A visual attention model based on hierarchical spiking neural networks

QingXiang Wu a,b,*, T.M. McGinnity b, Liam Maguire b, Rongtai Cai a, Meigui Chen a

a School of Physics and Optoelectronics Technology, Fujian Normal University, Fuzhou 350007, China
b Intelligent Systems Research Centre, School of Computing and Intelligent Systems, University of Ulster at Magee, Londonderry, BT48 7JL Northern Ireland, UK

ARTICLE INFO

Keywords:
Visual attention
Spiking neural network
Receptive field
Visual system

ABSTRACT

Based on the information processing functionalities of spiking neurons, hierarchical spiking neural networks are proposed to simulate visual attention. Using spiking neural networks inspired by the visual system, an image can be decomposed into multiple visual image components. Based on specific visual image components and image features, a visual attention system is proposed to extract attention areas according to top–down volition-controlled signals. The hierarchical spiking neural networks are constructed with a conductance-based integrate-and-fire neuron model and a set of specific receptive fields in different levels. The simulation algorithm and properties of the networks are detailed in this paper. Simulation results show that the attention system is able to perform visual attention of objects based on specific image components or features, and a demonstration shows how the attention system can detect a house in a visual image. Using the proposed saliency index, attention areas of interest can be extracted from spike rate maps of multiple visual pathways, such as ON/OFF colour pathways. According to this visual attention principle, the visual image processing system can quickly focus on specific areas while ignoring other areas.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The biological brain, with its huge number of neurons, displays powerful functionality in information processing, vision, reasoning, and other intelligent behaviours. Spiking neurons are regarded as essential components in the neural networks in the brain. Scientists working in neuroscience have made significant progress over the past decades with respect to the physiological and synaptic properties of individual neurons and some specific local neuronal circuits. For example, it is known that mammalian retinas contain approximately 55 distinct cell types, each with a different function [1]. A retinal cell type which responds to upward motion has been identified in [2]. Results in [3] demonstrate that information for segmenting scenes by relative motion is represented as early as the visual cortex V1. The results in [4] indicate that the rapid detection of categorical information in natural scenes is mediated by a category-specific biasing mechanism in an object-selective cortex that operates in parallel across the visual field, and biases information processing in favour of objects belonging to the target object category. Retinal ganglion cells convey the visual image from the eye to the brain [5]. Neurobiologists have found that various receptive fields exist in the visual cortex and play different roles [56]. Visual Attention enables the visual system to process potentially important objects by selectively increasing the activity of sensory neurons that represent the relevant locations and features of the environment [7]. Given the typical complexity of the visual environment, the ability to selectively attend to certain locations, while ignoring others, is crucial for reducing the amount of visual information to manageable levels in a computation system and for optimizing behavioural performance and response times. In the search for efficient visual processing models, visual attention has been studied by researchers along two main directions [13]. The first one is a bottom–up approach in which a saliency map is obtained by finding salient positions based on a low level representation of an image. The second one is a top–down approach in which an attention map is obtained by volition-controlled signals from a high level in the brain. Biological findings show that bottom–up controlled attention and top–down controlled attention are processed in different areas in the brain, while their integration is mostly processed in the prefrontal cortex [14,15]. Neural mechanisms of visual attention, in particular how top–down feedback highlights relevant locations, have been studied in [8]. The results in [8] show that the lateral intraparietal area (LIP) feedback can account for attention-enhanced visual cortex MT responses. The top–down flow of visual spatial attention signals from the parietal to occipital cortex has been studied in [9]. Biological findings show that the visual system can use feedback signals to highlight the relevant locations. However, the exact

* Corresponding author at: School of Physics and Optoelectronics Technology, Fujian Normal University, Fuzhou 350007, China. Tel.: +86 591 2286558; fax: +86 591 22867373.
E-mail addresses: q.wu@ulster.ac.uk, qxwu@fjnu.edu.cn (Q. Wu), tm.mcginnity@ulster.ac.uk (T.M. McGinnity), lp.maguire@ulster.ac.uk (L. Maguire), rtcia@fjnu.edu.cn (R. Cai), mchen@fjnu.edu.cn (M. Chen).

0925-2312/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.01.046
Please cite this article as: Q. Wu, et al., A visual attention model based on hierarchical spiking neural networks, Neurocomputing (2012), http://dx.doi.org/10.1016/j.neucom.2012.01.046
neuronal circuits remain unclear. Furthermore it is not at all certain how the principles of visual attention based on spiking neural networks can be applied to artificial systems, and the research literature is limited in these respects. In particular there are relatively few explanations of visual attention using spiking neural networks in the literature. In this paper we explore the application of a spiking neuron model and spiking neural networks to visual attention. The principles and simulation algorithms are presented in detail.

The remainder of the paper is organized as follows. The proposed attention system based on top–down control signals is detailed in Section 2. Attention based on prior knowledge is described in Section 3. Attention based on a saliency index is described in Section 4. Comparative results are shown in Section 5. Finally we discuss the proposed model and its performance in Section 6.

2. A spiking neural network model for visual attention

In order to provide an attention model based on top–down volition-controlled signals and explain behaviours of the network of spiking neurons in the model, a new attention system is proposed as in Fig. 1. This system uses spiking-rate maps from the retina-inspired models proposed in [16] and the visual pathway processing models in [17], combined with a new top–down volition-controlled model to obtain attention areas. A description of the proposed model is now presented.

The architecture of the proposed attention system has three levels. In the first level (the extraction level), features are extracted using the retina-inspired SNN models, which are detailed in [10,16,17], for example, edge detection SNN [10], colour segmentation SNN [16] and other retina inspired models. In the second level (the decomposition level), low level image components are decomposed in multiple visual pathways. The components can be vertical lines, horizontal lines, or corners which are decomposed from edge maps [10]. They can also be colour ON/OFF maps [16] or other visual pathways [18]. In the third level (the attention level), a new volition-controlled model is used to select and combine different image components to form an ‘attention area map’. Many researchers are exploring how to obtain saliency maps [19–21]. The advantage is that a saliency map can be obtained from an image using a bottom–up approach without prior knowledge. However, in many cases we hope to focus on specific areas that match our prior knowledge. In other words, the desired attention area is volition-controlled by specific consciousness. This attention area can focus on a set of specific features related to object types. It is supposed that volition-controlled signals come from a high level in the brain and contain the feature set upon which attention should be attended.

