Vision Coincidence Detection with STDP Adaptation for Object Recognition and Depth analysis

Zhijun Yang, Alan F. Murray

IMNS, School of Engineering and Electronics
The University of Edinburgh, Edinburgh, EH9 3JL, UK
Emails: {Zhijun.Yang, Alan.Murray}@ee.ed.ac.uk

Abstract

A cognitive vision neuronal network based on leaky integrate-and-fire (LIF) neurons is proposed for object recognition and depth analysis. In this network every LIF neuron is able to capture the edge flowing through it and record the temporal information. If the neuron issues a spike, the temporal information will be encoded by the time constant of the spike potential and transferred to its successor neuron through synapses. The successor neuron, on reception of the spike, will check whether that edge arrives at its sensor. In the case that both events synchronise the successor neuron will fire to confirm the correct edge propagation. Meanwhile, in the process the spike-timing-dependent plasticity (STDP) is employed to achieve the suitable synapse efficacies to reject spurious edge propagation. On recognition of the effective CMOS realisation of LIF neuron, our model aims to be a biologically inspired neuromorphic system amenable to aVLSI implementation.

1 Introduction

Early cognitive vision is one of the most important perceptive functions in the mammalian cerebral cortex. In cortical computation object recognition and the extraction of distance between observer and object, i.e., the depth information, are closely related intelligent activities. Although one of the most popular approaches for depth perception from motion parallax is to use binocular stereopsis [1][2][3], it is a common sense that people can still perceive the depth information with even the single eyes. Therefore, some more intriguing mechanism must be underlying the early cognitive vision functionalities with which the single eyes, instead of the cooperation of two or more eyes, are good enough to receive sufficient information for cortical computing. During information processing in the visual cortex, many bioelectrical signals are exchanged between cortical neurons, across their synaptic interconnections. When modelling visual cortex processing a dynamic scene can be modelled as an optical flow field, which can, in turn, be mapped on to a neuronal network. A possible, simple network structure has neurons placed along axes arranged radially from the optical flow field centre with only nearest-neighbour, on-axis connections. The characteristics of a dynamic scene can thus be reconstructed by local computation in neurons based upon the optical flow field [4][5]. On the other hand, recent advances in Neuroscience suggest that spike-timing-dependent plasticity (STDP) could be the intrinsic mechanism underlying synaptic plasticity in cerebral visual cortex region to coordinate pre- and postsynaptic neuronal activities within a critical time window to conduct information storage and processing [6][7][8]. This discovery provides an opportunity to develop a biologically plausible cognitive model for aVLSI implementation.

In this work we have modelled a large-scale leaky integrate-and-fire (LIF) neuronal network in the same radial network arrangement. The model uses edge-sensitive pixels to capture the position and speed of movement of features in a moving scene. Subsequently, object identification is achieved and depth information is recovered. In this example, we have restricted movement to a straight line, with a constant relative speed between a moving cognitive model and the static scene. The basic component of our network is a simplified LIF neuron with two excitatory afferent synapses and one efferent synapse. In operation a neuron in the network encodes the edge propagation by using the decay rate of dynamic membrane threshold and excitatory postsynaptic potential (EPSP) to associate the neuron's response with synchronised or unsynchronised inputs of its two afferent synapses. We will show that this LIF neuron component is fundamentally a leakage coincidence detector with a window mechanism. The window size is determined by the afferent synapse efficacy which is adapted by STDP rule. By incorporating adaptation this novel LIF-based network aims to be a neuromorphic sys-
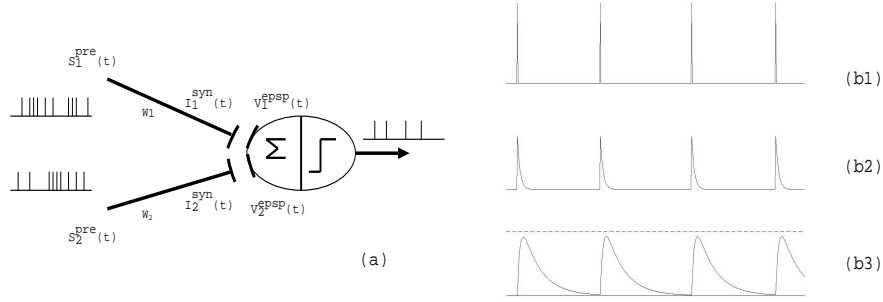
Figure 1: A schematic circuit diagram and simulation pulse stream propogation in one afferent pathway. (a) Two synapses converge to an IF neuron. (b1) Poisson δ-function current pulse stream propogates from the dendritic input terminals $S_i^{pre}(t)$ to (b2) dendritic output terminals $I_i^{syn}(t)$, and (b3) induce the EPSPs $V_i^{epsp}(t)$, where dashed line represents the neuron threshold.

tem capable of emulating some functionalities of retina and visual cortex.

