An Oscillatory Neural Network Model of Object Selection and Segmentation in the Visual Scene

Y.B. Kazanovich[1], R.M. Borisyuk[1,2], V. Tikhanoff[3], A. Cangelosi[3],

[1]*Institute of Mathematical Problems in Biology, Russian Academy of Sciences, Pushchino, Russia*

[2]*Centre for Theoretical and Computational Neuroscience, University of Plymouth, Plymouth, UK*

[3]*School of Computing, Communications and Electronic Engineering, University of Plymouth, Plymouth, UK.*

Abstract

An oscillatory neural network model is presented that allows selection of a specified object from the image. The processing of the image is divided into two stages. The first stage implements contour extraction by applying traditional image processing algorithms. The result is raw contours accompanied by the noise and some spurious objects. During the second stage the searched object is segmented from the image, its boundaries are determined, and the noise is suppressed. The second stage is implemented by a two layer network of phase oscillators controlled by a central oscillator. The extraction of an object is made in terms of the temporal correlation hypothesis. The oscillators coding the selected object form an assembly with coherent activity which also runs in phase with the central oscillator. Other oscillators do not show synchrony with this assembly. The work of the model is illustrated on an example of the image used in robotics.

1. Introduction

Extraction of a certain object from the image is a traditional problem in computer vision and robotics. It is also attracts attention of psychologists and neurobiologists who are interested in understanding the psychological and neurobiological mechanisms underlying visual object selection, in particular, how attention determines the result of selection. The problem of object selection is closely related to the problem of image segmentation because the selected object should be segmented from other objects in the image and from the background. This task may be relatively easy if the image contains objects which are isolated and located on a background whose optical characteristics are homogenous and essentially different from those of the searched object. In real images objects can overlap and the background can be non-homogenous which makes the problem of segmentation rather difficult.

Despite the fact that humans use more or less similar intuitive strategies for object selection and segmentation, it is hardly possible to invent a formal and universal measure of segmentation quality. It is clear that segmentation depends of the context, previous experience, and internal aims that are far beyond the information contained in the image itself. Computational methods that are used in this field are based mostly on intuition and common sense. Usually the procedure is divided into two stages. At the initial segmentation stage, some parts of the searched object are segmented basing on optical characteristics of these parts. At the recognition stage, a complete object is composed from its parts using stored memory and logical analysis. These stages can be iteratively repeated to improve the

results of selection and recognition. It is assumed that the computational procedures should be robust in the presence of noise and natural variation of objects and the background. Those methods are preferable that can be adapted to a larger class of images and different types of searched objects through supervised or unsupervised learning.

In the last years a lot of investigations have been made to clear out how the problem of object selection and segmentation is solved by the brain. It is known that different types of features (such as geometrical, spectral and motion characteristics of objects) that are simultaneously present in the visual stimulus are initially processed in different parts of the cortex and only later in associative areas of the cortex they are combined into representation of individual objects. In this relation, the questions arise: a) how the brain is able to keep the information about associations between individual features and objects to which these features belong and b) what is the mechanism that implements feature binding?

The theory that tries to answer these questions is based on the so-called temporal correlation hypothesis (TCH) (Malsburg, 1981; Singer and Gray, 1995) which states that the features of a single object are coded (binded) by coherent neural activity while there is no correlation of the activity corresponding to different object. Note that segmentation of an object can be considered in the frames of the binding problem. Suppose that an image is represented in the form of optical parameters of its pixels. If these parameters are included in the list of features, then attribution of these features to particular objects will result in segmentation of objects from each other.

Besides feature binding, another cognitive function plays an important role in image processing. This is attention which is used to select a particular object from the image. The experiments show that attention operates in a similar way as binding: if attention is focused on an object, this results in increasing coherence in the activity of those neurons that represent this object in the cortex (Steinmetz et al., 2000; Fries et al., 2001; Fries et al., 2002; Doesburg et al., 2008). Attention can be directed to a particular area of the image (spatial attention) or to some features of the object (object-based attention). If it is known that a particular area of the image or particular features belongs to the searched object, then the whole object can be restored since all its features are coded by the same coherent activity as any of its part.

The TCH suits well to modelling in terms of oscillatory neural networks (see reviews Ritz and Sejnowski, 1997; Wang, 2005). A general idea that is present in most models of binding is to use lateral synchronizing connections to obtain coherent activity of neurons representing a single object and to use long-range desynchronizing or inhibitory connections to make incoherent the activity of neurons representing different objects. Another idea is to set the connection strengths between the neurons in such a way that neighbouring neurons tend to work coherently if image areas located in their receptive fields have similar optical characteristics. If both ideas are combined, one can expect that in-phase working clusters of neurons will appear in the network as a result of its evolution, and the segment of the image that corresponds to each cluster will have similar or slowly changing optical characteristics.

A large variety of models of object selection and segmentation based on synchronization of neural activity have appeared in the last years (Wang, 1999; Wang and Terman, 1997; Chen et al., 2000; Chen and Wang, 2002; Broussard et al., 1999; Labbi et al., 2001; Borisyuk and Kazanovich, 2004; Palm and Knoblauch, 2005; Buhmann et al., 2005; Ursino et al., 2003; Ursino and La Cara, 2004a; Zhao and Macau, 2001; Zhao et al., 2003; Zhao et al., 2004).

They differ by the degree to which biological facts are taken into account, by the type of processed images, by the mechanisms of functioning, and by the results of application. Some authors try to closely follow the experimental results, other are more interested in practical tasks of image processing. The models that work with grey-scale or coloured images are usually built of neurons or neural oscillators whose receptive fields are represented by pixels of the image. Multilayer constructions are used if pixels are characterized by a set of features (e.g., spectral components of the colour). The most advanced models working with real images are reported to give the results that are comparable or even exceed those obtained by traditional image processing methods (Chen and Wang, 2002). Unfortunately, the best results are obtained for those processing algorithms that are rather complex and do not have support in experimental evidence. Moreover, in many cases the results critically depend on the parameter values.

In some papers the problem of segmentation is considered separately from the problems of attention and object selection. Other papers include object selection in their functionality but in this case consecutive selection of all objects present in the visual scene is implemented. In this work we suggest a model that combines a particular object selection with segmentation of this object from other objects and the background.

The whole procedure is divided into two stages. At the first stage, traditional image processing technique is used to extract contours of objects. The only restriction on the algorithms used at this stage is that they should have evident neural implementation and be able to process the image in parallel. The result of the first stage is raw contours with noise and with some number of spurious objects. At the second stage the searched object is segmented from the image, its boundaries are determined, and the noise and spurious objects are suppressed. This is done by an oscillatory neural network composed of phase oscillators.

The network has two layers whose activity is controlled by a special central oscillator (CO) that plays the role of the central executive of the attention system (Cowan, 1988; Baddeley, 1996). The oscillators in the layers are called peripheral oscillators (POs). The first layer fulfils the synchronization according to the TCH using the contours obtained at the first stage as restrictors for synchronization spread outside the border of the searched object. The second layer transforms the raw image into the final results of segmentation.

The central oscillator is used to select a particular object in the focus of attention. It is assumed that the focus of attention is represented by those POs that work in-phase with the CO. The dynamics of the model are organised in such a way that the CO can only synchronize with the assembly of oscillators that represents the searched object. The CO also keeps oscillators representing other objects and the background outside of the focus of attention.

In simulations we use the image that was created for experiments with robots in the real world (Tikhanoff et al. 2008a; 2008b). The image is shown in Figure 1. There are four balls of different colour in the image that are targets for robot's manipulations. The robot should be able to select a ball of a predefined colour and to pick it by its hand. So the visual system of the robot should provide information about the position and the boundaries of the ball to the mechanical system that controls the movements of robot's hand. The model presented here was developed to solve this practical task. But the methods used in the model are universal for any coloured images where the searched object differs from other objects by its colour.

