ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Visual focus with spiking neurons

Sylvain Chevallier[1,2] and Philippe Tarroux[2,3]

1- Université Paris-Sud XI
Orsay - France

2- LIMSI - CNRS UPR 3251
Orsay - France

3- École Normale Supérieure
Paris - France

{sylvain.chevallier,philippe.tarroux}@limsi.fr

Abstract. Attentional focusing can be implemented with a neural field [1], which uses a discharge rate code. As an alternative, we propose in the present work an implementation based on spiking neurons. Such implementation will allow to investigate the possible contribution of a spike time based code with a network of leaky integrate-and-fire neurons. The network is able to detect and to focus on a stimulus even in the presence of distractors. Experimental data show that this behavior is very robust to noise. This process implements an early visual attention mechanism.

1 Introduction

The attentional process is a key concept for biological vision efficiency [2] and understanding such process brings new solutions and approaches in artificial vision (see [3] for a review). In a recent work, Rougier & Vitay [1] described a model of attention based on the emergent properties of a neural population, which proved experimentally to be very robust.

Since neural fields rely on a rate code rather than on a precise spike-timing code, we propose here an implementation of the neural population described in [1] with a spiking neural network. Spiking models discretize the information as only spikes are propagated and thus we will be able to investigate the contribution of this process. The neural model and the network are described in Sect. 2. We evaluate the robustness of our network with noise and distractors. Results are detailed in Sect. 3.

2 Network description

Rougier *et al.* use the Continuum Neural Field Theory (CNFT) to describe the activity of their neural population. The CNFT can characterize the dynamic of pattern formation for a neural field with lateral inhibition. The authors use the dynamic properties of this neural population to focus and follow an input stimulus. Their experiments rely on the spatio-temporal continuity of the stimulus during the simulation.

Instead of using a rate code, as in neural field, the network described here is composed of Leaky Integrate-and-Fire (LIF) neurons that emit their spikes at a

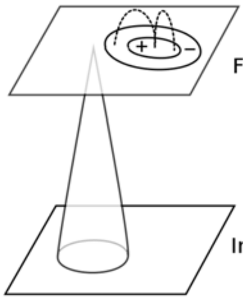

Fig. 1: Network structure: two 2D neural maps, the *Input map* convert input image in spike trains send to *Focus map* through a Gaussian connection. *Focus map* has self connections which implement a lateral inhibition.

precise time. The membrane potential V_i of neuron i is given by the following differential equation:

$$\begin{cases} \tau \dot{V}_i = g_{\text{leak}}(V_i - E_{\text{leak}}) + \text{PSP}_i(t) + \text{I}(t), \text{ if } V \leq \vartheta \\ \text{spike and reset } V \text{ otherwise} \end{cases} \tag{1}$$

where τ is the membrane time constant, g_{leak} is the membrane leak conductance, ϑ is the threshold and E_{leak} is the membrane resting potential [4]. $\text{I}(t)$ represents the influence of a external input current (see [4]). The $\text{PSP}(t)$ is the synaptic input function, describing the influence of incoming spikes on membrane potential. As in [5], there is no synaptic conductance in our model. Formally, outgoing PSPs from neuron j are denoted by:

$$S_j(t) = \sum_f \delta(t - t_j^{(f)} + d_j) \tag{2}$$

where $\delta(x)$ is the Dirac distribution, with $\delta(x) = 0$ for $x \neq 0$ and $\int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1$, $t_j^{(f)}$ is the spike emission time and d_j the synaptic delay. The influence of incoming PSPs on membrane potential is given by the simple relation:

$$\text{PSP}_i(t) = \sum_j w_{i,j} S_j(t) \tag{3}$$

The network is a set of two neural map (2D neural layer): an Input map (IM) and a Focus map (FM), see Fig. 1. Depending on the choice of g_{leak} and E_{leak} values, a spiking neuron can either integrate the information over a predefined

