Visual Attention Shift based on Image Segmentation Using Neurodynamic System

Lijuan Duan[1], Chunpeng Wu[1], Faming Fang[1], Jun Miao[2], Yuanhua Qiao[3] and Jian Li[1]

[1] College of Computer Science and Technology, Beijing University of Technology
Beijing, 100124, China

[2] Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences
Beijing, 100190, China

[3] College of Applied Science, Beijing University of Technology
Beijing, 100124, China

Abstract

A method of predicting visual attention shift is proposed based on image segmentation using neurodynamic system in this paper. The input image is mapped to a neural oscillator network. Each oscillator corresponding to a pixel is modeled by means of simplified Wilson-Cowan equations, and is coupled with its 8-nearest neighbors. Then the image is segmented by classifying the oscillation curves of the excitatory groups of all the oscillators. The classifier is constructed based on features of frequency, offset, phase and amplitude of the curves. The visual attention shift between the regions on the image is predicted according to the saliency strength of each region. Referring to the mechanism of winner-take-all competition, the saliency of a region is the aggregation of the dissimilarities between this region and all the other ones. Experimental results on images show the effectiveness of our method.

Keywords: *Neural oscillation, Coupling method, Image segmentation, Saliency, Visual attention shift.*

1. Introduction

Human vision system is able to select salient information among mass visual input to focus on. This selective attention mechanism enables us to efficiently understand the visual scenes without forming a complete, detailed representation of our surroundings [1]. When performing a visual task, according the mechanism of winner-take-all competition [2], the most salient region is attended first, and then our attention will shift to the less salient regions successively due to the adaptability of the visual system, while the attended regions may also be attended again after a few seconds. Computationally modeling such mechanism has become a popular research topic in recent years [3-5].

In this paper, a method of predicting visual attention shift is proposed based on image segmentation using neurodynamic system. In previous studies [2, 6], visual attention is often modeled to shift between pixels according the saliency strength of these pixels, i.e., these methods do not consider the semantic information in the image. We think commonly that visual attention should shift between meaningful regions on an image, and these regions should correspond to an object or at least part of an object. Therefore, in order to label the input image with meaningful regions, we apply our simplified Wilson-Cowan equations which we proposed in [7] to image segmentation. Wilson-Cowan equations [8] are based on the assumption that the features of an object are grouped based on the temporal correlation of neural activities [9]. Thus neurons that fire in synchronization would signal features of the same object, and groups desynchronized from each other represent different objects. Experimental observations [10] of the visual cortex of animals show that synchronization indeed exists in spatially remote columns and phase-locking can also occur between the striate cortex and extrastriate cortex, between the two striate cortices of the two brain hemisphere, and across the sensorimotor cortex. These findings have concentrated the attention of many researchers on the use of neural oscillators such as Wilson-Cowan oscillators [11].

The remainder of the paper is organized as follows: In Section 2, we firstly stated the framework of our visual attention shift method in details. Then we demonstrate our experimental results in Section 3. The summary is given in Section 4.

2. Proposed Method

2.1 Neural Oscillation and Synchronization

To segment a gray image by using neurodynamic system, we let the input image correspond to a neural oscillator network in our method. Therefore, each pixel in the image is mapped to a neural oscillator in the network, and the intensity of each pixel is considered to be the external input to the corresponding oscillator.

We describe an oscillator by means of simplified Wilson-Cowan equations. Such a model consists of two non-linear ordinary differential equations representing the interactions between two populations of neurons that are distinguished by the fact that their synapses are either excitatory or inhibitory. Thus, each oscillator consists of a feedback loop between an excitatory groups x_i and inhibitory groups y_i that obey the Equation (1):

$$\begin{aligned} \frac{dx_i}{dt} &= -r_1 x_i + r_1 H(ax_i - cy_i + I_i - \phi_x) \\ \frac{dy_i}{dt} &= -r_2 y_i + r_2 H(bx_i - dy_i - \phi_y) \end{aligned} . \quad (1)$$