2 Model

In order to achieve a balance between the image resolution and simulation time (hardware cost), we choose empirically 400 axes radially arranged from the optical flow field centre. The number of LIF neurons allocated to every axis is the same and depends on the resolution of edge-sensitive pixel array. For a 512×512 pixel array 48 neurons are aligned on every axis with the hyperbolically increased distance from the flow field centre.

2.1 Single neuron and coincidence of inputs

Each neuron in the model has two identical excitatory synapses configured to accept pulsed inputs. The synapses connecting the previous neighbouring neuron and the pixel sensor are referred to as the *flow synapse* and the *receptive synapse*, respectively. For analytical simplicity we make two assumptions. (i) No delay is associated with input pulse propagation within the synapses. (ii) The IF neuron fires only in response to EPSPs induced by a pair of synaptic inputs from the two different synapses. No individual input pulse or consecutive pulses from a same synapse can activate the neuron. Fig.1 (a) is a schematic diagram of such LIF neuron, and fig.1$(b1) \sim (b3)$ are the different stages as a pulse stream propagates from a previous neuron to current one.

The sub-threshold membrane status when such a pair is presented is:-

$$
\begin{cases}
\tau_{mem}\frac{dv(t)}{dt} = v_m - v(t) + r_m \sum_i I_i^{syn} \\
\tau_{syn}\frac{dI_i^{syn}(t)}{dt} = -I_i^{syn}(t) + w_i \sum_j \delta(t - t_j)
\end{cases}
\tag{1}
$$

where τ_{mem} and τ_{syn} are the membrane and synaptic time constants respectively, r_m is the membrane resistance, w_i is the weight of a synapse, $\delta(t - t_j)$ is a δ-function dendritic input at time t_j, $I_i^{syn}(t)$ is the current pulse at the post-dendrite terminal, $v(t)$ is the EPSP induced by $I_i^{syn}(t)$, in our case $i = 1, 2$.

The firing threshold of an IF neuron is a combination of an exponentially decaying threshold $V(t) = ae^{-\frac{t}{\tau_{th}}}$ and an intersecting constant threshold $V(t) = V_T$. The use of such a dynamic threshold for pattern formation is not biologically implausible [9][10]. The threshold equation is then:-

$$
V_{th} =
\begin{cases}
ae^{-\frac{t}{\tau_{th}}} & \text{if } 0 \le t \le -\tau_{th}\ln\frac{V_T}{a} \\
V_T & \text{if } -\tau_{th}\ln\frac{V_T}{a} < t
\end{cases}
\tag{2}
$$

where a is the threshold value at $t = 0$, the time when the previous neighbouring neuron fires, τ_{th} is the exponential threshold decay rate, V_T is the constant threshold value. $-\tau_{th}\ln\frac{V_T}{a}$ is a time instant when the predicted feature arrives at the neuron. If the neuron is depolarised by its *flow synapse* at time t, we would expect it to be further depolarised by its *receptive synapse* in a time range of $[t - \tau_{th}\ln\frac{V_T}{a} - \Delta t, t - \tau_{th}\ln\frac{V_T}{a} + \Delta t]$, where Δt is the response window to be determined by the adaptation mechanism. If the depolarisation from the *receptive synapse* is within the expected response window then the neuron issues a spike and its membrane potential is reset.
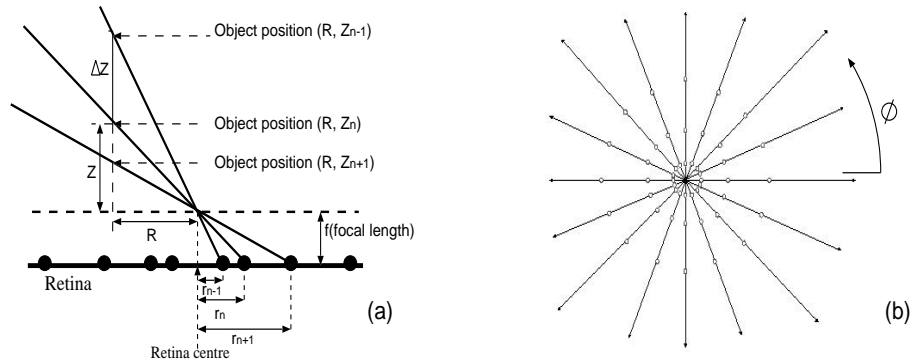
Figure 2: The geometrical projection of a point on to a retina through the light rays (side view), the dark circles are neurons (a), and the schematic layout of a retina, the small circles are neurons (top view) (b).