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

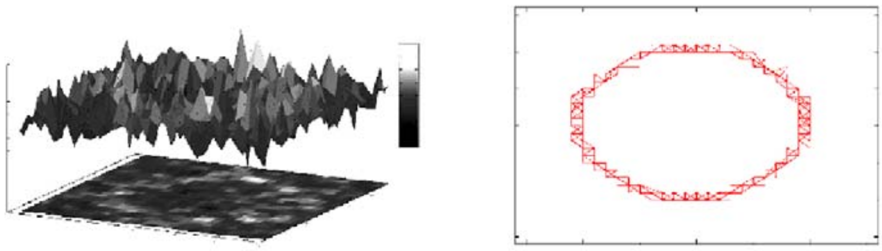

Fig. 2: Left: Mean activity of the Input map when the background is noisy ($\sigma = 0.5$). The activity along the circular path of the stimulus is slightly higher as it can be seen on the bottom projection. Right: Path followed by the FM activity centroid with a background noise ($\sigma = 0.5$).

temporal window or act as a synchrony detector, i.e. emitting spikes when inputs are condensed in a small period of time. IM neurons behave as integrator and FM neurons as synchrony detectors.

The IM translates an input image into spike trains. Each neuron is associated with a pixel, i.e. the pixel luminance determines the input term I(t) of the corresponding neuron. Each IM neuron is connected to FM neurons through a Gaussian mask. A mask is a static weight matrix and defines a generic projection from one neural map to another. The weight matrix values of the Gaussian mask can be viewed as the parameters of a Gaussian image filter.

The FM is self-connected with a difference of Gaussian (DoG, also known as Mexican hat) connection mask, which excites adjacent neighbors and inhibits distant ones. This self-connection alone is not sufficient to maintain a self-sustained activity. For this purpose, it needs the IM spikes to keep an ongoing activity. At time t, the stimulus generates an activity on FM and, when at $t+\Delta t$ the stimulus has moved, the activity on FM follows as long as the stimulus stays in the excited part of this activity.

The simulation is handled in our clock-based simulator. This simulator process only active neurons, i.e. neurons integrating PSPs [6]. The simulator version used in this paper is sequential and is not distributed.

3 Experimental results

We use an experimental set similar to [1]: a stimulus follows a circular path on a 30x30 pixels input image with either noise or distractors in the background. The stimulus is a Gaussian patch with an amplitude of 1. Distractors are exact copies of the stimulus but they lack spatio-temporal continuity, as they constantly appear and disappear in random places without following a continuous path. The added noise is assumed to be independent and drawn from a zero-mean

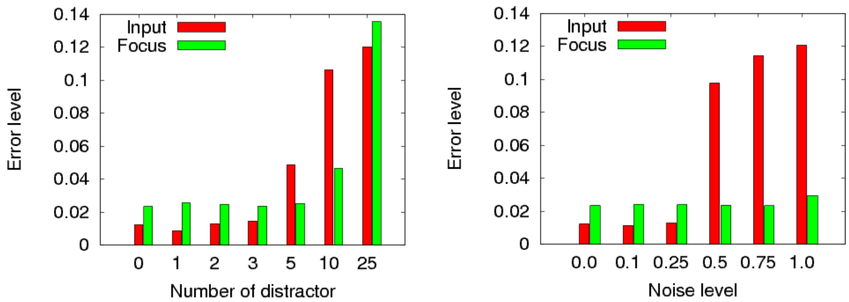

Fig. 3: Left: error level when the stimulus is presented along with distractors (0, 1, 2, 3, 5, 10 and 25). The dark red bars are for Input map and the light green ones are for the FM. Right: error level when the input image is noisy, the x-axis display the different values of σ (0.0, 0.1, 0.25, 0.5, 0.75 and 1.0). Same color coding as the left part.

