Neural Networks 22 (2009) 707–719

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2009 Special Issue

A neural model of selective attention and object segmentation in the visual scene: An approach based on partial synchronization and star-like architecture of connections☆

Roman Borisyuk[a,b,*], Yakov Kazanovich[b,c], David Chik[a], Vadim Tikhanoff[a], Angelo Cangelosi[a]

[a] School of Computing and Mathematics, University of Plymouth, Plymouth PL4 8AA, UK

[b] Institute of Mathematical Problems in Biology, Russian Academy of Sciences, Pushchino, Moscow Region, 142290, Russia

[c] State University of Pushchino, Pushchino, Moscow Region, 142290, Russia

ARTICLE INFO

Article history:
Received 5 May 2009
Received in revised form 30 May 2009
Accepted 25 June 2009

Keywords:
Selective visual attention
Object segmentation
Oscillatory neural network
Synchronization

ABSTRACT

A brain-inspired computational system is presented that allows sequential selection and processing of objects from a visual scene. The system is comprised of three modules. **The selective attention** module is designed as a network of spiking neurons of the Hodgkin–Huxley type with star-like connections between the central unit and peripheral elements. The attention focus is represented by those peripheral neurons that generate spikes synchronously with the central neuron while the activity of other peripheral neurons is suppressed. Such dynamics corresponds to the partial synchronization mode. It is shown that peripheral neurons with higher firing rates are preferentially drawn into partial synchronization. We show that local excitatory connections facilitate synchronization, while local inhibitory connections help distinguishing between two groups of peripheral neurons with similar intrinsic frequencies. The module automatically scans a visual scene and sequentially selects regions of interest for detailed processing and object segmentation. The **contour extraction module** implements standard image processing algorithms for contour extraction. The module computes raw contours of objects accompanied by noise and some spurious inclusions. At the next stage, the **object segmentation module** designed as a network of phase oscillators is used for precise determination of object boundaries and noise suppression. This module has a star-like architecture of connections. The segmented object is represented by a group of peripheral oscillators working in the regime of partial synchronization with the central oscillator. The functioning of each module is illustrated by an example of processing of the visual scene taken from a visual stream of a robot camera.

1. Introduction

Extraction of a certain object from the image is a traditional problem in computer vision and robotics. It also attracts the attention of psychologists and neurobiologists who are interested in understanding 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 on 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, based on optical characteristics of these parts. At the recognition stage, a complete object is composed from

☆ This paper is an extended version of the IJCNN 2009 conference proceedings paper [Borisyuk, R., Chik, D., & Kazanovich, Y. (2009). Partial synchronization of neural activity and information processing. In *Proceedings of IJCNN 2009*. #668].

* Corresponding address: School of Computing and Mathematics, University of Plymouth, Plymouth PL4 8AA, UK. Tel.: +44 1752 584949; fax: +44 1752 233349.

E-mail addresses: r.borisyuk@plymouth.ac.uk (R. Borisyuk), kazanovichyakov@gmail.com (Y. Kazanovich), david.chik@plymouth.ac.uk (D. Chik), Vadim.Tikhanoff@plymouth.ac.uk (V. Tikhanoff), A.cangelosi@plymouth.ac.uk (A. Cangelosi).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.06.047