The paper has the following structure. In Section 2, algorithms for contour extraction from coloured images are described. In Section 3 we present the results of extraction of contours. In Section 4 an oscillatory network for image processing according to the TCH is presented. The results of simulations are shown in Section 5. Section 6 is devoted to the discussion of the results and comparison with other models.

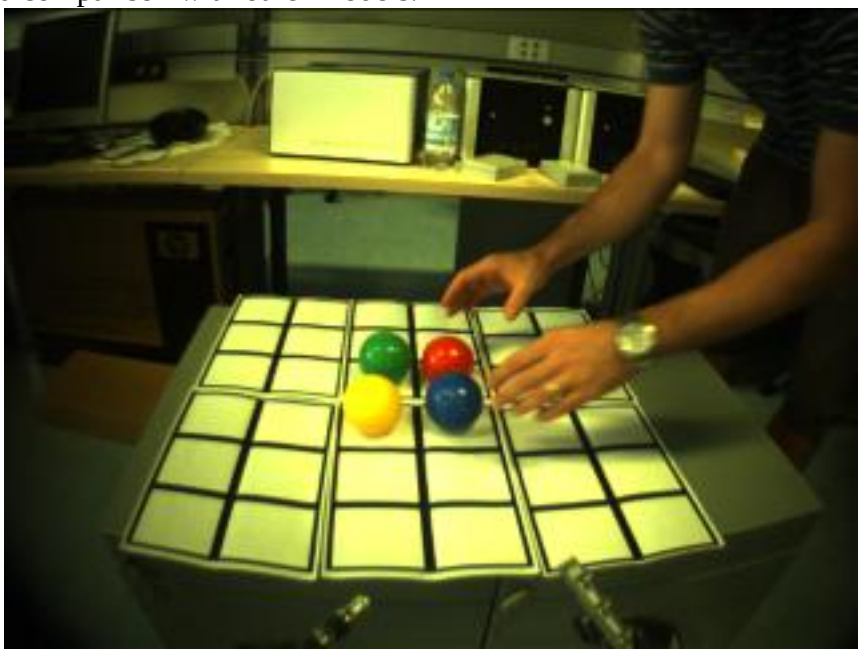

Figure 1. Original image of size 480×640 pixels. The final aim is to select from the image a ball of a predefined colour and to detect its boundaries.

2. Contour extraction

The human visual system is very efficient in detecting contours. In most cases it surpasses artificial systems in the solution of this task, though errors may appear if complex textures are present in the image or if the image is contaminated by strong noise. There may be different explanations for this efficiency but at least one of the reasons is that human visual system contains special neurons and neural structures that react to edges, that is to an abrupt change of some optical characteristics of the image, such as intensity or colour. Many algorithms of contour detection try to reproduce this ability of human vision by computing spatial derivatives of some functions determined in the pixels of the image.

Let *F(x, y)* be a function determined on the discrete plain where the image is located, (x, y) are the coordinates of a pixel on this plane. Different functions *F* can be used for contour extraction. In the case of grey-scale images intensities *I(x, y)* are used. In the case of coloured images the role of *F* can be played by the intensity of a component of the spectrum (e.g. red, green, or blue). Since R, G, B components are correlated (in the sense that if the intensity changes, all these components will change accordingly), they are often transformed to another set of parameters. It can be a linear transformation, e.g. for the parameter sets YIQ and YUV, or nonlinear transformation, e.g. for the parameter set HSI (Cheng et al., 2001). In the next section we specify the parameter set that has been used in our simulations.

Let $p = (x, y)$ and $\bar{g} = grad\, F(p)$. A pixel *p* can be considered as belonging to a contour if the following conditions are fulfilled:

$$|\bar{g}| > C_1 > 0, \quad (1)$$

$$\nabla^2_{\bar{g}}(F(p-\delta\bar{g}))\,\nabla^2_{\bar{g}}(F(p+\delta\bar{g})) < 0, \quad (2)$$

$$\nabla^3_{\bar{g}}(F(p)) < C_2 < 0. \quad (3)$$

Here $\nabla^2_{\bar{g}}$ and $\nabla^3_{\bar{g}}$ denote the second and third derivatives along the direction $\bar{g}$, respectively, δ is a small parameter, C_1 and C_2 are constant parameters. Formulas (1-3) have the following meaning. Formula (1) states that the function *F* should be "steep" in the neighbourhood of *p*. Formula (2) states that the second derivative of *F* in direction $\bar{g}$ should change sign at *p*. Formula (3) states that the third derivative of *F* in direction $\bar{g}$ should be negative and below some threshold. We use expressions (1-3) as a definition of a contour point.

In the computation of spatial derivatives for real images the following difficulties must be overcome. Firstly, the derivatives have to be computed using the function *F* that is determined on a discrete set of points. Secondly, the results of computations should be made robust in the presence of noise. Third, the computation results will strongly depend on the scales at which the derivative operators are applied (Lindeberg, 1998; Sumengen and Manjunath 2005), therefore different scales should be used in computations.

To decrease the influence of the noise we use the operator

$$A_{S_r}(p) = \frac{1}{|S_r|}\sum_{q\in S_r} F(q), \quad (4)$$

where S_r is a square of size r with the pixel p in the centre, $|S_r|$ is the number of pixels in S_r. Thus operator (4) averages the values of *F* in the square neighbourhood S_r of the pixel *p*. By this operator the function *F* is transformed to the function

$$G(p) = F(p) + \alpha A_{S_{r_1}}(p) - \beta A_{S_{r_2}}(p), \quad (r_2 > r_1 \geq 3, \alpha > 0, \beta > 0), \quad (5)$$

where α and β are weighting coefficients. The result of transformation (5) is similar to processing the image by a DoG filter. It averages the noise and makes slopes a bit steeper.

The derivative operators are applied to the function G. Consider first the case when the derivatives should be computed along the direction e_x parallel to the axis *x*. Let S_r be a square with the centre in the pixel *p*; R_{rl} be a rectangle of size $r \times l$ adjoined to S_r from the right side as it is shown in Figure 2a; R^1_{rl} and R^2_{rl} be rectangles of size $r \times l$ adjoined to S_r from the left and right sides, respectively as is shown in Figure 2b.

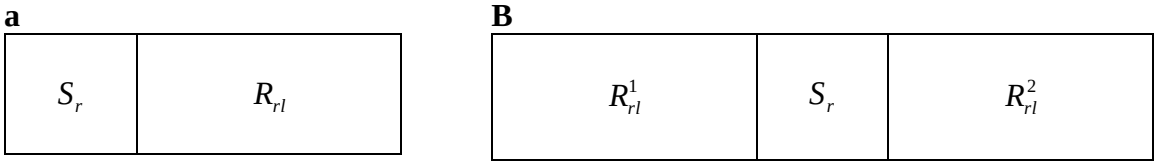

Figure 2. Computation of derivatives: a) areas used for the computation of the first derivative; b) areas used for the computation of the second and third derivative.

Denote

$$A_{R_{rl}}(p) = \frac{1}{|R_{lr}|} \sum_{q \in R_{rl}} F(q) \tag{6}$$

the average value of *F(p)* in the rectangle R_{rl}. The derivatives are computed as

$$\nabla^1_{e_x}(F(p)) = A_{R_{rl}} - A_{S_r}, \tag{7}$$

$$\nabla^2_{e_x}(F(p)) = A_{R^1_{rl}} + A_{R^2_{rl}} - 2A_{S_r} \tag{8}$$

$$\nabla^3_{e_x}(F(p)) = \nabla^2_{e_x}(F(p+\delta)) - \nabla^2_{e_x}(F(p-\delta)) \tag{9},$$

where δ is a small parameter (in computations we put $\delta = (r+1)/2$). Averaging used in (6-9) helps to suppresses the noise.

