Visual Selection with Feature Contrast-Based Inhibition in a Network of Integrate and Fire Neurons

Marcos G. Quiles, Fabricio Breve, Roseli A. F. Romero, and Liang Zhao
Institute of Mathematics and Computer Science (ICMC)
University of São Paulo (USP)
São Carlos, SP
{quiles, fabricio, rafrance, zhao} @ icmc.usp.br

Abstract

In this paper a visual selection mechanism based on an integrate and fire neural network is proposed for selecting objects in a given visual scene. In comparison to other visual selection approaches, our model is able to capture attention of objects in complex forms, including those linearly non-separable, and also processes a combination of features of an input scene, such as intensity, color and orientation. Moreover, computer simulations show that the model produce results similar to those observed in natural vision systems.

1. Introduction

Attention is an important mechanism used by biological systems to reduce the amount information obtained from the environment. By selecting just part of the input data available on the retina, the brain is able to focus its limited computational capacity on a specific task while ignoring irrelevant information [1, 2]. This process seems to optimize the search procedure by selecting a number of possible candidate images and feature subsets which can be used in more complex and specialized tasks such as object recognition [3].

Visual attention is mainly generated by a combination of two processes: information from the retina and early visual cortical areas (called *bottom-up attention* or *scene dependent attention*) and feedback signals from areas outside of the visual cortex (called *top-down attention* or *task dependent attention*) [4, 5].

Most of the bottom-up visual attention models are related to the concept of a Saliency Map [4]. In those models, the first stage of processing is responsible to decompose the input image into a set of feature maps. After that, a saliency map is generated by a combination of those feature maps. The saliency map is a topographical map which represents, by a scalar quantity, all salient points over the entire input visual stimulus [5, 4]. The main purpose of saliency map is to guide a selection mechanism to deliver the focus of attention to a specific region of the image.

The selection mechanism of several models is based on a Winner-Take-All (WTA) neural network, where just one neuron is activated by means of a competition among all neurons in the network. As a result, the winner is a single neuron representing a point or a very small area in the given scene but not a whole object or a whole component. In this way, it is not possible to delivery the focus of attention to complex form objects such as linearly non-separable objects. Moreover, substantial data from biology has supported that the selection is not only delivery to an specific region of the visual input but to entire objects or regions [1, 6]. It has been suggested that the visual system perform a type of preattentive segmentation before delivering the attention to a region of interest. Thus, to build a visual selection model where the competitive process is performed by objects and not by single neurons is apparently interesting. For this purpose, the visual attention model must have a selection mechanism and, at the same time, a way for representing objects.

von der Malsburg [7] proposed a mechanism of temporal correlation as a representational framework. This theory suggested that objects are represented by the temporal correlation of the firing activities of spatially distributed neurons coding different features of an object. Inspired by biological findings and von der Masburg's brain correlation theory, Wang and his collaborators have developed oscillatory correlation theory [8, 9, 10], which can be described by the following rule: neurons which process different features of the same object are synchronized, while neurons which code different objects are desynchronized. There are two basic mechanisms working simultaneously in each oscillatory correlation model: synchronization and desynchro-

978-0-7695-3304-9/08 $25.00 © 2008 IEEE
DOI 10.1109/ICNC.2008.209

nization. The former serves to group neurons into objects while the latter serves to distinguish one group of synchronized neurons (an object) from another. Oscillatory correlation theory has been extended and successfully applied to various tasks of scene analysis, such as image segmentation, motion determination, auditory signal segregation, and perception ([11] and references there in). Oscillatory correlation models have also been used to perform object selection [6]. This model [6] besides providing a temporal segmentation of the objects in the input image also provides a mechanism to perform object selection where the largest object keeps firing while the other remain silent. Although this model provides an interesting mechanism of competition among objects, it just considers the size as a salient feature.

In this paper, we propose an oscillatory correlation model for visual selection built on a network of Integrate and Fire neurons (I&F) with cooperative short-range connections and competitive long-range connections. In our model, as the system runs, each group of neurons representing an object of a visual input is synchronized due to the cooperative connections among neighbor neurons. At the same time, a competition mechanism is introduced by long-range connections among neurons. By means of such a competition mechanism, firing frequencies of neurons representing the salient object are increased, while frequencies of those neurons representing background objects are decreased. As a result, the neurons representing the salient object will keep firing and other neurons will slow down their firing activities until becoming silent. Another novelty of the proposed model is that a combination of several visual attributes, such as intensity, contrast of colors and orientations are considered. These features are among the most relevant features used by the visual system to guide the search for a visual target [12] and biological findings show that the perceptual system might encode contrast of features rather than the absolute level of them [13].