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 recent years a number of investigations have been made to clarify how the problem of object selection and segmentation is solved by the brain. There are significant reasons to consider synchrony as a basic dynamical regime when solving these tasks. Experimental recordings show that the coherence of spiking plays a major role in the control of attention (Fell, Fernandez, Klaver, Elger, & Fries, 2003; Fries, Schroeder, Roelfsema, Singer, & Engel, 2002; Singer, 1999; Steinmetz et al., 2000). These experimental facts are considered as supporting the Temporal Correlation Hypothesis (TCH) (Gray, 1999; Malsburg van der, 1981; Malsburg van der, 2001), which states that increased synchrony can reinforce the impact of spikes on subsequent cortical areas. The TCH is well suited to modelling in terms of oscillatory neural networks (see the reviews Ritz and Sejnowski (1997) and Wang (2005)). A general idea that is present in most models of feature 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 recent years (Borisyuk & Kazanovich, 2004; Broussard, Rogers, Oxley, & Tarr, 1999; Buhmann, Lange, & Ramacher, 2005; Chen & Wang, 2002; Chen, Wang, & Liu, 2000; Labbi, Milanese, & Bosch, 2001; Palm & Knoblauch, 2005; Ursino & La Cara, 2004a; Ursino, La Cara, & Sarti, 2003; Wang, 1999; Wang & Terman, 1997; Zhao & Macau, 2001; Zhao, de Carvalho, & Li, 2004; Zhao, Furukawa, & de Carvalho, 2003). 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 experimental results, while others 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 the 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 results that are comparable to or even exceed those obtained by traditional image processing methods (Chen & 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 parameter values.

Our goal is to develop a biologically inspired model of visual information processing and to use this model to control movement of the robot arm. Right now the processing is based purely on optical characteristics of the image and does not attract pattern recognition and memory retrieval. Available neurobiological data on functioning of the visual system are used to constrain our modelling. However, the experimental data are not complete and there are many open questions on neural mechanisms of functioning of the visual system. Therefore, we use an approach that is a combination of biologically inspired neural network models with some artificial algorithms which nevertheless allow biological implementation.

The model developed in this paper allows selection of objects from a visual scene, their processing and segmentation. It consists of three modules. The selective attention module scans a visual scene and selects a region of interest (ROI) which can include one or several objects. Two other modules are used for detailed analyses of the selected ROI and object segmentation. The selected ROI is used as an input to the contour extraction module which finds contours of all objects in the ROI. The object segmentation module uses the contour information to segment some object.

In the first and third modules we use a star-like architecture of connections which includes a central unit which has feedforward and feedback connections with multiple peripheral elements. The functioning of the network is based on the regime of partial synchronization between the central unit and some group of peripheral elements (Borisyuk & Kazanovich, 2003, 2004, 2006). For example, in the module of selective attention it is assumed that the focus of attention is represented by those peripheral elements which work coherently with the central unit. In this module the central unit also suppresses the activity of all unattended peripheral elements.

To confirm the concept, the functioning of each module is illustrated using an example of processing of a visual scene taken from the visual stream of a robot camera (Tikhanoff et al., 2008a; Tikhanoff, Cangelosi, Tani, & Metta, 2008b). Some components of the system such as automatic selection of image features and their phase-frequency coding are under development. The performance of the complete system as well as the comparison with other existing models will be reported in a separate publication.

(A) A *selective attention module* (see Section 2) is a biologically realistic model of spiking elements of the Hodgkin–Huxley type that allows consequent selection of images from a visual scene. The module consists of a layer of peripheral neurons (PNs) and a central unit which includes two central neurons, CN1 and CN2, with global feedforward and feedback connections. The attention focus is formed by those PNs that generate spikes synchronously with CN1. The activity of other peripheral neurons is suppressed. Attention switching is based on the synaptic plasticity of connections from CN2 to the PNs. The strength of inhibition to a group of partially synchronized PNs increases and the activity of the group decays. Thus, another group becomes active and partially synchronized with the central neuron CN1. It is shown that PNs with higher firing rates have a priority in being synchronized with the central neuron. This property establishes an order for object selection—an object that is coded by the fastest PNs will be selected first. We have found that local excitatory connections between PNs facilitate synchronization, while local inhibitory connections between PNs help distinguishing between two groups of peripheral neurons with similar intrinsic frequencies. The module automatically scans the visual scene and sequentially selects regions of interest for processing and object segmentation.

