Oscillatory Neural Network for Image Segmentation with Biased Competition for Attention

Tapani Raiko and Harri Valpola

Abstract We study the emergent properties of an artificial neural network which combines segmentation by oscillations and biased competition for perceptual processing. The aim is to progress in image segmentation by mimicking abstractly the way how the cerebral cortex works. In our model, the neurons associated with features belonging to an object start to oscillate synchronously, while competing objects oscillate with an opposing phase. The emergent properties of the network are confirmed by experiments with artificial image data.

1 Introduction

The success of animals depends on their ability to extract relevant information from their environment. Since animals have got very good at this during evolution, engineers have a lot to learn from the computational principles underlying the perceptual abilities of animals.

In mammals, the cerebral cortex is largely responsible for interpretation and extraction of relevant information. The perceptual abilities of cerebral cortex include *learning* to make relevant sensory discriminations, *segmentation* of objects and attentional *selection* of relevant objects.

In this article, we focus on segmentation. However, it is important to recognise that segmentation cannot be isolated from the other components. For example, when segmenting objects in a visual scene, one typically has to be able to recognise the individual objects to know which parts go with which object. On the other hand,

Tapani Raiko

Aalto University, Department of Information and Computer Science, Helsinki, Finland and Zen-Robotics Ltd., Helsinki, Finland, e-mail: `tapani.raiko@tkk.fi`

Harri Valpola

Aalto University, Department of Biomedical Engineering and Computational Science, Helsinki, Finland and ZenRobotics Ltd., Helsinki, Finland, e-mail: `harri.valpola@tkk.fi`

recognition usually requires one to segment out the object first. It is therefore impossible to achieve good performance by neatly separating the process into two consecutive stages of segmentation and recognition. Nevertheless, this is what is usually done in machine vision: first segment the object and then recognise it.

The cerebral cortex seems to take a different approach: segmentation and recognition are combined into an iterative, dynamical process. The goal of this article is to study such a process from an engineering perspective. We combine bottom-up biological inspiration with top-down demands and restrictions of the engineering problem of interpreting and extracting useful information for the environment.

The rest of the article is structured as follows. Section 2 gives a background on segmentation by coupled oscillators and on biased competition model for attention and learning. In Section 3 we list the desired emergent properties that we would like to have in our model. We start from high-level phenomena and work our way towards details. Section 4 gives the definition of a model that has such properties, which is then experimentally confirmed in Section 5, followed by discussion in Section 6.

2 Background

In this section, we will first introduce Gestalt principles which form the basis for segmentation. Next, we discuss segmentation by coupled oscillators. Finally, we introduce biased competition model for attention and learning. Our goal in this paper is to apply segmentation with coupled oscillators to models which use biased competition for perception and learning.

2.1 Gestalt principles

When we humans see a new object, we may not know its identity but we can nevertheless tell what is part of the object and what is not. In other words, we are able to segment out an object without having seen it before.

In perceptual psychology, the rules of the organisation of perceptual scenes are called Gestalt principles [9]. Psychologists have identified several principles, such as proximity, common fate, similarity, continuity, closure, symmetry and convexity. The Gestalt principle of continuity is illustrated in Figure 1a, where the human visual system groups some of the line segments to form a circle.

