International Journal of Neural Systems, Vol. 16, No. 3 (2006) 151–162

World Scientific
www.worldscientific.com

AN ADAPTIVE VISUAL NEURONAL MODEL IMPLEMENTING COMPETITIVE, TEMPORALLY ASYMMETRIC HEBBIAN LEARNING

ZHIJUN YANG[*,†], KATHERINE L. CAMERON[†], ALAN F. MURRAY[†,‡] and VASIN BOONSOBHAK[†]

[*]*Department of Computer Science, Nanjing Normal University, Nanjing 210097, China*

[†]*Institute for Integrated Micro and Nano Systems (IMNS), School of Engineering and Electronics, University of Edinburgh, Edinburgh EH9 3JL, UK*

[‡]*Alan.murray@ee.ed.ac.uk*

Received 28 March 2005
Revised 5 March 2006
Accepted 9 March 2006

A novel depth-from-motion vision model based on leaky integrate-and-fire (I&F) neurons incorporates the implications of recent neurophysiological findings into an algorithm for object discovery and depth analysis. Pulse-coupled I&F neurons capture the edges in an optical flow field and the associated time of travel of those edges is encoded as the neuron parameters, mainly the time constant of the membrane potential and synaptic weight. Correlations between spikes and their timing thus code depth in the visual field. Neurons have multiple output synapses connecting to neighbouring neurons with an initial Gaussian weight distribution. A temporally asymmetric learning rule is used to adapt the synaptic weights online, during which competitive behaviour emerges between the different input synapses of a neuron. It is shown that the competition mechanism can further improve the model performance. After training, the weights of synapses sourced from a neuron do not display a Gaussian distribution, having adapted to encode features of the scenes to which they have been exposed.

Keywords: Spike timing dependent plasticity; early vision model; integrate and fire neuron; adaptation; image analysis; object depth recovery.

1. Introduction

Synchronised spiking neuron activity has been observed in the cerebral cortex and in the hippocampus, and postulated to be essential for animal and human cognitive behaviour.[1,2] Experiments have demonstrated that blind patients can see a pattern of light, which corresponds to the pattern on an array of synchronised electrodes directly imposed on to their visual cortex.[3] Analytical approaches[4–6] and further experiments[7,8] have therefore been undertaken to investigate whether neural populations can propagate time-locked or synchronous spikes for neocortical information processing. Some such studies link synchronisation to object representation in the visual cortex.[8] This paper draws directly upon recent experiments on activity-dependent firing behaviour in mammalian hippocampal pyramidal cells *in vitro*[9–11] and tectal cells of Xenopus tadpoles[12] which indicates that neocortical neurons encode and process biological information by coordinating postsynaptic spikes with presynaptic inputs within a critical (millisecond-scale) time window. This temporal relationship between stimulus and response can be generated by adaptation of input synaptic weights, via a family of algorithms known as Temporally Asymmetric Hebbian Learning (TAH) or Spike-Timing-Dependent (synaptic) Plasticity (STDP).

Modelling studies, many of which employ integrate-and-fire (I&F) model neurons, have demonstrated the ability of neural networks to emulate

some neocortical phenomena. Furthermore, STDP adaptation in networks of neurons has been shown to improve temporal sequence learning, propagation and prediction.[13–18] The synfire chains model, for example, consisting of groups of feedforward I&F neurons, is able to propagate synchronised spike volleys through neuron pools without loss of synchrony.[5,19] The Suri model also shows that synchrony of spike propagation can be enhanced by STDP in a recurrent I&F network.[17] These models are fundamentally related to the hierarchical cortex model originated by Hubel and Wiesel[20,21] in the sense that their structure has an input layer which receives the input spike volleys and then transmits them into the processing layers.

Recent studies of the retinotopic mapping of the visual cortex have revealed that different areas of the visual cortex may be stimulated by different retinal sites,[22] possibly due to the fact that receptive fields lie on a continuum across the surface of the visual cortex.[23] This hypothesis provides a novel way of understanding the relationship between the external stimulus of a moving scene and cortical responses without a hierarchical cortex structure. Rao and Sejnowski presented a biophysical model capable of dealing with moving input stimulus without hierarchical structures.[16] However, the model is focused on simulating detailed neuronal dynamics rather than modelling visual cortical functionalities.

Based on Wörgötter *et al.*'s optical flow algorithm[24] and our previous study,[25,26] we continue to explore event-driven neocortical circuitry for object discovery and associated depth estimation in a real scene. We have been shown that the temporal sequence of edges may be predicted and the prediction accuracy improved through STDP adaptation of interconnecting synapses. Given that in our previous model, a neuron has only one synapse connecting to its preceding neuron, it is natural to ask how the model will perform if a neuron has additional synaptic connections to different neurons on the preceding concentric layer. This paper investigates whether a vision model based on a more extensively-connected I&F neural net, with an additional competition mechanism between the input adaptive synapses, will result in a different (better or worse) performance. The motivation of such a model is twofold. Firstly, a multi-connected neural network is almost certainly more biologically plausible. Secondly, earlier experiments showed that a neuron may predict an incoming edge wrongly while its neighbouring neurons on the same concentric layer are able to make correct predictions on a correlated edge. If they can be caused to "compete" to control the state of their successor neuron to cancel out the spurious prediction, then the model performance should be improved. It is this hypothesis that we set out to test.

The basic component of the model is a simplified I&F model neuron with multiple excitatory input synapses. Each I&F neuron has one synapse activated by an edge sensitive image sensor, to capture moving edges, representing intensity differences, in a dynamic scene. The remaining input synapses receive pulses from neighbouring neurons. Through adaptive synaptic weights and appropriate decay rate of the dynamic membrane threshold, a neuron is able to encode the edge propagation.[25,26] A spiking "window" mechanism exists within which a neuron can fire a spike provided that its membrane capacitor is charged above its dynamic threshold. The window size is determined through competition of input synaptic weights which are adapted by STDP rule.