The rest of the paper is organized as follows. Section 2 is devoted to the model description. Section 3 presents computer simulation results and Section 4 concludes the paper.

2. Model Description

The visual selection model presented in this paper is formed by a 2D network of I&F neurons with two types of connections: excitatory short-range connections and inhibitory long-range connections. Excitatory connections are employed to synchronize group of neurons representing a coherent object. Inhibitory connections are responsible to desynchronize different groups of neurons (segmentation) and also to inhibit background objects permitting the salient object to be highlighted.

The network of I&F neurons is defined by the following equation:

$$\frac{dv_i}{dt} = -v_i + I_i(t) + E_i(t) - Y_i(t) \tag{1}$$

where v_i is the neuron potential, I_i defines the external stimulation responsible for controlling the firing frequency of neuron i, $E_i(t)$ defines the excitatory coupling and $Y_i(t)$ defines the inhibitory coupling among neurons. The neuron i fires every time that $v_i \geq \theta$, where θ is the firing threshold. Without the coupling terms $E_i(t)$ and $Y_i(t)$, and taking $I_i(t)$ as a constant, Eq. (1) is a standard I&F neuron.

The excitatory coupling term $E_i(t)$ is defined by:

$$E_i(t) = \sum_{j \in \Delta_i} \omega_{ij} \delta(t - t_j) \tag{2}$$

where δ is the Dirac delta function, t_j represents the instant when neuron j fires, Δ_i defines the excitatory cooperation neighborhood of neuron i, which consists of the 8 nearest neighbors of neuron i. ω_{ij} is the excitatory coupling strength between neurons i and j and it is defined by:

$$\omega_{ij} = \frac{c_E}{|\Delta_i|} \tag{3}$$

where $c_E \in [0, 1]$ is a constant and $|\Delta_i|$ is the number of neurons in Δ_i. If we take the excitatory connection as a graph, $|\Delta_i|$ represents the degree of neuron i.

The inhibitory coupling term is defined by:

$$Y_i(t) = \sum_{j \in \Lambda_i} \sigma_{ij} \delta(t - t_j) \tag{4}$$

where Λ_i defines the competition neighborhood of neuron i and σ_{ij} is the inhibitory coupling strength between neurons i and j, which is defined by the following equation:

$$\sigma_{ij} = c_Y \exp\left(-\sum_k c_{ij}^k f_{ij}^k\right) \tag{5}$$

where $c_Y \in [0, 1]$ is a constant, f_{ij}^k and c_{ij}^k ($k \in [1, n]$) represent the contrast of the feature k (intensity, color or orientation) and its weight, respectively. The contrast is defined as the absolute difference of the feature k between neuron i and j, i.e.,

$$f_{ij}^k = |f_i^k - f_j^k| \tag{6}$$

In order to perform the visual selection of an object we introduce a mechanism to control the firing frequency of each neuron. The frequency is increased if the neuron is part of the salient object otherwise it is decreased. This is realized by modeling the $I_i(t)$ term of Eq. (1). Each time a neuron i fires, it increases the value of its own $I_i(t)$ by the following equation:

$$I_i(t) = c_I(I_{max} - I_i(t-1)) \tag{7}$$

where c_I and I_{max} are constants which defines the potential gain factor and the maximum value that $I_i(t)$ can hold, respectively.

To decrease the value of $I_i(t)$ we consider the following equation:

$$I_j(t) = c_Y(I_{min} - I_j(t-1)) \quad (8)$$

where c_Y is the same inhibitory coupling strength constant defined in Eq. (5) and I_{min} is the minimum value that $I_j(t)$ can hold. Each time a neuron j receives an inhibition signal from a neuron i, its $I_j(t)$ value is decreased according to Eq. (8). Considering the firing threshold $\theta = 1$, if $I_j(t) < 1$ the neuron j stopping firing and remains silent. Due to these properties, only the neurons representing the salient object continue firing while the others are slowed down until becoming silent.