2.2 Adaptation

We aim to make this neuromorphic reformulation of the depth-recovery algorithm adaptive, and thus robust against inevitable inaccuracy in circuits and fabrication processes. We use the TAH learning rule, which responds to spike synchrony, or near-synchrony. The adaptation of each neuron in the artificial retina potentiates or depresses the *receptive synapse* weight by a small amount according to whether the depolarisation from the *receptive synapse* is before or after the predicted time instant. The *flow synapse* weight is held constant for the model simplification. The adaptation scheme is:-

$$
f(w) = \begin{cases} A_+ e^{\frac{t-t_{pred}}{\tau_+}} & \text{if } t < t_{pred} \\ -A_- e^{-\frac{t-t_{pred}}{\tau_-}} & \text{if } t > t_{pred} \\ 0 & \text{if } t = t_{pred} \end{cases} \quad (3)
$$

where f(w) is the synaptic weight change, $\{A_+, \tau_+\}$ and $\{A_-, \tau_-\}$ are the initial amplitude and decay constant of potentiation and depression, respectively and t_{pred} is the predicted time instant. In an initial stage, the untrained weight of each neuron is set to a value corresponding to a very narrow window size, suitable for imperfection-free circuitry only. This idealised set-point corresponds to the origin in the TAH learning curve. As the neural and pixel circuits are not, of course, actually ideal, the TAH process will modify the window size.

2.3 Network - the retinotopic application

Recently Wörgötter et al. [4] devised a feature-based approach to the recovery of depth information from radial flow fields. The two key steps in the algorithm are: (1) calculation of the object coordinates in the environment; (2) a predictive mechanism to compare the predicted and actual time of arrival of an edge feature for pattern recognition and to improve noise rejection. A temporal tolerance or "window" is associated with neurons, which are arranged radially in the focal plane. At an individual pixel/neuron, a previously calculated edge is accepted if the predicted and actual arrival time of that edge feature are within the neuron's tolerance window. Otherwise the edge coordinates are rejected as spurious. We implement this mechanism as a spike-based scheme. Physiologically, we assume that the scenes will be mapped onto an optical flow field by the retina, and the optical flow information will be further processed by the cerebral visual cortex. By adopting the aforementioned model we can artificially encode the vision activities with a large-scale neuronal network and, further design an artificial vision device through the phenomenological model, which is also amenable to aVLSI implementation.

We consider a dynamic scene in which an observer moves along the optical axis towards the target objects at a constant speed. The geometrical light projection of such motion can be displayed by fig.2. The retina has pixel sensor neurons aligned along the radially arranged axes. The neurons on an axis are separated by a hyperbolically increasing distance $d_n = sinh(\frac{n}{\alpha})$, where n is the neuron order and α is a constant, chosen empirically according to be suitable for the pixel resolution. For a suitable α value, higher sensor density can be achieved by interpolating additional LIF neurons between the already arranged ones provided that different neurons have different corresponding sensors. The distance from each neuron to the retina centre is a hyperbolic function of its order referred to as $r_n = g(n)$.

The derivation of the depth information for a neuron is conceptually straightforward [4]. In cylindrical polar coordinates we have,

Scene-Retina Parameter	Value	Neuron Parameter	Value
Sequence	600 frames	Membrane τ	20ms
Resolution	512x512 pixel	Synapse τ	5ms
Step	1cm/frame	Adapt curve τ	20ms
Axes number	400	Rest potential	-70mv
Neuron/Axis	48	Reset potential	-70mv
Retina radius	250 pixel	Constant threshold	-56mv

Table 1: Simulation parameters for object recognition and depth recovery

$$\begin{pmatrix} R \\ Z \end{pmatrix}_\phi = \Delta Z \begin{pmatrix} \frac{r_n}{f(k_n - 1)} \\ \frac{1}{k_n - 1} \end{pmatrix}_\phi \qquad (4)$$

where R, Z, ϕ are the cylindrical polar coordinate components determining the actual position of an object in a $3-$D space, $k_n = \frac{r_n}{r_{n-1}}$, f is the focal length, ΔZ is the previous moving distance with projection from neuron $n - 1$ to n. It is obvious that, except ΔZ, the right hand side of this polar coordinate equation is fixed for each neuron. Assuming that the ego/object motion speed is constant and known, then ΔZ is available through a step counter to compute the time a light ray needs to travel between the corresponding neurons, the depth information is hence readily recovered.

