Object Selection by Oscillatory Correlation

DeLiang L. Wang

Department of Computer and Information Science and Center for Cognitive Science
The Ohio State University, Columbus, Ohio 43210, USA
dwang@cis.ohio-state.edu

Abstract

One of the classical topics in neural networks is winner-take-all (WTA), which has been widely used in unsupervised (competitive) learning, cortical processing, and attentional control. With global connectivity WTA networks, however, do not encode spatial relations in the input, and thus cannot support sensory and perceptual processing where spatial relationships are important. We propose a new architecture that maintains spatial relations. This selection network builds on oscillatory dynamics and slow inhibition. In an input scene with many objects, the network selects the largest object. The system can be easily adjusted to select several largest objects, which then alternate in time. The network is applied successfully to select the most salient objects in real images.

1. Introduction

Winner-take-all networks constitute a basic and large class of neural networks (see [11] for a review). A WTA network in response to an input pattern produces only one active neuron with highest input. The candidates in WTA competition are single neurons, corresponding to local representations. WTA dynamics has been applied to a variety of tasks, such as unsupervised (competitive) learning, pattern recognition, and cortical processing. It also underlies neural theories of selective visual attention [2][6].

The perceptual environment, on the other hand, consists of patterned stimuli, which correspond to object structures, and objects appear to be the basic units of perception. Experimental data suggest that objects act as wholes in neural competition [1]. Due to local representations, existing WTA networks are limited in addressing perceptual information processing. This is demonstrated in current models of selective attention, which do not capture object-level selection because WTA provides only a winning *pixel* (or location), not a winning *object*. In fact, recent psychophysical evidence suggests that visual attention is surface (region) based [5].

We study object-level selection using oscillatory dynamics. In particular, we use LEGION (locally excitatory globally inhibitory oscillator network) dynamics for object formation [7][10], i.e. to group similar pixels into an object and segregate dissimilar regions into different objects, as LEGION has been rigorouly shown to be capable of this task. This unique ability of LEGION is derived from the representation of oscillatory correlation [4][8] and rapid synchrony and desynchrony characterizing its dynamics. By introducing a new variable into each oscillator and a slow inhibitory mechanism, we show that the resulting network exhibits object selection. Based on time complexity analysis we argue that the stage of perceptual organization performed by LEGION is essential for efficient image analysis. Our integrated network is applied to identify the most salient objects in real images.

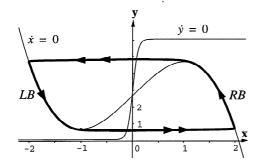## 2. Model description and analysis

The building block of our object selection network is a single oscillator i, which is defined as a feedback loop between an excitatory element x_i and an inhibitory element y_i:

$$\dot{x}_i = 3x_i - x_i^3 + 2 - y_i + I_i H(r_i - C z_s) + S_i + \rho \quad \text{(1a)}$$

$$\dot{y}_i = \varepsilon (\gamma (1 + tanh(x_i / \beta)) - y_i) \quad \text{(1b)}$$

where I_i represents external stimulation to the oscillator, H stands for the Heaviside step function, and S_i represents the overall input from other oscillators. ρ denotes Gaussian noise. The parameter ε is a small positive number. Hence (1) corresponds to a typical relaxation oscillator, similar to the van der Pol oscillator, and ε induces two time scales: a fast one dictated by the x variable and a slow one dictated by the y variable. The x-nullcline is a cubic curve, while the y-nullcline is a sigmoid function. If $I > 0$ and $H = 1$, these curves intersect along the middle branch of the cubic, and (1) is oscillatory (see Fig. 1A). The periodic orbit alternates between *silent* and *active* phases, corresponding to the left branch (LB) and the right branch (RB) of the cubic, respectively. Otherwise, the nullclines of (1) intersect at a stable fixed point, and the system produces no oscillation. β is a parameter, and γ controls the time ratio of RB to LB. Except for the Heaviside function, (1) is the same as defined previously [7].
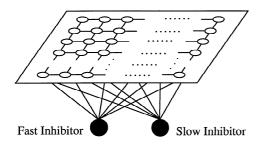
A

B

Fast Inhibitor ● ● Slow Inhibitor

Figure 1. A. Nullclines and limit cycle trajectory of a single oscillator. The oscillation is shown with a bold curve. B. Diagram of a two-dimensional selection network. An oscillator is indicated by an open circle, and two global inhibitors are indicated by the filled circles.