Since computation in the brain is based on spiking neurons, our attention model is simulated using spiking neural networks. In order to represent the simulation algorithms for this model, let A represent a set features obtained by retina-inspired spiking neural networks, A={a1, a2, a3,...}={Edge, Colour, Texture,...}. Let B represent a set of components of a visual image, B={b1, b2, b3,...}={Vertical line, Horizontal line, 45 degree line, colour Red-ON, ...}. Let P(x,y) represent a pixel in the input image with dimension N×M (i.e. x=1, 2, 3,..., N; y=1, 2, 3,..., M). Retina inspired spiking neural network models are used to extract features, a feature map is represented by a spike train array generated from a neuron array with the same dimension as the input image. The output of the neuron array generates a spike train array Sa(x,y), where a∈A, x=1, 2, 3,..., N, y=1, 2, 3,..., M.

\[ S_{a}(x,y)(t) = \begin{cases} 1 & \text{if output neuron } a(x,y) \text{ fires at time } t \\ 0 & \text{if output neuron } a(x,y) \text{ does not fire at time } t \end{cases} \] (1)

For example, Sa(edge(x,y)) represents a spike train array of an edge map for the input image. The algorithm for obtaining the spike train array can be found in [10]. The spike train array Sa(x,y)(t) can be regarded as feature information that is transferred from the extraction level to the decomposition level. The spiking neural network models in the decomposition level receive the information and decompose them into visual image components. For example, the edge map can be decomposed to corners or lines with different orientations. This algorithm is detailed in Section 3.

A colour image can be decomposed to different colour ON/OFF pathways. These visual image components are also represented by spike train arrays that are generated by corresponding output neuron arrays in the decomposition level. Let b∈B represent a visual image component in a visual pathway. The spike train array in the pathway is represented as follows.

\[ S_{b}(x,y)(t) = \begin{cases} 1 & \text{if output neuron } b(x,y) \text{ fires at time } t \\ 0 & \text{if output neuron } b(x,y) \text{ does not fire at time } t \end{cases} \] (2)

The spike train arrays from the decomposition level are regarded as inputs for the proposed attention model in this paper. The attention model receives the spike train arrays and then combines with volition-controlled signals to generate attention areas. It is very obvious that such attention areas are selected by the volition-controlled signals and spike train arrays of visual image components. Using different volition-controlled signals, different interest areas can be attended.

Let Sc(x,y) represent a spike train array from the output neuron array Nc(x,y), x=1, 2, 3,..., N, y=1, 2, 3,..., M.

\[ S_{c}(x,y)(t) = \begin{cases} 1 & \text{if output neuron } c(x,y) \text{ fires at time } t \\ 0 & \text{if output neuron } c(x,y) \text{ does not fire at time } t \end{cases} \] (3)

![Fig. 1. Top–down attention system based on spiking neural networks.](image-url)
The attention area map is represented by a spike rate map $I_{c(x,y)}$ as follows.

$$I_{c(x,y)} = \frac{1}{T} \sum_{t=\text{int}}^{t=\text{end}} S_{c(x,y)}(t)$$

where $T$ is a time window. The high spike rate areas represent interesting areas. The high spike rate areas represent areas of interest, while low spike rate areas are ignored. We use the conductance-based integrate and fire neuron model since its behaviour is very similar to the Hodgkin and Huxley neuron model while its computational efficiency is greatly improved.

Let $P(x,y)$ represent a pixel value of the input image. The function of the neural networks in the extraction level is to transform the pixel values to spike train arrays $S_{(x,y)}$ corresponding to specific features. The spiking neural network models have been proposed in [10,11,17]. The function of the neural networks in the decomposition level is to transform the spike train arrays $S_{(x,y)}$ (a $\in A$) of specific features to spike train arrays $S_{(x,y)}$ ($b \in B$) of visual image components. A general simulation algorithm can be expressed as follows. Neuron $b(x,y)$ in the decomposition level receives spike trains from neurons within a receptive field in the extraction level. These cause changes in the conductance of both excitatory and inhibitory synapses of neuron $b(x,y)$ as shown in the following equations.

$$\frac{g_{ex}^{(x,y)}(t)}{dt} = -\frac{1}{\tau_{ex}} g_{ex}^{(x,y)}(t) + \sum_{(x',y') \in RF_{ex}} w_{ex}^{(x',y')}(x,y)q_{ex}^{(x',y')}$$

$$\frac{g_{ih}^{(x,y)}(t)}{dt} = -\frac{1}{\tau_{ih}} g_{ih}^{(x,y)}(t) + \sum_{(x',y') \in RF_{ih}} w_{ih}^{(x',y')}(x,y)q_{ih}^{(x',y')}$$

where $q_{ex}$ and $q_{ih}$ represent the peak conductance for excitatory and inhibitory synapses respectively, $g_{ex}^{(x,y)}(t)$ and $g_{ih}^{(x,y)}(t)$ represent the conductance of excitatory and inhibitory synapses of neuron $N_{b(x,y)}$ respectively, and decay with time constants $\tau_{ex}$ and $\tau_{ih}$ respectively.

