A Visual Selection Mechanism Based on Network of Chaotic Wilson-Cowan Oscillators

Marcos Quiles, Fabricio Breve, Liang Zhao, and Roseli Romero
Department of Computer Science
Institute of Mathematics and Computer Science
University of São Paulo - São Carlos, SP, Brazil
$\{$*quiles, fabricio, zhao, rafrance*$\}$*@icmc.usp.br.*

Abstract

In this paper, a Visual Selection Mechanism based on a lattice of coupled chaotic Wilson-Cowan oscillators is proposed. The oscillators representing each object in a given visual scene are synchronized to produce a chaotic trajectory. Cooperation and competition mechanisms are also introduced to accelerate oscillating frequency of the salient object as well as to slow down other objects in the same scene. The model can not only discriminate each object among others in a given visual scene, but also deliver the focus of attention to the salient object. In comparison to other visual selection approaches, this model presents at least two new features. First, it is able to highlight objects in complex forms, including those that are non-linear separable. Second, oscillators representing the salient object will jump from chaotic phase to periodic phase. This behavior matches well to biological experiments on pattern recognition of rabbit. Computer simulations are performed and the results show that the proposed model is promising as a Selection Mechanism embedded in a Visual Attention System.

1. Introduction

Visual attention is an efficient way that the biological systems, such as humans, have developed to address the reduction of provided visual information. This reduction is an essential mechanism due to the limited processing capacity of the neuronal hardware. Attention appears to optimize the search procedure by selecting a number of possible candidate image and feature subsets which can be used in tasks such as recognition [17]. It is an efficient mechanism to break down complex tasks, such as scene understanding, into a series of small localized computational tasks [7]. According to Tsotsos et al. [14], intermediate and higher visual processes seem to select part of the sensory information received from the world and use just these selected data in a further processing. Visual attention is also responsible for reducing the combinatorial explosion resulting from the analysis of all incoming sensory information and possible image relationships [11, 13] and for identifying the part of the visual input where the processing is performed at the same time which irrelevant visual information is suppressed [3].

Visual attention is generated by a combination of information from the retina and early visual cortical areas (bottom-up attention - scene dependent) as well as feedback signals from areas outside of the visual cortex (top-down attention - task dependent) [8, 9]. Bottom-up attention is formed by simple features extracted from the image, such as intensity, stereo disparity, color, orientation, and others [8]. All this information is combined to create a saliency map which represents the conspicuity points in the visual input. The top-down attention signals are responsible for modulating the competition of all the points generated by the saliency map. This information can be, for example, a visual search for a specific object or features into the visual input.

By using these two mechanisms: bottom-up and top-down mechanisms, several models have been proposed and can be divided in two different approaches. The first approach belongs to computational neuroscience area, where the computational models are realistic implementations of the biological systems and they are used to simulate and understand biological systems [4, 5, 8]. The second approach is related to computer vision, where the models are developed to reduce the amount of incoming data by selecting only part of the visual information for further processing improving the performance or efficiency of the system [3, 9, 10, 13, 14, 17]. The model introduced in this paper belongs to the second approach where computer vision tasks are taken into account.

0-7695-2976-3/07 $25.00 © 2007 IEEE
DOI 10.1109/ISDA.2007.74

Most of the bottom-up visual attention models are related to the concept of a Saliency Map [8]. In these 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 [8, 9]. The main purpose of this map is to guide a selection mechanism, which is responsible for delivering the focus of attention in a specific (or most conspicuity) region of the image.

According to Tsotsos et al. [14], the mechanisms used by the visual attention appear to have, among others, the following components: a Selection mechanism to select a region of interest in the visual input; selection and extraction of the features from the input signal; the control of information through the system; and a shifting mechanism responsible to change the focus of attention among the several conspicuity point over the image.