This paper is organised as follows. The optical flow algorithm, together with the details of the visual plane architecture, are described in Sec. 2. Section 3 provides numerical simulation results of the model performance on a real image sequence. This is followed by a brief discussion and conclusions regarding the model properties, simulation outcomes and their biological links in Sec. 4.

2. Methods

The I&F neurons are distributed on concentric layers of a visual plane, and along axes arranged radially from the optical flow field centre with connections to their nearest-neighbours on the preceding concentric layer (see Fig. 1). Each of the neurons receives input stimuli from its corresponding image sensor. When a point edge is moving with a constant speed toward the visual plane, it passes space locations I, $I+1$, $I+2$, and is thus projected onto neurons number i, $i+1$ and $i+2$ on the visual plane. The distances between the neighbouring neurons, i to $i+1$, and $i+1$ to $i+2$ are arranged in the way that the time of travel of the projected edge from neuron i to $i+1$ is equal to that from neuron $i+1$ to $i+2$.

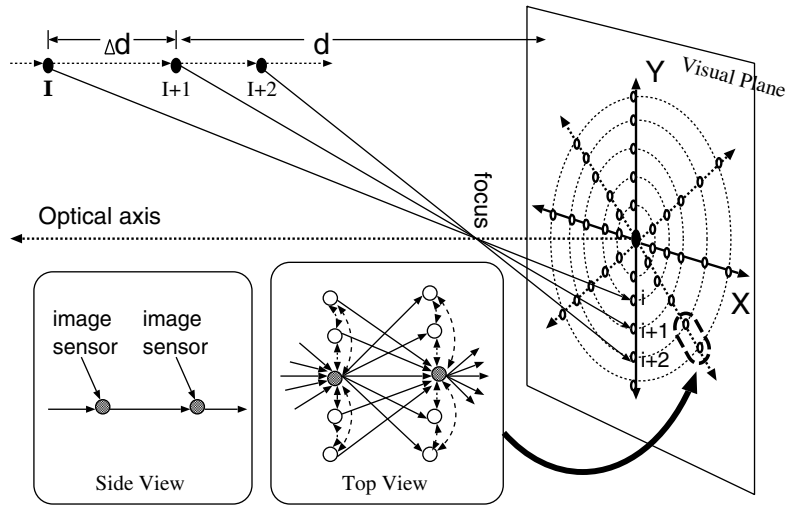


Fig. 1. The architecture of the visual plane model and schematic projection of objects onto the visual plane. A point object (solid circle) starts to move from location I to $I+1$ and then $I+2$, these 3 points are projected onto neurons i, $i+1$, $i+2$, respectively (open circles), on the visual plane. The neurons on the visual plane are arranged on concentric circular layers. Two insets give details to the connecting structure of two neighbouring neurons (dark circles) on one axis.

The derivation of the depth information for a neuron is a straightforward geometrical calculation.[24] In a 3-D space the depth of a point object to the visual plane is given by

$$d = \frac{\Delta d}{k_i - 1} \tag{1}$$

where Δd is the distance moved in the 3-D space from when the edge activates neuron i until the edge reaches neuron $i+1$, $k_i = \frac{r_{i+1}}{r_i}$, r_i is the distance from neuron i to the visual plane center. The only element on the right hand side of this polar coordinate equation that is not fixed for each neuron is Δd. Assuming that the ego/object velocity is constant and known, Δd is then available and depth information can be recovered readily. Once the depth has been recovered, a prediction of the time that the edge will reach $i+2$ can be calculated. The edge is only accepted as genuine if neuron $i+2$ fires within a given time window of this prediction. If neuron $i+2$ fires due to the arrival of an edge outside the time window, then that edge is rejected as a noisy signal.

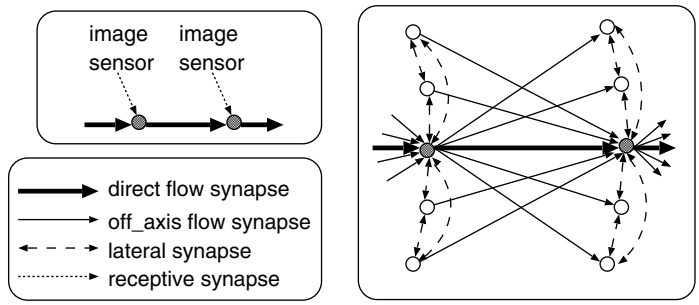
2.1. *Visual plane model*

As it is shown in Fig. 1, every neuron in the visual plane model receives inputs from three different sources: its neighbouring neurons on the preceding concentric layer, the image sensor, and its neighbouring neurons on the same concentric layer. The synapses connecting these different sources are referred to as *flow*, *receptive* and *lateral* synapses, respectively. Furthermore, the *flow synapses* are composed of *direct flow synapses* (i.e., on-axis connections) and *off-axis flow synapses* (see Fig. 2).

This visual plane model has a different architecture from that proposed previously[25,26] in that neurons are now interconnected by two additional forms of synapses, i.e., the *lateral synapses* and *off-axis flow synapses*, whose functions will be described in detail later. In this new model, edges transmitted by the *off-axis flow synapses* are correlated with those transmitted by the corresponding *direct flow synapse* such that the *off-axis flow synapses* can widen the effective "capture area" of edges from the preceding concentric layer and improve rejection of spurious single-point effects. The *lateral synapses* are inhibitory which cancel out the excitatory effects of the relative *off-axis flow synapses* to allow the latter to contribute to adaptation for a better time prediction, without exciting the target neuron.