(B) A *contour extraction module* (see Section 3) has been developed to process selected regions. It implements some algorithms for raw contour extraction based on a traditional approach used in computer vision and artificial neural networks. This approach combines Gabor filters with derivatives along the gradient direction for different scales (Broussard et al., 1999; Huang, Jiao, & Jia, 2008; Lindeberg, 1998; Petkov & Subramanian, 2007; Sumengen & Manjunath, 2005). The originality of our approach is in a special combination of these methods and orientation on predefined colours of searched objects. The only restriction on the algorithms used in this module is that they should have evident neural implementation and be able to process the image in parallel. The operations used by the algorithms of module B such as filtering and determination of the contrast of some optical characteristics are known from experimental studies of the brain. The result of

object processing is a raw contour with noise and, possibly, with some number of spurious inclusions. Similar algorithms have been realized in neural networks (Broussard et al., 1999; Huang et al., 2008; Petkov & Subramanian, 2007; Ursino & La Cara, 2004b), but in most cases additional contextual modulation has been added to improve the contours and to fill the gaps in them.

(C) The *object segmentation module* (see Section 4) comprises a star-like architecture of coupled phase-oscillators to find the precise boundary of the selected object, to suppress noise and spurious inclusions, and to segment the object from the image. The network has two layers whose activity is controlled by a special central oscillator (CO). The oscillators in the layers are called peripheral oscillators (POs). The first layer fulfils the synchronization between a group of POs and the CO according to the TCH using the contours obtained by module B as restrictors for synchronization spreading outside the border of the segmented object. The second layer transforms the raw image into the final result of segmentation. The segmented object is represented by a group of POs working in the regime of partial synchronization with the CO.

Section 5 is devoted to a discussion of the results and comparison with other models. Also we discuss further development of the system and its implementation in the robot visual system.

2. Module A: Selective attention model

2.1. Model description

The two-layer architecture of the model connections is shown in Fig. 1. The choice of the architecture is inspired by the idea of the central executive (Baddeley, 1996; Cowan, 1988; Shallice, 2002) responsible for the attention control of the working memory. In reality, the central executive is assumed to be a complex network in the forebrain area with not necessarily direct connections to the primary zones of the cortex. In the model, the central executive and its interaction with feature detectors are represented in a radically simplified form. In particular, direct feedforward connections to the central neurons are used to simulate convergent projections from lower layers of information processing to higher brain areas responsible for implementation of associative memory and attention (Damasio, 1989). The architecture with a central element has been applied to model various cognitive phenomena (Corchs & Deco, 2001; Wang & Terman, 1995). Here we use a visual attention model developed in Chik, Borisyuk, and Kazanovich (2009). In this model, PNs represent feature detectors in areas V1, V2, and V4 of the neocortex that are activated by an input image (the external stimulus). Each PN receives a signal (an external current) which is determined by optical characteristics of the pixel whose location in the image grid is identical to the location of the PN. The upper layer consists of two central neurons, CN1 and CN2. CN1 enables partial synchronization to be formed with a selected subset of PNs (it will be shown later). CN2 controls the shift of partial synchronization from one subset of PNs to another. This allows CN1 to consecutively synchronize its firing with different assemblies of PNs. Both CN1 and CN2 send inhibitory signals to all PNs, but excitatory input signals from the PNs are received by CN1 only. Neuron CN2 is an independent oscillator. The oscillations of CN2 are endogenous; they can be caused by intrinsic neuronal properties, network properties, background neural activities, etc. We assume that the connection strengths between CN1 and all PNs are constant and universal. There are no connections between CN1 and CN2.

The strength of inhibition from CN1 is fixed, but the inhibitory signals from CN2 vary over time. We suppose that there is short-term plasticity in the strength of the connection from CN2 to each PN. If both neurons have high activity for a prolonged time, the inhibitory synapse will be strengthened, while if either one of them has low activity the synaptic strength will decay (Fitzpatrick, Akopian, & Walsh, 2001; Zucker & Regehr, 2002). The PNs that are currently selected and synchronized by CN1 may be suppressed by CN2. This provides CN1 the opportunity to change the focus of attention by synchronizing its activity with another assembly of PNs. The inhibitory influence from CN2 to the PNs implements inhibition of return—attention is biased against returning to previously attended stimuli (Klein, 1988; Takeda & Yagi, 2000).