$w_{ex}^{(x',y')}(x,y)$ and $w_{ih}^{(x',y')}(x,y)$ represent the weight matrices of excitatory and inhibitory synapses from neuron $N_{b(x',y')}$ to neuron $N_{b(x,y)}$ respectively. $RF_{ex/ih}$ represents the receptive field of neuron $N_{b(x,y)}$. According to the conductance-based integrate-and-fire neuron model, the potential of neuron $b(x,y)$ is governed by the following equation.

$$\frac{c}{dt}V_{b(x,y)}(t) = g_{i}(E_{i} - V_{b(x,y)}(t)) + \frac{g_{ex}^{(x,y)}(t)}{A_{ex}}(E_{ex} - V_{b(x,y)}(t)) + \frac{g_{ih}^{(x,y)}(t)}{A_{ih}}(E_{ih} - V_{b(x,y)}(t))$$

where $c$ represents the capacitance of the membrane, $g_{i}$ represents the conductance of the membrane, $E_{ex}$ and $E_{ih}$ are the reverse potential for excitatory and inhibitory synapses respectively, $A_{ex}$ is the membrane surface area connected to an excitatory synapse, $A_{ih}$ is the membrane surface area connected to an inhibitory synapse. If the potential $V_{b(x,y)}(t)$ increases and reaches a threshold $\theta$, the neuron $N_{b(x,y)}$ generates a spike and moves into the refractory period for a time $\tau_{ref}$. After the refractory period the neuron can generate another spike. Therefore, the spike train array $S_{b(x,y)}(t)$ is obtained.

In the attention level, neuron $N_{c(x,y)}$ receives spike trains from neurons within a receptive field in the decomposition level as well as volition-controlled signals. The synapse conductance is governed by the following equations.

$$\frac{g_{ex}^{(x,y)}(t)}{dt} = -\frac{1}{\tau_{ex}} g_{ex}^{(x,y)}(t) + \sum_{b \in B} w_{ex}^{(x',y')}(x,y)D(b)q_{ex}^{(b)}$$

$$\frac{g_{ih}^{(x,y)}(t)}{dt} = -\frac{1}{\tau_{ih}} g_{ih}^{(x,y)}(t) + \sum_{b \in B} w_{ih}^{(x',y')}(x,y)D(b)q_{ih}^{(b)}$$

where $q_{ex}$ and $q_{ih}$ represent the peak conductance for excitatory and inhibitory synapses respectively corresponding to visual image component $b$ (e.g., vertical lines), $g_{ex}^{(x,y)}(t)$ and $g_{ih}^{(x,y)}(t)$ represent conductance for excitatory and inhibitory synapses of neuron $N_{c(x,y)}$ in the attention level and decay with time constants $\tau_{ex}$ and $\tau_{ih}$ respectively. $w_{ex}^{(x',y')}(t)$ and $w_{ih}^{(x',y')}(t)$ represent weight matrices of excitatory and inhibitory synapses from neuron $N_{b(x',y')}$ to neuron $N_{c(x,y)}$ respectively, $RF_{c(x,y)}$ represents the receptive field of neuron $N_{c(x,y)}$. Note that $D(b)$ is controlled by volition-controlled signals, i.e. which can be determined by prior knowledge or a learning algorithm. For example, vertical and horizontal lines are features of a house. This can be set $D(b) = 1$ for visual image component $b \in \{'vertical lines', 'Horizontal lines'\}$, otherwise $D(b) = 0$ for visual image component $b \in \{other\}$. Details of an example application are described in Section 3.

Based on the conductance-based integrate-and-fire neuron model, the potential of neuron $c(x,y)$ is governed by the following equation.

$$\frac{c}{dt}V_{c(x,y)}(t) = g_{e}(E_{e} - V_{c(x,y)}(t)) + \frac{g_{ex}^{(x,y)}(t)}{A_{ex}}(E_{ex} - V_{c(x,y)}(t)) + \frac{g_{ih}^{(x,y)}(t)}{A_{ih}}(E_{ih} - V_{c(x,y)}(t))$$

If the potential $V_{c(x,y)}(t)$ reaches a threshold $\theta$, neuron $N_{c(x,y)}$ generates a spike, and a spike train array $S_{c(x,y)}(t)$ is obtained, whose spike rate map represents attention areas. It is obvious that the spike rate map is determined by both the spike train arrays of visual image components $S_{(x,y)}(t)$ and the volition-controlled signal $D(b)$ in Eqs. (8) and (9), $w_{ex}^{(x',y')}$ and $w_{ih}^{(x',y')}$ are synapse strength matrices form excitatory centre-ON and inhibitory around-OFF receptive fields for neuron $c(x,y)$. The expressions are as follows.

$$w_{ex}^{(x',y')} = \begin{cases} \exp(-\frac{(x'-x)^2 + (y'-y)^2}{\sigma^2}) & \text{if } \sqrt{(x'-x)^2 + (y'-y)^2} \leq R_{RFc} \\ 0 & \text{if } \sqrt{(x'-x)^2 + (y'-y)^2} > R_{RFc} \end{cases} \tag{11}$$

$$w_{ih}^{(x',y')} = \begin{cases} \exp(-\frac{(x'-x)^2 + (y'-y)^2}{\sigma^2}) & \text{if } \sqrt{(x'-x)^2 + (y'-y)^2} > R_{RFc} \\ 0 & \text{if } \sqrt{(x'-x)^2 + (y'-y)^2} \leq R_{RFc} \end{cases} \tag{12}$$

where $(x',y')$ is the position of neuron $c(x,y)$ corresponding to the centre of the receptive field, $(x,y)$ is a neuron position within the receptive field, $R_{RFc}$ is the radius of the receptive field, $\sigma$ is a constant for determining the radius of centre ON, $\lambda$ is a constant for determining the radius around-OFF (i.e. the centre ON is represented by excitatory synapses and the around-OFF is represented by inhibitory synapses). These strength matrices enable the output neuron attention level array to combine the effects of all neighbourhood neurons to form an attention area. The parameters are satisfied $\lambda \leq \lambda \leq R_{RFc}$. According to the behaviours of a spiking neuron group in such receptive field, a competitive mechanism is performed. The activities of neurons in the centre-ON field strengthen each other, but the neurons in the around-OFF field are inhibited. If $R_{RFc} - \lambda$ is large, the activated neuron group can inhibit neurons far away from the group. If the $R_{RFc}$ is sufficiently large to cover the whole image, Eqs. (11) and (12) can be simplified as follows.

$$w_{ex}^{(x',y')} = \begin{cases} \exp(-\frac{(x'-x)^2 + (y'-y)^2}{\sigma^2}) & \text{if } \sqrt{(x'-x)^2 + (y'-y)^2} \leq R_{0} \\ 0 & \text{if } \sqrt{(x'-x)^2 + (y'-y)^2} > R_{0} \end{cases} \tag{13}$$

$$w_{ih}^{(x',y')} = e^{-\sigma^2(x'-x)^2 + (y'-y)^2} \text{for all neurons } (x',y') (x' = 1,...,N, y' = 1,...,M) \tag{14}$$
where $\varepsilon$ is a small number within $[0, 1]$ for all inhibitory synapses from neuron $b(x', y')$ to $c(x, y)$. This means that the centre-ON field with a radius $R_b$ is composed of excitatory synapses and synapses outside of the area are inhibitory synapses. Using this pair of weight matrices, the network can concentrate on the interesting areas and inhibit other areas.

