A Neuromorphic Depth-From-Motion Vision Model With STDP Adaptation

Zhijun Yang, Alan Murray, *Senior Member, IEEE*, Florentin Wörgötter, Katherine Cameron, *Student Member, IEEE*, and Vasin Boonsobhak, *Student Member, IEEE*

***Abstract*—We propose a simplified depth-from-motion vision model based on leaky integrate-and-fire (LIF) neurons for edge detection and two-dimensional depth recovery. In the model, every LIF neuron is able to detect the irradiance edges passing through its receptive field in an optical flow field, and respond to the detection by firing a spike when the neuron's firing criterion is satisfied. If a neuron fires a spike, the time-of-travel of the spike-associated edge is transferred as the prediction information to the next synapse-linked neuron to determine its state. Correlations between input spikes and their timing thus encode depth in the visual field. The adaptation of synapses mediated by spike-timing-dependent plasticity is used to improve the algorithm's robustness against inaccuracy caused by spurious edge propagation. The algorithm is characterized on both artificial and real image sequences. The implementation of the algorithm in analog very large scale integrated (aVLSI) circuitry is also discussed.**

***Index Terms*—Integrate-and-fire neurons, monocular depth recovery, neuromorphic vision model, synaptic plasticity.**

I. Introduction

DISPARATE elements of the primate visual pathway, comprising, largely, the retina, thalamic relay neuronal network and visual cortex, cooperate to give rise to perceptive mapping of external stimuli. These perceptual properties have been attributed to complex brain activity at the cellular and network level [1]–[4]. In addition, experimental evidence on the activity-dependent firing behaviors of mammalian hippocampal pyramidal cells *in vitro* [5]–[7], tectal cells of a type of frog's visual system [8] and cortical neurons of rats [9] indicate that hippocampal, retinotectal and cortical neurons encode and process biological information by coordinating postsynaptic spikes with presynaptic inputs within a critical time window. The pre- and post-synaptic spike timing can thus decide the modification of input synaptic strengths, now called spike-timing-dependent plasticity (STDP). STDP suggests that long-term strengthening of synapses occurs if presynaptic action potentials precede postsynaptic firing by no more than about 50 ms, and long-term weakening of synapses occurs if presynaptic action potentials follow postsynaptic spikes in the same time scale. This temporal correlation-based adaptation is speculated as a major learning rule in the nervous system (for reviews see [10], [11]).

Extensive theoretical and computational studies, based on the experimental results, have characterized the behavior of single neurons [12]–[15] or neuronal populations [16]–[20] in response to sensory stimuli. Many of these models are based on the widely used leaky integrate-and-fire (LIF) neuron model which, despite its simplicity, is capable of emulating some of the complex behavior of cortical spiking neurons [21]–[23]. Meanwhile, the effects of synaptic changes by STDP have become well established [24]–[26]. There are, however, few studies in which LIF networks with STDP adaptation mimic the biological perceptual functions for practical applications. Considering that vision processing is one of the most important functions of the primate brain, we explore in this paper one possible approach that mimics the early visual system.

Animal and human vision is able to detect moving irradiance edges and infer their associated depth in the image. It is well established that neurons in the primary visual cortex (area V1), a major part of the early visual cortex of primates, are able to detect edges in the visual field through their receptive fields [27], [28]. These neurons in V1 are retinotopically organized, following a roughly polar coordinate system [29]–[31]. Recent experiments also demonstrate that visual cortical neurons in V1 have the computational power to encode information about the three-dimensional (3-D) position of a stimulus in space [32], [33]. The outputs of V1 neurons contribute to the stereoscopic depth perception in the higher visual cortical areas [34], [35]. Although binocular stereopsis [36]–[38] is the most obvious way to recover depth, we can infer depth information using a single eye. In a dynamic environment, motion is a crucial cue in monocular depth recovery.

In applications of machine vision with motion, the scene is often modeled as an optical flow field, which may, in turn, be mapped onto a neuronal network [39]. One such simple network structure has neurons placed along axes arranged radially to the optical flow field center with only nearest-neighbor, on axis connections [40]. We aim to mimic the early visual processing in V1 with a simplified, continuous time vision model. The model uses a large-scale LIF neuronal network for edge detection and associated depth inference in a dynamic scene. The placement of LIF neurons in a polar coordinate arrangement bears some resemblance to the organization of V1 neurons. The characteristics of a dynamic scene are then reconstructed through local spike-timing computations along these radii.

Manuscript received July 13, 2004; revised August 18, 2005. This work is supported by EPSRC grant GR/R74567/01. This work is part of a joint EPSRC project involving the Universities of Edinburgh, Oxford, and Stirling.

Z. Yang is with the Department of Computer Science, Nanjing Normal University, Nanjing 210097, China (e-mail: zhijun.yang@ed.ac.uk).

A. Murray, K. Cameron, and V. Boonsobhak are with the School of Engineering and Electronics, The University of Edinburgh, Edinburgh EH9 3JL, U.K.

F. Wörgötter is with the Bernstein Centre of Computational Neuroscience, Göttingen 37099, Germany.

Digital Object Identifier 10.1109/TNN.2006.871711

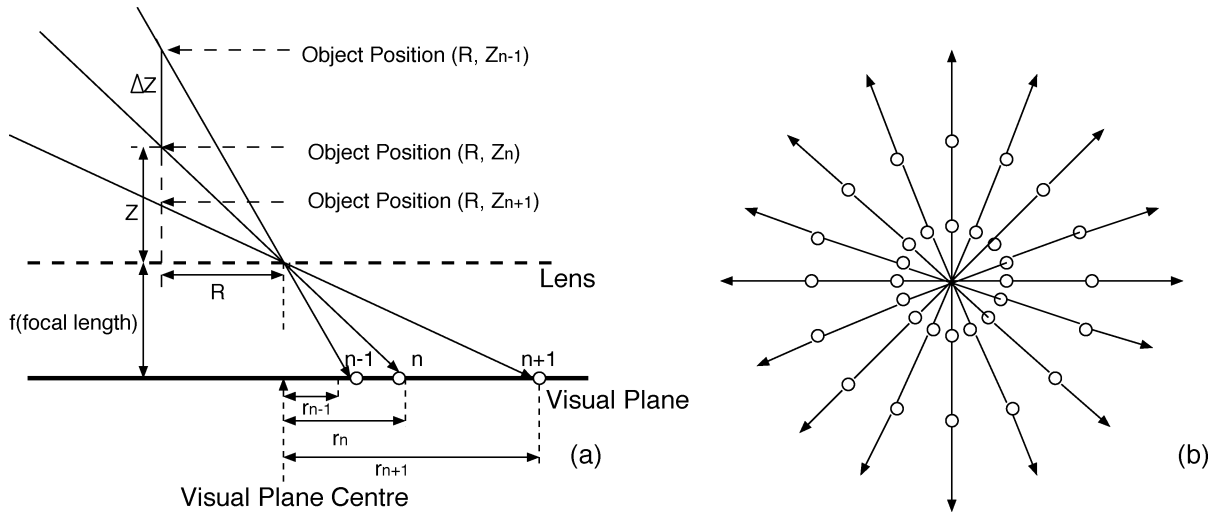

Fig. 1. Geometrical projection of a point object onto a visual plane of the proposed neuronal model via the light rays. (a) Side view, a 3-D point is described by the polar coordinate (R, Z), R, Z are the distance from the point to the optical axis and lens, respectively. Z is the point depth which is usually lost in the 3-D to 2-D mapping. (b) Top view, a schematic layout of the visual plane, the small circles are neurons.

For computational simplicity, a neuron has only one synapse connecting with the photoreceptor adapted using an STDP rule. The advantage of employing STDP in this study is twofold. First, spike-timing related plasticity is a possible mechanism for pairing pre- and postsynaptic activities in a population of pulse-coded spiking neurons. In our algorithm, the postsynaptic firing depends on the timing of two excitatory inputs. Therefore, we will show that STDP is also able to optimize this form of vision model by adapting synapses online. As a result, neurons can process genuine edges while rejecting spurious edges that result from noise or the effects of an ill-conditioned optical flow field. Second, STDP can be implemented in analog very large scale integrated (aVLSI) circuits [41]–[43], and is capable of compensating for the inevitable transistor mismatch in aVLSI chip fabrication [44].

In summary, our neuron is fundamentally a simple detector of coincidence within a time window. The window size is determined by the neuron's dynamic threshold and input synaptic strength, which is adapted by the STDP rule. STDP adaptation improves the performance of this novel LIF-based network. It is also a neuromorphic system that mimics some aspects of the early visual pathway, especially the area V1 properties, in a new and interesting form. The experimental results will inform the development of aVLSI circuitry in subsequent works.

II. Monocular Depth Inference

Recently, Wörgötter *et al.* [40] devised a feature-based approach to monocular depth recovery from radial flow fields. The two key steps in their algorithm are: 1) calculation of the object coordinates in the environment and 2) a predictive mechanism to compare the predicted and actual time-of-arrival of an edge to improve noise rejection.

In their algorithm, the ego or object motion is restricted to a straight line along the optical axis with constant velocity. Neurons are aligned along the radial axes from the visual plane center to the periphery (see Fig. 1 for the geometric light ray projection and neuron layout). Upon the projection of a new edge at a neuron, say $n-1$, neuron $n-1$ simply passes the gray level of that new edge to its succeeding neuron n, where it is saved. After some time, when the edge arrives at neuron n, the neuron compares its gray level with the saved one. If both levels match, neuron n treats the edge as having been successfully "confirmed once" and calculates the time-of-travel of the edge from neuron $n-1$ to n. From a known velocity, the distance that has been passed within the time-of-travel can be computed by neuron n, i.e., ΔZ in Fig. 1(a). Meanwhile, based on the temporal history of that edge, neuron n makes a prediction as to when neuron $n+1$ can expect the edge. This predicted time is transferred to neuron $n+1$ together with the gray level, and saved. On receiving the edge, neuron $n+1$ then compares it with the saved one. If both gray levels match, neuron $n+1$ continues to compare the predicted time and the actual time of its counter. If both predicted and actual time-of-travels match within a chosen tolerance, then the edge is treated as confirmed twice and is thus a more reliable edge. In the same way, neuron $n+1$ estimates the prediction time for neuron $n+2$ and enables the prediction-confirmation process to be repeated along the axis of neurons.

Clearly, a temporal tolerance or "window" is associated with the neurons. At an individual neuron, an edge is accepted if the predicted and actual time-of-travel of that edge are within the neuron's tolerance window. Otherwise the edge is rejected as spurious. Based on this monocular model, we present an early visual model incorporating STDP adaptation by using LIF neurons. We assume that the scenes will be mapped onto a visual plane to form an optical flow field through the retina. The optical flow field has edges flowing radially along axes of neurons, it comprises the spatio-temporal cues of objects in the three-dimensional space.