According to the input image pattern, the connections among neurons are set. The excitatory connections are created between neurons with similar input features, it means, if $f_{ij}^k = |f_i^k - f_j^k| < \theta^k$ the neurons i and j are connected, this process is done for the 8 nearest neighbors j of the neuron i. Based on these cooperative connections, neighbor neurons with similar inputs will synchronize and the trajectory of these neurons will represent a unique object of the input image.

The long-range inhibitory connections are defined based on the contrast of the input features (Eqn. 6). For example, if two neurons i and j have similar attributes, the f_{ij}^k term will hold a small value and due to the negative exponential of Eqn. (5), the inhibition between both will have a high value. The negative exponential of Eqn. (5) has the function of creating an inhibition signal which is high just when both neurons have almost the same features; otherwise, they will not inhibit each other or will do it with a small strength. Thus, the inhibitory signal is responsible for implementing the contrast into our model. It is worth noting that this inhibitory signal acts even among neighbor neurons, but, due to the excitatory connections which have a stronger weight, those neurons remain synchronized.

3. Computer Simulations

Given a static color image as input, each selected feature is computed using the three features provided by the image pixels: F_R, F_G, F_B, or red, green, and blue channels. Based on this three features intensity and local orientations are extracted.

The intensity is computed as $F_I = (F_R + F_G + F_G)/3$. The local orientations are obtained from F_I by means of the application of a laplacian filter followed by a convolution with four spatial masks with orientation 0^o, 45^o, 90^o, and 135^o. The neural network is set and the neurons are integrated using a fourth-order Runge-Kutta Method. For the excitatory connections just the intensity of pixels are consider in all experiments. For the inhibitory connections, all features are taken into account and the following weights are assumed for each feature: $c_{ij}^i = 1.0$; $c_{ij}^r = 1.0$; $c_{ij}^g = 1.0$; $c_{ij}^b = 1.0$; $c_{ij}^{0^o} = 0.25$; $c_{ij}^{45^o} = 0.25$; $c_{ij}^{90^o} = 0.25$; $c_{ij}^{135^o} = 0.25$. The network parameters are set as follow: $I_0 = 1.1$; $I_{Min} = 0.9$; $I_{Max} \in \{1.6; 2.0\}$; $c_E = 0,2$; $c_Y = \frac{0.03}{N}$, where N is the number of neurons in the network; $\theta^k = 0.1$.

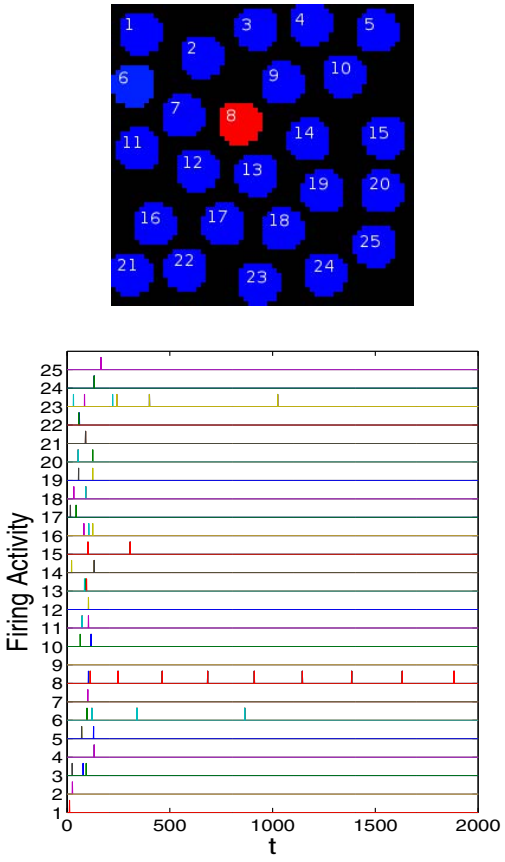

Figure 1. (a) Input image; (b) Temporal activities of neuron blocks. Each trace in the figure corresponds to a pulse train of an object in the input pattern.

