# A Spike-Based Saccadic Recognition System

Matthias Oster, Patrick Lichtsteiner, Tobias Delbrück, and Shih-Chii Liu
Institute of Neuroinformatics
University of Zürich and ETH Zürich
Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
Email: mao,patrick,tobi,shih@ini.phys.ethz.ch

*Abstract*— We present a spike-based saccadic recognition system that uses a temporal-derivative silicon retina on a pan-tilt unit and an aVLSI multi-neuron classifier with a time-to-first-spike output coding. By using the spike information during the last 150 ms of a saccadic movement, we generate a reliable, sparse stimulus representation of image patches. We describe a novel classification scheme where the retinal spikes during this time influence the time-to-first spike of classifier neurons which receive the same constant input current. The preferred pattern of the neuron is stored in the synaptic connectivity between the retina and the classifier neuron. We demonstrate the robustness and real-time performance of this recognition scheme on a saccadic system which uses analog VLSI components.

## I. Introduction

Artificial neural architectures can benefit from the consideration of the properties of a specific stimulus representation. Here we describe a spike-based saccadic recognition system that uses a stimulus representation based on the information in the spikes from a retina at the end of a saccade. Saccades are fast eye movements that shift the fovea to a new image location. We use the information contained in the spikes towards the end of a saccade to generate a reliable and sparse stimulus representation that is largely independent of the magnitude of the saccade.

Next, we design a classifier that is suitable for recognizing this type of stimulus representation. We use single neurons that are held in reset during a saccade and then receive spiking input from the retinal pixels towards the end of the saccade. The retina input modulates a constant input current to all neurons. By using a time-to-first-spike coding, a winner-take-all circuit then easily detects the neuron with the strongest input. This classifier is robust with respect to the number of spikes received and reaches its decision in a time that corresponds to the end of the saccade.

We implemented this system in mixed-signal hardware using analog VLSI technology and an asynchronous event-based communication protocol, so that we can demonstrate the speed and robustness of a system that uses asynchronous signalling and components with mismatch (that is, the neurons and synapses). The core components are a temporal-derivative retina chip that encodes temporal contrast changes in spike output [1], and a multi-neuron chip which contains the neurons of the classifier [3]. Spikes are transmitted between the two chips using an asynchronous, event-based real-time communication protocol (address-event representation, AER), implemented by a framework of digital AER components [2]. The

synaptic connectivity is implemented through a programmable connection table. A software framework monitors spikes, programs the connection table, and controls the movements of a pan-tilt unit to generate saccades. The framework for the hardware and software has been described in [5]. The retina chip has a resolution of 64x64 pixels, the multi-neuron chip has 256 neurons with eight synaptic types per neuron. The maximum spike bandwidth between chips is 1.2 MSpikes/s and the transmission spike latency is 2 $\mu$s.

The overview of the system is described in Fig. 1. A software control unit generates the required movement commands for the pan-tilt unit to perform a saccade. At the same time it resets all the classifier neurons by stimulating a global inhibitory neuron and suppressing inputs from the retina. Shortly before the end of the saccade, suppression is released and spikes from the retina stimulate the classifier neurons through the synaptic connection table.
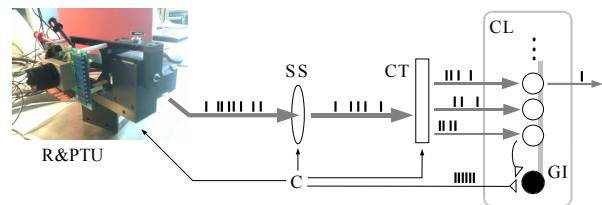


Fig. 1. Setup overview. The temporal-derivative retina (R) is mounted on a pan-tilt unit (PTU). Its output is directed through saccadic suppression (SS) and a connection table (CT) to the classifier chip (CL). Retina and classifier chips are implemented in analog VLSI technology while spikes are transmitted through a real-time digital hardware framework. The operation of software control unit (C), and the role of the global inhibitory neuron (GI) are described in the text.

Although the framework allows online learning and changing the connectivity table dynamically, we do not use this capability in the present work. We focus on the properties of the stimulus representation at the end of the saccade and the classifier network. To demonstrate the performance of the classifier, we compute the synaptic connectivity necessary for recognizing patches from a test image.