We consider a similar dynamic scene in which an observer moves along the optical axis toward the target objects at a constant velocity. The velocity of the projected edges on the visual plane increases with the distance from the visual plane center. The neurons in the visual plane are therefore arranged on the radial axes such that the distance between the consecutive pairs of neurons increases hyperbolically with increasing radius. With

TABLE I
SCENE AND RETINA DATA FOR EDGE DETECTION AND DEPTH INFERENCE

Scene Parameter	**Value**
Resolution	512x512 pixels
Gray level	8 bits
Edge contrast	0.10
Velocity	Constant value
Direction	Optical axis
Retina Parameter	**Value**
Visual plane α	4.025
Axes number	400
Neuron per axis	48 (artificial test)
Neuron per axis	74 (real test)
Retina radius	250 pixels

this arrangement, the time-of-travel of an edge between the consecutive pairs of neurons is approximately the same. Thus, the distance from each neuron to the visual plane center is

$$r_n = \sinh\left(\frac{n}{\alpha}\right) \tag{1}$$

where n is the neuron order which is a nature number increasing from the first neuron located near the visual plane center, and α determines the pixel resolution. After the placement of neurons in the visual plane, higher pixel resolution, if necessary, can be achieved by interpolating additional LIF units between existing neurons.

The depth information for a neuron is derived from simple geometric calculations [40]. In cylindrical polar coordinates we have

$$Z = \frac{\Delta Z}{k_n - 1} \tag{2}$$

where Z is the depth of a point object in the 3-D space, ΔZ is the distance moved from when the point activates neuron $n-1$ until it reaches neuron n, see Fig. 1(a), $k_n = r_n/r_{n-1}$. The only part of the right-hand side of this depth equation that is not fixed for each neuron is ΔZ. Assuming that the ego or object motion velocity is constant and known, then ΔZ is also available by counting the time-of-travel of the projected edge between two successive neurons, and depth information can thus be recovered easily.

III. MODEL

We chose to use 400 identical radial axes to achieve a balance between image resolution and simulation time (and, ultimately, hardware cost). The LIF neurons have only nearest-neighbor, on axis connections. In the model we use a 512×512 pixel array which records the input image sequence at the same resolution. Forty-eight neurons are aligned on every axis for the simulation involving artificial data and 74 neurons per axis for the real image sequences. An edge is detected if a pixel has more than 10% irradiance difference compared to its surrounding pixels (see parameters in Table I).

A. Spiking Neuronal Model

Each neuron in this simplified model has three excitatory synapses that receive pulsed inputs. The synapses that carry a pulse from the preceding neuron and from the photoreceptor are referred to as the *flow synapse* and *receptive synapse*, respectively, see solid arrows in Fig. 2. The third input is used simply to start edge propagation via an interneuron, as described later. We make two assumptions for analytical and simulation simplicity: 1) there are no pulse transmission delays in the *flow synapses* between the LIF neurons and 2) the LIF neuron can not issue a spike unless its membrane is depolarized by both *flow* and *receptive* synapses.

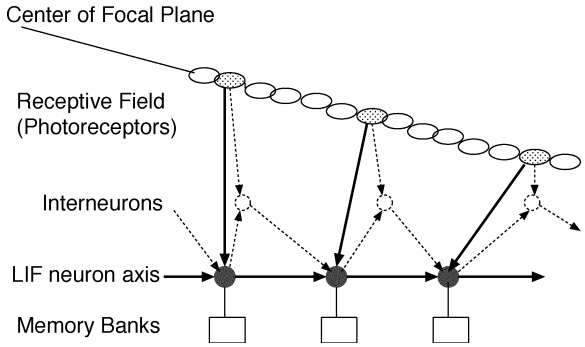

Fig. 2. Architecture of an axis of LIF neurons of the vision model. The photoreceptors align radially in the receptive field, they transfer the irradiance edge inputs to electrical pulse stimuli for the LIF neuron layer, the dotted photoreceptors have an actual connection with a corresponding neuron. A neuron is a basic computational unit with a memory. There is a layer of interneurons between the layers of photoreceptors and LIF neurons.

The second criterion shows that an appropriate correlation between both *flow* and *receptive* synaptic inputs is a necessary condition for a LIF neuron to fire. A LIF neuron is a homogeneous leaky integrator. After detecting a new edge, its subthreshold membrane state in the presence of spikes from both *flow* and *receptive* synapses is

$$C_i^{\text{mem}} \frac{dv_i(t)}{dt} = g_i^{\text{mem}} \left(v_{\text{rest}} - v_i(t)\right) + \sum_{j=0}^{1} I_j^{\text{syn}}(t) \tag{3}$$

where C_i^{mem} and g_i^{mem} are the membrane capacitance and conductance of neuron i, respectively. The last term of the right-hand side denotes summation of two excitatory postsynaptic currents (EPSCs). In the absence of the EPSCs and interneuron input, the membrane potential $v_i(t)$ will drift to its resting value v_{rest}. As the EPSCs drive the membrane potential above its threshold V_i^{th}, neuron i produces a short action potential and resets its potential to v_{reset}. We set $V_i^{\text{th}} > v_{\text{reset}} \geq v_{\text{rest}}$, and absolutely refractory period $t_{\text{ref}} = 0$. If neuron i detects a new edge, its interneuron is activated and emits a spike to the succeeding neuron $i+1$. In that case, there is no EPSC from the *flow* synapse to neuron $i+1$, which is, instead, depolarized by the interneuron.

The EPSC for both *flow* and *receptive* synapses, in our model, is a step followed by a rapid exponential decay [17]

$$\tau_{\text{syn}} \frac{dI_j^{\text{syn}}(t)}{dt} = -I_j^{\text{syn}}(t) + w_j S_j(t) \tag{4}$$

where τ_{syn} is the synaptic time constant, w_j is the strength of the *flow* or *receptive* synapse with $j = 0$ for *flow* and $j = 1$ for

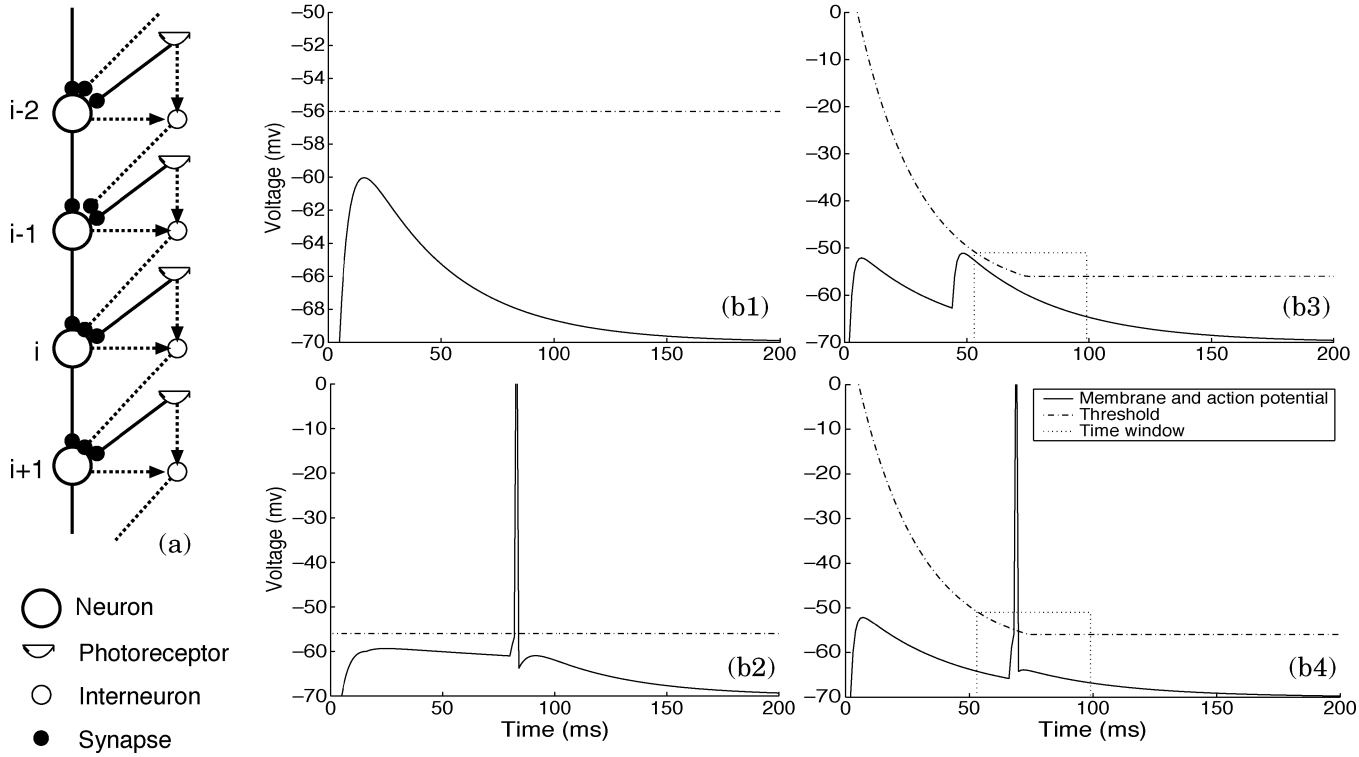

Fig. 3. Schematic diagram of an axis of spiking neurons and the possible membrane states. (a) Chain of neurons. Three synapses converge to a LIF neuron which has two outputs. (b1) EPSP on the membrane of neuron $i-2$ caused by a spike from the *receptive synapse* representing the arrival of a new edge. (b2) Plateau membrane potential of neuron $i-1$ caused by activation of the interneuron of neuron $i-2$ and the subsequent firing due to the arrival of an edge at the photoreceptor; the spike encodes the time-of-travel of the edge starting from the onset of the plateau potential. (b3) Neuron i first receives a spike from its *flow synapse*, decoding the prediction of its firing in, say 76 ms; if an edge arrives at the photoreceptor but out of a tolerance time window of [53, 99], then neuron i does not fire. (b4) In a similar case, an edge arriving at 64 ms activates neuron i.

receptive synapse. $S_j(t)$ is a series of input spikes to synapse j that signal edges occurring at times t_j^k in the optical flow

$$S_j(t) = \sum_k \delta\left(t - t_j^k\right). \tag{5}$$

The concept of a dynamic threshold for pattern formation is biologically plausible [45], [46] and has already been used for synchronous spike generation in a neuronal network [18]. In our design, the threshold of a LIF neuron, i, is the greater of an exponentially decaying threshold $V_i^{\mathrm{th1}}(t) = a_i \exp(-(t/\tau_i^{\mathrm{th}}))$ and an intersecting constant threshold $V_i^{\mathrm{th2}}(t) = V_T$

$$V_i^{\mathrm{th}}(t) = max\left(V_T, a_i \exp\left(-\frac{t}{\tau_i^{\mathrm{th}}}\right)\right). \tag{6}$$

Here, a_i is the threshold value at $t = 0$, the time when the preceding neuron $i-1$ fires, τ_i^{th} is the exponential threshold decay rate and V_T is the constant threshold value.