The computation of the derivatives along any other direction can be made by rotating the rectangle that combines S_r and R_{rl} or S_r, R^1_{rl}, and R^2_{rl} around the centre of the pixel *p* on the angle ϕ. In particular, the gradient at the pixel *p* is computed as a vector *e* along which the first derivative takes the highest value, that is

$$grad\, F(p) = \{e: \quad \nabla^1_e (F(p)) = \max_{\phi \in (0,2\pi)} \nabla^1_{e_\phi}(F(p))\}, \tag{10}$$

where e_ϕ is a vector that is rotated relative to e_x on the angle ϕ.

Note that different scales can be taken into account by varying the parameter *l* (the width of the rectangles R_{rl}, R^1_{rl}, and R^2_{rl}). In our computations we vary *l* in some range and check that for each value of *l* conditions (1-3) are fulfilled. Only in this case the pixel *p* is declared to be a contour point.

3. An example of contour extraction

The results of contour extraction crucially depend on the set of functions *F(p)* that are used in computations. For coloured images usually HSI (hue-saturation-intensity) colour space is used because in these parameters colour information is separated from intensity, which is important for having stable results for different illuminations. Since our task was to find an object of a particular colour, it was reasonable to use a priory information about these colours in selecting feature space.

According to our task four colours of the balls are of interest, that is red, green, blue, and yellow. In addition, the light-yellow colour of the table is useful since it plays the role of the background for the balls. Each of these colours is coded in RGB space as a vector c = (r, g, b). The components of this vector were computed by averaging RGB values in a small region chosen inside of each ball and inside of the square on the table. Thus we have got 5 vectors C_i $(i = 1,...,5)$ which have been used as templates to compare with the colours of pixels in the image.

Let c_j be the colour of the *j*th pixel in the image. We measure the distance between C_i and c_j by the angle

$$d_{ij} = A\cos\left(\frac{(C_i + U, c_j + U)}{|C_i + U|\,|c_j + U|}\right). \qquad (11)$$

The vector U = (u, u, u) (in computations u = 20) is added to the last formula to avoid small values in the denominator if $|c_j|$ is small. The advantage of this way of measuring the distance between colours is that this measure does not significantly depend on illumination.

By using d_{ij} five functions $F_i(p_j) = d_{ij}$ (i = 1,…,5) were formed. The results of their processing according to formulas (1-10) are presented in Figure 3.

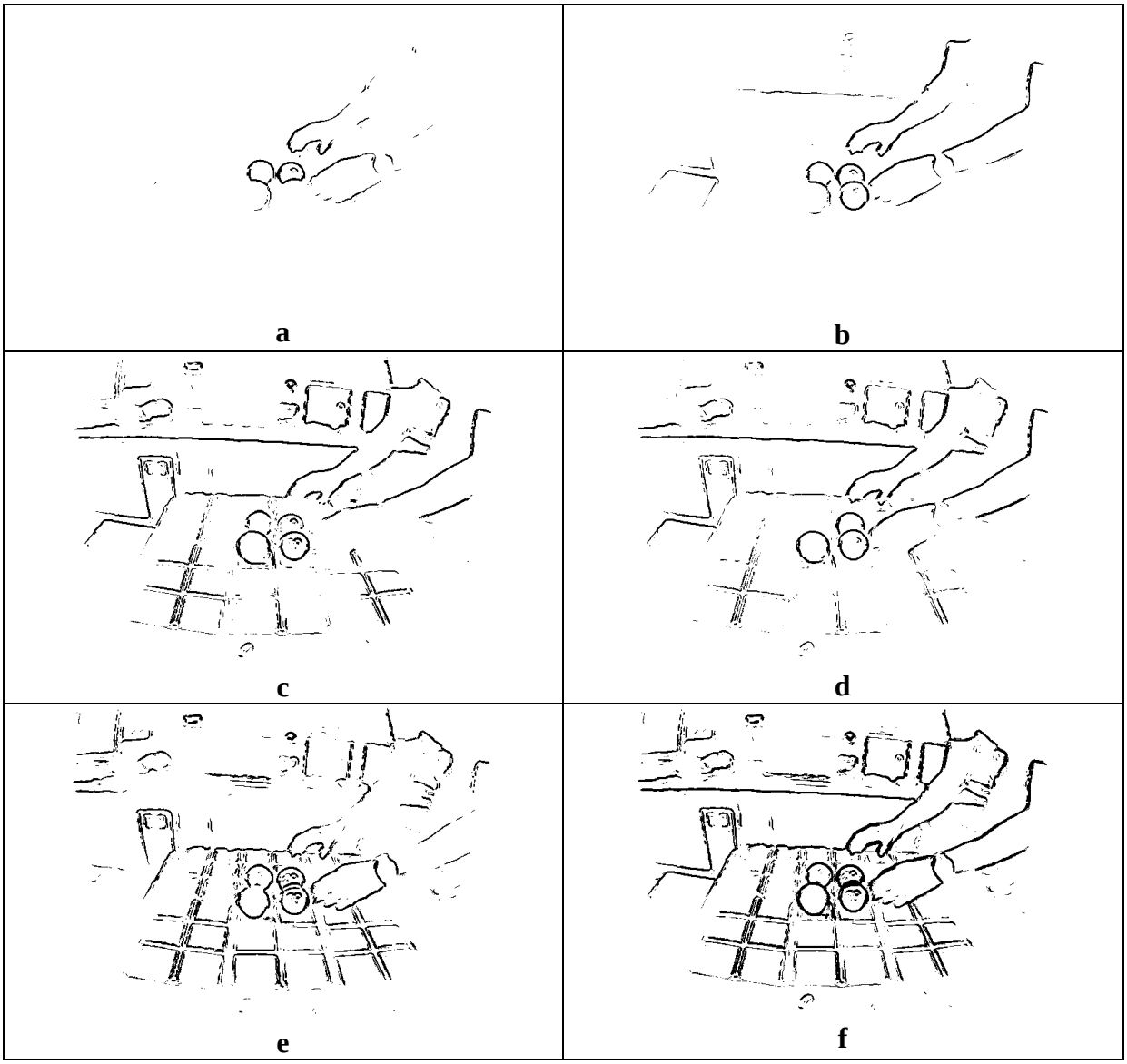

Figure 3. Contour extraction by using the templates for green (a), red (b), yellow (c), blue (d), and light-yellow (e). The final contour is shown in (f).

Frames (a)-(e) of Figure 3 show the results of contour extraction by using different templates. The last frame (f) shows the final result of contour extraction which is obtained by superposing all contour points of the frames (a)-(e). Figure 4 shows the fragment of the image that contains the balls (Figure 4a) and the contour points in this fragment (Figure 4b).

a

B

Figure 4. A fragment of the original image (a) and its contour points (b).

It is seen that the contours are noisy and many noisy fragments are present in the image. This is a typical result when contours are computed by using local filters. Therefore additional processing is needed to improve the results of contour extraction and to select a given object.

4. An oscillatory neural network for object selection

In this section we describe an oscillatory neural network that selects an object from the visual scene and focuses attention on this object. The selection is made in terms of the TCH, that is a selected object is coded by the activity of a synchronous assembly *A* of oscillators while oscillators coding other objects and the background work incoherently relative to *A*. Focusing attention on the object implies that *A* works coherently with a central oscillator and this coherence is used to identify the oscillators from *A*. As input information, the raw contours are provided to the network and also a small square *S* is given to mark an object that should be selected. The square *S* is located inside the boundaries of the selected object and is used as an initial attractor of attention. Starting from *S*, attention then spreads to the whole object. The contour points are used to restrict the spread of attention outside the borders of the object.

Since all computational procedures in the network are formulated in terms of synchronization / desynchronization, it is reasonable to use phase oscillators as the elements of the network [Kuramoto, 1984; Kazanovich and Borisyuk, 1999; Borisyuk and Kazanovich, 2003]. The activity of such an oscillator is described by a single variable, the phase of oscillations, and the interaction of these oscillators is described in terms of phase-locking. In our model we use synchronization that is induced by local connections. This type of synchronization has been studied in a number of papers [Sakaguchi et al., 1987; Daido, 1988; Strogatzt and Mirollo, 1988]. The main conclusion of the studies is that the value of the interaction coefficient should increase in order to synchronize a network of increasing size in 2D space. In other words, large interaction coefficient are needed to synchronize large networks of phase oscillators.