Although several models of bottom-up visual attention have been proposed, the Selection mechanism used by these models are implemented by a Winner-Take-All (WTA) neural network where just one neuron is activated but not the entire object becomes salient. For example, in the model proposed in [9], when a neuron receives the focus of attention, a circle with a fixed radius is considered to be the region of attention of the model. In this case, it is not possible to deliver the attention to complex objects that are non-linear separable. To deal with this limitation, new object selection mechanism should be developed.

von der Malsburg [15] 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. A natural way of encoding temporal correlation is by using synchronization of oscillators where each oscillator encodes some features of an object [12, 16, 18, 19]. Inspired from the biological findings and von der Masburg's brain correlation theory, Wang and his collaborators have developed oscillatory correlation theory for scene segmentation [1, 2, 12, 20], which can be described by the following rule: the 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 desynchronization. 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 ([18] and references there in). Another way to model oscillatory correlation and to achieve unlimited capacity of segmentation (number of objects can be segmented in a given scene) is to utilize the properties of chaos and chaotic synchronization [6, 21, 22]. In the model proposed by Zhao et. al. [22], a large number of locally coupled chaotic oscillators can be synchronized, so that each object in a given scene is represented by a synchronized chaotic trajectory in the corresponding network. Consequently, all such chaotic trajectories can be easily separated by the high sensitivity to initial conditions, which is the hallmark of chaos, and the fact that a chaotic trajectory is dense in its invariant set. With this procedure, the authors claim that the model has unbounded capacity of object segmentation.

In this paper, we construct a chaotic oscillatory correlation network for object selection. Our model is based on a network of coupled chaotic Wilson-Cowan oscillators [22]. Such a network is used to create a selection mechanism where one of several objects is highlighted (receives the focus of attention). As the system runs, each group of neurons representing an object of a visual input is synchronized and produce a unique chaotic trajectory. At the same time, a competition mechanism is also introduced, where synchronized neurons cooperate each other to accelerate their firing frequencies and slow down other neurons with different oscillating activities. Finally, the most salient object will jump to a high frequency periodic oscillating phase, while all other objects will be silent. In this paper, we consider pixel intensity as visual attention clue, i.e., the brightest object is considered as salient object.

The rest of the paper is organized as follows. In section 2, the segmentation process performed by a network of Wilson-Cowan oscillators is described and the proposed model are presented. In Section 3, the results obtained through simulation of the proposed model applied to synthetic images are showed. Finally, in Section 4, conclusions are drawn.

2. Model Description

In this section, we first present a network of coupled chaotic Wilson-Cowan oscillators for scene segmentation. Then we introduce a new selection mechanism, embedded in the segmentation network, to accomplish visual attention task.

2.1. Scene Segmentation using Coupled Wilson-Cowan Oscillators

The model is a two dimensional network governed by the following equations:

$$
\begin{aligned}
\dot{x}_{i,j} &= -ax_{i,j} + G(cx_{i,j} + ey_{i,j} + I_{i,j} - \theta_x) + k\Delta x_{i,j} \\
\dot{y}_{i,j} &= -by_{i,j} + G(dx_{i,j} + fy_{i,j} - \theta_y) + k\Delta y_{i,j} \qquad (1) \\
G(v) &= \frac{1}{1+e^{-(v/T)}}
\end{aligned}
$$

where (i, j) is a lattice point with $1 \leq i \leq N, 1 \leq j \leq M$. k is the coupling strength. $\Delta x_{i,j}$ and $\Delta y_{i,j}$ are coupling terms among excitatory units and inhibitory units, respectively. They are defined by

$$
\begin{aligned}
\Delta v_{i,j} = \; & \gamma_{i-1,j-1;i,j}(v_{i-1,j-1} - v_{i,j}) + \\
& \gamma_{i-1,j;i,j}(v_{i-1,j} - v_{i,j}) + \\
& \gamma_{i-1,j+1;i,j}(v_{i-1,j+1} - v_{i,j}) + \\
& \gamma_{i,j-1;i,j}(v_{i,j-1} - v_{i,j}) + \\
& \gamma_{i,j+1;i,j}(v_{i,j+1} - v_{i,j}) + \qquad (2) \\
& \gamma_{i+1,j-1;i,j}(v_{i+1,j-1} - v_{i,j}) + \\
& \gamma_{i+1,j;i,j}(v_{i+1,j} - v_{i,j}) + \\
& \gamma_{i+1,j+1;i,j}(v_{i+1,j+1} - v_{i,j})
\end{aligned}
$$

where

$$
\gamma_{i,j;p,q} = \begin{cases} 1, & \text{if element } (i,j) \text{ is coupled to } (p,q), \\ 0, & \text{otherwise.} \end{cases} \qquad (3)
$$