The intersection between $a_i \exp(-(t/\tau_i^{\mathrm{th}}))$ and V_T, t_i^{pred}, is the time when neuron $i-1$ predicts that an edge will reach neuron i. Since at that instant, $a_i \exp(-(t/\tau_i^{\mathrm{th}})) = V_T$, we thus have

$$t_i^{\mathrm{pred}} = -\tau_i^{\mathrm{th}} \ln \frac{V_T}{a_i} \tag{7}$$

and the threshold of a spiking neuron can be described in terms of t_i^{pred}

$$V_i^{\mathrm{th}}(t) = \begin{cases} a_i \exp\left(-\frac{t}{\tau_i^{\mathrm{th}}}\right), & \text{if } 0 \le t \le t_i^{\mathrm{pred}} \\ V_T, & \text{if } t > t_i^{\mathrm{pred}}. \end{cases} \tag{8}$$

An edge flowing along an axis of neurons is shown in Fig. 3. The interneurons are used to initialize the algorithm in the presence of new edges. If a new edge arrives at a photoreceptor connecting with neuron $i-2$, then that neuron does not fire because it has not been depolarized by its *flow synapse*. In this case, the corresponding interneuron is activated and, hence, emits a spike to stimulate the succeeding neuron $i-1$. Upon receiving a stimulus from the interneuron, the membrane of neuron $i-1$ is depolarized to a plateau potential. This neuron will fire a spike whenever it receives a stimulus from its *receptive synapse*. If neuron i is depolarized by its *flow synapse* at time t, it is expected to be further depolarized by its *receptive synapse* within a time range of $(t + t_i^{\mathrm{pred}} \pm \Delta t_i)$, where Δt_i is the tolerance window size of neuron i to be determined by the adaptation mechanism. If the depolarization from the *receptive synapse* is within the expected window, neuron i spikes and resets its membrane potential. The relative edge is then confirmed as a genuine feature by neuron i, and continues to flow along the axis.

t_i^{pred} is taken as the center of the tolerance window within which neuron i should receive an edge. We define the maximum and minimum tolerance window length as 60% and 10% of t_i^{pred}, respectively. Thus, we have $\Delta t_i^{\max} = 0.3 t_i^{\mathrm{pred}}$, $\Delta t_i^{\min} =$

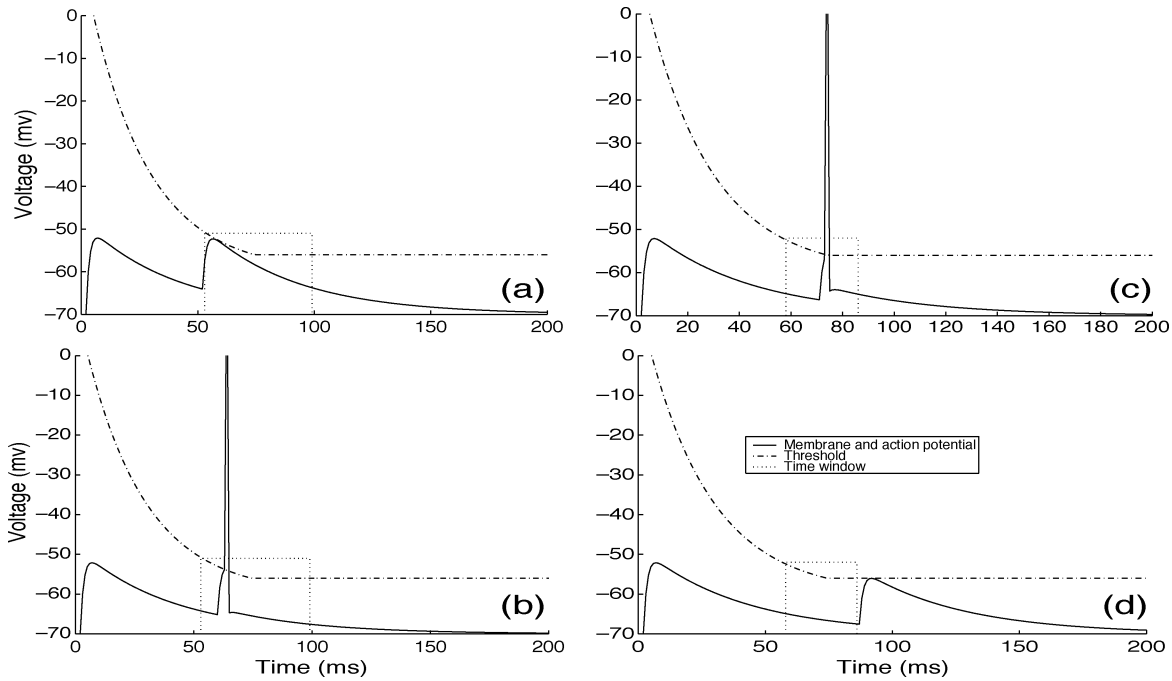

Fig. 4. Membrane dynamics of a neuron shows the temporal correlation of stimuli and response, and the effect of the STDP mechanism. The dynamic threshold consists of an exponential part and a constant part, they intersect at the prediction time when the neuron expects an edge from its *receptive synapse*. The first membrane depolarization is caused by the *flow synapse* spike, and the second by the edge from the *receptive synapse*. If the edge comes within a time window, then the neuron spikes. (a) Edge arrives at 52 ms. (b) Edge arrives at 60 ms. (c) Edge arrives at 71 ms. (d) Edge arrives at 87 ms. A time window of [53, 99] is applied for (a)–(b), and [58, 86] for (c)–(d). The prediction time for all figures is 75 ms. The test parameters are, the prediction parameter pair $\{\tau_{\rm th}, w_{\rm recept}\} = \{26\ {\rm ms}, 13.91\}$ for (a) and (b), $\{26\ {\rm ms}, 13.20\}$ for (c) and (d), the *flow synapse* $w_{flow} = 20$, the membrane time constant $\tau_{\rm mem} = 40\ {\rm ms}$, the synapse time constant $\tau_{\rm syn} = 2\ {\rm ms}$, the constant threshold $V_T = -56\ {\rm mv}$, the neuron reset potential $v_{\rm reset} = -65\ {\rm mv}$, and the resting potential $v_{\rm rest} = -70\ {\rm mv}$.

$0.05t_i^{\rm pred}$, the maximum possible firing window of neuron i is $(t + 0.7t_i^{\rm pred}, t + 1.3t_i^{\rm pred})$ and the minimum window is $(t + 0.95t_i^{\rm pred}, t + 1.05t_i^{\rm pred})$. For simulation simplicity, we only have the *receptive synapse* adaptable. An optimized window size for neuron i can then be determined by STDP adaptation of the *receptive synapse* (see Section III-C). If neuron i receives an edge from its photoreceptor before the predicted time $t_i^{\rm pred}$, then its *receptive synapse* is increased and the window expanded. Otherwise its *receptive synapse* is depressed and the window is narrowed. The selection of appropriate exponential threshold decay time constants $\tau_i^{\rm th}$ and *receptive synaptic* strengths are critical for the model. These two parameters are chosen and encoded by the preceding neuron $i - 1$ based on the prediction of time-of-travel of an edge. We simulate the spiking activities of a LIF neuron with a dynamic threshold first and then choose suitable values of $\tau_i^{\rm th}$ and *receptive synaptic* strength to form the prediction parameter pairs, $\{\tau_i^{\rm th}, w_{\rm recept}\}$, for each possible predicted time in the prediction range. These parameters then determine the learning dynamics of the full vision model. Fig. 4 shows an example of the selected prediction parameter pairs, $\{\tau_i^{\rm th}, w_{\rm recept}\}$, with arrival of a sensory spike outside (a) and (d) or inside (b) and (c) a tolerance window whose size has been adjusted by changing the *receptive synaptic* strength alone.

B. Vision Algorithm Based on Spiking Neurons

Each model neuron unit has five major components; an edge-sensitive photoreceptor, a LIF neuron, *flow* and *receptive* synapses, and a dynamic short-term memory. The memory has the same basic function as the membrane capacitor of a LIF neuron in terms of recording a spike event. Besides, we can use it to memorize some characteristics of a spike event such as an edge's gray level to facilitate software simulation. A LIF neuron also has a counter for the time course of the dynamic membrane threshold, and an additional input from an interneuron whose spiking status is determined by the states of the preceding LIF neuron and associated photoreceptor (Figs. 2 and 3). Image sequences with 512×512 pixel arrays (resolution) are used as the inputs.

Since the LIF neurons are *sparsely* allocated along radial axes, some photoreceptors have no connections with the LIF neurons, see the unfilled circles in the receptive field in Fig. 2. These photoreceptors are defined as the invalid photoreceptors. Accordingly, those photoreceptors connecting with LIF neurons are defined as the valid photoreceptors. In an initial period represented by a series of image frames starting from the first frame, there may be an edge, namely e1, projected onto the invalid photoreceptors which are located between two valid ones, p1 and p2. This edge has never passed p1, and it will pass p2. Before it arrives at p2, it is possible that the LIF neuron of p2 has had a plateau membrane potential which is induced by a new edge arriving at p1. In this case it is clear that, when edge e1 arrives at p2, it will, erroneously, force the LIF neuron of p2 to fire a spike based on the plateau potential which is not induced by e1 itself. We, hence, define the edges like e1 as the intermediate edges, and those edges which make the LIF neuron fire based on the plateau potential induced by themselves as the normal edges. After the algorithm starts, an initial self-organization stage can be seen. The convergence of depth inference occurs as soon as

all the intermediate edges become normal ones. We offer a proof of convergence of the algorithm in an Appendix.

When an edge, either normal or intermediate, arrives at the photoreceptor of neuron i, it induces an EPSP on the neuron's membrane. The sum of this EPSP, together with any EPSP induced by the *flow synapse*, determines the neuron's activity. If the total membrane potential is above the neuron's dynamic threshold, then the neuron emits a spike encoding the predicted time-of-travel after which that edge should arrive at the subsequent neuron, $i+1$.