We assume a homogeneous visual model with notionally identical neurons which are leaky integrators whose subthreshold membrane state is

$$C_{mem}\frac{dv(t)}{dt} = g_{mem}(v_{rest} - v(t)) + \sum_i I_i^{flow}(t) - \sum_j I_j^{lateral}(t) + I^{recept}(t) \tag{2}$$



(a) (b)

Fig. 2. A close view of four types of inter-neural connections in Fig. 1. (a) Side view of the connection structure of two neurons on an axis. (b) Top view of the connection structure of two neurons on an axis.

where C_{mem} and g_{mem} are the membrane capacitance and conductance of the neuron, respectively. The last three terms of right hand side denote the sum of input currents from the *flow*, *lateral* and *receptive synapses*, respectively. In the absence of all input currents, the membrane potential $v(t)$ will drift to its resting value v_{rest}. As the inputs drive the membrane potential above its threshold V_{th}, the neuron produces a short action potential and resets its potential to v_{reset}. In simulation we set $V_{th} > v_{reset} \geq v_{rest}$, and the refractory period $t_{ref} = 0$. When the synaptic time constants of all synapse types are equal, the dynamics of the three input current types are described by

$$\begin{cases} \tau_{syn}\frac{dI_i^{flow}(t)}{dt} = -I_i^{flow}(t) + w_i^{flow}S_i^{flow}(t) \\ \tau_{syn}\frac{dI_j^{lateral}(t)}{dt} = -I_j^{lateral}(t) - w_j^{lateral}S_j^{lateral}(t) \\ \tau_{syn}\frac{dI^{recept}(t)}{dt} = -I^{recept}(t) + w^{recept}S(t) \end{cases} \tag{3}$$

where $S(t)$ is the temporal sequence of spikes to the *receptive synapse*, whose weight is fixed in this study, $S_i^{flow}(t)$ and $S_i^{lateral}(t)$ are those to the *flow* and *lateral synapses* of neuron i. All spikes are modelled as events with a duration of zero. When a neuron fires, it issues a spike simultaneously to its on-axis neighbour and to the nearest off-axis neighbours on the next concentric layer, namely the on-axis neuron and off-axis neurons.

The weights of the *off-axis flow synapses* of neuron j are initialised to a Gaussian function of distance[6] with regard to the weight of the *direct flow synapse* of neuron j. Thus the weight of a synapse

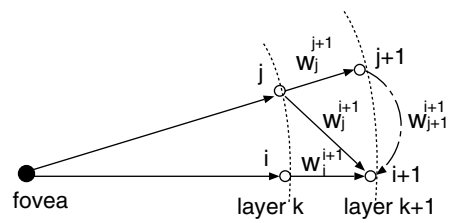


Fig. 3. Inter-neural connections in two neuron axes. Open circles represent neurons. Neuron $j+1$ is neuron j's *on-axis* neuron, $i+1$ is j's *off-axis* neuron, and vice versa. The synapses sourced from a neuron in the untrained visual net have a Gaussian weight distribution, $w_j^{j+1} > w_j^{i+1} \geq 0$. Lateral inhibition (dashed arrow line) with weight w_{j+1}^{i+1} acts to cancel out the activation effect of the *off-axis flow synapse* w_j^{i+1} to neuron $i+1$, i.e., $w_{j+1}^{i+1} = -w_j^{i+1}$.

sourced from neuron j to any interconnnected neurons, say $i+1$, on the succeeding concentric layer has a general description, $w_j^{i+1} = \frac{w_j^{j+1}}{\sqrt{2\pi}\sigma} exp \times \left(-\frac{|d_j^{i+1} - d_j^{j+1}|^2}{2\sigma^2}\right)$ (see Fig. 3). Here d_j^{j+1} and d_j^{i+1} are the Euclidean distances from a firing neuron to its on-axis neuron and to one of its off-axis neurons, respectively. If the Gaussian function has mean $\mu = 0$ and standard deviation $\sigma = \frac{1}{\sqrt{2\pi}}$, then we have

$$\begin{cases} w_j^{j+1} = w_j \\ w_j^{i+1} = w_j exp\Big(-\pi(d_j^{i+1} - d_j^{j+1})^2\Big) \end{cases} \tag{4}$$

From the above equation it is clear that the strengths of the *off-axis flow synapses* of a spiking neuron are defined with respect to its *direct flow synapse*. By using $w_j^{i+1} = -w_{j+1}^{i+1}$, the EPSP state of neuron $i+1$ is actually determined by the *direct flow synaptic* weight w_i^{i+1}. In simulation we set the

receptive synaptic weight as a constant dimensionless value of 20 for simplicity, and the initial *direct flow synaptic* weight $w_j = 15.8$. The model uses a biologically plausible, dynamic threshold for pattern formation.[27,28] The threshold of an I&F neuron, i, is the greater of an exponentially decaying threshold $V_i^{th1}(t) = a_i exp(-\frac{t}{\tau_i^{th}})$ and an intersecting constant threshold $V_i^{th2}(t) = V_T$, i.e.,

$$V_i^{th}(t) = max\left(a_i\ exp\left(-\frac{t}{\tau_i^{th}}\right), V_T\right) \tag{5}$$

where a_i is the membrane threshold value at $t = 0$, the time when the preceding neuron $i-1$ fires, τ_i^{th} is the exponential membrane threshold decay rate and V_T is the constant membrane threshold value. The exponentially decaying membrane threshold represents a relative refractory period[29] to refrain neuron i from firing immediately after the activation of neuron $i-1$. In our model design, the intersection between $a_i exp(-\frac{t}{\tau_i^{th}})$ and V_T, t_i^{pred}, is the time when neuron $i-1$ predicts, through its experience of edge arrival (see eqn.7), that an edge will reach neuron i. Since at that instant, $a_i exp(-\frac{t}{\tau_i^{th}}) = V_T$. We thus have

$$t_i^{pred} = -\tau_i^{th} \ln \frac{V_T}{a_i} \tag{6}$$

If a neuron is de-polarised by its *direct flow synapse* at time t, we would expect it to be further de-polarised by its *receptive synapse* in a time window $(t+t_{pred} \pm \Delta t)$, where Δt is the window length to be determined by the adaptation mechanism. If the depolarisation from the *receptive synapse* is within the expected response window with respect to the spike from the *direct flow synapse*, then the neuron spikes and resets its membrane potential (see Ref. 26 also for details).