What makes the Gestalt principles interesting in the current context is that they can be learnt from data. In neural terms, the Gestalt principles can be implemented by giving positive connections between certain neurons in one area and some other neurons in an adjacent area. Learning the connections can be based on simple correlations found in the data. For example, features responding to lines of certain orientation in one part of the visual field are more probably co-activated with fea-
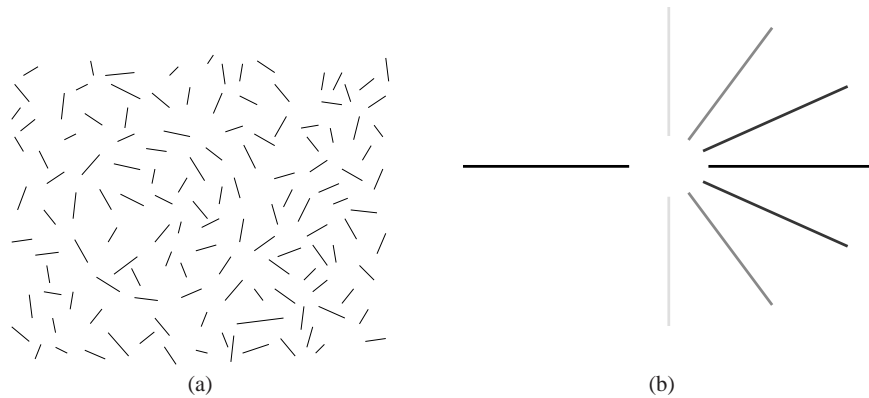
(a)   (b)

Fig. 1 a) Because of the Gestalt principles, a circle is perceived rather than some other grouping of the lines. b) Gestalt grouping of the neurons. The lines are features coded by different neurons. The shades of gray illustrate the connection strengths between the neurons on the right and the neuron on the left, darker meaning stronger. The lateral connections are stronger when the Gestalt principle is better fulfilled.

tures of similar orientation in some other part of the visual field. This mechanism is illustrated in Fig. 1b.

These "neural" Gestalt rules can be learnt from the data and they operate on the level on individual feature-coding neurons. The principle is therefore applicable to any modality and also between modalities unlike, for example, many segmentation procedures that make use of the spatial structure of visual images. Moreover, the neural Gestalt rules can be learnt locally and in parallel. In the visual domain this means that the local correlations found in familiar objects generalise to new objects which have different overall shapes but nevertheless obey the same local correlations.

2.2 Segmentation by coupled oscillators

In the cerebral cortex, the representation of objects is distributed. For instance, colour and movement are represented in different visual cortical areas in humans and other primates. Moreover, the brain represents the objects in many levels of abstraction. For instance, the cerebral cortex can recognise and represent the identity (lion or hammer), category (predator or tool) and emotional significance (dangerous or useful) of objects.

Since there are typically multiple objects present in the world, the brain needs to represent which feature belong together. Von der Malsburg [8] suggested that the brain would achieve this by a temporal code, by synchronising the neural firing

of the neurons which represent features belonging together. Such object-specific synchronisation was found experimentally by Gray and Singer [5] and Eckhorn et al [4].

From an engineering perspective, segmentation by synchronisation of coupled oscillators seems to be an attractive option because it should be able to work in a hierarchical neural network. Such networks are used in practice when recognising objects. It should therefore be possible to combine segmentation and recognition into an iterative process that solves both problems at the same time.

Modelling work has shown that coupled oscillators are able to synchronise their firing under suitable conditions (for a review, see e.g. [6]). This property has also been used for segmentation in many models (e.g., [12, 1]; for a recent review, see [10]).

2.3 Biased competition model for attention and learning

Based on psychophysical experiments, contextual (predominantly top-down) biasing of local lateral competition had been proposed as a model of covert attention in humans [3]. Usher and Niebur [11] then suggested a computational model for biased competition that has been shown to replicate many attentional phenomena, for instance both bottom-up and top-down aspects of attention [2].

Deco and Rolls [2] also showed that it is possible to learn the weights for contextual biasing by the mechanism outlined in Sec. 2.1. In other words, the neural Gestalt rules can be applied in a relatively straight-forward manner to implement selective attention.

Yli-Krekola [13] combined biased competition model with competetive learning. This model not only supports attention but forms a good basis for learning hierarchical feature representations suitable for categorising objects. The model already includes the lateral connections which can be used for segmentation. However, the model selects just one object. This makes it difficult to correctly segment objects if there are features which could belong to more than one object. For segmentation, it would usually be better that each feature is assigned to the object where it fits best. Such an assignment is in practice possible only if the model represents several objects at once or sequentially.