The network consists of two layers L_1 and L_2 of size $N = M \times L$ (thus N is the number of oscillators in a layer). Local connections in each layer are represented by the connections with 8 nearest neighbours (the number of connections can be less than 8 for oscillators on the

external boundary of the layer). There are bottom-up connections of a local type between the layers, that is each oscillator in L_2 receives the inputs from a square in L_1 of size $Q \times Q$ (in computations Q = 7). All local connections and bottom-up connections are of a synchronizing type. There is also a central oscillator *C* that interacts with oscillators of L_1. It receives synchronizing inputs from oscillators in *S*. It also sends desynchronising signals to all oscillators in L_1 except those in *S*. The synchronising input to *S* allows focusing attention on *S*. Desynchronizing signals from *C* to L_1 are used to desynchronize the oscillators that are in and out of the attention focus, respectively.

The oscillators of L_1 that correspond to contour points are supposed to be silent, that is they do not participate in network dynamics. Thus, in L_1 object boundaries are impenetrable for the spread of synchronization. There are no restrictions on the spread of synchronization in L_2, therefore oscillators in L_2 are involved in some type of synchronization according to the interaction with their neighbours in L_2 and L_1.

The natural frequencies of oscillators in L_1 are constant. They are randomly distributed in some interval $(\omega_{\min}, \omega_{\max})$ for all oscillators except those in *S*. The natural frequencies of oscillators in *S* are distributed in the range $(\omega_{\min} + s, \omega_{\max} + s)$ for some $s > 0$. A shift to higher frequencies for oscillators in the focus of attention is introduced to reflect experimental data that demonstrate higher activity of neurons that represent attended objects (Motter, 1993; Roelfsema et al., 1998; Kanwisher & Wojciulik, 2000). The initial values of natural frequencies of oscillators in L_2 are distributed in $(\omega_{\min}, \omega_{\max})$, but in contrast to L_1 the natural frequencies of oscillators in L_2 adapt to their current values.

Consider an oscillator *P* in L_2. Let *G* be the neighbourhood of *P* in L_1. If all oscillators in G belong to the same object, their work will be coherent, therefore *P* will be phase-locked by these oscillators and will work in-phase with them. If oscillators in G belong to different objects (this happens when P corresponds to a contour point or locates near the boundary separating different objects), they will compete for the synchronization with *P*. The larger is a synchronous assembly of oscillators in *G*, the greater is the chance that it will win the competition and that *P* will work coherently with this assembly. Also, local interactions in L_2 influence on the result of the competition smoothing the boundaries of segmented regions in L_2.

Adaptation of the natural frequency is also applied to the central oscillator *C*. Since the only synchronizing signals comes to *C* from the oscillators in S, after some short transitionally process C will work in-phase with these oscillators.

Equations for network dynamics have the following form.

$$\frac{d\theta_0}{dt} = \omega_0 + \frac{w_1}{|S|}\sum_{i \in S} g(\theta_i^1 - \theta_0), \tag{12}$$

$$\frac{d\theta_i^1}{dt} = \omega_i^1 - w_2(t)\sin(\theta_0 - \theta_i^1) + \frac{w_3(t)}{|N_i|}\sum_{j \in N_i}\sin(\theta_j^1 - \theta_i^1),\ (i = 1,..., N), \tag{13}$$

$$\frac{d\theta_i^2}{dt} = \omega_i^2 + \frac{w_4}{|N_i|}\sum_{k\in N_k}\sin(\theta_k^2 - \theta_i^2) + \frac{1}{|G_i|}\sum_{j\in G_i} w_{5j}\sin(\theta_j^1 - \theta_i^2),\ (i = 1,..., N), \tag{14}$$

$$\frac{d\omega_0}{dt} = \alpha(\frac{d\theta_0}{dt} - \omega_0), \tag{15}$$

$$\frac{d\omega_i^2}{dt} = \beta(\frac{d\theta_i^2}{dt} - \omega_i^2),\ (i = 1,..., N). \tag{16}$$

Equations (12-14) determine the dynamics of oscillator phases, equations (15-16) determine the dynamics of natural frequencies. The following notation is used in (12-16): θ_0 is the phase of the central oscillator *C*; θ_i^1, θ_i^2 are the phases of oscillators in layers L_1 and L_2, respectively; ω_0 is the natural frequency of *C*; ω_i^1, ω_i^2 are the natural frequencies of oscillators in the layers L_1 and L_2, respectively; $|\cdot|$ denote the number of elements in the corresponding set; $w_1, w_2.(t), w_3(t), w_4, w_{5j}$ are positive interaction parameters (w_1, w_4 are constants, $w_2.(t), w_3(t)$ depend on time, w_{5j} exponentially decays with the distance between the pixels *i* and *j*). The derivatives in the left part of (12-14) describe the current values of oscillator frequencies. According to (15-16), the natural frequencies of *C* and of the oscillators in L_2 are adapted to the current frequencies of these oscillators with rates α and β, respectively.

All interactions in the network are implemented by the function $\sin(x)$ except the interaction between *C* and *S* which is determined by the function $g(x)$ whose extremums are located near zero (Figure 5). Such interaction function accelerates computations and makes phase differences between *C* and oscillators in *S* smaller.

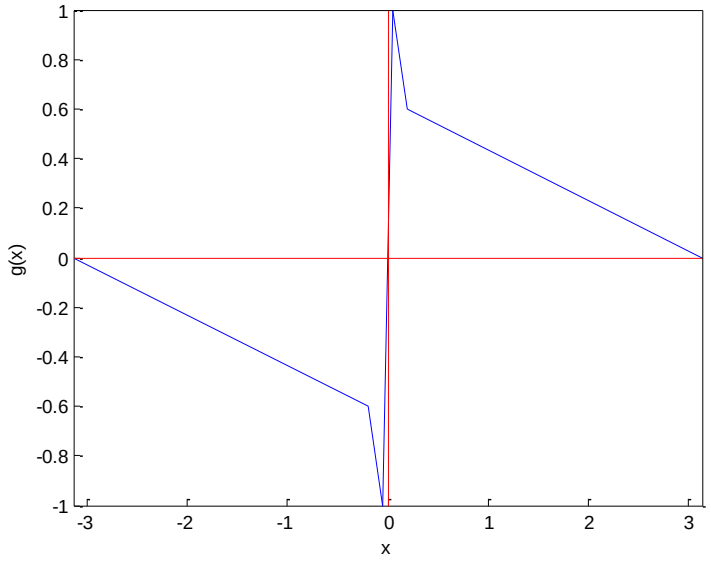

Figure 5. Graph of the function *g* (blue lines). Axes are shown by red lines.

The processes of synchronization / desynchronization in the network are controlled by the interplay between the variables $w_2.(t)$ and $w_3(t)$. The desynchronizing signals should not

prevent synchronization in the focus of attention. For this purpose, $w_2.(t)$ is made linearly decreasing to zero with time. In contrast, the values of $w_3(t)$ are made increasing as t^2 with time until the maximum value $w_{3\max}$ is reached. Making interaction in L_1 gradually stronger with time allows fast spread of synchronization on the whole object that should be selected in the focus of attention.

5. An example of object selection

We illustrate the model operation using the picture presented in Figure 4b. This image is of size 100×140. We restricted the processing to the fragment of Figure 1 (note that this fragment contains all target balls) because it reduces computation time but does not make computations easier in any other respect.

Initial phases of all oscillators were distributed randomly in the range (0, 0.2). Initial natural frequencies of all oscillators in layers L_1 and L_2 were distributed in the range (4, 5) which corresponds to the gamma frequency range 40 Hz – 50- Hz if the time unit is defined as $0.1/2\pi$ sec. The initial natural frequency of the central oscillator was 6.