2.2. *A vision prediction-confirmation scheme based on multi-connected spiking neurons*

When an edge arrives at an image sensor connecting to the *receptive synapse* of neuron $i-1$, it induces an EPSP on the neuron membrane. The sum of this EPSP, together with the EPSP induced by the *direct flow synapse*, determines the neuron's activity. If the total membrane potential is above the neuron's dynamic threshold, then the neuron issues a spike to inform its succeeding neuron, i, that an edge will arrive.

Meanwhile, neuron $i-1$ is able to estimate the predicted time of travel of the edge, t_i^{pred}, for neuron i. This temporal information is defined by the temporal history of the edge arriving at neuron $i-1$, which can be described with the recursive equation

$$t_i^{pred} = \frac{t_{i-1}^{actual} + C t_{i-1}^{pred}}{C+1} \tag{7}$$

where t_{i-1}^{pred} and t_{i-1}^{actual} are the predicted and actual time of an edge arriving at neuron $i-1$ from $i-2$ respectively. C is the number of "confirmation times" that an edge has been correctly predicted as it propagates along the corresponding axis to neuron $i-1$.

Although in our model neuron the excitatory effect of *off-axis flow synapses* is cancelled out by the inhibitory *lateral synapses*, *off-axis flow synapses* still have influence on the postsynaptic neuron parameters. That means, the predicted time of travel of an edge, now T_i^{pred}, is decided by spikes from both *direct flow synapse* and *off-axis flow synapses*

$$T_i^{pred} = \frac{w_{i-1}^i}{w_{total}} t_i^{pred} + \sum_{offaxis} \frac{w_{offaxis}^i}{w_{total}} t_i^{offaxis,pred} \tag{8}$$

where

- *offaxis* denotes the set of *off-axis flow synapses* which contribute to the modification of the predicted time of travel of an edge
- w_{i-1}^i is the synaptic weight from neuron $i-1$ to i
- $w_{offaxis}^i$ is the synaptic weight from one of the off-axis neurons on the preceding concentric layer to neuron i
- w_{total} is the sum of all flow synaptic weights which contribute to that neuron's spike generation
- $t_i^{offaxis,pred}$ is the predicted time of travel of an edge sourced from one of the off-axis neurons on the preceding layer to i, which equals to the time from the off-axis neuron to its on-axis neuron.

The consequence of Eq. (8) is that a neuron's spiking activity depends upon all the states of the preceding neurons that are connected to it. It is then clear that Eq. (8) is a general case of Eq. (7) by incorporating off-axis synaptic inputs. Combining Eqs. (6)

and (8), it can be seen that all neurons that are "upstream" from a receiving neuron have an influence on its parameters

$$-\tau_i^{th} \ln \frac{V_T}{a_i} = \frac{w_{i-1}^i}{w_{total}} t_i^{pred} + \sum_{offaxis} \frac{w_{offaxis}^i}{w_{total}} t_i^{offaxis,pred} \quad (9)$$

If we substitute t_i^{pred} for the recursive Eq. (7), we can obtain a solution to the I&F neuron's parameter pair, $\{\tau_i^{th}, w_{i-1}^i\}$, and therefore the visual plane model, including the dynamic threshold mechanism of its neurons, can be fulfilled.

2.3. *Adaptation*

In this study, STDP adaptation is used to render the model robust against inaccuracy, and hence to improve its depth-estimation performance. Inaccuracies may be caused by the collective behaviour of edge flow, by non-ideal positioning of image sensors as an array on the visual plane and by mismatches that are bound to result from fabrication tolerances in a silicon process. We use an STDP learning rule, that responds to spike synchrony within a time window.[30] The adaptation of each neuron potentiates or depresses its *flow synapses* by a small amount according to whether the depolarisation from a corresponding *flow synapse* is before or after the weighted prediction time instant. The *receptive synapse* weight is held constant. The adaptation of any individual flow synapse k converging to neuron i is

$$\Delta w_k = \begin{cases} A_+ e^{\frac{t_k^{pred} - T_i^{pred}}{\tau_+}} & \text{if } t_k^{pred} < T_i^{pred} \\ -A_- e^{-\frac{t_k^{pred} - T_i^{pred}}{\tau_-}} & \text{if } t_k^{pred} > T_i^{pred} \\ 0 & \text{if } t_k^{pred} = T_i^{pred} \end{cases} \quad (10)$$

where Δw_k is the synaptic weight change of synapse k, which belongs to a set of synapses converging to neuron i, this set consists of one *direct flow synapse* from neuron $i-1$ to neuron i, and all *off-axis flow synapses* converging to neuron i. $\{A_+, \tau_+\}$ and $\{A_-, \tau_-\}$ are the amplitude and decay constant of potentiation and depression, respectively. Initially, the untrained flow synaptic weights of each neuron are set to random values within the dynamic range of the neuron's spiking window. During learning, the STDP process adapts the effective window through modifying the flow synaptic weights.

3. Simulation Results

The proposed model is tested using a real image sequence. In the experiment, object edges excite the I&F neurons, each of which transmits a spike through one *direct flow synapse* and eight *off-axis flow synapses*. The parameters chosen for this study are shown in Table 1. All of the neuron parameters except the initial amplitudes of learning curves $A_+ = 0.05$ and $A_- = -0.08$ are chosen to be similar to neurobiological exemplars. We choose A_+ and A_- to be larger than those used in an earlier study,[30] as our I&F neurons are considerably sparser. We use Euler integration as the numerical method[32] and the time step is updated with each image frame.

In the scene, a book, a set of pliers, a maze board, a toy pendulum and a badminton racquet are arranged 1205, 1150, 2424, 1102, 2424 millimeters away, respectively, from the initial camera plane. The lighting is primarily daylight, with two 150 W fluorescent lamps fixed on the ceiling and an additional 500 W studio light source to improve the recording contrast. A manually focused progressive scan NTSC CCD camera (model CV-M7 of JAI corporation, Japan) with Pentax Cosmicar 16 mm TV lens

Table 1. Simulation parameters for object recognition and depth inference. The neuron parameters are adopted from Ref. 30.