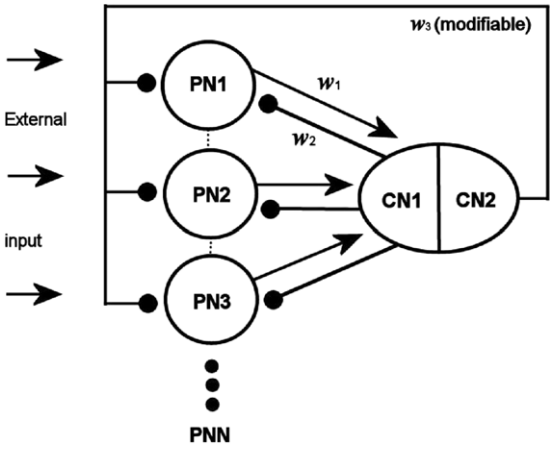

Fig. 1. Connection architecture of the central–peripheral neural block. PN1, PN2, PN3, …, are peripheral neurons encoding the features of the input image. CN1 and CN2 are central neurons which control the attention focus. Excitatory connections are shown by arrows, and inhibitory connections are shown by lines with black circles at their ends. The PNs form a rectangular grid which has the same size as the external image. Optionally, there can be local connections between neighbouring PNs.

The Hodgkin–Huxley model (Hodgkin & Huxley, 1952) is employed for each neuron. It is described by the following equations:

$$\frac{dV}{dt} = -I_{ion} + I_{ext} - I_{syn}, \tag{1}$$

$$\frac{dX}{dt} = A_X(V)(1 - X) - B_X(V)X, \quad X \in \{m, h, n\}, \tag{2}$$

where V is the membrane potential of the neuron; I_{ion} denotes the total ionic current, which is the sum of potassium, sodium and leakage currents; m, h, n are the gating variables of the ionic channels.

The description of variables and specifications of all parameter values can be found in Chik et al. (2009).

External stimulation of a PN is represented by the current I_{ext} that combines a constant component of the current and the Gaussian noise with the mean equal to zero and a small standard deviation. This current is supra-threshold, inducing an intrinsic firing rate for each PN. We assume that different features of the input image (e.g. different colours) are coded by different values of I_{ext} (hence PNs utilize frequency coding).

It is not necessary for CN1 to receive an external current to generate spikes. If the excitatory inputs from the PNs to CN1 are strong enough, they will force CN1 to fire. If CN1 already has an intrinsic firing rate, then the excitatory input will increase its firing rate even further. In contrast, CN2 requires an external current, otherwise it would be silent. In our simulations, the standard external input to CN1 is 5 mA and that to CN2 is 30 mA.

The interaction between neurons is expressed by the synaptic current I_{syn}. In our model, CN1 receives excitatory synaptic inputs from all PNs, while CN2 does not receive any synaptic input from other units of the network (it is assumed that CN2 receives an "external" current from the areas within the brain that are

not directly involved in the functioning of the attention system). The PNs receive inhibitory inputs from both CN1 and CN2. The conductance-based equations describing both excitatory and inhibitory synaptic currents can be found in Chik et al. (2009). The conductance is modulated by an alpha-function, which depends on the time interval from the previous spike of the pre-synaptic neuron.

The modifiable inhibitory connection strength from CN2 to the ith PN representing the short-term plasticity is determined by the following expression:

$$w_i(t) = \begin{cases} \tilde{w}, & T_{H,i} \le t \le T_{H,i} + \Delta h \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $\tilde{w}$ is the saturation value of plasticity, $T_{H,i}$ marks the onset time of plasticity, and Δh represents the duration of plasticity. Let $T_{R,i}$ be the time of the previous reset of the plastic connection from CN2 to the ith PN, i.e. the last moment when $w_i(t)$ jumped to zero. The onset time of plasticity $T_{H,i} > T_{R,i}$ is determined as the first time moment that satisfies the following equality:

$$\int_{T_{R,i}}^{T_{H,i}} \Theta(V_i(t) - \upsilon) \cdot \Theta(V_{CN2}(t) - \upsilon) \mathrm{d}t = \frac{1}{\varepsilon}, \tag{4}$$