An advantage of this attention model is that prior knowledge can be simply applied to obtain interest areas related to specific features. Since the model is based on multiple visual image components, prior knowledge is used to set $D(b)$ to control the pathways. $D(b)=1$ means that the information can be transferred through the pathway. $D(b)=0$ means that the information cannot be transferred through the pathway. If prior knowledge shows that an object contains some key visual image components, the interest areas can be obtained using this model through setting the corresponding $D(b)$ to 1 or 0. On the other hand, the visual pathways can also be selected using a saliency index of a visual image component map.

**Definition 1.** Let $r(x, y)$ represent a spike rate map calculated from a spike train array, $x=1, 2, 3, .., N; y=1, 2, 3, .., M$. A saliency index, denoted as $SI$, is defined by the following expression.

$$SI = \frac{1}{4} \sum_{x=1}^{N} \sum_{y=1}^{M} (\mu_{xy} - U), \text{sum for all}(x, y) \text{ with } (\mu_{xy} - U) > 0$$

$$K = \sum_{x=1}^{N} \sum_{y=1}^{M} \text{sign}(\mu_{xy} - U), \text{sum for all}(x, y) \text{ with } (\mu_{xy} - U) > 0$$

where $SI$ is an average value of $(\mu_{xy} - U)$ for all pixels $(x, y)$ with $(\mu_{xy} - U) > 0$. $K$ is the number of pixels for all $(x, y)$ with $(\mu_{xy} - U) > 0$. $U$ is an average spike rate over the whole spike rate map with dimension $(N \times M)$. $\mu_{xy}$ is a Gaussian-weighted spike rate within the receptive field $RF_{(x,y)}$ centred at $(x, y)$ within a circle with radius $R_s$. $K$ is calculated as follows.

$$U = \frac{1}{NM} \sum_{x=1}^{N} \sum_{y=1}^{M} r(x, y)/r_{\text{max}}$$

$$\mu_{xy} = \frac{\sum_{(x', y') \in RF_{xy}} F_{xy}(x', y') r(x', y')}{\sum_{(x', y') \in RF_{xy}} F_{xy}(x', y') r_{\text{max}}}$$

$$F_{xy}(x', y') = \begin{cases} e^{-\frac{(x-x')^2+(y-y')^2}{2\delta^2}} & \text{if } \sqrt{(x-x')^2+(y-y')^2} \leq R_s \\ 0 & \text{if } \sqrt{(x-x')^2+(y-y')^2} > R_s \end{cases}$$

where $R_s$ is the radius of the receptive field $RF_{(x,y)}$, $\delta$ is a constant, and $r_{\text{max}}$ is the maximal value of $r(x, y)$ among all $(x, y)$. Since $\mu_{xy}$ is an average spike rate in a local area weighted by the Gaussian-distribution $F_{xy}(x', y')$, the value depends on the spike rates of neighbourhood neurons around location $(x, y)$ and in particular the neurons close to the centre $(x, y)$. If the value of $\mu_{xy}$ is large, the neurons around location $(x, y)$ have high spike rates. Therefore, the value of $(\mu_{xy} - U)$ indicates that a group of neurons centred at $(x, y)$ has spike rates higher than average spike rate. The saliency index $SI$ is an average of $(\mu_{xy} - U)$ for all the pixels with $(\mu_{xy} - U) > 0$. If $SI$ is large, there are some neuron groups with higher spike rates than other neuron groups. If $SI$ is 0, all neurons in the array have similar spike rates. Let $SI(b)$ represent saliency index for spike rate map $p_{b}(x, y)$ from visual pathway $b$. Using this index, the pathway $b^*$, which corresponds to maximum of saliency index, can be automatically selected as follows.

$$b^* = \arg\max_{b \in \mathcal{B}}(SI(b)),$$

where $\mathcal{B} = \{b_1, b_2, b_3, \ldots\} = \{\text{Vertical line, Horizontal line, 45 degree line, colour Red-ON, \ldots}\}$. Set $D(b') = 1$ and $D(b) = 0(b \in B - b^*)$ in the attention model. The attention areas corresponding to the $b^*$ pathway can be obtained. A demonstration of this algorithm is presented in Section 4.

### 3. Attention based on prior knowledge

#### 3.1. Spiking neural network pathways of vertical and horizontal lines

Based on the proposed attention model, a case study is provided in this section. We consider the case of paying attention to houses in an image. From prior knowledge, we know that houses generally exhibit key features that are represented as horizontal and vertical lines. If the network focuses on the horizontal and vertical lines and the signals from the pathways of horizontal and vertical lines are regarded as feedback to the early visual cortex layers, the region containing the house can be strengthened relative to objects in other areas of the image. As the basic processing units in the brain are spiking neurons, the research question is what is the exact spiking neural network for this attention performance in the brain? Based on spiking neuron models and a top–down signal feedback mechanism, a spiking neural network is proposed in Fig. 2 which can be used to extract the region containing houses from a visual image. Suppose that a visual image is presented to the retina and an edge spike rate map [10] is obtained as the input for the network in Fig. 2. The decomposition level contains two pathways i.e. $\mathcal{B} = \{\text{Horizontal}, \text{Vertical}\} = \{h, v\}$, (‘h’ stands for ‘horizontal’ and ‘v’ for ‘vertical’). The neuron array for the horizontal pathway is labelled as $N_h$ and has the same size as the input neuron array. Each neuron has a receptive field corresponding to a horizontal synapse strength matrix $W^h$. The neuron array for the vertical pathway is labelled as $N_v$ and has also the same size as input neuron array.