Scene-retina parameter	Value	Neuron parameter	Value
Resolution	512×512 pixels	Membrane τ	20 ms
Focal parameter	4.025	Synapse τ	5 ms
Axes number	400	Rest potential	−70 mv
Neuron/Axis	74	Reset potential	−70 mv
Retina radius	250 pixel	Constant threshold	−56 mv

is mounted on a small board which is pulled by a small motor toward the objects. Before testing, the white balance of the camera is adjusted empirically. No other special procedure is adopted to adjust the camera's optics. The real scene is recorded in 250 frames with camera motion speed of approximately 1.1 mm per frame. The image frames are saved in bitmap format with a resolution of 1276×1016 and 24 bits of colour. The recording speed is 24 frames per second. The image frames are thereafter preprocessed to fit the final resolution of 512×512 and transferred to a SUN Blade 1500 workstation for analysis.

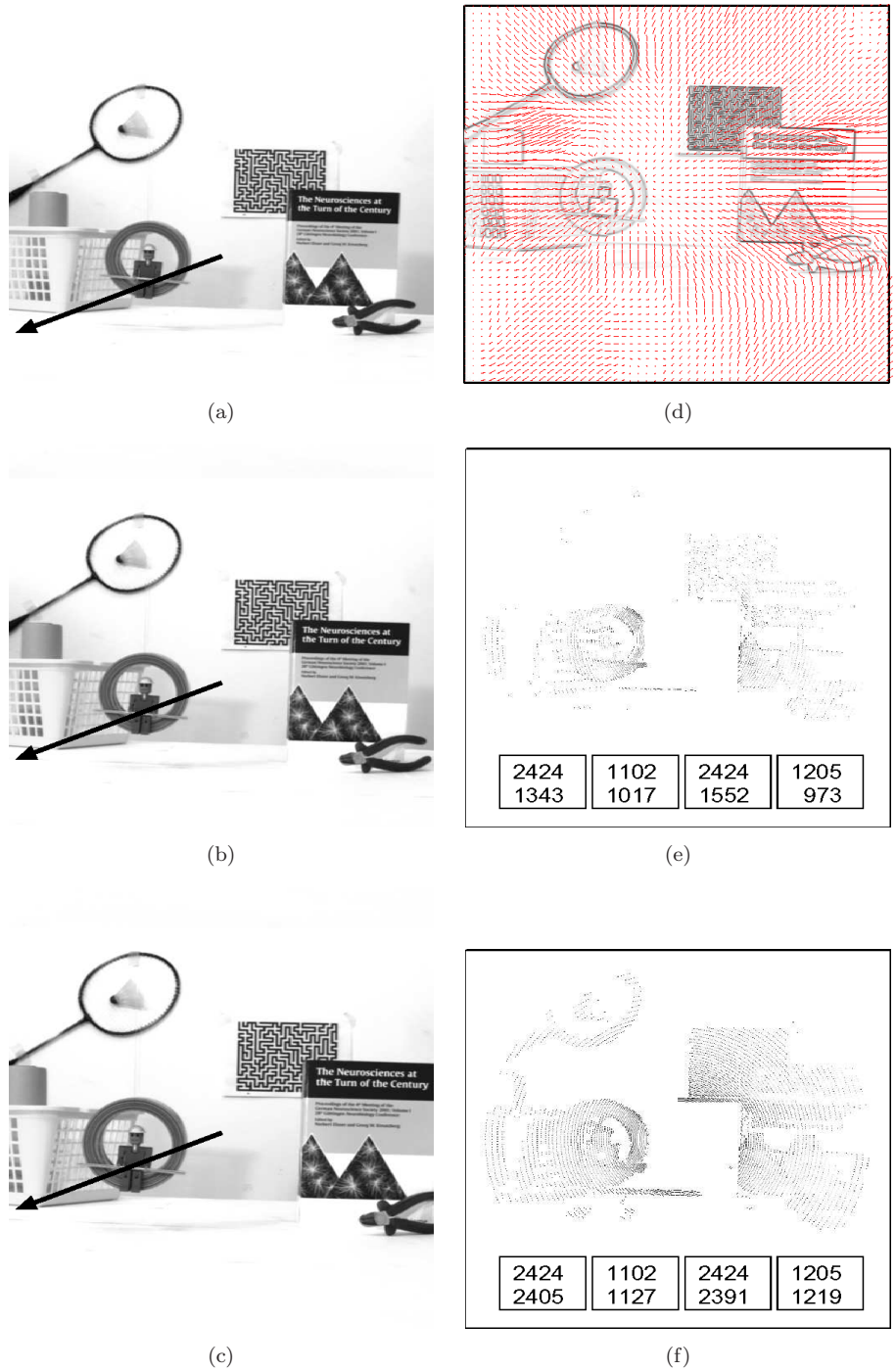
During camera movement, edges flow along the visual plane axes. Initially, in Fig. 4(a), all I&F activations are caused by the receptive synapses only and new events are transmitted, subsequently, to the neighbouring neurons via the *flow synapses*. As movement continues, any neuron activated by both the correlated *receptive* and *flow* synaptic inputs issues a spike. When a new edge is detected its presence at the current neuron is recorded. After this edge has been confirmed the position of the first detecting neuron is added to the pixel map as a black dot, see Figs. 4(e) & (f). The black dots are known as effective pixels.

The algorithm undergoes a "self-organisation" period as it converges to a stable stage, after which the number of the effective pixels remains approximately unchanged (this follows the method in Ref. 26). The stimuli from the *receptive synapses* begin to depolarise the membrane of the neurons after their image sensors detect optical edges. If the EPSPs induced by both the *receptive* and *flow* synapses of a neuron sum within the firing window to produce an activity that exceeds the combined threshold, then the neuron is activated, thus confirming the identification of the associated edge. The source neuron of the confirmed edge is then included in the pixel map and depth information can be calculated by the firing neuron according to Eq. (1). As further movement occurs in the scene and the confirmed edge continues to flow along the retinal axis it will be re-confirmed repeatedly as an ever more reliable representation of the edge.[24] In the pixel map of Fig. 4(f), the geometrical shape of objects is clear, and depth information can be recovered as described in Sec. 2. We use the arithmetic average of the depth values computed by a group of neurons to recover an edge's depth information. This simple method can help to reduce the impact of inevitable noise.