where V_i is the membrane potential of the ith PN, V_{CN2} is the membrane potential of CN2, υ is the threshold value for detecting the generation of a spike, ε is a parameter between 0 and 1, and $\Theta(x)$ is a binary function of either 0 or 1. By multiplying two binary functions, the integral in formula (4) measures how often the spikes of the ith PN and CN2 coincide in time. This can be considered as an approximation of Hebbian-type learning. As soon as this value reaches the threshold $1/\varepsilon$, the connection strength jumps up to the saturation level $w_i(t) = \tilde{w}$. The period of saturation lasts for Δh ms. After that time the value of the connection strength $w_i(t)$ decays back to zero and the process of connection modification begins a new cycle. The decay of the connection strength can be motivated by a limited store of releasable neurotransmitters such that the synapse will be disabled after prolonged activation (Brager, Capogna, & Thompson, 2002; Sudhof, 2000).

During the simulation, all equations are numerically integrated using the fourth-order Runge–Kutta method with time step 0.01 ms. We found that the period of spikes may range from 10 ms (100 Hz) to 100 ms (10 Hz). We defined a window of 2 ms as a criterion for synchronization.

2.2. Results of model simulations

The intensive study of dynamical regimes and synchronization was undertaken in Chik et al. (2009). The PNs were arranged into two groups (A and B), with one group having a stronger external input and consequently a higher frequency of spike generation. The external currents were uniformly distributed in the range of ±10% around the mean values, and the mean values of external currents for groups A and B are I_1 and I_2 respectively ($I_1 > I_2$). The neuron CN1 received an external current of 5 mA, which is lower than both I_1 and I_2.

In the regime of partial synchronization the neurons of one group generate spikes coherently with CN1 while the neurons of the other group do not generate spikes. In the presence of noise, we define a criterion that if more than 70% of the spikes of a PN coincide with the spikes of CN1 then we say that these neurons are synchronized. Fig. 2 shows the spiking activities when the neurons of group A (PN1–PN5) are in partial synchronization with CN1 (the influence of CN2 on PNs is switched off, $w_i(t) = \tilde{w}$). For neurons of group B (PN6–PN10), the fluctuations of the membrane potential are synchronized with the spikes of CN1, but the amplitude of these fluctuations is too small to generate a spike (except for some occasional spikes). The regime of partial synchronization appears if the inhibitory influence of CN1 on the PNs is strong enough (but not too strong to shut down the firing of the PNs). We interpret this situation as the concentration of attention on a stimulus represented by the neurons of group A.

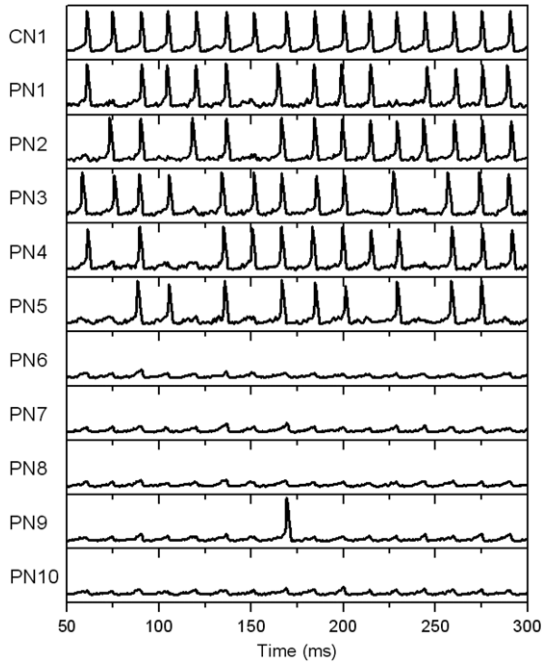

Fig. 2. Spiking activity during partial synchronization. The panels show the potential traces of different neurons. CN1 is the central neuron; PN1–PN5 are five (out of 100) neurons selected randomly from group A; PN6–PN10 are five neurons from group B. The input current to CN1 is 5 mA, $I_1 = 16$ mA, $I_2 = 6$ mA. The connection strengths are $w_1 = 0.1$, $w_2 = 0.5$, $w_3 = 0.1$. The standard deviation of noise in the external currents is 0.3.