Both x_i and y_i are interpreted as the proportion of active excitatory and inhibitory neurons respectively, which are supposed to be continuous variables and their values may code the information processed by these populations. Especially, the state $x_i = 0$ and $y_i = 0$ represents a background activity which correspond to the background in an image. The parameters in Equation (1) are as follows: a is the strength of the self-excitatory connection, d is the strength of the self-inhibitory connection, b is the strength of the coupling from x to y, and c is the strength of the coupling from y to x. Both ϕ_x and ϕ_y are thresholds, r_1 and r_2 modify the rate of change of the x and y group respectively. Figure 1(a) shows the model of an oscillator. All the above parameters are non-negative, and I_i is external input to the oscillator in position i which corresponds to a pixel in the image. $H()$ is a sigmoid activation function defined as in Equation (2):

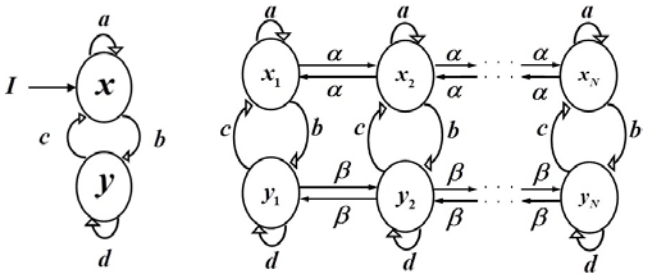
$$H(z) = \frac{1}{1 + e^{-z/T}} . \quad (2)$$

T is a parameter that sets the central slope of the sigmoid relationship.

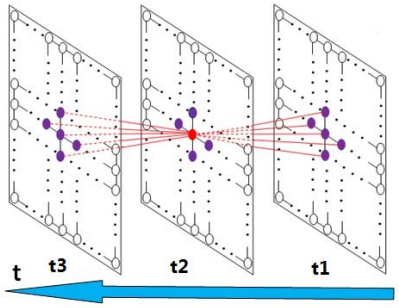
We locally couple the above simplified Wilson-Cowan oscillators as in Equation (3):

$$\begin{aligned} \frac{dx_i}{dt} &= -r_1 x_i + r_1 H(ax_i - cy_i + I_i - \phi_x) + \alpha \Delta x_i \\ \frac{dy_i}{dt} &= -r_2 y_i + r_2 H(bx_i - dy_i - \phi_y) + \beta \Delta y_i \end{aligned} . \quad (3)$$

α and β represent the strength of the connection between neurons. Δx_i and Δy_i represent coupling terms from all other adjacent oscillators in the neural network. An open chain of coupled oscillators is shown in Figure 1(b) which is preferable for 1-D neural network. Since an image corresponds to a 2-D network in our method, we couple each oscillator with its 8-nearest neighbors. The coupling method in our model is illustrated in Figure 1(c). In Figure 1(c), each ellipse represents an oscillator and all the oscillators enclosed by a rectangle represent the neural oscillator network corresponding to the input image. In Figure 1(c), for the red oscillator at t_2, its coupling strength is related to five purple oscillators (including itself) at last moment t_1. And the coupling strength of these five oscillators at t_3 are related to the red oscillator at t_2. Note that for better illustration, each oscillator only connects with 4 other ones in Figure 1(c). Thus Δx_i and



(a) (b)



(c)

Figure 1. (a) A single oscillator. (b) An open chain of coupled oscillators. (c) The coupling method in our model. For the red oscillator at t_2, its coupling strength is related to five purple oscillators (including itself) at last moment t_1. And the coupling strength of these five purple oscillators at t_3 are related to the red oscillator at t_2.

Δy_i in Equation (3) are computed as in Equation (4):

$$\begin{aligned} \Delta x_i &= \sum_{j \in N(x_i)} r_j (x_j - x_i) \\ \Delta y_i &= \sum_{k \in N(y_i)} r_k (y_k - y_k) \end{aligned} \quad . \tag{4}$$

$N(x_i)$ and $N(y_i)$ are the 8-nearest neighbors of x_i and y_i respectively. The weight r_j and r_k is determined as in Equation (5):

$$r_{ij} = \begin{cases} 1 & |I_i - I_j| < \phi \\ 0 & otherwise \end{cases} \quad . \tag{5}$$

ϕ is a threshold to decide whether two adjacent oscillators couple with each other.