Figure 6 shows the process of selection and segmentation of the green ball in the time interval from 1 to 8. Upper row of frames shows evolution of phases in layer L_1, lower row of frames shows evolution of phases in layer L_2. In each frame the colour of a pixel reflects the phase difference between the corresponding oscillator in one of the layers and the central oscillator. Phase differences are scaled in the range (0, 256) so that darker pixels correspond to the lower phase difference.

The process of synchronization starts from the square *S* that is clearly seen at the moment $t = 1$. Gradually the synchronization is spreads to the whole object that should be selected in the focus of attention. In parallel, the phases of other oscillators in the network tend to be different from the phase of the central oscillator (and hence different from the phases of oscillators in the focus of attention). The noise that is present in layer L_1 is suppressed in layer L_2 and all contour points are distributed between the selected object and other objects in the image. It can be seen that the final boundary of the selected object smoothes the defects in the original contour. This is obtained due to the local interaction between oscillators in L_2.

The process of selection and segmentation of other target objects (red, yellow, and blue balls) is shown in Figures 7-9.

Note that the process of segmentation of a target object takes 8-10 time units which corresponds to 130-160 msec which is in agreement with experimental findings on times needed for attention focusing.

6. Discussion

A two stage approach to selection of an object from the visual scene and its segmentation from other objects and the background has been suggested. The first stage is devoted to contour extraction and is described in terms of an artificial vision algorithm. At the second stage selection of an object marked by a small square and its segmentation are realized by an oscillatory neural network. Segmentation and focusing attention on a particular object are

fulfilled in terms of the temporal correlation hypothesis (TCH). Oscillators representing an objects work coherently and in addition oscillators representing an object in the focus of attention work coherently with the central oscillator.

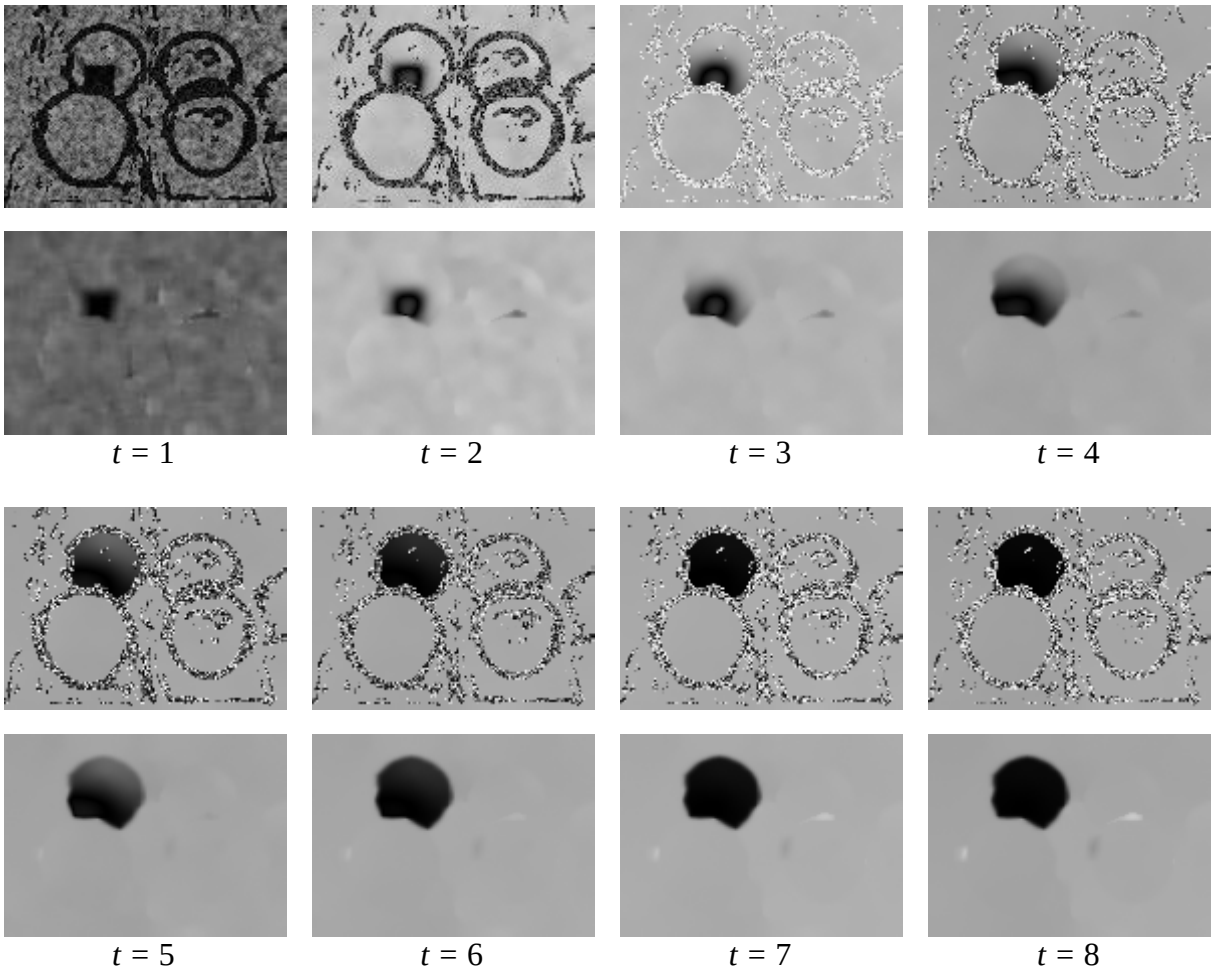

Figure 6. Selection and segmentation of the green ball. Each pixel in frames shows the difference between the phase of the central oscillator and the phase of an oscillator in layer L_1 (top frames) and L_2 (bottom frames). Phase differences $(0,2\pi)$ are scaled in the range (0, 256). Zero phase difference corresponds to black colour.

In developing the first stage we mostly followed a traditional approach to contour extraction. Gabor filters, filters of DoG type, and derivatives along the gradient direction for different scales have been widely used for this purpose both in artificial vision and neural network models (Lindenberg, 1998; Broussard et al., 1999; Sumengen and Manjunath, 2005; Petkov and Subramanian, 2007; Huang et al., 2008). The originality of our approach is in special combination of these methods and orientation on predefined colours of searched objects. Though the first stage is represented as an artificial vision algorithm, its neural implementation is possible. The operations used by the algorithm such as filtering and determination of the contrast of some optical characteristics are known from the experimental studies of the brain. Similar algorithms have been realised in neural networks (Broussard et al., 1999; Ursino and La Cara, 2004b; Petkov and Subramanian, 2007; Huang et al., 2008), but in most cases additional contextual modulation has been added to improve the contours and to fill the gaps in them.