Without consider the coupling terms, eqn. (1) represents a Wilson-Cowan neural oscillator, which has been widely used in neural network modelling [18, 22]. It is a feedback loop between an excitatory unit (x) and an inhibitory unit (y). $I_{i,j}$ is an external stimulation received by oscillator (i, j). If $I_{i,j}$ is a constant, no chaos can appear since it is a two-dimensional continuous flow. In order to get a chaotic oscillator, the external stimulation is defined as a periodic function: $I_{i,j}(t) = A_{i,j}cos(t)$, where $A_{i,j}$ is the amplitude of external stimulation. In all simulations of this paper, $A_{i,j}$ is considered as a bifurcation parameter, which receives gray level of an input pixel.

The segmentation strategy is described below. Considering a scene image containing several objects. The network is organized that each element corresponds to a pixel of the image and a proper parameter of each oscillator is chosen to encode the gray level of the corresponding pixel. As the system runs, the neurons self-organize according to a predefined similarity criterion, such that the connections between pairs of neighboring oscillators with similar gray level will be maintained, while those connections between oscillators of different gray level will be cut. Consequently, all neurons belonging to the same segment (object) will be synchronized to form a unique trajectory, then each object is represented by a synchronized chaotic orbit. Following the

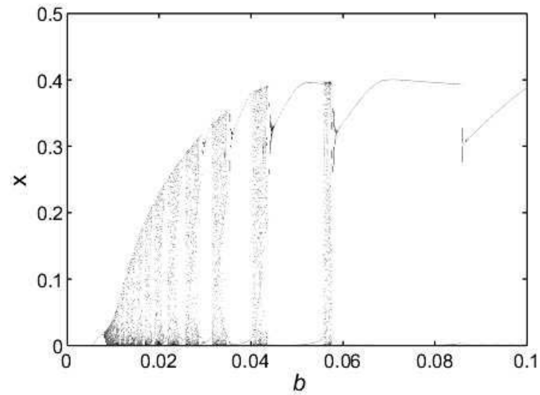

Figure 1. Bifurcation diagram of periodically driven Wilson-Cowan oscillators by varying parameter b. The stepsize $\Delta b = 0.0001$

high sensitivity to initial conditions and the dense properties of chaos, all such synchronized chaotic trajectories will be mutually different in time. In this way, objects in a given scene can be segmented.

2.2. Network of Wilson-Cowan Oscillators as a Visual Selection System

The proposed model can be seen as an adaptation of the model described above with a new major feature: a mechanism to select a salient object by increasing the firing frequency of its corresponding oscillators, and by decreasing the firing frequency of the oscillators corresponding to the other objects. This feature satisfies the following essential requirements of a Visual Selection system:

- considering as input one or a combination of features (Saliency Map), the neural network must highlight (select) the region of the image where the focus of attention should be directed;
- all other locations of the visual input must be suppressed by the system in order to keep the focus of attention on just one of the active objects.

The spiking frequency of Wilson-Cowan oscillators can be controlled by changing the parameter b in equation (1). So, we first analyzed the bifurcation diagram of periodically driven Wilson-Cowan oscillators by varying the parameter b, as shown in Fig. 1. From this figure, we see that when $b \leq 0.005$ there is nearly no oscillation. As b increases we see sessions of periodic windows and chaotic behavior. When b is small the chaotic behavior is predominant, but as b increases the periodic behavior becomes more frequent.