Weight-independent STDP[30] is included to enhance the model's adaptation, as described in Sec. 2.3 above. When weight modifications are independent of the weight value, a bimodal weight distribution emerges from the learning process with weights reaching either the maximum or the minimum hard limits of synaptic strength due to the competition between synapses. This balanced form of bimodal weight distribution helps to stabilise the output rate.[30,31] Figure 5(a) shows a comparison of performance between 3 different vision models, **A**, **B** and **C** in Table 2.

In **Model A**, the *off-axis flow synapses* compete with the *direct flow synapse* to control the postsynaptic neuron parameters. This model therefore takes better account of the collective behaviour of multiple inputs with similar spatiotemporal characteristics. The competition potentiates the more reliable synapses while depressing the less reliable ones according to Eq. (10). It is clear that, after an initial self-organisation period, the performance of the 3 models has become stable. With adaptation added in **Models A** and **B**, it was expected that a significantly improved performance would be measured. **Model B** confirms this by identifying approximately 35% more effective pixels than **Model C**, which is a basic model without adaptation and off-axis connections. **Model A** shows consistently superior performance to that of **Model B** by adding a further 7% improvement in the total number of the effective pixels. For the real image sequence, with the ideal case of 4421 edges projected on the neurons of the visual plane at the beginning, after convergence **Model C** can identify a total number of 2734 (62%) edges, **Model B** identifies 3689 (83%) edges, and **Model A** identifies 3875 (88%) edges.

In order to explain this further performance improvement, the weight distributions of the *direct* and *off-axis flow synapses* are shown in Fig. 5(b)–(d). Without loss of generality, all the weights of both the *direct* and *off-axis flow synapses* are initialised to their maximum possible values so that their expressions can be normalised to 1. The synapses of one axis of neurons, the 230th axis in this study, are used for analysis. After the initial period the weights of the *direct flow synapses* of some neurons are driven to the minimum hard limit while the others stick to the



(a) (d)

(b) (e)

(c) (f)

Fig. 4. The pixel maps of a real scene. (a)–(c) are the snapshots of image frames at the start, 60 and 200 steps respectively. The arrow represents the 230th axis which will be studied. (d) is the still contour of the first frame without running the spike-based algorithm, the short lines signifies the strength and direction of the optical flow in the 2D space. (e) and (f) are the pixel maps. The depth boxes correspond to the badminton racquet, toy pendulum, maze board and book, respectively, from left to right. Upper values in the boxes are the actual original depth of each object, lower values are the estimated depth, both described in millimeters.