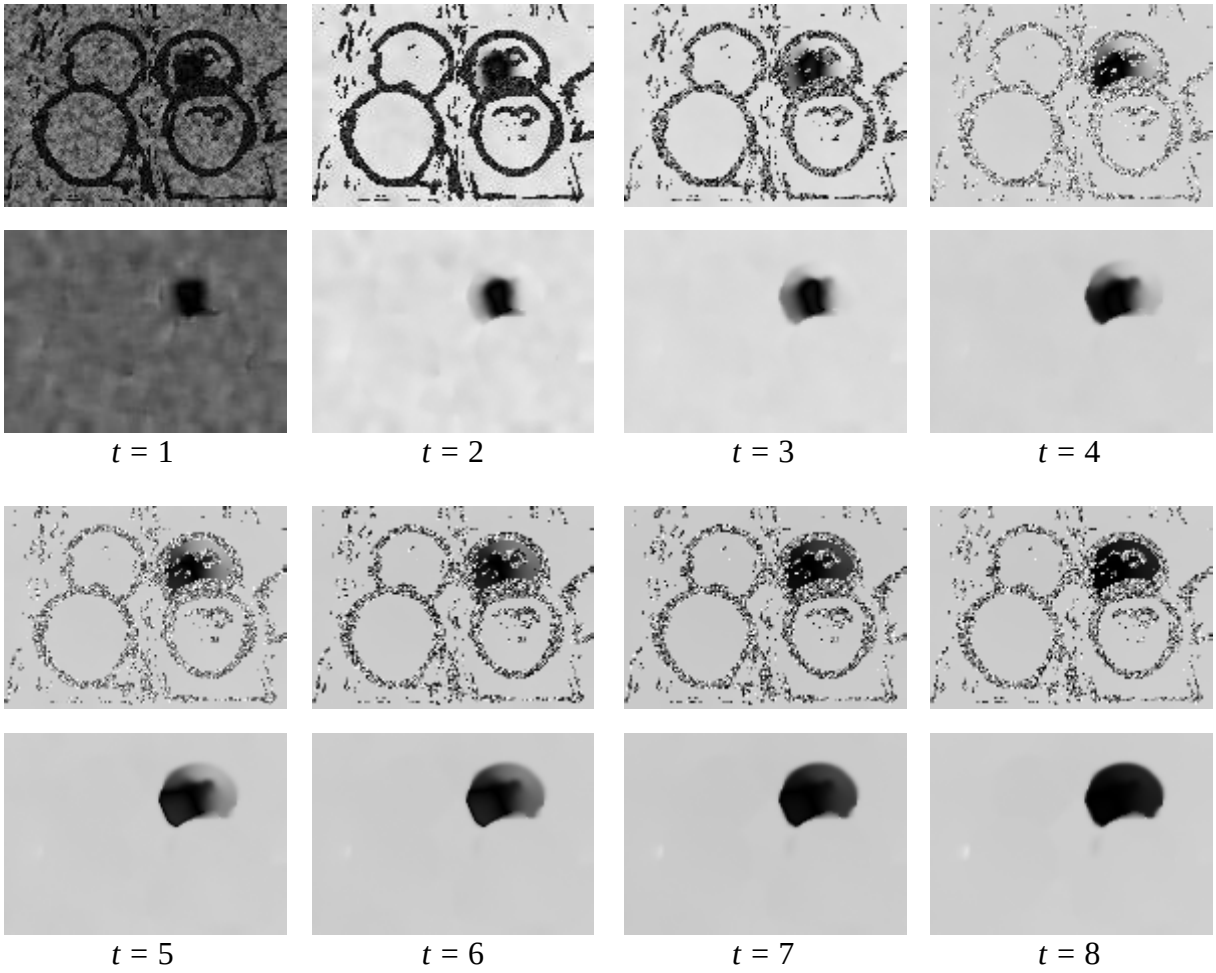

Figure 7. Selection and segmentation of the red ball.

A combination of artificial vision algorithms and an oscillatory neural network may look inappropriate for modelling the visual system, but in fact it may be closer to the reality than attempts to describe the work of this system by a single mechanism. Simple principles based on image filtering and rate coding seems to be most efficient on the earlier stages of processing while oscillatory mechanisms may participate in such cognitive functions as feature binding and attention. Our aim was to find how these different principles may interact in the solution of the problems of selection and segmentation. Our results show that oscillatory neural networks can be most useful in improving the raw information obtained during contour extraction.

The TCH assumes that synchronization that binds oscillations representing a visual object must be stopped at the boundaries of this object. There are two ways of realising this demand. The first one is implemented in the model LEGION (locally excitatory globally inhibitory oscillatory network) (Chen et al., 2000; Chen and Wang, 2002). LEGION is a single layer network built of Van der Pol type oscillators controlled by a central inhibitory unit. The process of synchronization in LEGION starts from oscillators that are called leaders and that are definitely located inside of meaningful objects. Then the synchronization is spread to the boundaries of objects and is stopped there due to a proper modification of local connection weights and input signals. Unfortunately, the principles of LEGION operation (especially the algorithm of connection weights modification) are rather complex and poorly justified from a biological point of view. LEGION can simultaneously process only a relatively low number

of objects, and errors in binding object features may appear (sometimes different objects are bound together by the process of synchronization).

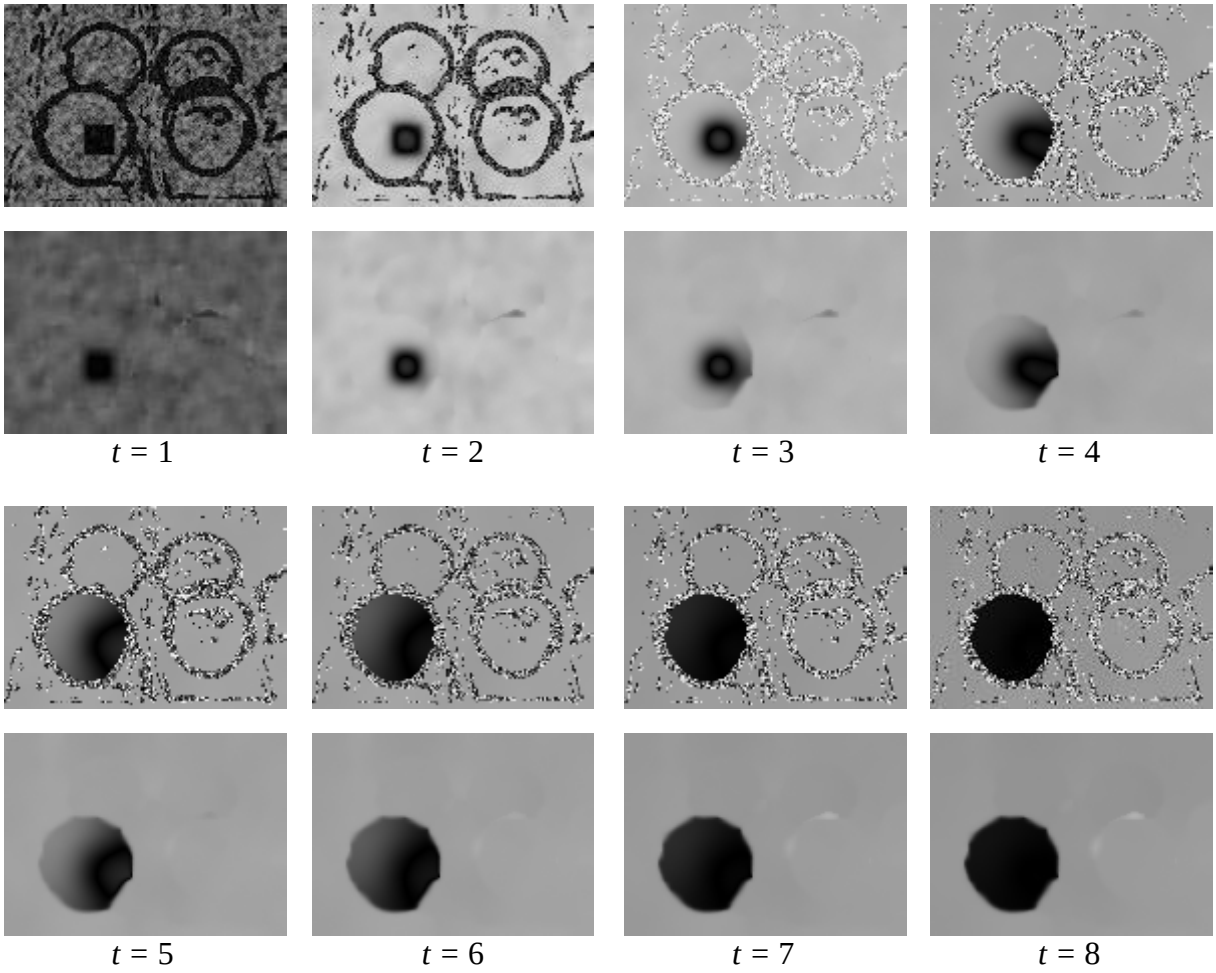

Figure 8. Selection and segmentation of the yellow ball.