Gaussian distribution with different variance levels σ. Consequently, the $\mathrm{I}(t)$ term of Eq. 1 can be express as:

$$\mathrm{I}(t) = G(x, y; \theta(t)) + \eta \tag{4}$$
$$\eta \sim N(0, \sigma) \tag{5}$$

where $G(x, y; \theta(t))$ describes the circular path of the Gaussian patch.

The network is bootstrapped with an input image containing only the stimulus (as in [1]) until the first spikes appear on the FM (usually 20 computations steps). Then, the stimulus begins to move in a noisy background or among distractors. Each image is presented during 10 computations steps.

To validate the performance of the network, we compute the distance between the stimulus center and the centroid of the activity. The activity centroid is defined as the centroid of all emitted spikes by the FM during the integration steps of the image presentation (see Fig. 2-right and 3).

The errors of FM remain low even in the presence of high background noise, as shown in Fig. 2-left. The IM is more sensitive to noise or distractor presence whereas the FM keeps a constant error level and is robust to perturbation. One can note that on Fig. 3, in the conditions of small number of distractors or low noise level, the error of FM is higher than the error on IM. This effect is a result of the discretization induced by the small size of the input image and it is reduced when the image resolution is increased.

The implementation used in this paper is quite fast since one computation step takes from 1 to 9 milliseconds, depending on the overall activity of the network. Thus, the network can process up to 100 images per second on a Intel Core2Duo at 1,8 GHz. These values are not so surprising because our network is small, roughly 2,000 neurons, but when the image size is increased the network scales well [7].

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

4 Conclusion

The presented network of LIF neurons is able to focus on a target stimulus and stays focused even when the stimulus moves. Noisy background or distractors have only a small influence on the behavior of the network. Our choice of a neural model without synaptic conductance prove to be efficient for achieving fast computation time.

The experimental results are close to those observed with a neural field [1]. However, some important differences can be emphasized, such as the filtering done by the Input map neurons. The lowest luminance values are not processed, due to the discrete nature of spiking neurons. Only above threshold information is propagated into the network. This reduces the computational load of the overall network.

Another difference concerns the complexity: a neural field with DoG lateral connections implements a $O(n)$ algorithm for each spatial position considered in the discretized equations[1]. Our network is also in $O(n)$ (see [8] for a complexity analysis of clock driven algorithms), but only the active neurons are processed. Only a small set of neurons are active (i.e. process PSP) at each computation step, thanks to the threshold mechanism of the Input neurons.

This network implements an early visual attentional mechanism and can be used in more complex architectures. We work on a larger spiking neural network [6], which uses this focus process to implement covert attention [7]. This network seems suitable for a real-time implementation in robotic frameworks.

References

[1] Nicolas Rougier and Julien Vitay. Emergence of attention within a neural population. *Neural Networks*, 19(5):573–581, 2006.

[2] John K. Tsotsos. Analysing vision at the complexity level. *Behavioral and Brain Sciences*, 13(1):423–469, 1990.

[3] Laurent Itti, Geraint Rees, and John K. Tsotsos, editors. *Neurobiology of Attention*, chapter Models of Bottom-Up Attention and Saliency, pages 576–582. Elsevier, San Diego, CA, January 2005.

[4] Wulfram Gerstner and Werner Kistler. *Spiking Neuron Models: Single Neurons, Population, Plasticity*, chapter 4.1.1. Cambridge University Press, New York, NY, USA, 2002.

[5] Nicolas Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8:183–208, 2000.

[6] Sylvain Chevallier, Philippe Tarroux, and Hélène Paugam-Moisy. Saliency extraction with a distributed spiking neural network. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks*, pages 209–214, Bruges, Belgium, 2006.

[7] Sylvain Chevallier and Philippe Tarroux. Covert attention with a spiking neural network. In *6th International Conference on Computer Vision Systems*, Santorini, Greece, 2008.

[8] Romain Brette. Exact simulation of integrate-and-fire models with synaptic conductances. *Neural Computation*, 18(8):2004–2027, 2006.

[1] for a n x n connectivity filter and only if this filter is separable.