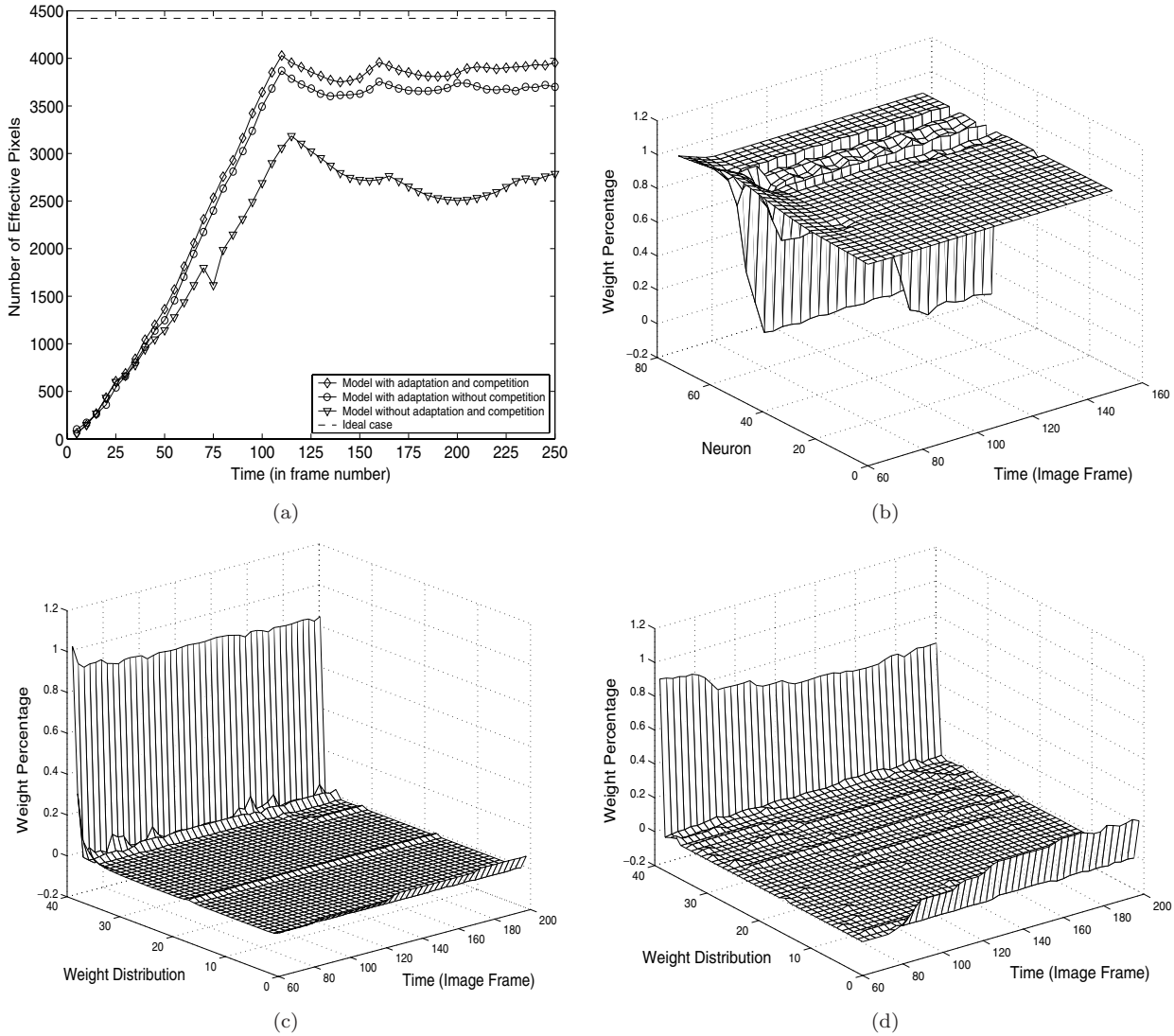


Fig. 5. Characteristics of the model performance and weight distribution. (a) A comparison of the performance of three different vision models. All the models experience a convergence stage and remain approximately stable despite the fluctuation. The dashed line represents the optimal result in which *all* the effective pixels, i.e., the edges projected onto the neurons of the visual plane at the start, are shown in an ideal pixel map. (b) The normalised weight values of the *direct flow synapses* of the 230th axis. They are initialised to their maximum values. As adaptation proceeds, some attain their minimum possible values and remain there. Most of the *direct flow synapses* of the neurons on the first half axis retain their maximum values because no stimulus is detected and no adaptation occurs. (c) The weight distribution of the *off-axis flow synapses* of the 230th axis. The emergence of weights distributed at the minimum hard limit is the result of competition between the *direct* and *off-axis flow synapses*. (d) The weight distribution of the *direct flow synapses* of the same axis. More of these synapses are driven to their minimum limit.

maximum hard limit, see Fig. 5(b). Figure 5(d) shows the weight evolution of the *direct flow synapses* of the 230th axis from the maximum to minimum hard limit in 40 bins. As the adaptation drives weights of maximum hard limit to the minimum limit, the weight distribution gradually forms a stable, balanced bimodal format. The weight distribution of the *off-axis flow synapses* is shown in Fig. 5(c). Compared with Fig. 5(d), far fewer *off-axis flow synapses* are driven to their minimum hard limit after the initial period. This implies that a modest level of competition does occur between the *direct* and

Table 2. Characteristics of three different vision models.

	Model A	Model B	Model C
Off-axis synapses	√	×	×
Adaptation	√	√	×
Input competition	√	×	×

off-axis flow synapses. This allows us to explain the further performance improvement brought about by competition. Fundamentally, more competition leads to a more significant performance improvement because the competition mechanism strengthens synapses with the correct timing prediction, which then dominate the postsynaptic neuron parameters according to Eq. (9). The level of competition depends upon the image context and network structure. As an extreme example, in a scene with many feature curves whose curvature is the same as the curvature of the concentric neural layer, then multiple consecutive neurons on the same concentric layers can and will be activated simultaneously. They will then compete to determine the states of their succeeding neurons through their synapses. On the other hand, if a scene contains few feature curves that fall on or near the image sensor circles, the level of competition will be lower.

Normally, if an edge is correctly predicted and confirmed more than twice by the consecutive neurons on an axis, it can be treated as a genuine feature and hence its origin neuron location can be permanent in the pixel map. In this case the curves in Fig. 5(a) should be flat after the initial self-organisation period.

In order to test the influence of stochastic components in the optical flow field on performance, we have modified the edge inclusion criterion such that an edge is removed from a pixel map if it is judged by a neuron as a spurious edge, even if it has been confirmed by more than two neurons before. This criterion gives rise to the systematic fluctuation in the number of effective pixels in Fig. 5(a). The stochasticity can be caused by various reasons, such as tremble of the camera and disturbance of the scene context. However, our experiment shows that the stochastic fluctuation does not compromise the model's performance, and that, with the adaptation and competition mechanisms, the model is even more robust to such fluctuations.

4. Conclusions

We have introduced a new spike-based, adaptive visual plane algorithm that allows competition between on-axis and off-axis neurons to improve both the rate of acceptance of real edges and rejection of noise. We show that inter-synaptic competition results in a better model performance since the synapses carrying more "reliable" edge signals can win the competition to determine the activities of their postsynaptic neurons. It may therefore be desirable to have more synapses competing with their corresponding *direct flow synapse*. We are led to speculate that there are two factors which influence the level of competition in the model, i.e., the image context and neuron density on a visual plane. Clearly, the characteristics and curvature of the image can not be regarded as adjustable parameters. Neuron density seems, however, to be a major influence on the level of competition. If the neuron density is greater (i.e., there are many more axes), edges will inevitably create more off-axis signals and thus more competition.

When a neuron receives spike events from both the *direct* and *off-axis flow synapses*, weight-independent STDP included in the model introduces inter-synaptic competition to drive weights towards bimodal limits. Thus the initial Gaussian weight function of the synapses sourced from a neuron may not be sustainable.

The form of the weight distribution is also, clearly, affected by the characteristics of the features in the image. In the case of an "archery target" image of concentric circles, for example, competition will be maximised and the off-axis weight distribution tends to a balanced, bimodal form. In the case of an essentially random image, little competition will occur and the off-axis weights will remain at their initial (maximal) values.

In summary, we have presented a novel monocular vision model for object recognition and depth inference. As the underlying algorithm is based upon spike-timing and synchronisation, we have incorporated STDP adaptation to improve performance. In this study, adaptation includes lateral inter-neuron synapses, to allow information that is off the most obvious neuron axes to have a (restricted) influence on the fundamentally-axial computation. To our knowledge, this study is one of the very few works incorporating STDP adaptation, a new synap-

tic plasticity rule, into a vision model which is subsequently applied and measured in a vision case study. If the hard maximum and minimum limits are imposed on the synaptic weights, and weight modification is independent of its value, then STDP-induced competition will drive the weights to form a bimodal distribution. This weight redistribution helps representing the various information in a scene. We have demonstrated in our model that both the adaptation and competition can improve the model performance, although a law of "diminishing returns" inevitably applies as performance is driven towards 100%.