The predicted time-of-travel, t_{i+1}^{pred}, is defined by the temporal history of an edge arriving at neuron i, which can be described with the recursive equation

$$t_{i+1}^{\text{pred}} = \frac{t_i^{\text{actual}} + P t_i^{\text{pred}}}{P+1} \tag{9}$$

where t_i^{pred} and t_i^{actual} are the predicted and actual time-of-travel of an edge arriving at neuron i from $i-1$ respectively. P is the number of times that an edge has been correctly predicted as it propagates along the corresponding axis to neuron i. Upon receiving the prediction, neuron $i+1$ then has $t_{i+1}^{\text{pred}} = -\tau_{i+1}^{\text{th}} \ln(V_T/a_{i+1})$ to determine the values of $-\tau_{i+1}^{\text{th}}$ and a_{i+1} for an appropriate dynamic threshold.

If no spike is emitted by neuron i upon arrival of an edge from its photoreceptor, then there are two possible reasons: 1) the edge is a new one, and the neuron's membrane has not yet been facilitated by a spike stimulus from its *flow synapse* and 2) the edge has arrived too early, or too late and is outside the tolerance window.

In either case, the relative interneuron emits a spike to neuron $i+1$. The membrane of neuron $i+1$ is thus facilitated, and the facilitation results in a plateau potential on the membrane of neuron $i+1$ which will activate $i+1$ on receiving an edge, intermediate or normal, from the photoreceptor at any time. The plateau potential is sustainable for a maximum duration of the time-of-travel of an edge from neuron i to $i+1$.

C. Adaptation

We use STDP adaptation to improve this algorithm's robustness against inaccuracies caused by the collective behavior of edge flow, nonideal positioning of pixels as an array on the visual plane and inaccuracies from fabrication tolerances in an aVLSI implementation. The weight of the *receptive synapse* of each neuron is increased by a small amount if the *receptive synapse* fires before the predicted time instant and decreased if the *receptive synapse* fires late. The *flow synaptic* weight is held constant for simplicity. The adaptation scheme is

$$\Delta W_i(t) = \begin{cases} A_+ e^{\frac{t-t_i^{\text{pred}}}{\tau_+}}, & \text{if } t < t_i^{\text{pred}} \\ -A_- e^{-\frac{t-t_i^{\text{pred}}}{\tau_-}}, & \text{if } t > t_i^{\text{pred}} \\ 0, & \text{if } t = t_i^{\text{pred}} \end{cases} \tag{10}$$

where $\Delta W_i(t)$ is the *receptive synaptic* weight change of neuron i, $\{A_+, \tau_+\}$ and $\{A_-, \tau_-\}$ are the amplitude and decay constant of potentiation and depression, respectively and t_i^{pred} is the predicted time instant when neuron i should fire. The *receptive synaptic* weights are initialized to random values that correspond to a tolerance window width within the dynamic range, i.e., $(t + t_i^{\text{pred}} \pm \Delta t_i)$. If an edge arrives at the photoreceptor earlier than predicted, that edge has the potential to activate the neuron and the window is expanded to try to include such stimuli. If, however, an edge arrives later than predicted, it is likely to be uncorrelated with the predicted postsynaptic spike and the window is narrowed to exclude it. During the operation of the vision algorithm, the STDP process widens or narrows the tolerance window in order to include the correlated stimuli and exclude uncorrelated ones.

TABLE II
MODEL NEURON DATA FOR EDGE DETECTION AND DEPTH INFERENCE

Neuron Parameter	Value
Membrane τ	40ms
Synapse τ	5ms
Adaptive τ_+	5ms
Adaptive τ_-	5ms
Rest potential	-70mv
Reset potential	-70mv
Constant threshold	-56mv
Flow weight	20
Adaptive amplitude A_+	0.2
Adaptive amplitude A_-	-0.3

IV. SIMULATION RESULTS

This spike-based algorithm has been tested in both artificial and real environments. The neuron parameters are shown in Table II and we use 48 and 74 neurons per axis for the artificial and real images, respectively. All of the neuron parameters except the amplitudes of learning curves $A_+ = 0.2$ and $A_- = -0.3$ are chosen to be neurobiologically plausible. We choose A_+ and A_- to be larger than the values used in [24], as our pixel stimuli are considerably sparser. We use Euler integration as the numerical method [47] and the step is updated with each image frame.

A. Artificial Images

The artificial scene shown in Fig. 5 is constructed using OpenGL. Three objects, a cylinder, a cone and a sphere, are located at distances of 1500, 1100, and 700 cm, respectively, in front of a white background at 1600 cm. The observer is at a height of 1.6 m and moves toward the scene at a constant velocity of 1 cm per frame.

During motion, edges flow along the radial axes of neurons. Initially, in Fig. 5(a), the membrane of any LIF neuron whose photoreceptor detects a new edge is facilitated only by the *receptive synapse*. When a new edge is detected, the detecting neuron sends a spike representing the new edge to its neighboring neuron via the relative interneuron, which induces the plateau potential on the membrane of the neighboring neuron. The neighboring neuron then issues a spike when it receives a sensory edge at any time. The spike encodes and transmits the predicted time-of-travel of that sensory edge to the subsequent neuron on the axis. As the motion continues, any neuron activated by the correlated *receptive* and *flow* synaptic inputs issues

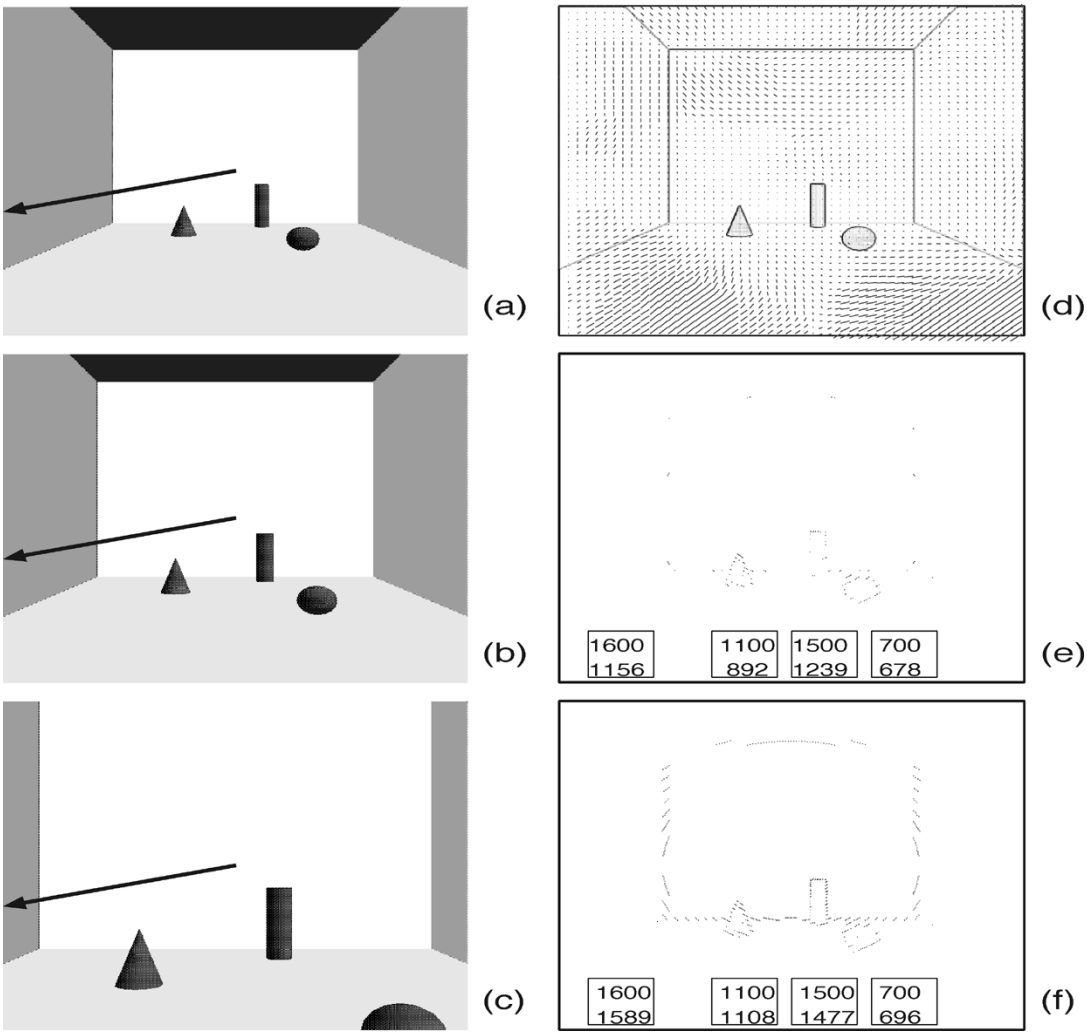

Fig. 5. Pixel maps of a hallway scene. (a)–(c) Snapshots of image frames at the start, 175 and 600 steps respectively, the arrow line is the approximate location of the 210th axis in each pixel map. (d) Still contour of the first frame without running the spike-based algorithm, the size of the short lines signifies the strength of optical flow in the 2-D space. (e)–(f) Pixel maps which are calculated by the algorithm from the start frame till the corresponding frame shown on the left. Each activated (black) pixel indicates the location of the first neuron which detected an edge and started the algorithm to encode the time-of-travel information for its subsequent neurons. If an edge flows reliably along an axis of neurons, that detecting neuron location can be recovered by a neuron far away from it. The depth boxes correspond to the depth inferred for the background frame, cone, cylinder and sphere, respectively, from left to right. Upper values in the boxes are the actual original depth of each object, lower values are estimated depth, both described in centimeters.

a spike, which represents the agreement of the actual and predicted time-of-travelsand, hence, confirms the correct prediction of the time-of-travel of the relative edge. After the confirmation, the position of the detecting neuron is shown with black dot in Fig. 5(e) and (f), which form the reconstructed images corresponding to the scene before motion. We define these dots as the "effective pixels" and the images of these dots as the pixel maps.

Fig. 5(e) and (f) are examples of pixel maps for the artificial image test. Few edges appear in the early pixel maps, shown in Fig. 5(e), while the model is self-organizating (see appendix for details). After the initial settling period, all intermediate edges have become the normal edges in the sense that any predicted time-of-travel of an edge, given by the preceding neuron i, matches the actual time-of-travel of an edge arriving at neuron $i+1$. It implies that an edge arriving at neuron i facilitates the membrane of neuron $i+1$ and then, after a time-of-travel, the *same* edge will cause neuron $i+1$ to fire. The detecting neuron of the confirmed edge is then included in the pixel map. The counter of the firing neuron records the time-of-travel of the edge between the successive neurons. As the velocity is known, ΔZ in (2) is available and the depth of that edge can be recovered.