Figure 1B shows a two dimensional selection network, with nearest neighbor coupling. In addition to the local connectivity, there are two global inhibitors, denoted by z_f and z_s for *fast* and *slow* inhibitors respectively. The coupling term S_i is given by

$$S_i = \sum_{k \in N(i)} W_{ik} H(x_k - \theta_x) - W_z H(z_f - 0.5) \quad (2)$$

where W_{ik} is the connection weight from oscillator k to oscillator i, $N(i)$ is the set of neighboring oscillators that connect to i, and the threshold θ_x lies between LB and RB. Connection weights to an oscillator are normalized to a constant, W_T. W_z is the weight of inhibition from z_f, whose activity is defined as

$$\dot{z}_f = \sum_k H(x_k) - z_f \quad (3)$$

Note that the Heaviside is 1 when k is on RB and 0 when k is on LB. Thus z_f approaches the number of oscillators in the active phase on the fast time scale.

Coming back to the Heaviside term in (1a), r_i is defined as

$$\dot{r}_i = -[r_i - z_f]^+ H(x_i - \theta_x) \quad (4)$$

Here the function $[v]^+ = v$ if $v \geq 0$ and $[v]^+ = 0$ otherwise. r_i changes its value only when the Heaviside is 1. When r_i is ready to change, it reduces to the value of the fast inhibitor on the fast time scale when $r_i > z_f$, otherwise it remains unchanged. Thus, r_i records the number of oscillators that are in the active phase when i is in the active phase. The activity of the slow inhibitor is defined as

$$\dot{z}_s = [\sum_k H(x_k) - z_s]^+ - \mu \, \varepsilon \, z_s \quad (5)$$

Here μ is a parameter. The small ε results in quick rise and slow decay for the slow inhibitor.

Let a *block* be a set of oscillators stimulated by a connected pattern. The system defined in (1)-(5) has been analyzed, and details can be found in [9]. The main conclusion is that after at most m cycles of oscillations only the largest block is oscillatory, where m is the number of patterns in the input image. Due to space limitations, only the following outline of the analysis can be given here.

(a) When the largest block jumps to the active phase, z_s approaches the size of the block, say s_M, quickly. z_s decays when the block is on LB. Recently, Linsay and Wang calculated the times that a block spends on RB and LB in the singular limit $\varepsilon \to 0$ [3]. In particular, the time on LB is $\tau_L = \ln((I_T + 4)/I)$, where $I_T = I + W_T - W_z$. According to (5), z_s decays to $s_M \exp(-\mu \, \tau_L)$. Thus, if we set C to $C_M = \exp(\mu \, \tau_L)$, then this largest block is just able to jump after it spends τ_L on LB and no other block can jump to the active phase. Notice that C_M is independent of the specific size of the block. This is an important property, as it implies that the parameter does not need to be adjusted with specific input images.

(b) The complications caused by the fact that the LEGION network has a natural segmentation capacity [10], fortunately, do not disrupt selective dynamics except that the largest block may take up to m cycles to suppress all other blocks.

(c) When $C > C_M$, the above behavior remains the same except that each cycle is longer. On the other hand, when $C < C_M$, more than one block may survive the selection process and oscillate.