![Fig. 2. Spiking neural network model for attention on vertical and horizontal lines.](Image)
Each neuron has a receptive field corresponding to a vertical synapse strength matrix $W_v$. Therefore, the spike rate map of the neuron arrays $N_h$ and $N_v$ represent horizontal and vertical lines respectively. In the attention level, the neuron array at the output is labelled as $N_d(x,y)$ with the same size as the edge spike rate map. Using prior knowledge of features of a house, we set $D_v^b(b) = 1$ for visual image component $b \in \{h,v\}$, otherwise $D_v^b(b) = 0$ for visual image component $b \in \{\text{other}\}$. Each neuron only receives spike trains from two receptive fields on the horizontal and vertical neuron arrays respectively. The two receptive fields have the same synapse strength matrix $W^a$, determined by Eqs. (13) and (14). If there is any horizontal line or vertical line around neuron $N_h(x,y)$, the neuron will be activated, and horizontal or vertical line areas can be obtained from the output layer. If these signals are regarded as feedback to an earlier layer of the visual system, objects in the area can be extracted and objects in other areas can be ignored.

3.2. Simulation algorithms of the line-based SNN

Suppose that only an edge map is considered in this case, i.e., $a=\text{edge}$ in Eq. (1). Let $e$ stand for $\text{edge}$. The spike train array for the edge map is $S_{\text{edge}}(t)$ which can be obtained using the spiking neural network in [10].

$$S_{\text{edge}}(t) = \begin{cases} 1 & \text{if neuron } (x,y) \text{ fires at time } t \\ 0 & \text{if neuron } (x,y) \text{ does not fire at time } t \end{cases} \quad (20)$$

The edge map is decomposed in two image components, horizontal and vertical lines. Horizontal and vertical lines are represented by two neuron arrays $N_{h(x,y)}(b=\{h,v\})$, $N_v(x,y)$ and $N_h(x,y)$ represent neuron arrays for horizontal and vertical lines respectively. $v_{h(x,y)}(t)$ and $v_{v(x,y)}(t)$ represent the membrane potential of the two neuron arrays, $g_{h(x,y)}^e(t)$ and $g_{h(x,y)}^i(t)$ represent conductances for excitatory and inhibitory synapses respectively for neuron $N_h(x,y)$. $g_{v(x,y)}^e(t)$ and $g_{v(x,y)}^i(t)$ represent conductance for excitatory and inhibitory synapses respectively for neuron $N_v(x,y)$. According to the conductance-based integrate-and-fire neuron model in Eqs. (5) and (6), we have

$$\frac{dV_{h(x,y)}^e}{dt} = -\frac{1}{\tau_h} \sum_{(x',y') \in \text{RF}_{h(x,y)}} W_{h(x',y')}^e S_{h(x',y')}(t) + g_{h(x,y)}^e(t) \left[ E_h - V_{h(x,y)}(t) \right]$$
$$\frac{dV_{h(x,y)}^i}{dt} = -\frac{1}{\tau_i} \sum_{(x',y') \in \text{RF}_{h(x,y)}} W_{h(x',y')}^i S_{h(x',y')}(t) + g_{h(x,y)}^i(t) \left[ E_i - V_{h(x,y)}(t) \right]$$

where $\tau_h$ and $\tau_i$ are time constants for excitatory and inhibitory synapses, $\tau_h$ is short for excitatory, and $\tau_i$ for inhibitory. If the neuron $(x,y)$ in the receptive field $\text{RF}_{h(x,y)}$ generates a spike, the conductance of excitatory and inhibitory synapses increases by the amount $g_{h(x,y)}^e$ and $g_{h(x,y)}^i$ respectively. If the neuron $(x,y)$ does not generate a spike at time $t$, $g_{h(x,y)}^e(t)$ and $g_{h(x,y)}^i(t)$ decay with time constants $\tau_h$ and $\tau_i$ respectively. The conductance changes lead to different currents that are injected to neurons $N_h(x,y)$ and $N_v(x,y)$. According to Eq. (7), the membrane potential $V_{h(x,y)}(t)$ and $V_{v(x,y)}(t)$ of neurons $N_h(x,y)$ and $N_v(x,y)$ are governed by the following equations.

$$\frac{dV_{h(x,y)}(t)}{dt} = \frac{g_{h(x,y)}^e(t)}{A_{h(x,y)}} \left[ E_h - V_{h(x,y)}(t) \right]$$

$$\frac{dV_{v(x,y)}(t)}{dt} = \frac{g_{v(x,y)}^e(t)}{A_{v(x,y)}} \left[ E_v - V_{v(x,y)}(t) \right]$$

where $E_h$ and $E_v$ are the reverse potential for excitatory and inhibitory synapses respectively, $\tau_h$ represents the capacitance of the membrane, $A_{h(x,y)}$ represents the conductance of the membrane, $A_{v(x,y)}$ is the membrane surface area connected to an excitatory synapse, $A_{h(x,y)}$ is the membrane surface area connected to an inhibitory synapse, $w_{h(x,y)}^e$ represents excitatory synapse strength in matrix $W^h$, and $w_{h(x,y)}^i$ represents inhibitory synapse strength in matrix $W^h$. If the edge in the receptive field matches the pattern in $W^h$, the neuron $N_h(x,y)$ receives a strong input current through the excitatory synapses and the membrane potential $V_{h(x,y)}(t)$ increases. When the potential reaches a threshold $v_{th}$, the neuron $N_h(x,y)$ generates a spike and moves into the refractory period for a time $\tau_{ref}$. After the refractory period, the neuron receives inputs and prepares to generate another spike. By analogy, $g_{v(x,y)}^e(t)$ represents excitatory synapse strength in matrix $W^v$, and $g_{v(x,y)}^i(t)$ represents inhibitory synapse strength in matrix $W^v$. If the edge in the receptive field matches the pattern in $W^v$, the neuron $N_v(x,y)$ receives a strong input current through the excitatory synapses and generates spikes. The matrix pair for synapse strength distribution of $W^h$ and $W^v$ is determined by the following equations.