Acknowledgment

This work is supported by EPSRC grant GR/R74567/01. We thank Florentin Wörgötter, Leslie Smith and Markus Dahlem for helpful discussions.

References

1. P. N. Steinmetz, A. Roy, P. J. Fitzgerald, S. S. Hsiao, K. O. Johnson and E. Niebur, Attention modulates synchronized neuronal firing in primate somatosensory cortex, *Nature* **404** (2000) 187–189.
2. E. Salinas and R. Romo, Neurobiology: A chorus line, *Nature* **404** (2000) 131–133.
3. W. H. Dobelle, M. G. Mladejowsky and J. P. Girvin, Artificial vision for the blind: Electrical stimulation of visual cortex offers hope for a functional prothesis, *Science* **183** (1974) 440–444.
4. M. Abeles, *Corticonics* (Cambridge University Press, Cambridge, 1991).
5. M. Diesmann, M. O. Gewaltig and A. Aertsen, Stable propagation of synchronous spiking in cortical neural networks, *Nature* **402** (1999) 529–533.
6. C. Mehring, U. Hehl, M. Kubo, M. Diesmann and A. Aertsen, Activity dynamics and propagation of synchronous spiking in locally connected random networks, *Biol. Cybern.* **88** (2003) 395–408.
7. Y. Ikegaya, G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster and R. Yuste, Synfire chains and cortical songs: Temporal modules of cortical activity, *Science* **304** (2004) 559–564.
8. R. Eckhorn, A. Gail, A. Bruns, A. Gabriel, B. Al-Shaikhli and M. Saam, Neural mechanisms of visual associative processing, *Acta Neurobiol. Exp.* **64** (2004) 239–252.
9. H. Markram, J. Lubke, M. Frotscher and B. Sakmann, Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs, *Science* **275** (1997) 213–215.
10. G. Q. Bi and M. M. Poo, Activity-induced synaptic modifications in hippocampal culture, dependence on spike timing, synaptic strength and cell type, *Journal of Neuroscience* **18** (1998) 10464–10472.
11. D. Debanne, B. H. Gahwiler and S. M. Thompson, Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures, *Journal of Physiology* **507** (1998) 237–247.
12. L. I. Zhang, H. W. Tao, C. E. Holt, W. A. Harris and M. M. Poo, A critical window for cooperation and competition among developing retinotectal synapses, *Nature* **395** (1998) 37–44.
13. L. F. Abbott and K. I. Blum, Functional significance of long-term potentiation for sequence learning and prediction, *Cereb. cortex* **6** (1996) 406–416.
14. W. Gerstner and L. F. Abbott, Learning navigational maps through potentiation and modulation of hippocampal place cells, *J. Comput. Neurosci.* **4** (1997) 79–94.
15. D. A. August and W. B. Levy, Temporal sequence compression by an integrate-and-fire model of hippocampal area CA3, *J. Comput. Neurosci.* **6** (1999) 71–90.
16. R. P. N. Rao and T. J. Sejnowski, Predictive sequence learning in recurrent neocortical circuits, *Advances in neural information processing systems* **12** (MIT press, 2000) 164–170.
17. R. E. Suri and T. J. Sejnowski, Spiking propagation synchronized by temporal asymmetric Hebbian learning, *Biol. Cybern.* **87** (2002) 440–445.
18. A. P. Shon, R. P. N. Rao and T. J. Sejnowski, Motion detection and prediction through spike-timing dependent plasticity, *Network: Comput. Neural Syst.* **15** (2004) 179–198.
19. M. Herrmann, J. A. Hertz and A. Prugel-Bennet, Analysis of synfire chains, *Network: Comput. Neural Syst.* **6** (1995) 403–414.
20. D. H. Hubel and T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, *J. Physiol.* **195** (1968) 215–243.
21. D. H. Hubel and T. N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol.* **160** (1962) 106–154.
22. J. S. Sunness, T. Liu and S. Yantis, Retinotopic mapping of the visual cortex using functional magnetic resonance imaging in a patient with central scotomas from atrophic macular degeneration, *Ophthalmology* **111** (2004) 1595–1598.
23. D. L. Ringach, Mapping receptive fields in primary visual cortex, *J. Physiol.* **558** (2004) 717–728.
24. F. Wörgötter, A. Cozzi and V. Gerdes, A parallel noise-robust algorithm to recover depth information from radial flow fields, *Neural Computation* **11** (1999) 381–416.
25. Z. J. Yang and A. F. Murray, A biologically plausible neuromorphic system for object recognition and depth analysis, in *Proc. Europ. Symp. Art. Neural Networks* (Brugge, Belgium, 2004), pp. 157–162.

26. Z. J. Yang, A. F. Murray, F. Wörgötter, K. L. Cameron and V. Boonsobhak, A neuromorphic depth-from-motion vision model with STDP adaptation, *IEEE Trans. Neural Networks* **17**(2) (March, 2006), in press.
27. D. H. Perkel, A computer program for simulating a network of interacting neurons. I. Organization and physiological assumptions, *Comput. Biomed. Res.* **9** (1976) 31–43.
28. R. Borisyuk, Oscillatory activity in the neural networks of spiking elements, *BioSystems* **67** (2002) 3–16.
29. R. Sarpeshkar, J. Kramer, G. Indiveri and C. Koch, Analog VLSI architectures for motion processing: From fundamental limits to system applications, *Proceedings IEEE* **84** (1996) 969–987.
30. S. Song, K. D. Miller and L. F. Abbott, Competitive Hebbian learning through spike-timing-dependent synaptic plasticity, *Nature Neuroscience* **3**(9) (2000) 919–926.
31. R. Kempter, W. Gerstner and J. L. van Hemmen, Intrinsic stabilization of output rates by spike-based Hebbian learning, *Neural Computation* **13** (2001) 2709–2741.
32. D. Hansel, G. Mato, C. Meunier and L. Neltner, On numerical simulations of integrate-and-fire neural networks, *Neural Computation* **10** (1998) 467–483.