All the parameters above are determined to make the differential equation in Equation (3) reach synchronization asymptotically, and we use 4th order Runge-Kutta method to find the iterative solution of this differential equation.

Figure 2 shows an input image and the corresponding oscillation curves of the excitatory groups of all the oscillators in the network. In Figure 2(b), except that the oscillators corresponding to background pixels in the input image become silent over time, the oscillation curves of all the other oscillators can be obviously clustered into 4 classes: the red arrow points to one class, the green arrow points to the other three classes. Consequently, the actually 4 objects in the input image can be reasonably segmented by classifying the neural oscillation curves. We will explain how to classify the oscillation curves automatically using the features extracted from the curves in the next subsection.

2.2 Image Segmentation by Classifying the Oscillation Curves

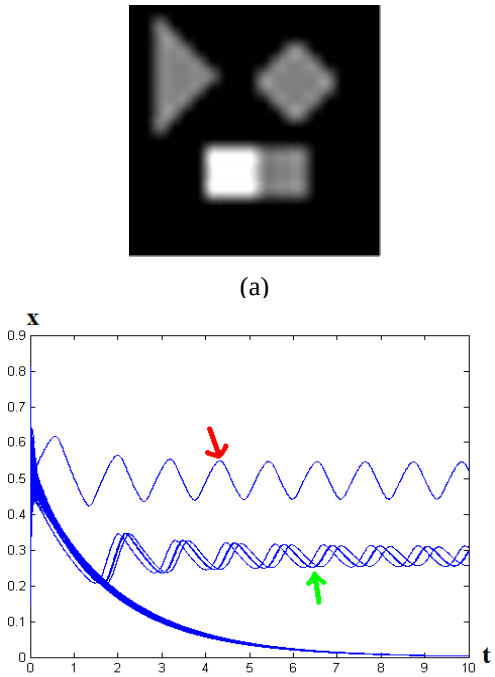
After observing and analyzing many oscillation curves, we assume that each oscillation curve generated by Equation (3) is the superposition of sine and cosine waves. Therefore, we fit Fourier curves to the data of oscillation curves as in Equation (6):

$$\hat{x} = m_0 + m_1 \cos(\omega \cdot t) + n_1 \sin(\omega \cdot t) \quad . \tag{6}$$

$\hat{x}$ is an approximation of the x in Equation (3), i.e., the strength of x corresponding to each moment t in Figure 2(b). And m_1, n_1, ω are parameters. Equation (6) can also be written as Equation (7):

$$\hat{x} = m_0 + \sqrt{m_1^2 + n_1^2} \sin(\omega t + \theta) \quad . \tag{7}$$

where $\theta = arctg(m_1 / n_1)$. In Equation (7), ω represents frequency, θ represents phase, $\sqrt{m_1^2 + n_1^2}$ represents amplitude, and m_0 represents offset from t-axis as shown in Figure 2.



(a)

(b)

Figure 2. Input image and neural oscillation. (a) An input image. (b) Oscillation curves of the excitatory groups of all the oscillators.

As stated in last subsection, in order to segment the input image, we classify all the corresponding oscillation curves by combining the above four features (frequency, phase, amplitude and offset) extracted from each curve. A four-layer classifier is constructed as shown in Figure 3, and each layer classifies the oscillators using one feature respectively. We define the concept of an *oscillator slice* as a group of oscillators with same class label who connect with each other on the neural network, so each oscillator slice can also be labeled a class which is the same as any oscillator of this slice. By setting a threshold according to the feature, one oscillator slice from last layer can be classified into two or more slices on current layer. Moreover, at each layer, maybe there are several oscillator slices with same class label. We argue that if these slices with same class label connect with each other, the combination of these slices corresponds to a meaningful region in the input image, therefore these slices do not need to be further classified. Figure 4 shows results of image segmentation by classifying the oscillation curves.