We can study the depth recovery process in more detail by concentrating on a single radial axis. In this simple example only one edge feature, the line where the left side wall and the back wall meet, will propagate along the chosen axis. The initial location of this edge feature is 16 m away from the camera. The depth information recovered by the neurons on the 210th axis is shown in Fig. 6(a) alongside the actual known depth of the edge in the image. The estimated depth matches the actual depth well, after an initial self-organizing process. The effect of the STDP mechanism can also be seen more explicitly through the distribution of the *receptive synapse* weights before and after adaptation. Initially weights are distributed randomly between maximum and minimum values [Fig. 6(b)]. After online STDP adaptation the initial uniform distribution converges toward the bimodal distribution of Fig. 6(c).

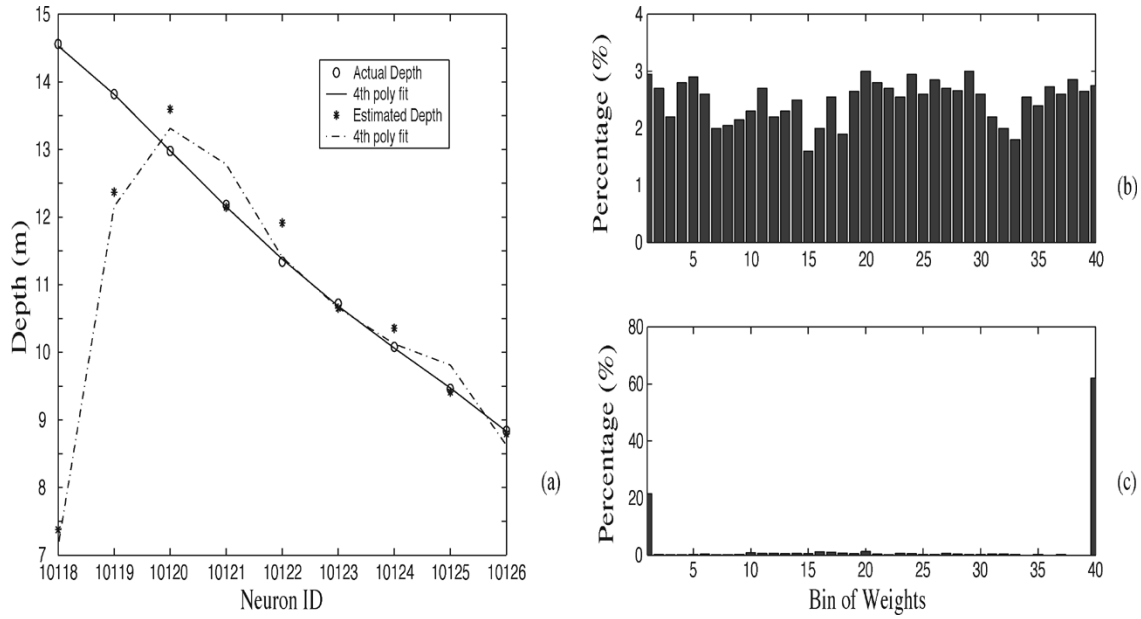

Fig. 6. Characteristics of the vision model operating on the hallway scene. (a) Comparison between the depth information estimated by the neurons on axis 210th and the corresponding actual depth, the data is fitted with a 4th degree polynomial function in a least-squares sense; the increasing order of neuron ID also represents the time course. (b) Initial random distribution of *receptive synapse* strengths in the weight range. (c) Bimodal distribution after STDP adaptation.

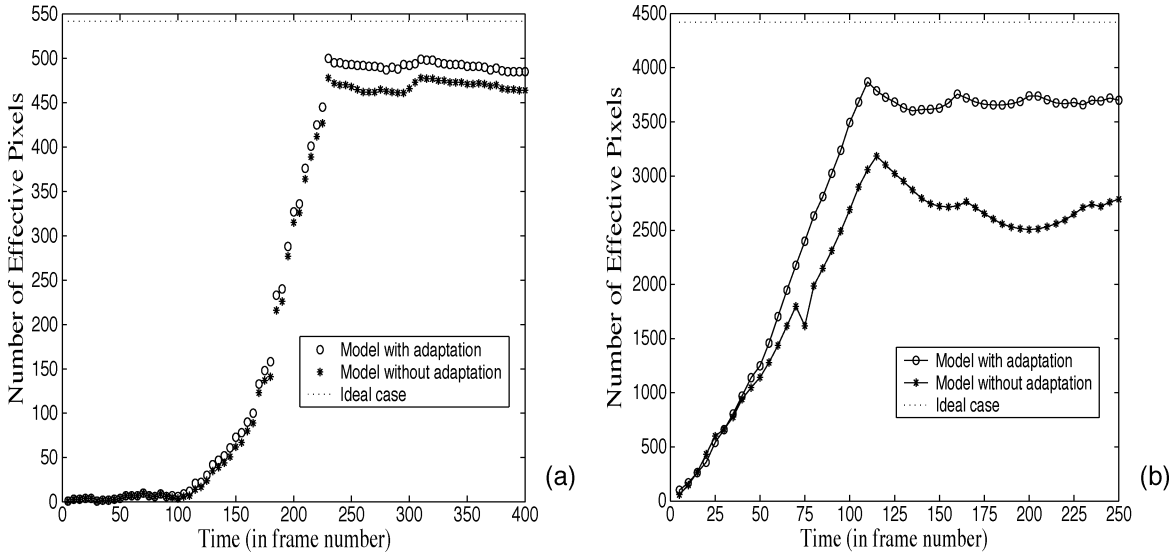

Fig. 7. Performance comparison of models with and without synaptic adaptation. The horizontal axis describes the time course with the number of image frames. The dotted lines in both figures are the ideal number of effective pixels. (a) Test on artificial images. The camera moves 1 cm/frame. Labels "o" and "*" represent the number of effective pixels detected by the models with and without adaptation, respectively, at a relative time instant. (b) Test on real images. The camera moves at 1.1 mm/frame with recording speed of 24 frames/s. Labels "o" and "*" represent the number of effective pixels detected by the models with and without adaptation, respectively.

The performance of the vision model can be shown quantitatively by comparing the number of effective pixels generated by the models with and without STDP adaptation. The comparison of performance of two models is shown in Fig. 7(a).

It is clear that, after the period of self-organization, both models reach a saturated stable state in which there are no significant changes in the number of edges detected by both models. The model with STDP adaptation is always superior. Both models take approximately 220 time units, equivalent to 2.2 m, to converge. After convergence, the model with adaptation has detected an average of 492 effective pixels, or 92.3%, out of a total sum of 533 effective pixels. The model without adaptation has detected an average of 468 effective pixels, or 87.8% out of the same sum. The slight fluctuation observable after self-organization can be quantified as the mean and standard deviation of the number of effective pixels, (492, 4) for the model with adaptation, and (468, 5) for the model without adaptation. Although the performance of a model with adaptation is always better than that without adaptation, the difference is not worthy of note in the artificial test. This is intuitively correct, as the artificial image is, effectively, "perfect" and noise-free. The irregularities that our STDP adaptation process aims to reduce are, therefore, essentially absent.

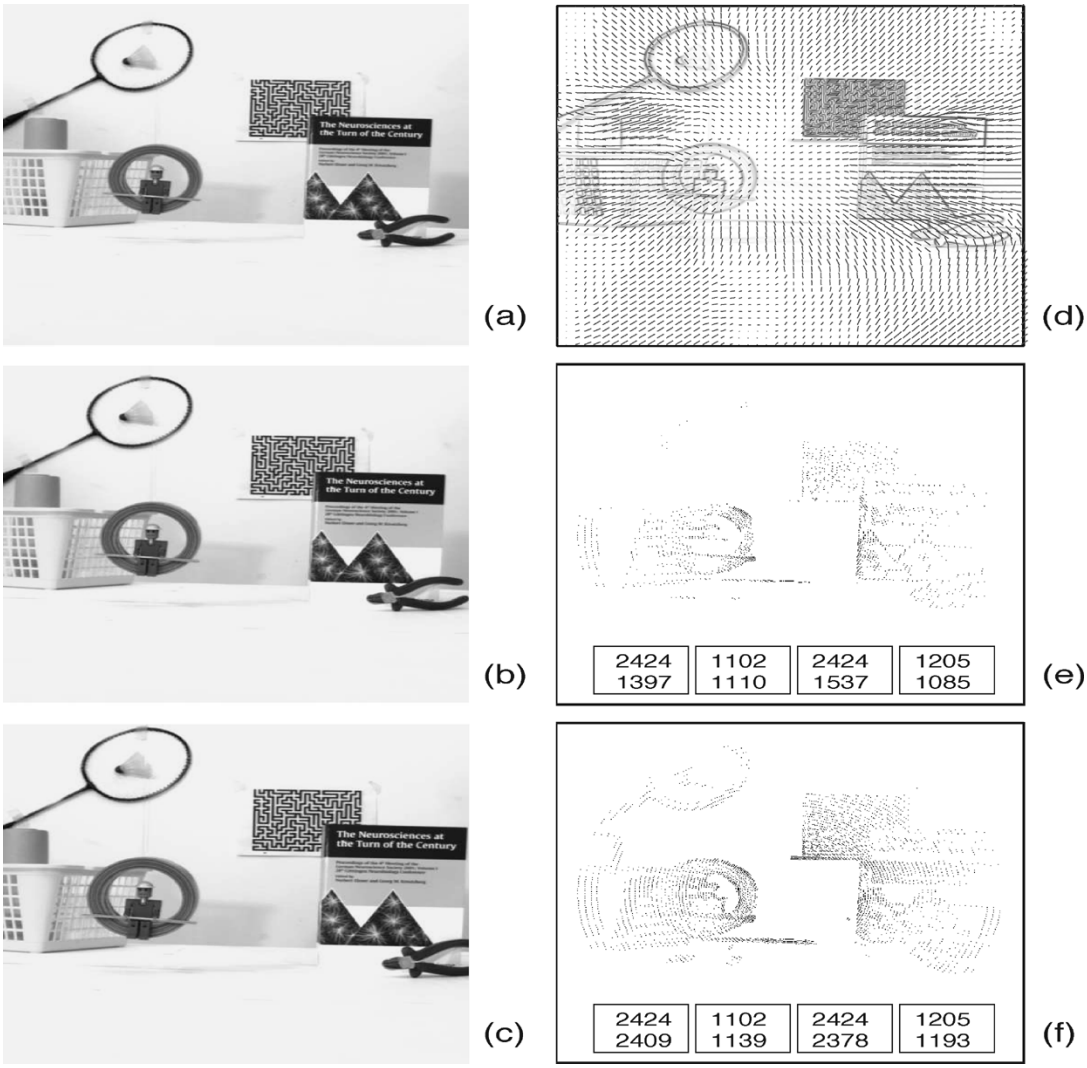

Fig. 8. Pixel maps of a real scene. (a)–(c) Snapshots of image frames at the start, 70 and 170 steps respectively. (d) Still contour of the first frame without running the spike-based algorithm, the size of the short lines signifies the strength of optical flow in the 2-D space. (e)–(f) Pixel maps. The depth boxes correspond to 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 estimated depth, both described in millimeters.