Yli-Krekola et al [14] added a habituation mechanism to the model. The purpose was to make the model switch between different objects and thus improve the segmentation capability. In this paper we further develop the model by replacing the habituation mechanism by oscillators. Essentially we are combining coupled-oscillator models with biased competition model. This development is a step towards models which would support learning, segmentation and selection of relevant objects in a hierarchical network.

3 Design for Emergence

Our goal is to combine the oscillator ideas (Section 2.2) with the biased competition model (Section 2.3). Both oscillations and competition are found on cerebral cortex so they should be compatible.

The overall structure of our network is such that there are so called areas that correspond to patches in the image. The areas get bottom-up input from the pixels. The areas should be connected to each other with local interactions only, that is, there is no hierarchy or global signals. The different areas should work in the same way, using the same algorithms.

We start from four high-level emergent properties and continue with requirements for one area that would lead to those properties.

3.1 Expected Emergent Properties

A1　The network should integrate information from local patches. When seeing an object, the local features belonging to it should start to oscillate. The local features strengthen each other if the features are compatible, for instance in the sense that an edge is continuous. These pairwise connections translate into the gain of the whole population corresponding to the object.

A2　When a scene contains many objects or many different interpretations, the object with globally highest gain emerges. The attention is thus drawn to the most obvious object.

A3　The oscillations of an object should synchronise internally and completely. Not only should pieces of objects synchronise to each other, but the synchrony should spread to the whole population.

A4　When there are multiple objects in a scene, they should desynchronise between each other. When using only local connections, this can of course only happen when the objects overlap. This should not be considered as restrictive as it might first sound. When building a hierarchical representation of the scene, objects do "overlap" on high-level representations with large receptive fields, even if they do not on the pixel level or on the local feature level.

3.2 Requirements for One Area

We hope to accomplish the above emergent high-level properties by incorporating the following low-level properties in a single area.

B1　Not many neurons (e.g., less than one percent) should be active at the same time. This is required for attention (A2).

B2   The activity or oscillation of a single neuron should not grow without limits regardless of the strength of the bottom-up input or lateral support.
B3   Without any lateral support from other areas, the neuron should not start to oscillate clearly.
B4   With constant lateral support, the oscillation should become clearer.
B5   With oscillating lateral support, the neuron should oscillate with the same frequency and phase as the lateral input.
B6   The amplitude of the oscillation should describe the bottom-up activation and the oscillation of the lateral support. The requirements B2–B6 are the basic requirements for the oscillatory system in general (A1).
B7   As the different neurons in an area compete for their time to become active (as required in B1), their oscillations should become antisynchronised, that is, the phase differences in lateral inputs should become even stronger in the activations. This should lead into global desynchronisation (A4).
B8   The change from one neuron being active to another being active, should happen quickly and completely. This should help in finding a clear separation between the segments (A3).

4 Implementation

We propose a model that has the above properties. Since the full model is rather complicated, we start by defining a single oscillator and show how it works, and only then expand it to the full model with several areas each consisting of several neurons.

4.1 Single Oscillator Model

We start by introducing a single oscillator (or neuron) consisting of a pair of signals, called activity $x(t)$ and inhibition $y(t)$. The pair works a bit like cos and sin functions that are each other's derivatives with one minus sign: $\cos'(t) = -\sin(t)$ and $\sin'(t) = \cos(t)$. The system is activated by a bottom-up input $u(t)$ and it has intrinsic Gaussian noise $n_k(t)$.

$$x(t+1) = \left[(1-\alpha_1)x(t) - \beta_1 y(t) + \gamma u(t) + \varepsilon_1 n_1(t) \right]_+ / \delta \tag{1}$$

$$y(t+1) = \left[(1-\alpha_2)y(t) + \beta_2 x(t) + \varepsilon_2 n_2(t) \right]_+ \tag{2}$$

$$n_k(t) \sim N(0,1) \tag{3}$$

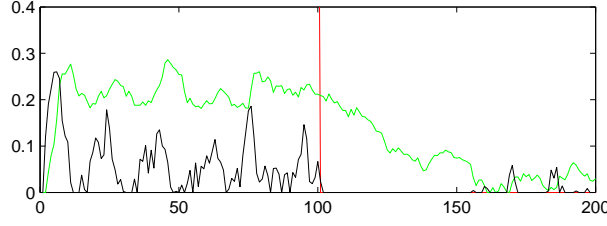$$[x]_+ = \max(0,x) = \frac{x+|x|}{2} \tag{4}$$