3 Simulation results

This spike-based algorithm is tested with an artificial environment in which there are three objects (cylinder, cone, sphere) located at different distances ($15m$, $11m$, $7m$) in front of a white background $16m$ away. The observer is 1.6m high and moves towards the scene with a constant speed. In this experiment, object edges are used as the stimulus to the IF neurons and depth maps are defined only at the edges of the objects. The test parameters are shown in table 1. All the neuron parameters except the initial amplitudes of learning curves $A_+ = 0.2$ and $A_- = -0.3$ are chosen with guidance from neurobiology. We choose A_+ and A_- to be larger than those ones used in an earlier study [11], as our pixel stimuli are much sparser. We use Euler integration as the numerical method and the update step is increased with each image frame.

The depth map is shown in fig.3. The three objects are identified clearly by the edge features flowing along the retina axis. At the outset, (fig.3a), all features that fall on the pixel sensors are new and consequently stimulate the neighbour neurons via their *flow synapses*. Few edges appear on the depth map at this stage.

After this initial settling period, as the actual light rays move from the original pixel sensors to the next ones, the stimuli from the *receptive synapses* begin to depolarise the neurons' membrane.

If the EPSPs from the two synapses to a neuron sum within the firing window to produce an activity that exceeds than the combined threshold, then the neuron is activated to confirm the identification of the associated edge feature. The confirmed feature is then included in the depth map. Meanwhile its depth information can be calculated by the firing neuron according to equation (4). As further movement occurs in the real scene and the confirmed feature continues to flow along the retinal axis it can be confirmed repeatedly to become an ever more reliable representation of the edge that it represents. In the depth map of fig.3f, the geometrical shape of objects is clear and depth information can be recovered as described above.

All the depth maps corresponding to the different image frames are boolean images in the sense that the simulated pixel sensors have only boolean outputs. Any neuron in the model can estimate the depth information provided that it has "seen" an edge and, therefore issue a spike. We use the neurons on the $210th$ axis in our model as an example to show the object depth recovery according to equation (4). In this example only one edge feature, the cross line of left side wall and back wall, will propagate along this axis. The initial location of this edge feature is 16 metres away from the camera. The depth information recovered by the neurons on the $210th$ axis is shown in fig.4(a). It is clear that the estimated depth matches well with the actual depth after an initial self-organising process. The effect of STDP mechanism can be shown explicitly through the distribution of *receptive synapse* efficacies before and after online adaptation. Initially we set all *receptive synapse* efficacies randomly between their maximum and minimum values. The normalised random *receptive synapse* efficacies lead to a uniform distribution. After online STDP adaptation the initial uniform distribution converges towards the bimodal distribution as shown in fig.4(b) and (c).
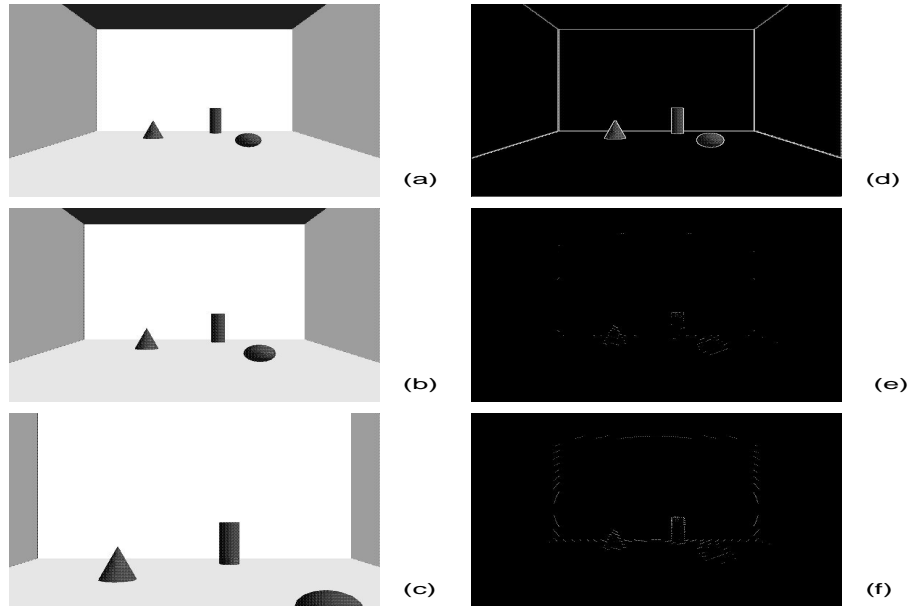
Figure 3: The depth maps of a hallway scene. $(a) \sim (c)$ are the snapshots of image frames at the start, 175 and 600 steps respectively. (d) is the still contour of the first frame without running the spike-based algorithm. (e) and (f) are the depth maps calculated from the start frame till the corresponding frame shown on the left with the algorithm.
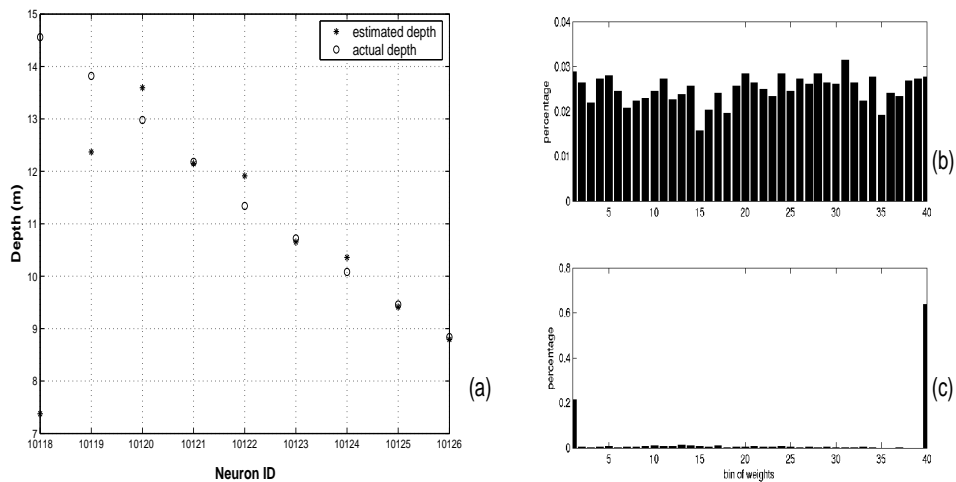