The spiking activity in Fig. 2 could be made less regular and more similar to the real spiking activity of neurons if we added more noise to the system. However, we wanted to illustrate the meaning of partial synchronization, which is clearer for regular spike trains.

An important result of these simulations is that selective attention (associated with partial synchronization) always favours a group with higher frequency. If group A has a higher intrinsic frequency ($I_1 > I_2$) then group A will be engaged in the regime of partial synchronization (and vice versa). This is also true if the number of groups is higher than two. In neurophysiology, higher salience of a stimulus is usually revealed in a higher level of neural activity representing this stimulus in the cortex (Morris, Friston, & Dolan, 1997). Therefore the model complies with the fact that the most salient stimulus should have the priority in being selected in the focus of attention.

Let us compare how the model performs sequential selection of objects in the focus of attention in the following three cases: (a) the model without local connections, (b) the model with excitatory local connections between PNs, (c) the model with inhibitory local connections between PNs. We use a still camera image of size 320 × 240 pixels that contains three objects: an orange, a pear, and a blue cloth. The objects are placed on an almost white background. A grid of 320 × 240 PNs was used in the simulation, providing a one-to-one correspondence between the PNs and the pixels of the image. We suppose that different colours are encoded as different intrinsic frequencies of the PNs. (Notice that this colour-based implementation can be easily modified to be saliency based). We converted the RGB values of the pixels into the values of the external current I_{ext} in the range 10–40 mA.

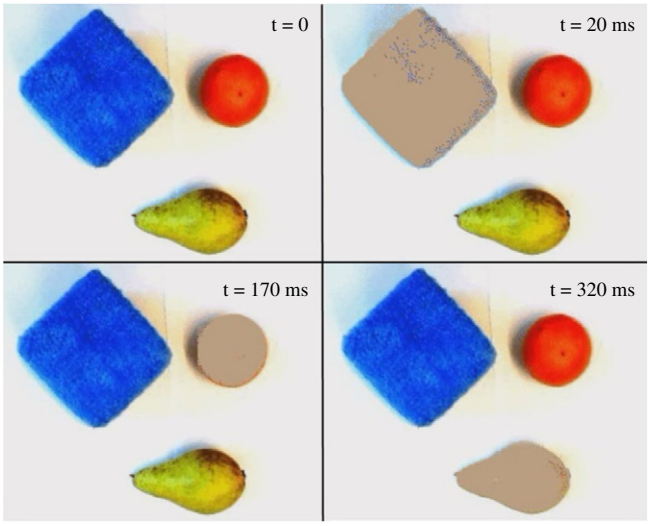

Fig. 3. Sequential selection of objects in a real image (the original AVI movie can be found at http://www.pion.ac.uk/binmam/index.htm). The PNs have one-to-one correspondence with image pixels. The fawn area indicates the location of active PNs. The colour values of the image pixels are converted to the external current values of the PNs. The input to CN1 is equal to 5 mA. The input to CN2 is equal to 30 mA. The parameters for connection weight modification are $w = 5$, $\upsilon = -10$ mV, $\Delta h = 650$ ms, and $\varepsilon = 0.16$/ms.

Without loss of generality, we suppose the blue colour yields a highest spiking frequency, followed by orange, then green. The white colour of the background is assumed to yield too small input currents for the PNs, therefore the PNs corresponding to the background are silent.

The results of simulations in case (a) are presented in Fig. 3. The firing frequencies are in the gamma range (around 40 Hz). An overlay is used on top of the image to mark the currently attended area: if a PN fires then the corresponding pixel is coloured in fawn. During the time interval 0–20 ms there is a transitory period when no selection of objects occurs. At $t = 20$ ms, the cloth is selected. At this moment the system is in the regime of partial synchronization between CN1 and the PNs located in the cloth area. Other PNs are all suppressed. The selection of the cloth continues until the moment $t = 170$ ms when the orange is selected. Then at $t = 320$ ms the pear is selected. After that the focus of attention returns back to the cloth, and so on.

