on Artificial Neural Networks (ESANN)*, pages 385–390, 2008.
- [11] S. Chevallier and P. Tarroux. Covert attention with a spiking neural network. In A. Gasteratos, M. Vincze, and J.K. Tsotsos, editors, *Int. Conf. on Computer Vision Systems (ICVS)*, volume 5008 of *Lecture Notes in Computer Science*, pages 56–65. Springer, 2008.