For all the simulations, the salient object was considered to be the one which shows the largest contrast to the background objects. This assumption matches well with biological findings supporting that the contrast of features is highly used by the visual perception system to perform visual searching tasks [12, 13]. Figure 1(a) shows the input image used to check our model with contrast of color. Here, the red object seems to pop-out from the background com-

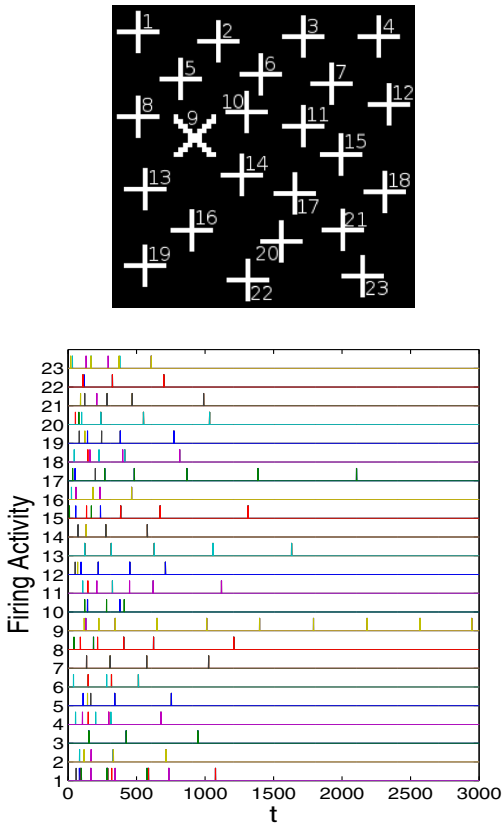

Figure 2. (a) Input image; (b) Temporal activities of neuron blocks. Each trace in the figure corresponds to a pulse train of an object in the input pattern.

posed of tones of blue objects becoming the salient one. In Figure 1(b) is presented the time series of the spiking activity of some neurons representing all the objects observed in the input image. Here we can see that, after some cycles, just one group of neurons remains activity, which is exactly the group of neurons that represents the red object.

The second simulation was performed using the Figure 2(a) where the orientation defines the most salient object. The result of this simulation can be seen in Figure 2(b), where the neurons which remain active represents the 'X' pattern. A more complex simulation was performed using Figure 3(a). In this figure, the salient object, it means, the object which contrast most with the others does not seem to pop-out from the background easily. It happens because there is no unique feature that defines this object. Here, the salient object is defined based on a conjunction of features (color and orientation), named *conjunction search*. For example, we can say that, first, our attention is directed to the red objects and then another process based on orientation guides our attention to the red 'X' object, which is the one whose shape/color differs most from the background objects. However, our network does not perform this type of serial selection for an specific feature but realizes a parallel competition among all presented features. It means, due to the differences between orientation and color of the red 'X' against the others, the red 'X' object is select because it receives, on average, less inhibitory signal than the other objects. In Figure 3(b), we can see that just one group of neurons remains active during the simulation and this group is the one that represents the red 'X' object.

A final experiment was performed in order to check the proposed model with real images. In Figure 4(a) we show the input image which consists of a some green objects and a orange fruit. Due to the background color similarity, the fruit seems to pop-out becoming the selected object. In Figure 4(b) we can see that after some iterations, just the neurons representing the orange fruit remains active.

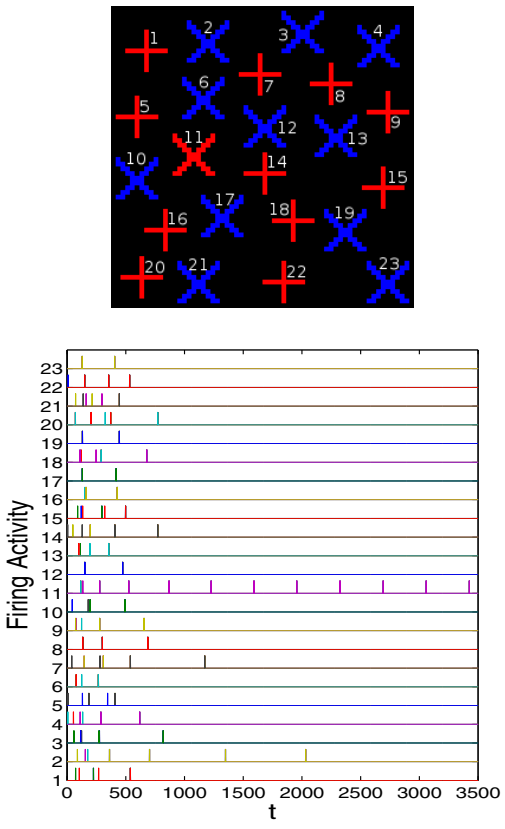

Figure 3. (a) Input image; (b) Temporal activities of neuron blocks. Each trace in the figure corresponds to a pulse train of an object in the input pattern.