B. Real Images

In a real scene (Fig. 8), a book, a set of pliers, a maze board, a toy pendulum and a badminton racquet are arranged 1205, 1150, 2424, 1102, and 2424 mm 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 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 optical and focus characteristics. The real scene is recorded in 250 frames with the camera motion speed approximately 1.1 mm/frame. The image frames are saved in a bitmap format with a resolution of 1276×1016 and 24 bits of color. The recording speed is 24 frames per second. The image frames are thereafter preprocessed to fit the final resolution of 512×512 and 8 bits of gray level, and transferred to a SUN Blade 100 workstation for analysis.

The contours of the toy pendulum and book are easily recognizable from the pixel maps in Fig. 8. However, as the distance between edges in the maze approximates that of the pixel layout resolution, significant detail cannot be seen. The late appearance of the badminton racquet is caused by its location which is in the (peripheral) area of low neuron and pixel density. Depth information is recovered, as before, from the image sequence and the object depths are shown in Fig. 8. These results demonstrate that the algorithm works under real world conditions, once again assuming stable optical flow.

The ubiquitous noise and disturbance of optical flow can prevent the individual spiking neurons in a population corresponding to an object from having the equal inferred depth. A reliable method is to use the arithmetic average of the depth values computed by a group of neurons to depict an object's depth information. After an initial convergent period, the actual and estimated depth values agree quite well, see Fig. 8(f).

A performance comparison of the model working on the real image sequence, with and without STDP adaptation, is displayed in Fig. 7(b). After convergence, the model with adaptation has detected an average of 3689 effective pixels, or 83.4%, out of a total sum of 4421 effective pixels. The model without adaptation has detected an average of 2734 effective pixels, or 61.8% out of the same sum. Clearly the STDP adaptation plays an important role in improving the model performance.

The fluctuations in the number of effective pixels are a result of the inclusion and deletion method of the detecting neurons in the pixel map. A detecting neuron is included in the pixel map as long as the predicted time-of-travel of the relative edge is confirmed; it is excluded if any further confirmation fails in the subsequent neurons. The imperfect behavior that STDP ameliorates in this real image includes effects of camera tremble and disturbances of the scene context. We describe the fluctuation quantitatively with the mean and standard deviation of the number of effective pixels, (3689, 55) for the model with adaptation, and (2734, 181) for the model without adaptation. Our experiment shows that the model with adaptation is remarkably more robust in the real scene, where there are some imperfections to deal with.

The tests on both artificial and real images show that the distribution of all *receptive synaptic* strengths will display the bimodal mode after adaptation; some of the *receptive synapses* tend to be depressed while the others potentiated. It implies that, given the radial neuron layout, the actual time-of-travel of edges between adjacent neurons may be consistently shorter (or longer) than the preceding neuron's prediction. This mismatch, namely layout mismatch, is caused by the inaccurate assignment of pixels (i.e., photoreceptor input) to the neurons during Cartesian to polar coordinate transform. This transform is necessary because the image sequences are in 512×512 Cartesian description while the visual plane uses polar coordinate to align neurons. Comparing the model performance with and without adaptation (Fig. 7), clearly STDP can compensate for this kind of mismatch remarkably well.

V. Discussion

Using delays and coincidences, neural timing networks can perform time-domain signal processing operations to compare, extract and separate temporal patterns [48]. Temporal coding characteristics of spiking neuronal networks have been implemented in silicon circuits [49] to simulate some primitive functions of the early visual system, e.g., the orientation selective response of V1 cells [50], [51], and common recurrent networks [52]. Following this research trend, this study aims to provide a spike-timing dependent neuromorphic early visual model that emulates the function of the retina and area V1, which is thought to be capable of detecting edges and initiating depth inference. The model also aims to point toward a neuromorphic VLSI realization of the early visual processing in a dynamic scene. We therefore discuss other approaches to aVLSI motion detection along with the biological plausibility and neuromorphic feasibility of our model.

A. Some Two-Dimensional Neuromorphic Visual Motion Models

There are two major classes of motion detection algorithms, i.e., token-based and intensity-based approaches [53]. The correlation algorithm, a subdivision of the intensity-based class, measures spatio-temporal correlations caused by moving objects or ego-motion. Since correlation is based on multiplication, the correlation algorithms show superior numerical stability. However, most of these algorithms are unsuitable for implementation in integrated circuits that compute extended optical flow fields, because they are tuned to a narrow velocity range and tend to be expensive in silicon area [54].

For example, Delbrück's correlation-based, velocity-tuned retina was built on a hexagonal neuron grid and incorporated delay lines in three directions [55]. If the velocity of a moving edge matched a pre-set delay time of the delay line, the output signals grow in strength; otherwise they shrink. As the tuning is very narrow, the sensor is fairly limited, and considerable chip area is required to measure motion over a large velocity range. Similar work was also reported [56], but with the same limitations and is, hence, not suitable for the monolithic implementation of dense velocity-sensing arrays [54].

An arguably more practical system was proposed by Kramer *et al.* [57], [58] in the time domain. They presented two different circuits to measure the time-of-travel of an edge between two adjacent photoreceptors. One of which, i.e., the facilitate-and-sample (FS) motion method, is similar in principle to the decaying threshold mechanism in our vision algorithm. An equivalent approach was also implemented by Etienne–Cummings *et al.* [59], who suggested that this form of motion sensor may be more compact and thus suitable for use in a dense neuromorphic visual array.

Similarly, our algorithm measures the time-of-travel of an edge between two photoreceptors. This kind of method is more attractive for implementation in compact arrays [54]. Moreover, our model includes global edge detection at system level, as well as depth inference. Meanwhile, the study incorporates the real time synaptic strength adaptation mechanism to deal with the ubiquitous noise and mismatch problems, and the model performance is favorably displayed. None of the above methods address these issues.

B. Biological Plausibility

In different species of animals, different mechanisms reduce the optical flow to, if possible, a single component. Airborne animals such as birds and flies, tend to fly in straight lines. Under these conditions, only forward motion exists and optical flow is reduced to its radial components. Edges in the visual plane then move radially outwards from the focus of expansion with a velocity proportional to the radial distance from the focus of expansion but inversely proportional to their depth in the scene [40]. Hence, the depth can be inferred from this edge expansion which contains the temporal cues, i.e., the time-of-travel between adjacent photoreceptor locations in a retinotopically organized vision model.

Our model, which aims to simulate early visual mechanisms without the mediation of higher visual organizations, can be

seen to imply significant simplification of reality. Nevertheless, due to the adoption of an error-tolerance window mediated by STDP adaptation, we hope to relax the restriction of 100% constant velocity without compromising the model performance unduly.

C. Neuromorphic Feasibility

Presently the preliminary circuit simulation using Hspice in Cadence EDA tool has shown the possibility of neuromorphic realization of our model. We have included four new elements, plateau membrane potential, dynamic membrane threshold, counter and short-term memory, to the traditional LIF neuron. Three of them, except the memory, are essential for the hardware implementation of the algorithm. The memory is introduced to memorize the facilitation, either decaying EPSP or plateau potential, caused by an edge arriving at the preceding neuron. Considering that a silicon LIF neuron has a membrane capacitor which serves as a short-term memory to integrate input currents, we can simply ignore the memory design in the silicon neuron without undermining the vision model. The algorithm remains convergent, the proof in the Appendix still holds.

The concept of plateau potential is used to initiate the edge flowing along an axis of neurons. It is motivated by the experimental and analytical works revealing that a plateau potential can be triggered by depolarization in preceding neurons [60], [61]. In simulation, we use the output of an interneuron to generate the plateau potential on the membrane of a LIF neuron i. In hardware the effect of the interneuron can be replaced by an AND logic with two inputs from the output of neuron $i-1$ and its associated photoreceptor, respectively. The output current of the AND logic charges the membrane capacitor of neuron i to an appropriate plateau voltage just below its (constant) threshold. Then neuron i can fire if it receives an edge from its *receptive synapse*.

When an edge is confirmed by a preceding neuron $i-1$, the membrane potential of neuron i is facilitated by its *flow synapse*. The dynamic threshold is used to refrain neuron i from firing earlier than expected. For instance, starting from the firing time of neuron $i-1$, neuron i should not be permitted to fire during an initial period to allow a time-of-travel of the relative edge. It is clear that the concept of dynamic threshold is closely related to the time-of-travel of an edge, which is measured by a counter. Suppose that the time-of-travel has been measured, in neuromorphic engineering the complementary metal-oxide semiconductor (CMOS) transistor circuits can be designed as a switch to restrict the firing window of a neuron. The window width can be adjusted by including the adaptation circuitry mediated by STDP.

The counter circuit of a neuron is the real concern in the neuromorphic implementation. As we introduced, many aVLSI circuits have been presented to measure the time-of-travels or velocities of edges between neurons during motion [55]–[59]. However, most of them are just the representation of the time durations or velocities in the sense that they have not used these values for further computing. In depth recovery we need the concrete time-of-travel information. Therefore, one option is that, for every neuron, we use an analog, calibrated current source charging a capacitor as the time counter activated by a universal clock pulse generator, which generates a constant number of spikes per second, acting like a frame generator in a video camera device. The voltage of the capacitor can thus represent the time-of-travel of an edge, and can be used for further computing.

In summary, the proposed numerical model provides a possible approach to implement a monocular edge detector in aVLSI hardware, and certainly some modifications are necessary to implement the new elements which are not the traditional parts of a LIF model neuron. The output of the photoreceptors is binary, i.e., black and white, where black values for edges. The test results show that STDP adaptation can minimize the effects of process mismatch in the context of a spike-timing processing scheme for visual scenes. Further works will be focused on the two dimensional design of visual plane, including, e.g., communication between the photoreceptor array and visual plane processing circuit module and measurement of time-of-travel of edges.