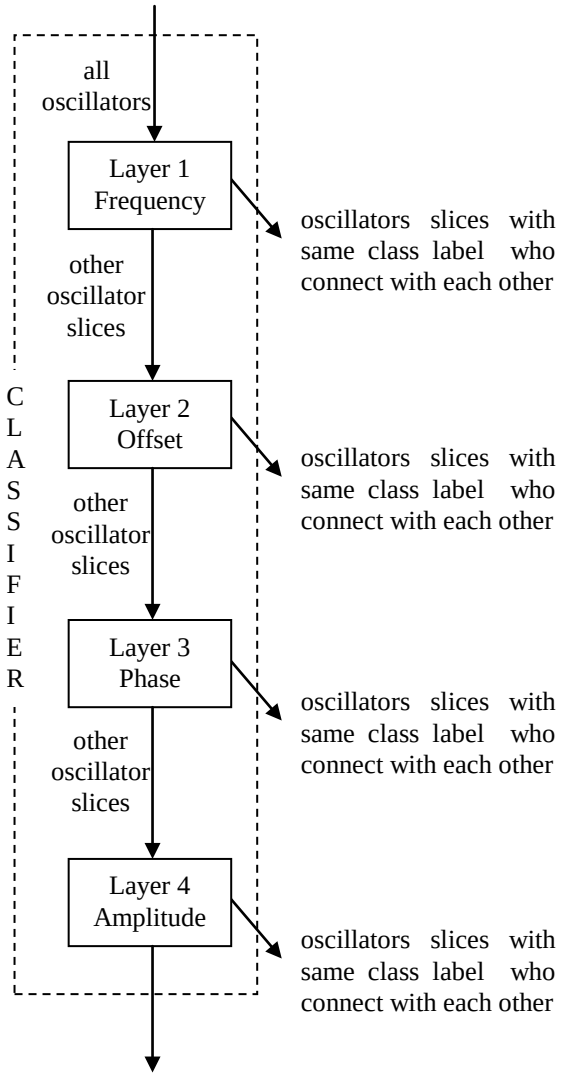
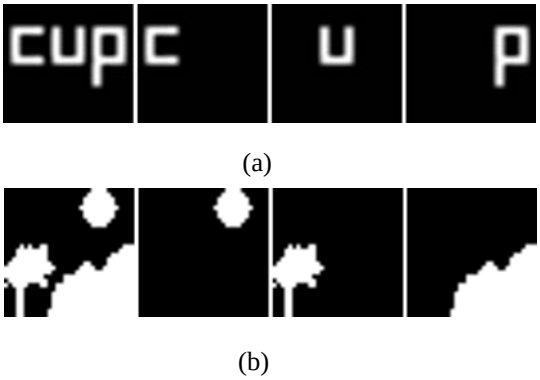


Figure 3. A four-layer classifier used for analyzing the neural oscillation curves.



(a)

(b)

Figure 4. Results of image segmentation by classifying neural oscillation curves. (a) An input image and the image segmentation results. (b): Another Input Image and the corresponding segmentation results.

2.3 Visual Attention Shift between Regions

According to the mechanism of winner-take-all (WTA) competition, visual attention shifts from the most salient region to the lease one. In our method, given an image, all the regions are labeled by classifying the neural oscillation curves based on simplified Wilson-Cowan equations. The saliency of each region is calculated in a local-global manner as follows: the dissimilarities between the current region and all the other regions of the image are calculated, and the aggregation of these dissimilarities is the saliency strength of the current region. If one region is more "irregular" measured by the above mentioned local-global method than other regions, this region is more salient. Given a gray image I with P regions labeled by the image segmentation method as stated above, we compute the average of intensity of each region as in Equation (8):

$$AveIntensity_i = \frac{1}{Num(i)} \sum_{(u,v)\in R(i)} I_{u,v} \quad (i = 1,2,...,P). \quad (8)$$