$$w_{h(x,y)}^e(t) = \begin{cases} -e^{\frac{(x'-x)^2}{2\sigma^2}} & \text{if } (x'-x)^2 + (y'-y)^2 \leq R_{h(x,y)}^e \\
0 & \text{if } (x'-x)^2 + (y'-y)^2 > R_{h(x,y)}^e \end{cases}$$

$$w_{h(x,y)}^i(t) = \begin{cases} -e^{\frac{(x'-x)^2}{2\sigma^2}} & \text{if } (x'-x)^2 + (y'-y)^2 \leq R_{h(x,y)}^i \\
0 & \text{if } (x'-x)^2 + (y'-y)^2 > R_{h(x,y)}^i \end{cases}$$

$$w_{v(x,y)}^e(t) = \begin{cases} -e^{\frac{(x'-x)^2}{2\sigma^2}} & \text{if } (x'-x)^2 + (y'-y)^2 \leq R_{v(x,y)}^e \\
0 & \text{if } (x'-x)^2 + (y'-y)^2 > R_{v(x,y)}^e \end{cases}$$

$$w_{v(x,y)}^i(t) = \begin{cases} -e^{\frac{(x'-x)^2}{2\sigma^2}} & \text{if } (x'-x)^2 + (y'-y)^2 \leq R_{v(x,y)}^i \\
0 & \text{if } (x'-x)^2 + (y'-y)^2 > R_{v(x,y)}^i \end{cases}$$
(9), we have

$$\frac{d g_{ox,y}^{ex}}{dt} = -\frac{1}{\tau_{ex}} g_{ox,y}^{ex}(t) + \sum_{b \in \{h,v\}} \sum_{(x',y') \in RF_{ox,y}} W_{ox,y}^{ex} S_{ox,y}^{ex}(t) D_{ox,y}^{ex}(b) q_{ox,y}^{ex},$$

$$\frac{d g_{ox,y}^{inh}}{dt} = -\frac{1}{\tau_{inh}} g_{ox,y}^{inh}(t) + \sum_{b \in \{h,v\}} \sum_{(x',y') \in RF_{ox,y}} W_{ox,y}^{inh} S_{ox,y}^{inh}(t) D_{ox,y}^{inh}(b) q_{ox,y}^{inh},$$

Because only horizontal and vertical line pathways are considered, $B=\{h,v\}$. $D_{ox,y}^{ex}(b)=1$ for visual image component $b \in \{h,v\}$, otherwise $D_{ox,y}^{ex}(b')=0$ for visual image component $b' \in \{\text{others}\}$. $q_{ox,y}^{ex}$ and $q_{ox,y}^{inh}$ are regarded as the peak conductance for neurons in the attention level in different pathways. According to Eqs. (13) and (14), the synapse strength matrix is determined by the following equations.

$$W_{ox,y}^{ex} = \begin{cases} e^{-\frac{(x'-x)^2+(y'-y)^2}{\sigma^2}} & \text{if } \sqrt{(x'-x)^2+(y'-y)^2} \leq R_0 \\ 0 & \text{if } \sqrt{(x'-x)^2+(y'-y)^2} > R_0 \end{cases}$$

$$W_{ox,y}^{inh} = \frac{r}{T} \text{ for all neurons } (x',y') \ (x'=1,...N, y'=1,...M)$$

we have

$$\frac{d g_{ox,y}^{ex}}{dt} = -\frac{1}{\tau_{ex}} g_{ox,y}^{ex}(t) + \sum_{(x',y') \in RF_{ox,y}} W_{ox,y}^{ex} S_{ox,y}^{ex}(t) q_{ox,y}^{ex}$$

$$\frac{d g_{ox,y}^{inh}}{dt} = -\frac{1}{\tau_{inh}} g_{ox,y}^{inh}(t) + \sum_{x'} \sum_{y'} W_{ox,y}^{inh} S_{ox,y}^{inh}(t) q_{ox,y}^{inh}(t).$$

Let $v_{ox,y}(t)$ represent the membrane potential of neuron $N_{o}(x,y)$. We have

$$c \frac{dv_{ox,y}(t)}{dt} = g_{in} (E_i - v_{ox,y}(t)) + \frac{g_{ex}}{\lambda_{ex}} (E_e - v_{ox,y}(t))$$

$$+ \frac{g_{inh}}{\lambda_{inh}} (E_{inh} - v_{ox,y}(t)).$$

Let $S_{ox,y}(t)$ represent a spike train generated by Neuron $N_{o}(x,y)$ in the attention level. The spike rate for neuron $N_{o}(x,y)$ is calculated by the following expression.

$$r_{ox,y} = \frac{1}{T} \sum_{t=1}^{T} S_{ox,y}(t)$$

By plotting this firing rate as an image with a colour bar, areas of horizontal and vertical lines are obtained. Suppose that pixel values of the input image are represented by $P(x,y)$ and the output signals are regarded as feedback to filter out the attention area as follows.

$$G(x,y) = P(x,y) r_{ox,y}/r_{max} (r_{max} = \max_{(x,y)} |r_{ox,y}| \text{ all pixels in the image})$$

Then the attention area image is obtained by a plot of image $G(x,y)$.  

Fig. 3. Simulation results for extraction of attention area.
3.3. Simulation results for the visual attention algorithm assuming prior knowledge

The proposed spiking neural network is combined with the edge detection network model in [10]. If a visual image as shown in Fig. 3(A) is presented to the edge detection network, edges of the image are reflected in the output neuron array, as in [10]. The spike trains from this neuron array are regarded as inputs for the proposed network in Fig. 2. In our simulation, the spike rate map of edges of the image in Fig. 3(A) is shown in Fig. 3(B). If the spike trains are transferred to horizontal and vertical line pathways according to the network architecture of Fig. 2, the spike rates of the horizontal and vertical line neuron arrays are shown in Fig. 3(C) and (D). The spike trains from horizontal and vertical line neuron arrays are transferred to the output layer. The spike rates of the output neuron array are shown in Fig. 3(E). The output firing rates can be regarded as feedback signals to obtain the attention areas that are based on the key features of horizontal and vertical lines. The image corresponding to the attention area is obtained in Fig. 3(F). It can be seen that the region around the building has been strengthened and other areas are ignored.