VI. Conclusion

This study presents a neuromorphic approach to a large-scale, asynchronous, spiking network that will perform useful computation on a real visual scene. We have explored an artificial vision model based on LIF neurons for edge detection and depth analysis. The postsynaptic LIF neuron's activity depends on the correlation of two synaptic inputs. We demonstrate that the internal neuron parameters determine a temporal tolerance window within which the effect of correlations between the neuron's inputs affects its output. The window size is adjustable through synaptic strengths and membrane threshold time constants which are directly related to circuit parameters, rendering the model amenable to aVLSI application. As STDP is a likely adaptation mechanism underlying the visual cortical activities, we have introduced synaptic weight adaptation based on STDP, leading to a useful improvement in the model performance. Our approach to using STDP to ameliorate real system imperfections and deal with real data is general and is likely to have applications in other domains where timing is an intrinsic part of computation, such as audition.

Appendix

Hereby, we prove our belief that the optical flow of edges on the visual plane of spiking neurons will always converge, i.e., after a possible self-organizing period an edge passes spiking neuron i will be surely recognized by neuron $i+1$ on the same axis in its traveling time of Δt_i between neurons i and $i+1$. For simplicity, in the proof we assume that the traveling time between neuron pairs is the same, i.e., $\Delta t_i = \Delta t_{i+1} = \Delta t$. For cases in which this assumption does not hold, the proof is still valid.

Proposition 1: In the proposed spiking neuron-based vision algorithm, the optical flow of edges in the visual plane of spiking neurons will always converge.

Proof: Without loss of generality, we have two assumptions: 1) in an ideal, continuous and real scene each edge will pass, rather than skip over, a neuron before it can reach the next

neuron on an axis. and 2) no edges can escape from the axis or no new edges can add into the axis during the propagation of edges which initially exist on an axis.

The simplest case is that, at the starting time of the algorithm, there is no edge on the invalid photoreceptors. If the valid photoreceptor of neuron i detects an edge for the first time, neuron $i+1$ has a plateau potential on its membrane and will be activated by any stimulus from its *receptive synapse*. It is clear that the edge coming from neuron i will be confirmed as a predicted edge at neuron $i+1$ in a period of Δt_i. This temporal interval information will be encoded by the spike issued to neuron $i+2$. Since all the consequent neurons on the same axis after neuron $i+1$ will sum up the *receptive synapse* induced EPSP with the *flow synapse* induced EPSP on their membrane, it is obvious that the edge sourced from neuron i will be consistently confirmed as the predicted edge as it propagates further from neuron $i+1$.

The more general case is that, after the algorithm starts, the valid photoreceptor of neuron i detects an edge for the first time whilst more edges are also detected by the invalid photoreceptors between neuron i and $i+1$ where $i=0,1,\ldots,k-1$ and k is the total neuron number. In this case the key is to prove that a plateau potential event in a neuron's memory will finally be matched by its initiator edge rather than by other edges from the *receptive synapse*. In order to achieve that, we assume a series of neurons $\{n_0,n_1,n_2,\ldots,n_{k-1}\}$ on an axis. $\{a,b\}_{n_i}$ represents an EPSP event on the membrane of neuron n_i. $a=-1$ for a plateau potential event, or a equals to any positive number for the time after which neuron n_i expects an EPSP from its *receptive synapse*. b is the time when the event itself is created on the membrane of neuron n_i. At the start of the algorithm, the membrane potentials of all neurons are at the resting value. If an edge arrives at the photoreceptor of n_i, we mark this time instant as Δt_i. As the algorithm runs, the EPSP events on the membrane of the sequent neurons will be $\{-1,\Delta t_i\}_{n_{i+1}}$, $\{\Delta t_{i+1}-\Delta t_i,\Delta t_{i+1}\}_{n_{i+2}}$, $\{-1,\Delta t_{i+2}\}_{n_{i+3}}$, $\{\Delta t_{i+3}-\Delta t_{i+2},\Delta t_{i+3}\}_{n_{i+4}}$, $\{-1,\Delta t_{i+4}\}_{n_{i+5}},\ldots,\{(-1)^{k+1}(\Delta t_{i+k}-\Delta t_{i+k-1})^{1-(-1)^k/2},\Delta t_{i+k}\}_{n_{i+k+1}}$, which satisfy

$$\begin{cases}\Delta t>\Delta t_{2n+1}>\Delta t_{2n}\\ \Delta t_{2n}>\Delta t_{2n-1}\end{cases} \tag{11}$$

where $n=k/2$ if k is even, and $n=k-1/2$ if k is odd. Thus, we know that the time sequence, when the *receptive synapse* induced EPSPs occur, $\{\Delta t_i\}, i=0,2,\ldots,k-1$, is a monotonically increasing sequence with the upper limit of Δt. Hence, we have

$$\lim_{i\to\infty}\Delta t_i=\Delta t. \tag{12}$$

Since every new edge passing the photoreceptor of neuron i will generate a plateau potential event to neuron $i+1$, the above equation means that gradually a plateau potential event on a neuron's membrane will finally be matched by its initiator edge, hence, the algorithm converges after a self-organizing period. ■

ACKNOWLEDGMENT

The authors would like to thank L. Smith and M. Dahlem for helpful discussions, N. Pugeault for real image sequence acquisition, and P. Hillman for help on preprocessing of the real image sequence

REFERENCES

[1] M. A. Arbib and P. Erdi, "Précis of neural organization: structure, function, and dynamics," *Behav. Brain Sci.*, vol. 23, pp. 513–571, 2000.

[2] U. S. Bhalla and R. Iyengar, "Emergent properties of networks of biological signaling pathways," *Science*, vol. 283, pp. 381–387, 1999.

[3] M. Tsodyks and C. Gilbert, "Neural networks and perceptual learning," *Nature*, vol. 431, pp. 775–781, 2004.

[4] K. Funke, Z. F. Kisvarday, M. Volgushev, and F. Wörgötter, "Integrating anatomy and physiology of the primary visual pathway: from lgn to cortex," in *Models of Neural Networks IV: Early Vision and Attention*, J. D. Cowan, E. Domany, and J. L. van Hemmen, Eds. New York: Springer-Verlag, 2002.

[5] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann, "Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps," *Science*, vol. 275, pp. 213–215, 1997.

[6] G. Q. Bi and M. M. Poo, "Activity-induced synaptic modifications in hippocampal culture, dependence on spike timing, synaptic strength and cell type," *J. Neuroscience*, vol. 18, pp. 10 464–10 472, 1998.

[7] D. Debanne, B. H. Gähwiler, and S. M. Thompson, "Long-term synaptic plasticity between pairs of individual ca3 pyramidal cells in rat hippocampal slice cultures," *J. Phys.*, vol. 507, pp. 237–247, 1998.

[8] 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*, vol. 395, pp. 37–44, 1998.

[9] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann, "Physiology and anatomy of synaptic connections between thick tufted pyramidal neurons in the developing rat neocortex," *J. Phys.*, vol. 500, pp. 409–440, 1997.

[10] P. D. Roberts and C. C. Bell, "Spike timing dependent synaptic plasticity in biological systems," *Biological Cybernetics*, vol. 87, no. 5–6, pp. 392–403, 2002.

[11] D. Yang and M. M. Poo, "Spike timing-dependent plasticity of neural circuits," *Neuron*, vol. 44, pp. 23–30, 2004.

[12] M. V. Tsodyks and H. Markram, "The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability," *Proc. Nat. Acad. Sci.*, vol. 94, pp. 719–723, 1997.

[13] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press, 2001.

[14] C. Koch, *Biophysics of Computation—Information Processing in Single Neurons*. Oxford, U.K.: Oxford Univ. Press, 1999.

[15] J. F. Feng and D. Brown, "Impact of correlated inputs on the output of the integrate-and-fire models," *Neural Comput.*, vol. 12, pp. 671–692, 2000.

[16] M. Usher, H. G. Schuster, and E. Niebur, "Dynamics of populations of integrate-and-fire neurons, partial synchronization and memory," *Neural Comput.*, vol. 5, pp. 570–586, 1993.

[17] J. J. Hopfield and A. V. M. Herz, "Rapid local synchronization of action potentials: toward computation with coupled integrate-and-fire networks," *Proc. Nat. Acad. Sci.*, vol. 92, pp. 6655–6662, 1995.

[18] W. Gerstner, "Populations of spiking neurons," in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds. Cambridge, MA: MIT Press, 1999.

[19] N. Brunel, "Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons," *J. Computat. Neurosci.*, vol. 8, pp. 183–208, 2000.

[20] N. Levy, D. Horn, I. Meilijson, and E. Ruppin, "Distributed synchrony in a cell assembly of spiking neurons," *Neural Netw.*, vol. 14, pp. 815–824, 2001.

[21] H. C. Tuckwell, *Introduction to Theoretical Neurobiology: Nonlinear and Stochastic Theories*. Cambridge, U.K.: Cambridge Univ. Press, 1988, vol. 2.

[22] L. F. Abbott and T. B. Kepler, "Model neurons: from hodgkin-huxley to hopfield," in *Statistical Mechanics of Neural Networks*, L. Garrido, Ed. New York: Springer-Verlag, 1990.

[23] T. W. Troyer and K. D. Miller, "Integrated-and-fire neurons matched to physiological f-i curves yield high input sensitivity and wide dynamic range," in *Computational Neuroscience: Trends in Research*, J. M. Bower, Ed. New York: Academic, 1997.