Fig. 2 An experiment with a single neuron. The time series of the activity $x(t)$ is shown in black, and the inhibition $y(t)$ in green. The bottom-up input $u(t)$ is one until t=100 and zero after that.

We use a nonlinearity $[\cdot]_+$ to keep all the signals positive. This is important to ensure that the negative terms such as the inhibition $-\beta_1 y(t)$ cannot activate the system. Parameters $\alpha, \beta, \ldots$ are positive numbers whose values are given in Table 4.2.

Figure 2 shows how the system responds to a bottom-up input that first activates the oscillator at time after $t = 0$, and then shuts it down after $t = 100$. As required in B3, the oscillation does not have a constant frequency, but it is heavily affected by noise instead. This will help the system to adapt to other signals when the model is enriched below. Note also that the response is true to the bottom-up input in the sense that when the input disappears, the activity $x(t)$ diminishes very quickly.

4.2 Full Model Definition

In the full model, each neuron $i = 1 \ldots n$ has five non-negative signals at time t, the bottom-up input $u_i(t)$, the output activity $x_i(t)$, the inhibition $y_i(t)$, the gain control $g_i(t)$, and the lateral support $v_i(t)$. The neurons are organised in areas $a = 1 \ldots A$ such that each neuron i belongs to exactly one area $a(i)$. $\mathbf{W}$ is an $n \times n$ non-negative weight matrix that has been learned separately.

The discrete time dynamics of the signals are as follows:

$$x_i(t+1) = \Big[(1-\alpha_1)x_i(t) - \beta_1 y_i(t) + (\zeta_1 v_i(t) + \gamma)u_i(t) \tag{5}$$

$$- \eta_1 \sum_{j \in a(i)} x_j(t) + \varepsilon_1 n_{1i}(t)\Big]_+ \Big/ [g_i(t+1) + \delta]$$

$$y_i(t+1) = \Big[(1-\alpha_2)y_i(t) + \beta_2 x_i(t) + \zeta_2 v_i(t)u_i(t) \tag{6}$$

$$- \eta_2 \sum_{j \in a(i)} y_j(t) - \theta \sum_{j \in a(i)} x_j(t) + \varepsilon_2 n_{2i}(t)\Big]_+$$

$$g_i(t+1) = (1-\alpha_3)g_i(t) + \alpha_3\left[(1-\kappa)x_i(t) + \kappa\frac{\sum_{j \in a(i)} x_j(t)}{|a(i)|}\right] \tag{7}$$

$$v_i(t) = \sum_j W_{ij} x_i(t) \qquad (8)$$

$$n_{ki}(t) \sim N(0,1) \qquad (9)$$

The nonlinearity $[\cdot]_+$ picks the positive part of the input. The bottom-up input feeds the activity especially if it has lateral support, see the term $(\zeta_1 v_i(t) + \gamma)u_i(t)$. A competition of activities within an area (B1) is produced by the term $-\eta_1 \sum_{j \in a(i)} x_j(t)$. The activities are saturated (B2) by dividing by the gain control signal plus constant, $g_i(t) + \delta$. The inhibition signal $y_i(t)$ has many similar terms as the activity itself. The gain control signal $g_i(t)$ is a running average of the activity of the single neuron and the average activity in the area. The lateral support signal $v_i(t)$ is only an auxiliary variable that collects the activities of other neurons mapped through the weight matrix $\mathbf{W}$. Weights $W_{ij} = 0$ whenever neurons i and j are in the same area.