Another approach was implemented in the papers (Ursino et al., 2003; Ursino et al., 2004a). The model developed in these papers is a two layer neural network. The first layer is build of traditional by-rate-coding neurons that fulfill contour extraction by image filtering. This layer operates similarly to the first stage of image processing in our model though more simple algorithms of contour extraction are used by Ursino and co-authors and the processing is restricted to grey-scale images.

The second layer of the model is nearly identical to LEGION. The difference is that the model works under constant values of connection strengths and the spread of synchronization is stopped at the boundaries that have been computed by the first layer. The processing implemented by the second layer is similar to the one that is fulfilled by the first layer of our model. The difference is that there is a second layer in our model that can correct the errors of contour extraction and inhibit noise. Moreover in our model there is a possibility to select a particular object in the focus of attention while in the model of Ursino and co-authors all objects that are present in the image are selected in some random sequence. Errors in binding are also possible in the model of Ursino and co-authors, while in our model such errors have not been observed, a single object is always selected in the focus of attention.

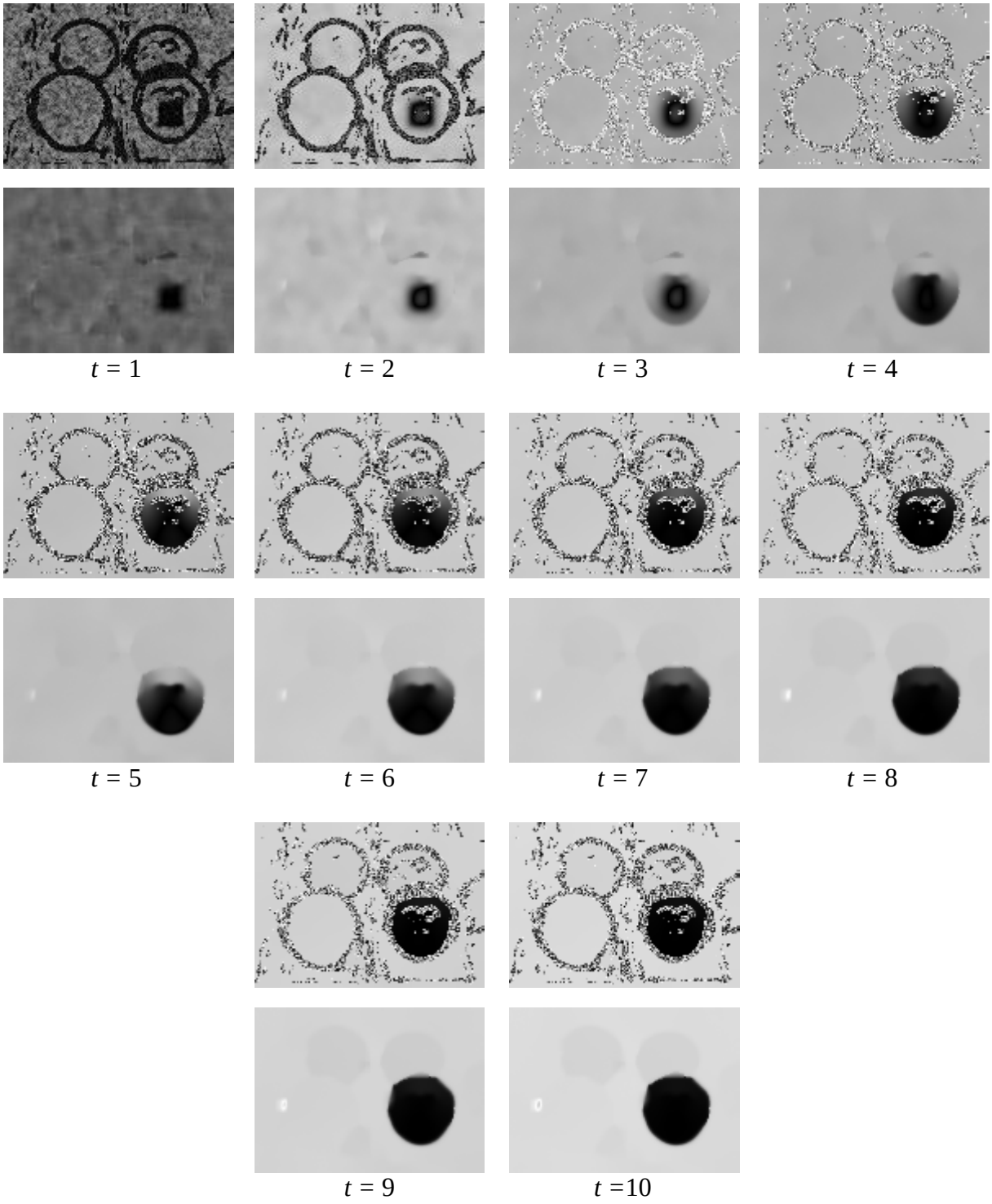

Figure 9. Selection and segmentation of the blue ball.

In our model, selection of an object is made by initially marking a small part (a square) of this object and including this part in the focus of attention. This type of marking can be attributed to spatial attention. Another possibility would be to initially concentrate attention on a set of features identifying a particular object. This would represent object-based attention. For example, colour can be used for this purpose. Average colour of a small segment of the object (e.g. a square) should be determined by using an image that contains this object. If a new image is presented, the pixels of this image that have this colour should

be initially taken in the focus of attention. More elaborate procedure would be to keep in memory not the average colour of a segment but the distribution of colours in the segment. Then in a new image the focus of attention should be initially concentrated on a segment with a similar distribution of colours. After attention is directed to a segment of the image all other image processing will go in the same way as in the case of spatial attention.

7. Conclusion

The problem of object selection and segmentation is important both for neurobiological research and robotics. From a biological point of view this problem is closely related to the problems of binding and attention, therefore computer models of these phenomena can help in understanding psychological experiments on visual search [Treisman and Sato, 1990]. On the other hand, reliable brain based devices for object selection are needed in robotics as a preliminary step to pattern recognition, visual scene understanding, and object manipulation.

Two kinds of models have been constructed to solve the problem of selection and segmentation. Some models are based on traditional approaches of computer vision or neural networks, others try to use the principles of synchronization provided by oscillatory neural networks. We suggest a compromise solution that combines traditional and oscillatory mechanisms of visual information processing.

Our paper as well as many other papers show that conventional approaches to contour extraction are efficient enough, and there is no need to attract oscillatory neural networks at this stage of processing. We restricted the computational procedures used at this stage to rather simple algorithms that take into account only local characteristics of the image. Better results could be obtained if some logical operations were added, but we tried to avoid any complications that would go beyond the known operations in the primary visual cortex. We have shown that synchronization principle is efficient in improving the results of contour extraction and binding the features of a searched object according to the temporal synchronization hypothesis. More simulations are needed to confirm our results on other types of images. Also automatic selection of the initial area for attention focusing should be added to the model. Note that not only stationary scenes but scenes with moving objects are suitable for processing by our model. In particular, object tracking can be easily added to the functionality of the model. The principles of our neural network functioning are universal enough to expect that it can be a helpful instrument in the solution of the problems of object selection and segmentation.

The main application area in which such an object selection algorithm might bring significant benefits if that of cognitive robotics and human-robot interaction. For example, ongoing research with humanoid robotic platform iCub (Metta et al., 2008) is centred on the integration of various cognitive capabilities. Some studies are focussing on the integration of vision, action and linguistic capabilities (Cangelosi et al., 2008), whilst others specifically address the selective attention of objects in response to the activation of sensorimotor properties of the objects (Tucker and Ellis, 2001). The proposed object selection algorithm provides a neurally-plausible modelling framework for the extension and integration of vision processing dynamics with other cognitive mechanisms. Future extension will specifically aim at the top-down and bottom-up contribution of sensorimotor representations in object selection.

Acknowledgement

This work was supported by the grants of EUcognition (Grant 0001645 for YK), UK EPSRC (Grant EP/D036364/1 for RB), the Russian Foundation of Basic Research (Grant 07-01-00218 for YK and RB), and EU FP7 Project ITALK (ICT-214668 for VT and AC).