The formation of the focus of attention is controlled by CN1. At any moment one assembly of PNs has higher frequency than the other two assemblies, and this assembly of neurons is selected first. The shift of attention to another object is controlled by CN2 which increases its inhibitory influence on the firing PNs and keeps this inhibition for some time. During this period inhibited PNs cannot generate spikes but they return back to the normal spiking regime after the inhibitory period has expired.

Let now a PN be synaptically connected with its surrounding neighbours (8 neighbours for a PN in the middle of the grid, while the PNs at the corners or boundaries have fewer neighbours). The synaptic current from neighbouring neurons is added to the equations of the PN's membrane potential in the same way as the synaptic currents from the central neurons.

Fig. 4 shows the results of simulations in case (b) (local connections are excitatory). In this case synchronization is facilitated. For example, if there is large noise in the external current such that only some part of the blue cloth is selected (Fig. 4(a)), providing local excitatory connections can spread the region of partial synchronization to the whole object (Fig. 4(b)). However, the PNs corresponding to the background of the image should be silent in order to avoid further propagation of neural activities.

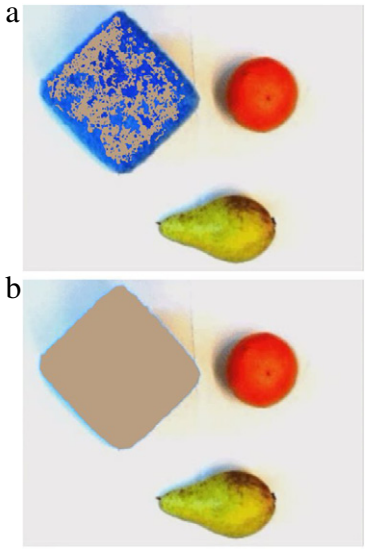

Fig. 4. The effect of excitatory local connections. (a) Without local connections only a part of neurons representing the blue cloth forms partial synchronization with CN1. (b) With excitatory local connections the whole cloth is included in the focus of attention.

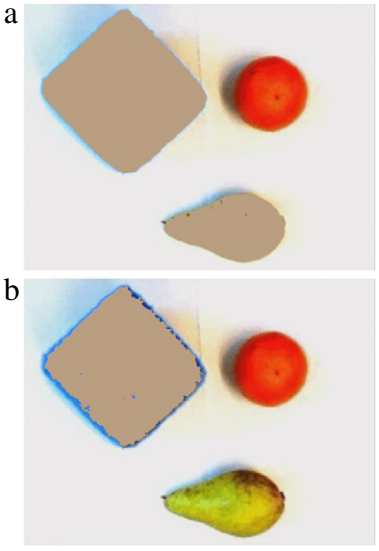

Fig. 5. The effect of inhibitory local connections. (a) Without local connections the neurons representing both the cloth and the pear form partial synchronization with CN1. (b) With inhibitory local connections, only the cloth is included in the focus of attention.

Fig. 5 shows the results of simulations in case (c) (local connections are inhibitory). In this case the ability of distinguishing between objects can be improved. For example, if the mean values of external currents for the blue cloth and the pear are similar (e.g. 17 mA and 12 mA, respectively), both objects will be selected (Fig. 5(a)): the PNs in both groups will be synchronized with CN1. By providing local inhibitory connections, the PNs corresponding to the pear will be suppressed (Fig. 5(b)). The system now focuses on only one object (the cloth).