[24] S. Song, K. D. Miller, and L. F. Abbott, "Competitive hebbian learning through spike-timing-dependent synaptic plasticity," *Nature Neurosci.*, vol. 3, no. 9, pp. 919–926, 2000.
[25] M. V. Rossum, G. Q. Bi, and G. G. Turrigiano, "Stable hebbian learning from spike-timing-dependent plasticity," *J. Neurosci.*, vol. 20, no. 23, pp. 8812–8821, 2000.
[26] R. Kempter, W. Gerstner, and J. L. van Hemmen, "Hebbian learning and spiking neurons," *Phys. Rev. E*, vol. 59, pp. 4498–4514, 1999.
[27] D. Marr, *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. San Francisco, CA: Freeman, 1982.
[28] V. Lamme, V. R. Rodriguez, and H. Spekreijse, "Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey," *Cerebral Cortex*, vol. 9, no. 4, pp. 406–413, 1999.
[29] G. Holmes, "The organization of the visual cortex in man," in *Proc. Roy. Soc. London Ser. B*, vol. 132, 1945, pp. 348–361.
[30] G. Holmes, "The representation of the visual field in human striate cortex," *Arch. Ophthalmol.*, vol. 109, pp. 816–824, 1991.
[31] S. A. Engel, G. H. Glover, and B. A. Wandell, "Retinotopic organization in human visual cortex and the spatial precision of functional mri," *Cerebral Cortex*, vol. 7, pp. 181–192, 1997.
[32] Y. Trotter and S. Celebrini, "Gaze direction controls response gain in primary visual-cortex neurons," *Nature*, vol. 398, pp. 239–242, 1999.
[33] Y. Trotter, S. Celebrini, B. Stricanne, S. Thorpe, and M. Imbert, "Modulation of neural stereoscopic processing in primate area v1 by the viewing distance," *Science*, vol. 257, pp. 1279–1281, 1992.
[34] B. T. Backus, D. J. Fleet, A. J. Parker, and D. J. Heeger, "Human cortical activity correlates with stereoscopic depth perception," *J. Neurophys.*, vol. 86, pp. 2054–2068, 2001.
[35] B. G. Cumming and A. J. Parker, "Responses of primary visual cortical neurons to binocular disparity without depth perception," *Nature*, vol. 389, pp. 280–283, 1997.
[36] N. Qian, "Binocular disparity and the perception of depth," *Neuron*, vol. 18, pp. 359–368, 1997.
[37] C. Zhou and H. Tao, "Dynamic depth recovery from unsynchronized video streams," in *Proc. Int. Conf. Comput. Vis.*, vol. 2, 2003, pp. 351–358.
[38] J. L. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, pp. 43–77, 1994.
[39] H. A. Mallot, *Computational Vision—Information Processing in Perception and Visual Behavior*. Cambridge, MA: MIT Press, 2001.
[40] F. Wörgötter, A. Cozzi, and V. Gerdes, "A parallel noise-robust algorithm to recover depth information from radial flow fields," *Neural Comput.*, vol. 11, pp. 381–416, 1999.
[41] P. Häfliger, M. Mahowald, and L. Watts, "A spike-based learning neuron in analog vlsi," in *Advances in Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 1996, vol. 9, pp. 692–698.
[42] A. Bofill, A. F. Murray, and D. P. Thompson, "Circuits for vlsi implementation of temporally-asymmetric hebbian learning," in *Advances in Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2002, vol. 9, pp. 1091–1098.
[43] A. Bofill and A. F. Murray, "Synchrony detection and amplification by silicon neurons with stdp synapses," *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1296–1304, May 2004.
[44] K. L. Cameron and A. F. Murray, "Can spike timing dependent plasticity compensate for process mismatch in neuromorphic analogue vlsi?," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 5, Vancouver, BC, Canada, 2004, pp. 748–751.
[45] D. H. Perkel, "A computer program for simulating a network of interacting neurons. i. organization and physiological assumptions," *Comput. Biomed. Res.*, vol. 9, pp. 31–43, 1976.
[46] R. Borisyuk, "Oscillatory activity in the neural networks of spiking elements," *BioSyst.*, vol. 67, pp. 3–16, 2002.
[47] D. Hansel, G. Mato, C. Meunier, and L. Neltner, "On numerical simulations of integrate-and-fire neural networks," *Neural Comput.*, vol. 10, pp. 467–483, 1998.
[48] P. A. Cariani, "Temporal codes and computations for sensory representation and scene analysis," *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1100–1111, May 2004.
[49] S. C. Liu and R. Douglas, "Temporal coding in a silicon network of integrate-and-fire neurons," *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1305–1314, May 2004.
[50] K. Shimonomura and T. Yagi, "A multichip avlsi system emulating orientation selectivity of primary visual cortical cells," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 972–979, Apr. 2005.
[51] S. C. Liu, J. Kramer, G. Indiveri, T. Delbrück, and R. Douglas, "Orientation-selective avlsi spiking neurons," in *Advances in Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2002, vol. 2, pp. 1123–1130.
[52] E. Chicca, D. Badoni, V. Dante, M. Andreagiovanni, G. Salina, L. Carota, S. Fusi, and P. Giudice, "A vlsi recurrent network of integrate-and-fire neurons connected by plastic synapses with long-term memory," *IEEE Trans. Neural Netw.*, vol. 14, no. 5, pp. 1297–1307, May 2003.
[53] R. R. Harrison and C. Koch, "A robust analog vlsi reichardt motion sensor," *Analog Integr. Circuits Signal Process.*, vol. 24, no. 3, pp. 213–229, 2000.
[54] R. Sarpeshkar, J. Kramer, G. Indiveri, and C. Koch, "Analog vlsi architectures for motion processing: from fundamental limits to system applications," *Proc. IEEE*, vol. 84, no. 7, pp. 969–987, Jul. 1996.
[55] T. Delbrück, "Silicon retina with correlation-based, velocity-tuned pixels," *IEEE Trans. Neural Netw.*, vol. 4, no. 3, pp. 529–541, May 1993.
[56] H. C. Jiang and C. Y. Wu, "A 2-D velocity- and direction-selective sensor with bjt-based silicon retina and temporal zero-crossing detector," *IEEE J. Solid-State Circuits*, vol. 34, no. 2, pp. 241–247, Feb. 1999.
[57] J. Kramer, "Compact integrated motion sensor with three-pixel interaction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 4, pp. 455–459, Apr. 1996.
[58] J. Kramer, R. Sarpeshkar, and C. Koch, "Pulse-based analog vlsi velocity sensors," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 44, no. 2, pp. 86–101, Feb. 1997.
[59] R. Etienne-Cummings, J. Van der Spiegel, and P. Mueller, "A focal plane visual motion measurement sensor," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 44, no. 1, pp. 55–66, Jan. 1997.
[60] R. E. Russo and J. Hounsgaard, "Short-term plasticity in turtle dorsal horn neurons mediated by l-type ca^{2+} channels," *Neuroscience*, vol. 61, pp. 191–197, 1994.
[61] V. Booth, J. Rinzel, and O. Kiehn, "A compartmental model of vertebrate motoneurons for ca^{2+}-dependent spiking and plateau potentials under pharmacological treatment," *J. Neurophys.*, vol. 78, pp. 3371–3385, 1997.

Zhijun Yang was born in Yangzhou, China. He received the B.Eng. degree from Central China University of Science and Technology, Wuhan, China, in 1989 and the M.Sc. degree from Nanjing University, Nanjing, China, in 1992, both in electrical engineering. He received the D.Sc. degree from the Department of Computer Science, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, with a dissertation on the theory and application of machine learning, in 1999.

He worked as an Analog Electrical Engineer in Nanjing Panda Electronics Co. Ltd., Nanjing, China, for four years, where he was responsible for video signal circuit design. After receiving his D. Sc. Degree, he was a Senior Software Engineer for Motorola Global Software Center, responsible for developing an embedded Java project for mobile computing. In 2001, he worked as a Research Associate with the Department of Physics, Heriot-Watt University, Edinburgh, U.K. Since 2002, he has worked at Department of Computer Science, Nanjing Normal University, Nanjing, China. Since December 2002, he has also been with the Institute for Integrated Micro and nano Systems (IMNS) of Edinburgh University as a Research Fellow. His research interests include machine learning, mixed signal VLSI circuit design, and bio-inspired software development.

Alan Murray (M'91–SM'94) was born in Edinburgh, U.K., in 1953. He received the B.Sc. degree (hons.) in physics in 1975 and the Ph.D. degree in solid state physics in 1978, both from the University of Edinburgh.

He worked for three years as a Research Physicist (two in Canada), and for three years as an Integrated Circuit Design Engineer. In 1984, he was appointed a Lecturer in electrical engineering at Edinburgh University, became a Reader in 1991, and Professor of Neural Electronics in 1994. He is interested in all aspects of neural computation and hardware issues and applications have been his primary research interest since 1985. In 1986, he developed the "pulse stream" method for neural integration. His interests have since widened to include all aspects of neural computation, particularly hardware-compatible learning schemes, probabilistic neural computation and neural forms that utilize the temporal—and noisy characteristics of analogue VLSI—as well as applications of hardware neural networks. He is also developing a new interest in the interface between silicon and neurobiology, along with colleagues in Biomedical Sciences and in Glasgow University, Glasgow, U.K. He has over 200 publications, including an undergraduate textbook and research texts on neural VLSI, applications of neural networks and noise in neural training (with Peter Edwards).

Dr. Murray is a Fellow of the Institute of Electrical Engineers, U.K., a member of the International Neural Network Society (INNS), and a Fellow of the Royal Society of Edinburgh.

Florentin Wörgötter is a Research Professor at the Bernstein Centre of Computational Neuroscience, Göttingen, Germany, and co-affiliated as a Professor at the Department of Psychology, University of Stirling, Stirling, U.K.

He has undertaken experimental neuroscience research since 1985 and theoretical neuroscience research since 1988. His main research goal is to bring perception and action together in artificial systems (computer vision systems, robots), building on the knowledge gained from theoretical neuroscience. His research is focused on the visual system (perception) and on biophysical machine learning (action). He developed a neuromorphic VLSI algorithm for stereoscopic depth analysis, which is now (as an FPGA implementation) used in several industrial applications. This broad interdisciplinary research has lead to publications in *Nature, Trends in the Neurosciences, Journal of Neuroscience*, and others. In addition, he has published textbook chapters on computational neuroscience of the primary visual pathway. His past research interests were focused on early vision problems; current research is now centerd on problems of data-fusion in the field of active vision as well as the integration of sensor and motor aspects.

Katherine Cameron (S'04) was born in Edinburgh, U.K., in 1979. She received the M.Eng. (hons.) degree in Electronics from the University of Edinburgh, Edinburgh, U.K., in 2002, where she is currently working toward the Ph.D. degree in neuromorphic engineering.

Her current research interests include bio-inspired engineering solutions to analogue computation imperfections, mixed-signal VLSI design and neural computation.

Vasin Boonsobhak (S'03) was born in Bangkok, Thailand, in 1975. He received the B.Eng. (hons.) degree in electrical engineering from Mahanakorn University of Technology, Bangkok, Thailand, in 1996 and the M.Sc. degree in analog and digital integrated circuit design from Imperial College, London, U.K., in 1998. He is currently working toward the Ph.D. degree in electrical engineering at the University of Edinburgh, Edinburgh, U.K.

Since 1996, he has been with the Department of Electronics Engineering, Mahanakorn University of Technology. His current research interests include mixed-mode VLSI systems for image processing, image sensors and asynchronous digital interfaces for interchip connectivity.