We used the parameter values given in Table 4.2. The parameters were found by tuning them by hand in order to achieve the desired properties described in the previous section. The signals were initialized to zeroes, that is, $x_i(1) = y_i(1) = g_i(1) = 0$.

α_1	α_2	α_3	β_1	β_2	γ	δ	ε_1	ε_2	ζ_1	ζ_2	η_1	η_2	θ	κ
0.4	0.05	0.01	0.2	0.2	0.05	0.3	0.01	0.01	0.4	0.12	0.09	0.09	0.02	0.7

**Table 1** Values of parameters used in the experiments.

5 Experiments

The experimental part includes two examples. In the first experiment, we show how a single area syncronises to a fixed lateral input. In the second experiment with artificial image data, we show how the system segments an image.

5.1 Experiment with a Single Area

In the first experiment, we used only a single area $a = 1$ and three neurons $i = 1, 2, 3$. The bottom-up input $u_i(t) = 1$ is a constant 1 for each neuron the whole time. Instead of modelling the lateral support with Equation (8), we used the signals $v_i(t)$ shown in the top-most subfigure of Figure 3, where two neurons get oscillating lateral support, but the third one does not.

The resulting signals are shown in Figure 3. The two neurons with oscillating lateral support start to oscillate strongly, whereas the third neuron only very mildly (see requirements B3–B6). Note that the active phases in $x_i(t)$ are better separated

Fig. 3 An experiment with a single area. Each subplot shows the signals as a function of time t, from top to bottom, lateral support $v_i(t)$, output activity $x_i(t)$, inhibition $y_i(t)$, and the gain control $g_i(t)$. Each colour corresponds to one neuron i. Note that the bottom-up input $u_i(t)$ is a constant 1 for all neurons.
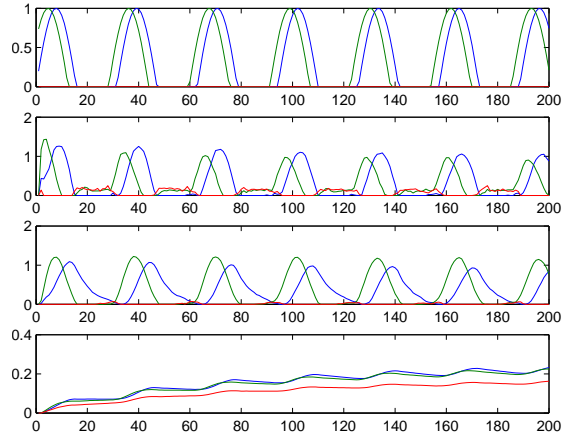
than in the lateral support $v_i(t)$ (B7). This is a good property that will help to separate the oscillations of separate object to have separate phases (A4).

5.2 Experiment in Image Segmentation

We generated artificial data for the image segmentation problem. Grey scale images of 30×30 pixels that each contain one object as shown in Figure 4. The images were divided into 5×5 areas of 10×10 pixels each with 5 pixel overlap in each direction. Because of overlapping areas, the grey scale values were multiplied with a kernel $\sqrt{(1+\cos(\phi))/2}$ in both horizontal and vertical directions where ϕ goes from $-\pi$ to π within the 10 pixel segment. The patches from each area of 10000 examples were pooled into one and 64 features were learned with k-means clustering. The features looked like 63 different curve segments and one empty feature.

We transformed the data into binary activations of the features such that exactly one of the 64 features i got $u_i(t) = 1$ for all t (the nearest prototype vector in k-means), and the others have $u_i(t) = 0$ for all t. The empty feature was then dropped out. Weights were learned by computing the correlation coefficient ρ_{ij} for each feature over the data set and setting $W_{ij} = \rho_{ij}^{0.7}$ when $\rho_{ij} \geq 0.06$ and $W_{ij} = 0$ otherwise. Using sparse connectivity is important for computational efficiency. Some of the weights are visualized in Figure 4.