The functioning of the system is also illustrated on an example of the image taken from a visual stream of a robot camera (Tikhanoff et al., 2008a, 2008b). The image is shown in Fig. 6. We will use the same image and its fragment to illustrate the functioning of each module. There are four balls of different colour in the image that are targets for the robot's manipulations. Our goal is to develop a system for the robot control that will allow

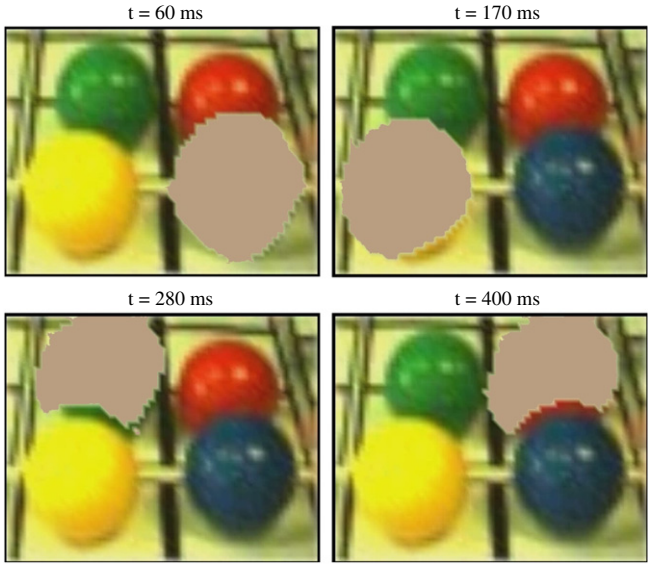

Fig. 8. Sequential selection of balls of different colours. We show here four frames taken at different time moments from the video clip (the video file is available from http://www.pion.ac.uk/binmam/ball.avi). The currently selected ball (active peripheral elements synchronously spiking with a central neuron) is shown by the fawn colour.

Fig. 6. Original image taken from the visual stream of the robot camera. This image is used to illustrate the functioning of the system. The final aim of selective attention and segmentation procedures is to select from the image a ball of a predefined colour and to detect its boundaries.

Fig. 7. A fragment of the original image shown in Fig. 6.

sequential selection of target objects. The robot should be able to pick up the selected ball by hand. The visual system of the robot should provide information about the position and the boundaries of the ball to the control system for accurate movement of the robot's hand. Although the system presented here was developed to solve this practical task, the methods used are universal for any coloured images where the searched object differs from other objects by its colour.

To illustrate the performance of the selective attention module, we cut a fragment from the image of Fig. 6 which contains four balls of different colours. The fragment is shown in Fig. 7. A grid of 61×46 PNs was used in the simulation, providing one-to-one correspondence between the PNs and the pixels of the fragment. We use the same approach of coding a colour by the external input to the neuron corresponding to the pixel which has been used above for the description of Fig. 3. The RGB value at each pixel is transformed to the value of external current of the corresponding PN. All current values are in the interval [10, 50] mA. The coding of colours was selected in such a way that the highest value of the current relates to blue colours, and decreasing values of the current correspond to yellow, green and red colours, respectively. In this simulation we use 24 local excitatory connections for each PN (8 nearest neighbours plus 16 neighbours of the second order). Such connections allow the system to fill small holes within an object. The synchronous activity initiated inside of an object can propagate in the direction of the object boundary. This activity propagation can be seen in the video clip (http://www.pion.ac.uk/ball.avi/). Fig. 8(a)–(d) shows the snapshots of the model simulation at times 60, 170, 280 and 400 ms, respectively. First the attention is focussed on the ROI which roughly corresponds to the blue ball (shown by the fawn colour), then it shifts to the ROIs corresponding to yellow, green, and red balls, respectively.

3. Module B: Contour extraction

3.1. Description of numerical procedures

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 of this efficiency but at least one of the reasons is that the 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 plane 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 the 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, Jiang, Sun, & Wang, 2001). In the next section we specify the parameter set that has been used in our simulations.

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

$$|\bar{g}| > C_1 > 0, \tag{5}$$

$$\nabla^2_{\bar{g}}(F(p - \delta\bar{g}))\ \nabla^2_{\bar{g}}(F(p + \delta\bar{g})) < 0, \tag{6}$$

$$\nabla^3_{\bar{g}}(F(p)) < C_2 < 0. \tag{7}$$

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, and C_1 and C_2 are constant parameters. Formulas (5)–(7) have the following meaning. Formula (5) states that the function F should be "steep"