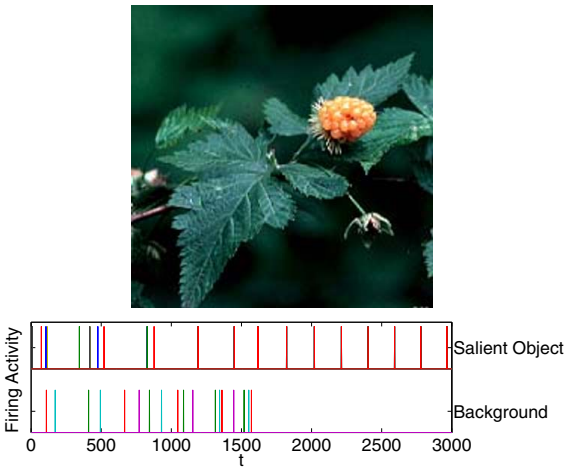

Figure 4. (a) Input image; (b) Temporal activities of neurons representing the background objects and the salient object.

4. Conclusions

This paper presents a visual attention mechanism realized by a network of integrate and fire neurons. The combination of contrast of features of an input image, such as intensity, color and orientation are considered to stimulate the corresponding neurons in the network. The local connections among neighbor neurons serve to synchronize neurons representing a coherent object in a given input image, while those long-range coupling terms have a function to select the salient object by inhibiting distracters, i.e., lower the contrast between an object and other part of the input image is, stronger the inhibitory signals the corresponding neurons receive.

Computer simulations show that the model is able to select the most salient object in an input image based on the contrast of features. The results obtained by our model matches well with biological experiments with humans considering the correct object selection though quantitative analysis considering the reaction time needed to select the salient object were still not taken into account.

Another interesting point of our modeling is related with the Eq. (5), where it is easy to insert top-down mechanism by controlling the weights of each input feature. In this way, one can facilitate the selection of specific group of features changing their respective weights.

5. Acknowledgment

This work is supported by the São Paulo State Research Foundation (FAPESP) and the Brazilian National Research Council (CNPq).

References

[1] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual Review of Neuroscience*, vol. 18, pp. 193–222, 1995.

[2] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artificial Intelligence*, vol. 78, pp. 507–545, 1995.

[3] D. Walther, U. Rutishauser, C. Cock, and P. Perona, "Selective visual attention enables learning and recognition of multiples objects in cluttered scenes," *Computer Vision and Image Understanding*, vol. 100, pp. 41–63, 2005.

[4] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, pp. 194–203, 2001.

[5] L. Itti, C. Koch, and E. Niebur, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11.

[6] D. Wang, "Object selection based on oscillatory correlation," *Neural Networks*, vol. 12, pp. 579–592, 1999.

[7] C. von der Malsburg, "The correlation theory of brain function," Internal report 81-2: Max-Planck Institute for Biophysical Chemistry, Göttingen, Germany, Tech. Rep., 1981.

[8] S. R. Campbell, D. L. Wang, and C. Jayaprakash, "Synchrony and desynchrony in integrate-and-fire oscillators," *Neural Computation*, vol. 11, pp. 1595–1619, 1999.

[9] D. Terman and D. Wang, "Global competition and local cooperation in a network of neural oscillators," *Physica D*, vol. 81, pp. 148–176, 1995.

[10] D. Wang and D. Terman, "Image segmentation based on oscillatory correlation," *Neural Computation*, vol. 9, pp. 805–836, 1997.

[11] D. Wang, "The time dimension for scene analysis," *IEEE Transactions on Neural Networks*, vol. 16, no. 6, pp. 1401–1426, 2005.

[12] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it ?" *Nature Review Neuroscience*, vol. 5, pp. 495–501, 2004.

[13] S. Yantis, "How visual salience wins the battle for awareness," *Nature Neuroscience*, vol. 8, no. 8, pp. 975–977, 2005.