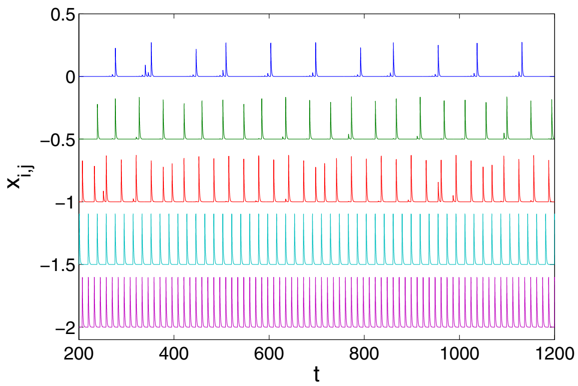

Figure 2. Temporal activities of oscillators with $b = 0.01$, 0.02, 0.034, 0.05 and 0.1 respectively. Vertical scale of second to fifth oscillators were shifted downwards by 0.5

In Fig. 2, we show the time series of a Wilson-Cowan oscillator varying b. From this figure we can notice that as b increases, the frequency of oscillation increases. When b takes a higher value (for example, the case of $b = 0.1$ in the figure), the oscillator not only fires more frequently, but also shows periodic resembling behavior.

In our model, we take advantage of these oscillatory changes to determine visual attention, which means that the synchronized oscillators corresponding to the salient object will present a periodic oscillation, while the oscillators corresponding to the other objects will become near silent.

In order to achieve this, we let the oscillators run with a fixed b parameter until they synchronize and the segmentation task can be performed. After that, whenever any oscillator fires, say oscillator (i, j), it will produce two types of signals:

- a reinforcement signal to itself and all other oscillators that fire in unison.
- an inhibitory signal to all oscillators that have different activities to (i, j).

The reinformecement/inhibitory behavior is defined by the following equation:

$$b_{p,q}(\tau) = b_{p,q}(\tau-1) + \frac{\alpha}{M(\tau)} \sum_{i,j \epsilon \Delta(\tau)} f(||x_{i,j} - x_{p,q}||) \quad (4)$$

$$b_{min} \leq b_{p,q}(\tau) \leq b_{max} \quad (5)$$

where (p, q) is neuron's index, τ is a time instant with at least one firing oscillator, $M(\tau)$ is the number of oscillators

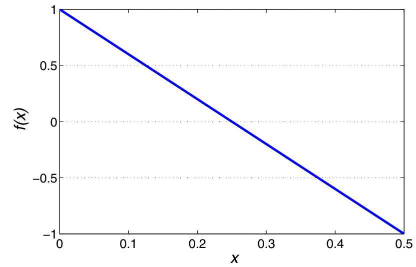

Figure 3. Illustration of $f(x)$. $a_1 = -4$ and $a_2 = 1$.

at the firing state in τ time, $\Delta(\tau)$ is the set of oscillators at the firing state in τ time, b_{min} and b_{max} are constant, $||x||$ is the norm of x, and

$$f(x) = a_1 x + a_2 \quad (6)$$

as shows by the graphic in Fig. 3.

$f(x)$ defines that each firing neuron, say neuron (i, j), may send a positive or a negative signal to another neuron, say (p, q), depending on the difference between them.

Each oscillator that receives a reinforcement signal will increase the value of parameter b accordingly, while the oscillator receiving an inhibitory signal will decrease the value of parameter b as well. The maximum value of b an oscillator can hold is b_{max}, while the minimum value is b_{min}. After some time, only the oscillators corresponding to the salient object will keep firing, increasing their activation frequency until they become periodic. The oscillators corresponding to the non-salient objects will gradually decrease their activation frequency, until they become silent. In our model, this object which remains active is the one that holds the focus of attention.

3. Computer Simulations

This section presents the simulation results performed on synthetic images in order to check the viability of the proposed model as a Selection Mechanism.

In all simulations of this paper, the following parameters are held constant at: $a = 1.0$, $c = 1.0$, $d = 0.6$, $e = -2.5$, $f = 0.0$, $\theta_x = 0.2$, $\theta_y = 0.15$, $T = 0.025$, $b_{min} = 0.002$, $b_{max} = 0.1$, $\alpha = 0.001$ and the initial value of b is 0.02. Such a configuration can guarantee that initially each oscillator is chaotic [22].