3. Computer simulation

We illustrate how object selection is performed by simulating a 50x50 selection network. We arbitrarily

1183

selected an input scene with three objects (designated by shape as the *sun*, a *tree*, and a *mountain*) as shown in Fig. 2A. The equations (1)-(5) were solved using the singular limit method [3], which is a fast numerical method for integrating relaxation oscillator networks in the singular limit $\varepsilon \to 0$. The following important parameters have values: $\gamma = 6.5$, $W_z = 1.5$, $W_T = 8.0$, and $\mu = 0.125$. For these parameter values, $C_M = 1.6407$. Thus, we choose $C = 1.64$, to be slightly less than C_M. The following simulation results have been confirmed using a fourth-order Runge-Kutta method with $\varepsilon = 0.02$.

A

B

C

Sun

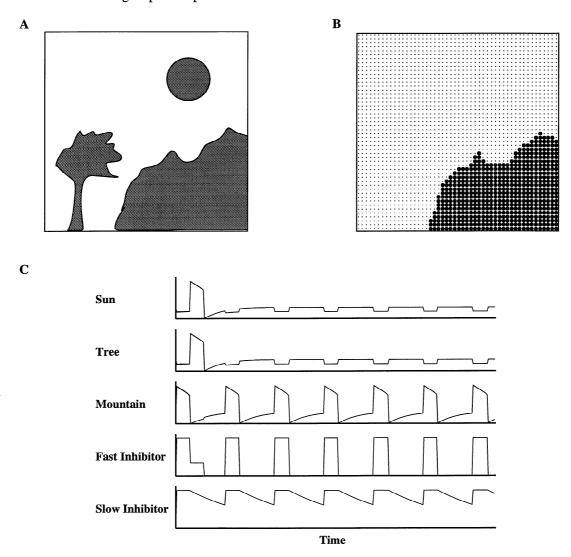Tree

Mountain

Fast Inhibitor

Slow Inhibitor

Time

Figure 2. A. Input scene with three patterns. B. A snapshot of the selection network taken several cycles after the beginning. Only one pattern (*mountain*) is selected. C. Temporal process of selection. The upper three traces show the activities of the three oscillator blocks. The lower two traces show the activities of the two global inhibitors.

Fig. 2B shows a snapshot after the system reaches stable limit cycles, when the only active block is the one corresponding to the *mountain* - the largest object in the scene. In the display, the diameter of each circle indicates the x activity of the corresponding oscillator. To display the entire competition process, Fig. 2C shows the temporal evolution of every stimulated oscillator, where the oscillators belonging to the same block are combined

in the display. The unstimulated oscillators never oscillate, hence excluded from the display. Note that numerical integration using the singular limit method makes an oscillator block appear to be perfectly synchronized. We note that when C is reduced to 0.3, the same network produces two winning objects: the *mountain* and the *tree*, consistent with the analysis.

1184

4. LEGION as a preprocessing layer

As stated in Sect. 2, the time complexity of the selection process is $O(m)$. This complexity implies a very slow process if there are many objects in a scene. For example, Fig. 3A shows the same input as Fig. 2A, with 20% noise added so that each uncovered box has a 20% chance of being covered (stimulated). With four nearest-neighbor coupling, Fig. 3A gives rise to hundreds of oscillator blocks. Thus, even though the selection network can eventually identify the largest block, it takes a very long time.

A

B

Figure 3. A. The input scene of Fig. 2A with 20% noise added. B. A snapshot of the two-stage selection network taken several cycles after the beginning.

As described in [10], a LEGION network with a lateral potential for each oscillator has the ability to remove small noisy regions *in parallel*. This consideration and psychophysical theories of preattentive processing lead us to suggest a two-stage process for object selection. The first stage corresponds to a LEGION layer, which organizes the input scene into a number of major regions and a background that is the rest of the scene. The second

stage corresponds to the selection layer of Fig. 1B. The connections from the LEGION layer to the selection layer are one-to-one correspondence. For input scenes with many insignificant regions, the two-layer network represents remarkable speedup for object selection.