For testing the segmentation, we introduced two objects into the same image, as shown in Figure 5. First we tried to feed the combination on the pixel level (top right subplot) to activate several features in each area. This led into a problem where a contour from a single object activates more than one feature, which start to compete with each other and oscillate in different phases. This issue is further discussed in Section 6. What we did instead, was to go around this problem and find (at most) one active feature for each object in each area, and combine the activations only
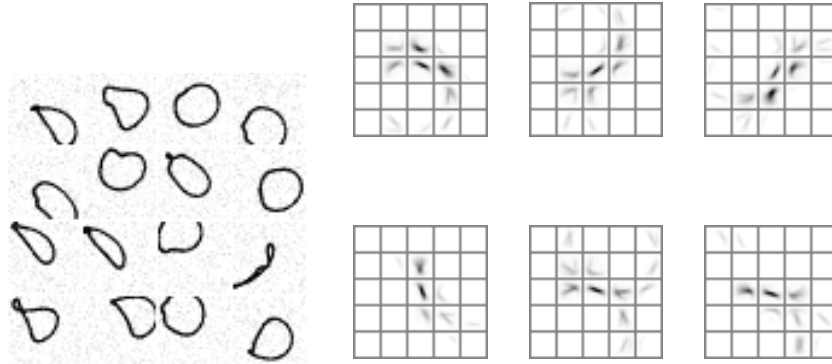
Fig. 4 Left: Example images from the training data. Right: Some of the learned weights. Each subfigure shows the sum of all the weights corresponding to the feature in the middle patch.

Fig. 5 Two objects are inserted into the same image as a test case for segmentation. The top row shows the generated and combined images, the middle row shows the activated features without the overlap, and the bottom row shows the activated features with overlap which is also the reconstruction of the images based on the feature activations.
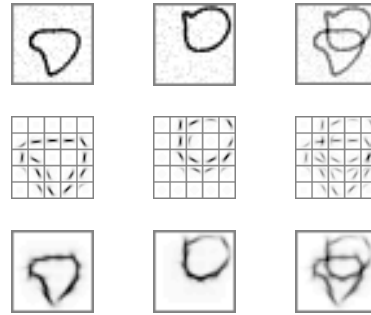
afterwards (see middle row in Figure 5). This way we ensured that similar features activated by just one contour do not start to compete with each other.

The results are shown in Figure 6. The activities caused by the two object start to oscillate with opposite phases (A1–A4) and the segmentation happens in just a few wavelengths of the system. The activities $x_i(t)$ at times $t = 101 \ldots 300$ were fed into non-negative matrix factorisation [7] in order to represent them compactly as a product of two time series and two sets of features. The two sets of features are shown in the result figure and they show the two objects well separated.

6 Conclusion and Discussion

We have shown preliminary but promising results of a system that combines oscillations for segmentation and biased competition for attention. The problem of finding the subset of features that best support each other to form objects is a very difficult one. Instead of making an exhaustive (exponentially difficult) search for
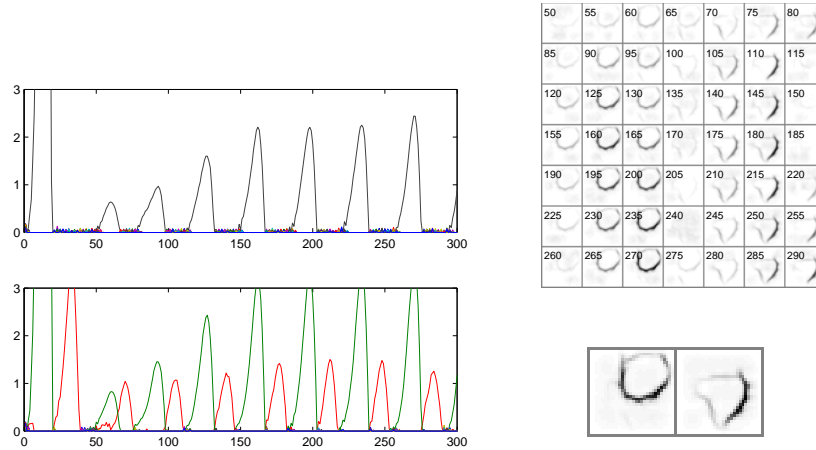
Fig. 6 Left: The activities $x_i(t)$ corresponding to the middle patch (top) and the patch below it (bottom) plotted as a function of t. Note that the largeness of the first peaks could be easily avoided with a reasonable non-zero initialisation for the gain control $g_i(1)$. Top right: The reconstruction of images based on the activities $x_i(t)$ at different times t. Bottom right: The segmentation result obtained from non-negative matrix factorisation of the signals $x_i(t)$.