With these parameter values, we perform some experiments with synthetic images. $I_{i,j}$ is defined based only in the intensity of the image pixels (grey level).

The first experiment was carried by using the artificial image shown in Fig. 4, which has two twisted spirals with

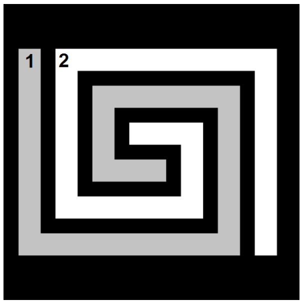

Figure 4. Artificial image with 2 linearly non-separable spirals (25×25 pixels).

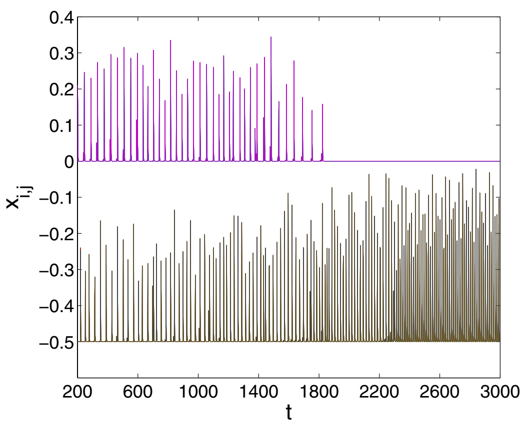

Figure 5. Temporal activities of oscillator blocks. Each trace in the figure (top-down) corresponds to an object (1-2) in the input pattern of Fig. 4. Vertical scale of the second objects is shifted downwards by 0.5. Segmentation and selection mechanism are activated.

2 different gray levels. It can be seen that the two spirals are *linearly nonseparable*. The coupling strength k is set to 20. Figure 5 shows the temporal activities of oscillator groups in the case that both the segmentation and the attention mechanisms are activated. It is possible to see that the proposed model is able to perform the requirement describe above, where initially the two objects are correctly segmented and after some running time, only one group of oscillators representing a spiral continues spiking, which means that the corresponding object is the one receiving the focus of attention. The highlighted spiral is the one of higher pixel intensity.

Our second experiment is performed by using the artificial image shown in Fig 6, which has 160×120 pixels containing 5 objects of different gray levels. The coupling

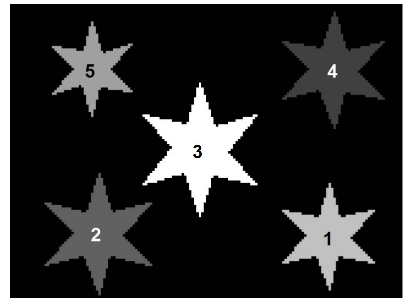

Figure 6. Artificial image with 5 objects (160×120 pixels)

strength we use here is $k = 5$. Figure 7 shows the temporal activities of the oscillators corresponding to each object with attention mechanism. Again, we see that the five objects are correctly segmented. After some running time, only one group of oscillators kept spiking, their temporal activities are represented by the third trace in the figure, which corresponds to the object in the middle of the image (Fig. 6). And again, the highlighted object is the brightest one.

In both experiments it is possible to observe that the firing frequency of non-highlighted objects decreases gradually until they become silent, while the firing frequency of highlighted object increases gradually until it becomes periodic, as we expected.

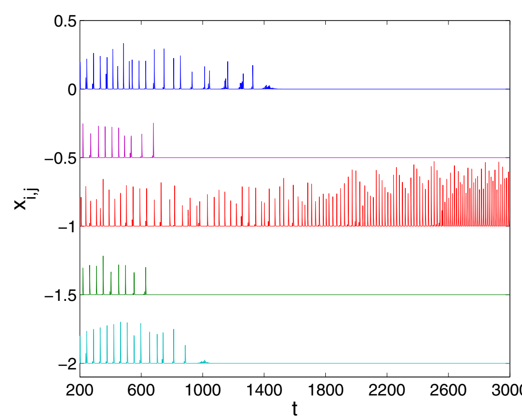

Figure 7. Temporal activities of oscillator blocks. Each trace (top-down) in the figure corresponds to an object (1-5) in the input pattern of Fig. 6. Vertical scale of second to fifth objects are shifted downwards by 0.5. Segmentation and selection mechanism are activated.