$Num(i)$ is the total number of pixels in region i, $R(i)$ represents all the coordinates of pixels in region i, and $I_{u,v}$ is the intensity of pixel (u,v) on image I. Then the saliency strength of each region is calculated as in Equation (9):

$$Saliency_i = \sum_{j=1, j\neq i}^{P} | AveIntensity_i - AveIntensity_j |. \quad (9)$$

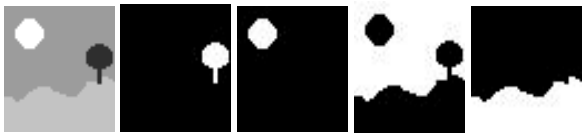
Then $Saliency_i (i = 1,2,...,P)$ are sorted by descending order, and visual attention are supposed to shift from the most salient region to the lease one. Figure 5(a) shows an input image, and Figure 5(b) – Figure 5(e) shows the whole process of visual attention shift predicted by our method. Note that the white region is currently attended in Figure 5(b) – Figure 5(e). The tree is attended firstly, and then the sun, the background, the hill.

3. Experimental Validation

To guarantee that the neurodynamic system described in Equation (3) reach synchronization asymptotically, the parameters in Equation (1) – Equation (5) are set as follows: $\alpha = 20$, $\beta = 14$, $r_1 = 1$, $r_2 = 1$, $\phi = 0.1$, $a = 1$, $b = 1$, $c = 2$, $d = 0.5$, $\phi_x = 0.2$, $\phi_y = 0.15$, $T = 0.025$. Note that all pixel values of the input image are normalized to [0, 1] before neural oscillation. We use 4th order Runge-Kutta method to find the iterative solution of the differential equations in Equation (3).

Figure 6 demonstrates the whole process of our method. Given a gray image as shown in Figure 6(a), we map this image to a neural oscillator network. Then Figure 6(c) illustrates the oscillation curves of the excitatory of all the

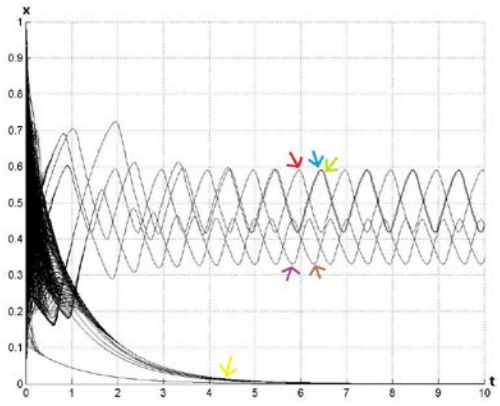
oscillators. The six color arrows point to six clusters of curves corresponding to six regions in the input image. Note that the blue and the green arrows actually point to two different clusters of curves although these two clusters are similar to each other. Figure 6(b) shows the course of visual attention shift predicted by our method. And the order is from upper left to lower right. The background is predicted to be attended firstly, and the jeep last.



(a) (b) (c) (d) (e)

Figure 5. Prediction of visual attention shift. (a): An input image. (b) – (e): the whole process of visual attention shift predicted by our method. Note that the white region is currently attended. The tree is attended firstly, and then the sun, the background, and the hill.

(a) (b)



(c)

Figure 6. (a) Input image. (b) The whole course of visual attention shift predicted by our method, and the order is from upper left to lower right. Note that the white region in each image is currently attended. (c) Neural oscillation curves of the excitatory groups of all the oscillators.

4. Conclusions

In this paper, a method of predicting visual attention shift has been proposed based on image segmentation using neurodynamic system. We think commonly that visual attention should shift between meaningful regions on an image, and these regions should correspond to an object or at least part of an object. Therefore, we apply our simplified Wilson-Cowan equations to image segmentation. The input image is mapped to a neural oscillator network. Then by analyzing the features of frequency, offset, phase and amplitude of the oscillation curves, the image is labeled with different regions. To determine the order of visual attention shift, we define the saliency strength of each region as the aggregation of dissimilarities between this region and all the other ones, and more salient region is predicted to be attended earlier.