such subsets, our approach solves it by using the emergent properties of oscillators. Experiments confirm that it takes only a few wave-lengths before the oscillations synchronise to find the objects.

We noticed an important problem in the implementation. Neurons in the same area that represent correlated features such as edges with only a slight difference in orientation or location, start to compete with each other. This causes problems since a single object will in general activate many of such features, which causes unwanted competition within a single object. In the current implementation we eliminated this problem in an artificial manner that cannot be used in a real problem. We still need to solve this problem by changing the competition mechanism such that it would take into account the correlations within an area.

Our current implementation may be excessively complex, for instance, it is very likely that some of the terms in the system dynamics described in Equations (5–9) could be left out by changing the other parameters. In future, we hope to make it as simple as possible. We would also like to set the parameter values automatically in order to achieve as fast and accurate segmentation as possible.

The computational complexity of the proposed system is very high even though we already use a sparse connection matrix **W**. There are many ways to make the system faster. Firstly, we could skip modelling the activations of neurons that do not get enough bottom-up input. This way, we would miss the phenomenon of false contours, though. Secondly, we could change the time-scale of the system such that one would need fewer time points to see the oscillations settle down. Thirdly, a parallel implementation of the system would be very efficient due to the design of local computations only.

A natural extension of the current work is to further bias the competition with connections from other parts of a larger network. For instance, if the system is looking for food, the food-like objects should draw the attention more easily than others.

References

[1] Choe Y, Miikkulainen R (1998) Self-organization and segmentation in a laterally connected orientation map of spiking neurons. Neurocomputing 21(1–3):139–158

[2] Deco G, Rolls ET (2004) A neurodynamical cortical model of visual attention and invariant object recognition. Vision Research 44:621–642

[3] Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annual Review of Neuroscience 18:193–222

[4] Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitboeck HJ (1989) Coherent oscillations: A mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. Biological Cybernetics 60:121–130

[5] Gray CM, Singer W (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. Proceedings of the National Academy of Sciences 86:1698–1702

[6] Hoppensteadt FC, Izhikevich EM (1997) Weakly Connected Neural Networks. Springer

[7] Hoyer PO (2004) Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research 5:1457–1469

[8] von der Malsburg C (1981) The correlation theory of brain function. Tech. Rep. Internal Report 81-2, MPI Biophysical Chemistry, Gttingen, Germany

[9] Todorovic D (2008) Gestalt principles. Scholarpedia 3(12):5345

[10] Ursino M, Magosso E, Cuppini C (2009) Recognition of abstract objects via neural oscillators: Interaction among topological organization, associative memory and gamma band synchronization. IEEE Transactions on Neural Networks 20(2):316–335

[11] Usher M, Niebur E (1996) Modeling the temporal dynamics of it neurons in visual search: A mechanism for top-down selective attention. Journal of Cognitive Neuroscience 8:311–327

[12] Wang DL, Terman D (1995) Locally excitatory globally inhibitory oscillator networks. IEEE Transactions on Neural Networks 6:283–286

[13] Yli-Krekola A (2007) A bio-inspired computational model of covert attention and learning. Master's thesis, Helsinki University of Technology, Finland

[14] Yli-Krekola A, Särelä J, Valpola H (2009) Selective attention improves learning. In: Proceedings of the International Conference on Artificial Neural networks (ICANN), pp 285–294