References

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5-28.

Borisyuk, R.M. and Kazanovich, Y.B. (2003). Oscillatory neural network model of attention focus formation and control. *BioSystems*, 71, 29-38.

Borisyuk, R. and Kazanovich, Y. (2004). Oscillatory model of attention-guided object selection and novelty detection. *Neural Networks*, 17, 899-915.

Broussard, R.P., Rogers, S.K., Oxley, M.E., and Tarr, G.L. (1999). Physiologically motivated image fusion for object detection using a pulse coupled neural network. *IEEE Trans. Neural Networks*, 10, 554-563.

Buhmann, J.M., Lange, T., and Ramacher, U. (2005). Image segmentation by networks of spiking neurons. *Neural Computation*, 17, 1010-1031.

Cangelosi, A., Belpaeme, T., Sandini, G., Metta, G., Fadiga, L., Sagerer, G., Rohlfing, K., Wrede, B., Nolfi, S., Parisi, D., Nehaniv, C., Dautenhahn, K., Saunders, J., Fischer, K., Tani, J., and Roy, D. (2008). The ITALK project: Integration and transfer of action and language knowledge. In: *Proceedings of Third ACM/IEEE International Conference on Human Robot Interaction (HRI 2008)*. Amsterdam, 12-15 March 2008.

Chen, K. and Wang, D.L. (2002). A dynamically coupled neural oscillator networks for image segmentation. *Neural Network*, 15, 423-439.

Chen, K., Wang, D.L., and Liu, X. (2000). Weight adaptation and oscillatory correlation for image segmentation. *IEEE Trans. Neural Networks*, 11, 1106-1123.

Cheng, H.D., Jiang, X.H., Sun, Y., and Wang, J. (2001). Color image segmentation: advances and prospects. *Pattern Recognition*, 34, 2259-2281.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163-191.

Daido, H. (1988). Lower critical dimension for population of oscillators with randomly distributed frequencies: a renormalization-group analysis. *Phys. Rev. Letters*, 61, 231-234.

Doesburg, S.M., Roggeveen, A.B., Kitajo, K., and Ward, L.M. (2008). Large-scale gamma-band phase synchronization and selective attention. *Cerebral Cortex*, 18, 386-396.

Fries, P., Reynolds, J., Rorie, A., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291, 1560-1563.

Fries, P., Schroeder, J-H., Roelfsema, P.R., Singer, W., and Engel, A.K. (2002). Oscillatory neural synchronization in primary visual cortex as a correlate of stimulus selection *J. Neurosci.*, 22, 3739-3754.

Huang, W., Jiao, L., and Jia, J. (2008). Modeling contextual modulation in the primary visual cortex. *Neural Networks*, 21, 1182-1196.

Kanwisher, N. and Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience*, 1, 91-100.

Kazanovich, Y.B. and Borisyuk, R.M. (1999). Dynamics of neural networks with a central element. *Neural Networks*, 12, 441-454.

Kuramoto, Y. (1984). Chemical oscillations, waves, and turbulence. Springer-Verlag, Berlin.

Labbi, A., Milanese, R., and Bosch, H. (2001). Visual object segmentation using FitzHugh-Nagumo oscillators. *Nonlinear Analysis: Theory, Methods & Applications*, 47, 5827-5838.

Li, J. and Gray, R.M. (2000). Image Segmentation and Compression Using Hidden Markov Models (The International Series in Engineering and Computer Science). - Klawer Academic Publishers.

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *Int. J. Comp. Vision*, 30, 117-154.

Malsburg, C. von der (1981). The correlation theory of brain function. *Internal report 81-2, Max-Plank Institute for Biophysical Chemistry (*reprinted in *Models of Neural Networks, E. Domany, J.L. van Hemmen, K. Schulten (Eds.).* pp. 95-119, Springer, New York, 1994).

Metta, G., Sandini, G., Vernon, D., Natale, L., and Nori, F. (2008). The iCub humanoid robot: an open platform for research in embodied cognition. In *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop (PerMIS'08), R. Madhavan and E.R. Messina (Eds.)*. Washington, DC, USA.

Motter, B.C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, 70, 909-919.

Pal, N.R. and Pal, S.K. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26, 1277-1294.

Palm, G. and Knoblauch, A. (2005). Scene segmentation through synchronization. In *Neurobiology of Attention, L. Itti, G. Rees, and J.K. Tsotsos. (Eds.),* pp. 618-623. Elsevier, San Diego, CA.

Petkov, N. and Subramanian, E. (2007). Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal Gabor filters with surround inhibition. *Biol. Cybern.*, 97, 423-439.

Ritz, R. and Sejnowski, T.J. (1997). Synchronous oscillatory activity in sensory systems: new vistas on mechanisms. *Curr. Opin. Neurobiol*., 7, 538-548.

Roelfsema, P.R., Lamme, V., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395, 376-381.

Sakaguchi, H., Shinomoto, S., and Kuramoto, Y. (1987). Local and global self-entrainment in oscillator lattices. *Progr. Theor. Phys.*, 77, 1005-1010.

Shapiro, L.G. and Stockman, G.C. (2001). Computer Vision. - New Jersey, Prentice-Hall.

Singer, W. and Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.*, 18, 555-586.

Steinmetz, P.N., Roy, A., Fitzgerald, P., Hsiao, S.S., Johnson, K.O., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404, 187-190.

Strogatzt, S. H. and Mirollo, R. E. (1988). Collective synchronisation in lattices of non-linear oscillators with randomness. *J. Phys. A: Math. Gen.*, 21, L699-L705.

Sumengen, B. and Manjunath, B.S. (2005). Multi-scale edge detection and image segmentation. *EUSIPCO-2005*, #207.

Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., and Nori, F. (2008a). An open-source simulator for cognitive robotics research: The prototype of the iCub humanoid robot simulator. In *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop (PerMIS'08), R. Madhavan and E.R. Messina (Eds.)*. Washington, DC, USA.

Tikhanoff, V., Cangelosi, A., Tani, J., and Metta G. (2008b). Towards language acquisition in autonomous robots. *Proceedings of ALIFE XI: International Conference on Artificial Life*, Winchester, UK.

Treisman, A. and Sato, S. (1990) Conjunction search revisited. *J. Exper. Psychol.: Human Perception and Performance*, 16, 459-478.

Tucker, M. and Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, 8, 769-800.

Ursino, M. and La Cara, G.E. (2004a). Modeling segmentation of a visual scene via neural oscillators: fragmentation, discovery of details and attention. *Network*, 15, 69-89.

Ursino, M. and La Cara, G.E. (2004b). A model of contextual interactions and contour detection in primary visual cortex. *Neural Networks*, 17, 719-735.

Ursino, M., La Cara, G.E., and Sarti, A. (2003). Binding and segmentation of multiple objects through neural oscillators inhibited by contour information. *Biol. Cybern.*, 89, 56-70.

Wang, D.L. (1999). Object selection based on oscillatory correlation. *Neural Networks*, 12, 579-592.

Wang, D.L. (2005). The time dimension for scene analysis. *IEEE Trans. Neural Networks*, 16, 1401-1426.

Wang, D.L. and Terman, D. (1997). Image segmentation based on oscillatory correlation. *Neural Computation*, 9, 805-836.

Zhang, A. (1996). A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29, 1235-1346.

Zhao, L., de Carvalho, A., and Li, Z. (2004). Pixel clustering by adaptive pixel moving and chaotic synchronization. *IEEE Trans. on Neural Networks*, 15, 1176- 1185.

Zhao, L., Furukawa, R.A., and de Carvalho, A. (2003). A network of coupled chaotic maps for adaptive multy-scale image segmentation. *Int. J. Neural Systems*, 13, 129-137.

Zhao, L. and Macau, E. (2001). A network of dynamically coupled chaotic maps for scene segmentation. *IEEE Trans. on Neural Networks*, 12, 1375-1385.

Zhu, S.C. and Yuille, Y. (1996). Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18, 884-900.