Acknowledgments

This research is partially sponsored by Natural Science Foundation of China (Nos.60702031, 60970087, 61070116 and 61070149), Beijing Natural Science Foundation(Nos.4072023 and 4102013), Beijing Municipal Education Committee (No.KM200610005012) and Beijing Municipal Foundation for Excellent Talents(No.20061D0501500211) and National Basic Research Program of China (Nos. 2007CB311100 and 2009CB320902).

References

[1] R. Rensink, K. O'Regan, and J. Clark. To see or not to see: the need for attention to perceive changes in scenes. *Psychological Sciences*, 1997.

[2] L. Itti, C. Koch and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transanctions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1998.

[3] T. Judd, K. Ehinger, F. Durand and A. Torralba. Learning to predict where humans look. *IEEE International Conference on Computer Vison (ICCV)*, 2009.

[4] W. Wang, Y. Wang, Q. Huang, and W. Gao. Measuring visual saliency by site entropy rate. *IEEE International Conference on Computer Vison and Pattern Recognition (CVPR)*, 2010.

[5] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-Aware saliency detection. *IEEE International Conference on Computer Vison and Pattern Recognition (CVPR)*, 2010.

[6] Raghu Raj, W. S. Geisler, Robert A. Frazor, and A. C. Bovik. Contrast statistics for foveated visual systems: fixation selection by minimizing contrast entropy. *J. Opt. Soc. Am. A*, 2005.

[7] Y. Meng, Y. Qiao, J. Miao, L. Duan, and F. Fang. Qualitative analysis in locally coupled neural oscillator network. *International Conference on Neural Networks (ICNN)*, 2009.

[8] H. R. Wilson and J.D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Jounal*, 1972.
[9] D. Wang. The time dimension for scene analysis. *IEEE Transactions on Neural Networks (TNN),* 2005.
[10] A. K. Engel, A. K. Kreiter, P. König, and W. Singer. Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proceedings of the National Academy of Sciences of the United Stated of America (PNAS)*, 1991.
[11] S. Campbell and D. Wang. Synchronization and desynchronization in a network of locally coupled Wilson-Cowan oscillators. *IEEE Transactions on Neural Networks (TNN)*, 1996.

Lijuan Duan received her Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences in 2003. She is currently an Associated Professor at the College of Computer Science and Technology, Beijing University of Technology, China. Her research interests include artificial intelligence, neural networks, neural information processing, image understanding and biological vision. She has published more than 30 research articles in refereed journals and proceedings on image retrieval, visual neural information coding, image segmentation, visual perception and cognition.

Chunpeng Wu received the B.S. degree in computer science from the College of Computer Science and Technology, Beijing University of Technology, Beijing, in 2008. He is currently a candidate for Master of Computer Science and Technology, Beijing University of Technology, Beijing. His research interest is visual saliency detection.

Faming Fang received the B.S. degree in computer science from the College of Computer Science and Technology, YanTai University, Shandong, in 2005. He is currently a candidate for Master of Computer Science and Technology, Beijing University of Technology, Beijing. His research interest is image segmentation based on neurodynamics.

Jun Miao received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2005. He is currently an Associated Professor at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing. His research interests include artificial intelligence, neural networks, neural information processing, image understanding, and biological vision. He has published more than 30 research articles in refereed journals and proceedings on face detection, visual neural networks, visual neural information coding, neural oscillation, image segmentation, visual perception and cognition.

Yuanhua Qiao received the Ph.D. degree of hydromechanics from the College of Life Science and Bioengineering in Beijing University of Technology, in 2005. She is currently an Associated Professor at The College of Applied Sciences, Beijing University of Technology, China. Her research interests include differential dynamics, neuron dynamics, pulse differential equations, mathematical model construction, biomathematics, neural networks and image understanding. She has published more than 15 research articles in refereed journals and proceedings in the field of biomathematics, Bioengineering, Image segmentation and visual perception.

Jian Li received the B.S. degree in 1984 from Beijing University of Technology. He is currently a Professor at the College of Computer Science and Technology, Beijing University of Technology, China. His research interest is signal processing and information security.