4. Conclusions

This paper presents a Visual Selection Mechanism based on a network composed of chaotic Wilson-Cowan oscillators. This mechanism can be seen as a part of a Visual Attention System, which is responsible for selecting one of several regions of interest into the visual input image. The proposed model utilize the properties of chaos and chaotic synchronization to discriminate the objects that compose the visual input and also included a inhibition mechanism which is responsible for highlighting the most salient object (in this work, the brightest one). Another interesting characteristic of the proposed model is its change of behavior when the object receives the focus of attention. In this case, the former chaotic behavior gives place to a trajectory with periodic behavior.

Computer simulations were performed in order check its viability as a selection mechanism and the results show that our model is a promising mechanism for computer vision systems.

As a future work we intend to introduce more features in our model, such as color, saturation, orientation, etc., in order to test our model using real images and compare our results with other models based on saliency maps. In addition, we will also verify the possibility of including some biasing mechanism to emulate top-down factors based on prior knowledge of the visual input, such as a memory holding a specific object.

5. Acknowledgment

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

6. References

[1] S. R. Campbell and D. Wang, "Synchronization and desynchronization in a network of locally coupled wilson-cowan oscillators". *IEEE Transactions on Neural Networks*, 1996, v.7, pp. 541–554.

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

[3] L. Carota, G. Indiveri, and V. Dante, "A software-hardware selective attention system". *Neurocomputing*, 2004, v. 58-60, pp. 647–653.

[4] G. Deco and E. T. Rolls, "Attention, short-term memory, and action selection: A unifying theory". *Progress in Neurobiology*, 2005, v. 76, pp. 236–256.

[5] G. Deco, E. T. Rolls, and J. Zihl, *Neurobiology of Attention*, "chapter 97 - A Neurodynamical Model of Visual Attention", pp. 593–599. Elsevier, Oxford, 2005.

[6] D. Hansel, "Synchronization and computation in a chaotic neural network". *Physical Review Letters*, 1992, v. 68(5), pp. 718–721.

[7] L. Itti, *Neurobiology of Attention*, "chapter 94 - Models of bottom-up attention and saliency", pp. 576–582. Elsevier, Oxford, 2005.

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

[9] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, v. 20(11), pp. 1254-1259.

[10] T. Jost, N. Ouerhani, R. von Wartburg, R. Müri, and H. Hügli, "Assessing the contribution of color in visual attention". *Computer Vision and Image Understanding*, 2005, v. 100, pp. 107–123.

[11] F. Shic and B. Scassellati, "A behavioral analysis of computational models of visual attention". *International Journal of Computer Vision*, 2007, v. 73(2), pp. 159–177.

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

[13] J. K. Tsotsos, "On the relative complexity of active vs. passive visual search". *International Journal of Computer Vision*, 1992, v. 7, pp. 127–141.

[14] 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*, 1995, v. 78, pp. 507–545.

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

[16] C. von der Malsburg and W. Schneider, "A neural cocktail-party processor". *Biological Cybernetics*, 1986, v. 54, pp. 29–40.

[17] 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*, 2005, v. 100, pp. 41–63.

[18] D. Wang, "The time dimension for scene analysis". *IEEE Transactions on Neural Networks*, 2005, v. 16(6), pp. 1401–1426.

[19] D. Wang and D. Terman, "Locally excitatory globally inhibitory oscillator networks". *IEEE Transactions on Neural Networks*, 1995, v. 6(1), pp. 283–286.

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

[21] L. Zhao and E. E. N. Macau, "A network of dynamically coupled chaotic maps for scene segmentation". *IEEE Transactions on Neural Networks*, 2001, v. 12(6), pp. 1375–1385.

[22] L. Zhao, E. E. N. Macau, and N. Omar, "Scene segmentation of the chaotic oscillator network". *International Journal of Bifurcation and Chaos*, 2000, v. 10(7), pp. 1697–1708.