Figure 4: The characteristics of the cognitive vision model operating on the hallway scene. (a) a comparison between the depth information estimated by the neurons on axis $210th$ and the corresponding actual depth; the increasing order of neuron ID also represents the time axis. (b) the initial uniform distribution of *receptive synapse* efficacies, and (c) the bimodal distribution after STDP adaptation.

4 Conclusion

In this paper we have explored a neuromorphic cognitive vision model consisting of LIF neurons for object recognition and depth analysis. The postsynaptic LIF neuron's activity is dependent on the correlation of two synaptic inputs. We demonstrate that the intrinsic neuronal circuit parameters will determine a temporal tolerance window within which the effect of correlated inputs can result in different postsynaptic neuron response. Afferent synchrony can thus be detected between the intercellular connections. The window size is adjustable through synaptic and membrane time constants which are directly related with the circuit parameters, hence render the model applicable in aVLSI application. Our study therefore provides an approach to the realisation of synaptic weight adaptation based on STDP, which is a possible adaptation mechanism underlying the visual cortical activities.

Acknowledgments

References

[1] Qian, N., *Binocular disparity and the perception of depth*, Neuron **18**:359-368, 1997.

[2] Zhou, C. & Tao, H., *Dynamic depth recovery from unsynchronized video streams*, Proc. IEEE Conf. on CVPR, 2003.

[3] Barron, J.L., Fleet, D. & Beauchemin, S., *Performance of optical flow techniques*, Int. J. Comp. Vis. **12**:43-77, 1994.

[4] Wörgötter, F., Cozzi, A. & Gerdes, V., *A parallel noise-robust algorithm to recover depth information from radial flow fields*, Neural Computation **11**:381-416, 1999.

[5] Dayan, P. & Abbott, L.F., *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, The MIT Press, 2001.

[6] Markram, H., Lubke, J., Frotscher M. & Sakmann B., *Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs*, Science **275**:213-215, 1997.

[7] Bi, G.Q. & Poo, M.M., *Activity-induced synaptic modifications in hippocampal culture, dependence on spike timing, synaptic strength and cell type*, Journal of Neuroscience **18**:10464-10472, 1998.

[8] Zhang, L.I., Tao, H.W., Holt, C.E., Harris, W.A. & Poo, M.M., *A critical window for cooperation and competition among developing retinotectal synapses*, Nature **395**:37-44, 1998.

[9] Perkel, D.H., *A computer program for simulating a network of interacting neurons. I. Organization and physiological assumptions*, Comput. Biomed. Res. **9**:31-43, 1976.

[10] Borisyuk, R., *Oscillatory activity in the neural networks of spiking elements*, BioSystems **67**:3-16, 2002.

[11] Song, S., Miller, K.D. & Abbott, L.F., *Competitive Hebbian learning through spike-timing-dependent synaptic plasticity*, Nature Neuroscience **3(9)**:919-926, 2000.