Coming back to the scene in Fig. 3A, the LEGION layer segments it into three major regions, corresponding to the *sun*, the *tree*, and the *mountain*, respectively, in at most 4 cycles [10]. After that, the selection layer takes one more cycle to select the *mountain* region, the largest object in the noisy scene, as shown in Fig. 3B.

5. Real images

To relate the previous study on LEGION [10], we use two images that have been processed by LEGION to illustrate object selection. These are two MRI (magnetic resonance imaging) images of the human head. The first is shown in Fig. 4A. When presented to the LEGION layer, the image is segmented to 21 regions plus a background as shown in Fig. 4B (see [10]). In the figure, gray levels indicate the phases of the oscillator blocks, resulting in a *gray map*. The background is indicated by scattered black areas. Like the preceding layer, the selection layer has eight nearest-neighbor coupling. The layer is integrated using the singular limit method with the following parameter values: $\gamma = 6.5$, $W_z = 0.7$, $W_T = 8.0$, and $\mu = 0.125$. For these parameter values, $C_M = 1.6558$. Thus, we choose $C = 1.65$. In the selection, we ignore the homogeneous background region, which is the largest (see Fig. 4B). Fig. 4C shows the output of the selection layer. The only selected region is the upper part of the brain, which is the largest segment. The second MRI image is shown in Fig. 4D. Again, the image is first segmented by the LEGION layer, yielding 17 regions plus a background. Fig. 4E shows the gray map of the segmented regions. The selection layer has the same configuration and the same parameter values as for processing the first MRI image. Fig. 4F shows the result of selection. The brain region, which is the largest segment from the first layer, is selected.

As discussed in [10], different parameter settings for the LEGION layer produce different segmentation results. For example, when the level of global inhibition is increased, the LEGION layer can further segment the upper part of the brain (Fig. 4B) into the cerebral cortex, the cerebellum, the callosum/fornix regions and its surrounding septum [10]. Regardless of parameter settings in the first layer, the selection layer works in the same way: it always selects the largest segment generated by the first layer.

6. Discussion

Compared with previous WTA models and competitive dynamics, our model offers the unique ability of object-level selection. To our knowledge, object selection has
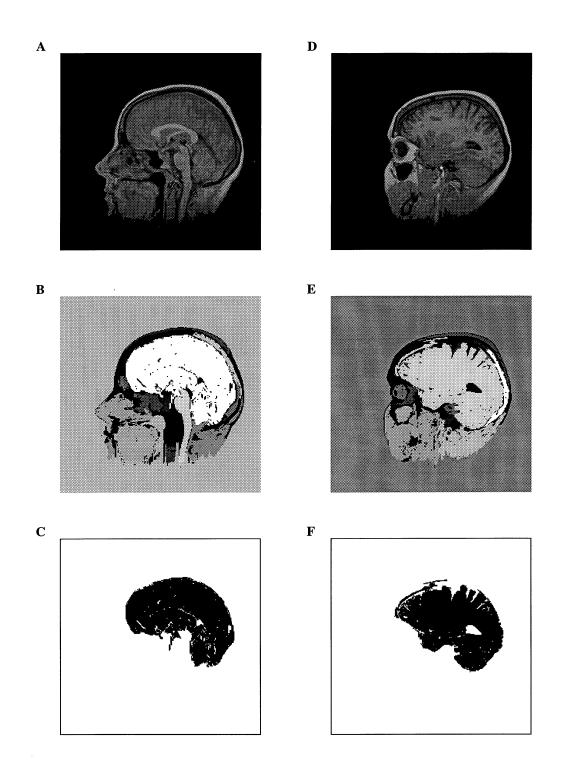
Figure 4. A. An MRI image consisting of 257x257 gray-level pixels. B. A gray map showing the segmentation result of the image in A by a 257x257 LEGION layer. C. A snapshot of the selection layer showing the result of selection. D. Another MRI image consisting of 257x257 pixels. E. A gray map showing the segmentation result of the image in D. F. A snapshot of the selection layer showing the result of selection.

not been addressed in the literature. On the other hand, considerable psychophysical and neurobiological data support the idea that competition is at the object level [1][5]. We point out that both neural oscillations and the basic network architecture are consistent with neurobiological data.