4. Attention based on a saliency index

In this section it is supposed that colour decomposition is obtained using the ON/OFF spiking neural networks in [16], i.e. a colour image has been decomposed into visual image components $B = \{\text{Red}_\text{ON}, \text{Red}_\text{OFF}, \text{Green}_\text{ON}, \text{Green}_\text{OFF}, \text{Blue}_\text{ON}, \text{Blue}_\text{OFF}, \text{White}_\text{ON}, \text{White}_\text{OFF}\}$. For simplicity, the set $B$ is rewritten as $B = \{R, RX, G, GX, B, BX, W, WX\}$, where $R$ stands for Red-ON, RX for Red-OFF, etc. According to the attention system in Fig. 1, we have eight spike trains of visual image components as follows.

$$S_{bx}(t) = \begin{cases} 1 & \text{if output neuron } b(x,y) \text{ fires at time } t \\ 0 & \text{if output neuron } b(x,y) \text{ does not fire at time } t \end{cases}$$

(38)

where $b \in B$, $b(x,y)$ is the output neuron array in the colour decomposition model in [16]. Let $r_{b(x,y)}$ represent a spike rate map calculated from a spike train array, $x = 1, 2, 3, ..., N; y = 1, 2, 3, ..., M$.

$$r_{bx}(y) = \frac{1}{T} \sum_{t=1}^{T} S_{bx}(t)$$

According to the definition in Eqs. (15)–(18), the saliency index $S(b)$ for each pathway $b$ can be obtained. According to Eq. (19), $b^* = \operatorname{argmax}(S(b))$.

$b^*$ is the pathway with maximal saliency index $S$, for example, considering the original image to the left of Fig. 4. This is decomposed into eight spike train arrays with spike rate maps shown in the right of Fig. 4. The saliency index, which is calculated from Eq. (15), for each visual pathway $b$ is listed in Table 1.

Table 1 shows that $b^* = 'BX' \quad (i.e. \text{Blue}_\text{OFF})$. We set $D\left(b^*\right) = 1$ and $D(b) = 0(b \in B - b^*)$ in the attention model. Therefore, only the visual image component from the BX pathway is selected to generate attention areas. According to Eqs. (8) and (9), the synapse conductance for attention neurons $c(x,y)$ is governed by the following equations.

$$\frac{d g_{ex}^{b(x,y)}(t)}{dt} = - \frac{1}{\tau_{ex}} g_{ex}^{b(x,y)}(t) + \sum_{b = BX, Y} \sum_{(x', y') \in R_{bx}} w_{ex}^{b(x', y')}(t) D(BX) g_{ex}^{b(x', y')}$$

Table 1. Saliency index $S(b)$ for each pathway.

<table>
<thead>
<tr>
<th>Photo</th>
<th>Saliency index $S(b)$ for each pathway</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>0.24</td>
<td>0.09</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Fig. 4. Spike rate maps for eight ON/OFF pathways.

Fig. 5. Attention Areas Obtained from pathway $b^*$.
The potential of neuron $c(x,y)$ is governed by the following equation.

$$
v_c(x,y) = g_{E, L}(v_c(x,y)) + g_{E, L}^e(x,y) \frac{E_{E, L}}{C_0/C_1} + g_{E, L}^h(x,y) \frac{E_{E, L}}{C_0/C_1}
$$

Let $S_{c(x,y)}$ represent the spike train array from the output neuron array $N_{c(x,y)}$, $x=1, 2, 3, \ldots, N$, $y=1, 2, 3, \ldots, M$.

$$
S_{c(x,y)}(t) = \begin{cases} 
1 & \text{if output neuron } c(x,y) \text{ fires at time } t \\
0 & \text{if output neuron } c(x,y) \text{ does not fire at time } t 
\end{cases}
$$

**Fig. 6.** Result compression.
The attention area map is represented by a spike rate map as follows.

\[ r_{i(x,y)}(t) = \frac{1}{T} \sum_{t=1}^{T} s_{i(x,y)}(t) \]

By plotting this firing rate as an image with a colour bar, attention areas corresponding to spike trains from the Blue_OFF pathway (i.e. \( b^p = 'B' \)) are obtained, as shown in Fig. 5.

5. Comparison of results

To the authors knowledge, there are no top–down attention models which are based on spiking neural networks. Therefore, we have compared the results of our proposed model with current attention models [19,20] which are not based on spiking neural networks. The important differences are that our proposed model, being based on SNNs, is more biologically plausible. The parameters of spiking neurons are based on data from real biological neurons [12]. The following values of parameters are set for all the spiking neurons in the proposed models. \( \nu_{th} = -60\,\text{mv} \), \( \nu_{reset} = -70\,\text{mv} \), \( E_{ex} = 0\,\text{mv} \), \( E_{in} = -75\,\text{mv} \), \( g = 1.0\,\text{muS/mm}^2 \), \( c_m = 10\,\text{nS/mm}^2 \), \( A_{syn} = 0.028953 \,\text{mm}^2 \), \( A_{inh} = 0.014103 \,\text{mm}^2 \), \( \tau_{ref} = 4\,\text{ms} \), \( \tau_{syn} = 4\,\text{ms} \). For some parameters, their values can be changed in a significant range, they can be set different values in different part of the spiking neural network. For example, \( \tau_{ref} \) is set to 6 ms for neurons in feature extraction level and decomposition level, but \( \tau_{ref} \) is set to 25 ms for attention level. \( A_{syn}, q_{ex}, q_{inh}^p, q_{inh}^c, q_{inh}^d, q_{inh}^s \) and \( q_{inh}^s \) can be adjusted from a large range for example [0.0001,1], which can be determined by trying and error method. Parameters for the receptive fields are set as follows. (1) For receptive fields of horizontal and vertical lines, we have \( \delta_{x} = 20, \delta_{y} = 2, \delta_{x} = 20, \delta_{y} = 20, \Delta = 2, R_{sp} = 15 \). (2) For receptive fields of neurons in attention level, \( \delta = 10, R_{sp} = 20, \tau = 0.04 \). (3) For saliency index, we set \( R_c = 3, \sigma = 2 \).