Our model can easily be extended to incorporate a decay mechanism so that the selection process shifts from one object to another. The idea of using a decay mechanism was proposed by Koch and Ullman [2] for producing attentional shifting. With such mechanism in place, our network can readily select the second largest object after the largest one is selected, and the third largest, etc. Thus, it implements a form of shifting in object selection. This occurs because slow inhibition decays if the slow inhibitor is not stimulated by any active block, and eventually the next largest block can overcome slow inhibition and oscillate.

Object selection studied so far is based on the size of an object. Obviously size is only one aspect of object saliency, and it is our intention to use size as a concrete way of embodying the saliency. In general, saliency includes other measures such as brightness, color, motion, texture, context, and familiarity. The selection framework described here might be extended to include other measures. For example, if motion should be an important measure, we can assign moving objects higher magnitudes in competition. This way, a smaller moving object can win over a large stationary object.

Oscillatory correlation is usually used as a mechanism of binding features into objects in the presence of multiple objects. Given that object selection usually (but not always) selects only one item, one might ask the question, why use oscillatory correlation in the first place? Our study reveals that the *process* of generating the selection is intimately related to oscillatory correlation, where multiple organizations participate in competition and grouping within each one is embodied by synchronization. Moreover, the slow inhibitor depends on oscillatory recharge from the winning object in order to prevent other organizations from being activated.

Conceptually, besides competition as used in all forms of WTA dynamics, our selection network includes cooperation between neighboring oscillators in the network. The cooperation between elements that likely form the same object gives rise to object selection, as opposed to pixel selection. In other words, the selection network incorporates *both* competition through global inhibition and cooperation through local excitation.

Acknowledgments

References

[1] R. Desimore and J. Duncan, "Neural mechanisms of selective visual attention", *Ann. Rev. Neurosci.*, **18**, 1995, pp. 193-222.

[2] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry", *Human Neurobiol.*, **4**, 1985, pp. 219-227.

[3] P.S. Linsay and D.L. Wang, "Fast numerical integration of relaxation oscillator networks based on singular limit solutions", *IEEE Trans. Neural Nets.*, in press.

[4] P. Milner, "A model for visual shape recognition," *Psychol. Rev.*, **81**, 1974, pp. 521-535.

[5] K. Nakayama, Z.J. He, and S. Shimojo, "Visual surface representation: A critical link between lower-level and higher-level vision," In S.M. Kosslyn and D.N. Osherson (Ed.), *An Invitation to Cognitive Science* (pp. 1-70). MIT Press, 1995.

[6] E. Niebur and C. Koch, "Control of selective visual attention", In D. Touretzky, M. C. Mozer, and M. E. Hasselmo (Ed.), *Advances in Neural Information Processing Systems* (pp. 802-808). MIT Press, 1996.

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

[8] C. von der Malsburg, "The correlation theory of brain function", Internal Rep. 81-2, Max-Planck-Institute for Biophysical Chemistry, 1981.

[9] D.L. Wang, "Object selection based on oscillatory correlation", Tech. Rep. 12/96 - TR 67, OSU Dept. of Computer and Information Science, 1996.

[10] D.L. Wang and D. Terman, "Image segmentation based on oscillatory correlation", *Neural Comput.*, **9**, 1997, pp. 805-836 (See errata in *Neural Comput.*, **9**, 1997, pp. 1623-1626).

[11] A.L. Yuille and D. Geiger, "Winner-take-all mechanisms", In M.A. Arbib (Ed.), *Handbook of Brain Theory and Neural Networks* (pp. 1056-1060). MIT Press, 1995.