The attention system was implemented using Matlab. The time window \( T \) was set to 200 ms, consistent with the processing time of the biological visual system. Simulation results were obtained within this time window. The six photos in Fig. 6, which are from [19], were used to test the proposed attention model. The first column shows original pictures; the second column shows the results of Itti’s model [20]. The third column shows the results of Multi-Scale model [19]; the fourth column shows the results of our proposed model. Attention maps in column 4 for Photos A, B, D and F are obtained using the saliency based attention. Attention maps in column 4 for Photos C and E are obtained using the attention approach based horizontal and vertical lines (prior knowledge).

6. Conclusion

Since learning mechanisms of artificial neural networks (ANN) have been improved in different ways (e.g. [23–25]), ANNs have been applied to image processing and pattern recognition [22,26]. The novelty of this paper is to simulate an attention model using biologically plausible SNNs instead of ANNs. Complex combination of receptive fields in conjunction with hierarchical structures of spiking neurons enable SNNs to perform very complicated computation tasks, learning tasks and intelligent behaviours in the human brain. This paper has proposed a visual attention system that integrates different spiking neural network models inspired by the visual system. Receptive fields in three levels have been described in detail. Multiple image components and features have been extracted by the hierarchical spiking neural networks. Based on these multiple visual pathways and top–down vigilance-controlled signals, attention areas with specific image components or features can be extracted using the proposed attention system that is composed of parallel spiking neuron networks. The system can perform visual attention to obtain attention areas for object types with specific image components or features, such as horizontal and vertical lines, colour features in ON/OFF visual pathways, etc. Simulations show that the results can be obtained within a time window 200 ms, which is consistent with the biological visual system. In order to select good visual image components for attention, a saliency index is defined. The index can indicate whether an image is well separated between interesting areas and non-interesting areas. This index can be used in the saliency-based attention approach, and can also be used to evaluate the comparison results from different attention models.

In this paper, the image components and features have been limited to consideration of edges, orientation lines, and eight ON/OFF colour visual pathways. Based on the proposed attention system, other complicated image features can also be added to obtain attention areas for more complicated objects. This model can be combined with other bio-inspired approaches such as tensor features [27] and scene classification [28,29]. These are topics of further study.

Acknowledgements

The authors gratefully acknowledge the fund from Research Council UK and the Centre of Excellence in Intelligent Systems (CoEIS) grant (InvestNI, IDF and University of Ulster). The authors also acknowledge the fund from the Natural Science Foundation of China (Grant no. 61179011) and the Natural Science Foundation of Fujian Province (Grant no. 2011J01340).

References

QingXiang Wu received the Ph.D. degree in intelligent systems from the University of Ulster in United Kingdom in 2005. He is a RCUK academic fellow in the School of Computing and Intelligent Systems in the University of Ulster. He has been a research fellow in School of Electronics, Electrical Engineering and Computer Science, Queen's University. He is a professor in School of Physics and OptoElectronics Technology, Fujian Normal University, China. His research interests include intelligent systems, robotics, spiking neural networks, machine learning, knowledge discovery, data mining, automatic disease diagnosis, and Chinese information processing. He has more than 80 publications on these topics in international journals, Chinese journals, and proceedings of international conferences.

Martin McGinnity holds the post of Professor of Intelligent Systems Engineering within the Faculty of Computing and Engineering, University of Ulster. He has a first class honours degree in physics, and a doctorate from the University of Durham, is a Fellow of the IET, senior member of the IEEE, and a Chartered Engineer. He is currently the Director of the Intelligent Systems Research Centre (ISRC), on the Magee Campus of the University, which encompasses the research activities of approximately 100 researchers. He is a former Director of both the University’s technology transfer company, Innovation Ulster, and a spin out company Flex Language Services. He is the author or co-author of over 250 research papers, and has been awarded both a Senior Distinguished Research Fellowship and a Distinguished Learning Support Fellowship by the University in recognition of his contribution to teaching and research. His current research interests are focussed on the study of computational intelligence in general, and the creation of bio-inspired intelligent computational systems, which explore and model biological information and signal processing. He is particularly interested in developing models of neural processing which accurately reflect the complexity of neural dynamics and brain coordination. He founded the computational neuroscience research team at the ISRC, with a focus on neurodegeneration (Alzheimer’s and MDD, major depressive disorder) and there is ongoing research in this team into supporting topics such as Diffusion Weighted MRI imaging, and EEG based brain computer interfacing. Professor McGinnity has attracted over £17milion to support his research in recent years, from a wide range of sources.

Liam P. Maguire received MEng and PhD degrees in Electrical and Electronic Engineering from the Queen’s University of Belfast, UK, in 1998 and 1991, respectively. He is currently Head of the School of Computing and Intelligent Systems at the University of Ulster. He is also the leader of the Bio-inspired Systems and Neuro-engineering research team within the Intelligent Systems Research Centre. His research interests are in two primary areas: fundamental research in bio-inspired intelligent systems (such as the development of computational effective spiking neural networks) and the application of existing intelligent techniques in different domains. He is the author of over 200 research papers including 70 journal papers. He has an established track record of securing research funding, approximately £1.1M since his appointment and has also supervised 15 PhD and 3 MPhil students to completion.

Rongtai Cai received the BS degree in Automation from Jilin University and PhD degree in Mechatronic Engineering from Chinese Academy of Science, China, in 2003 and 2008, respectively. He is a Lecturer in the School of Physics, Optics, Electronics and Information, Fujian Normal University, China. His research interests include intelligent computation, pattern recognition, machine learning (spiking neural networks, Bayesian method and so on) and their application in computer vision (such as visual object detection, tracking, recognition and medical image analysis). He has more than 30 publications on these topics in international journals, Chinese journals, and proceedings of international conferences.

Meigui Chen received the bachelor’s degree in Network Engineering from Fujian Normal University, China, in 2009. She is currently a post graduate student in the School of Physics and OptoElectronics Technology, Fujian Normal University, in China. Her research interests are medical image processing and spiking neural networks.