## II. Saccades

The human eye uses saccades to shift its foveal vision to objects of interest. We make a simplified model of a saccadic system by mounting the retina chip onto a pan-tilt unit with two stepper motors. The motors are aligned so that the two
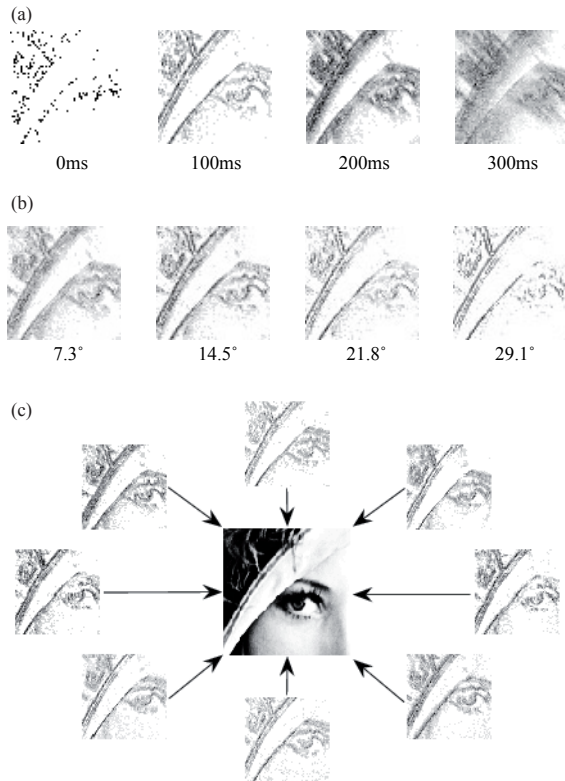
Fig. 2. Saccadic stimulus representation: (a) Effect of different time window lengths: retina spikes are only counted during a time window at the end of the saccade. The number below each figure indicates the time before the end of the saccade, when the suppression is released. Actual window times are slightly larger due to delays when transmitting the start command, as can be seen by the presence of few spikes for a window length of 0ms. (b) Effect of different saccade magnitudes: saccades were performed over different angular distances, from a starting point $45°$ up and to the left over the indicated angular distance to the center. (c) Effect of different directions: saccades were made in steps of $45°$ from an angular distance of $7.3°$ visual angle. The resulting spike count images are shown in the direction from which the saccade was made. The original image is shown in the center. Different directions show slightly different results, e.g. the vertical strand of hair is not visible during a vertical saccade.

axes of rotation intersect, resulting in pitch and yaw rotation. The center of the retina is mounted as close as possible perpendicular to these axes to reduce translational movements. The accelerations of the movements are adjusted so that the center of view of the retina moves in a straight line.

Although our pan-tilt unit can reach moderate speeds up to $150°/s$ (human saccades can reach up to $900°/s$ with a smooth velocity profile [7]), we discovered that smooth acceleration and deceleration are necessary to avoid vibrations of the platform. The retina moves in a straight line from the starting point of a saccade to the target location.

We use the retina in a foveal configuration by mounting a long lens (12mm focal length) that projects a visual field of $9.5°x9.5°$ onto the 64x64 pixels of the chip. The 512x512 'Lenna' image covered $45°$ of visual angle at a distance of 19cm.

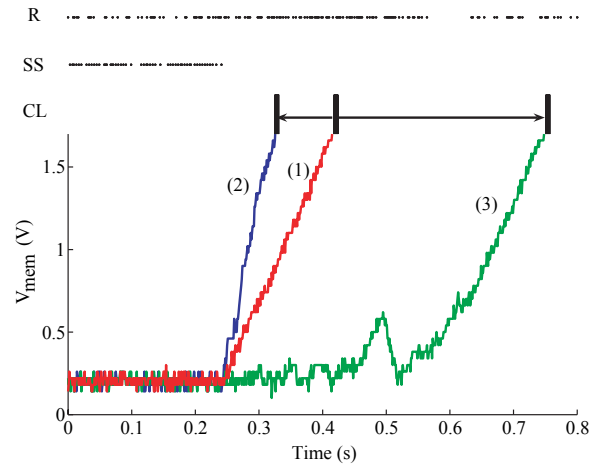It is generally assumed that the biological retina output is



Fig. 3. Time-to-first spike gives the identity of the winning classifier neuron. The top line shows the raster plot of spikes from the retina (R) to the classifier neurons. The second line shows the raster plot of spikes to the inhibitory neuron during the saccadic suppression (SS), thus resetting all classifier neurons during this time. Every dot corresponds to a spike. As soon as the suppression is released, the membrane voltage, $V_{mem}$ of every classifier neuron is charged by a constant input current until it reaches threshold (1). If a neuron receives more excitation or inhibition as input from the retina during this time, it will reach threshold earlier (2) or later (3) respectively.

suppressed in the lateral geniculate nucleus during a saccade, see e.g. [8] for a review. We do the same, but release the suppression before the end of the saccade, and we use the spike information available towards the end of the saccade. In this time window, a retina pixel is only stimulated by edges at the target position of the saccade. A count of spikes in the time window at the end of the saccade gives an image of the target. The number of spikes per pixel determines the contrast of an edge.

We adjusted the length of the time window until we found a clear image (Fig. 2a). This output represents a qualitative optimum between the number of spikes obtained and a smearing of the image. The optimum window length is 150ms which is 15% of of the duration of an average saccade.

We confirmed that the saccade mechanism described here produces a reliable stimulus that is only dependent on the image at the target location, and not on the distance or direction of the saccade (Figs. 2b and c). We do not claim that this stimulus representation is used in a biological system. In our system, a retinal pixel is responsive to changes in illumination with frequencies up to the kHz range and a typical saccade lasts about 1s. In the human eye, a cone is responsive to frequencies up to 55Hz and a saccade lasts in the range of tens of milliseconds [7].

### III. TIME-TO-FIRST-SPIKE CLASSIFIER

Using the retina spikes during a saccade as the input to a classifier means the classifier should have the following properties:

- it should integrate only over a time window at the end of the saccade before it makes a decision.

- the decision should not depend on the total number of spikes from the retina.

We fulfill these requirements by using a classifier where the retina spikes at the end of the saccade excite or inhibit classifier neurons that are all driven by a constant input current.

The classifier consists of an array of excitatory integrate-and-fire neurons and one global inhibitory neuron. Each excitatory neuron represents a stimulus class (Fig. 1). The global inhibitory neuron receives excitation from every array neuron and in turn inhibits all excitatory neurons. The excitatory neurons receive excitatory or inhibitory synaptic connections from the retina pixels. These connections are implemented using the spike-based connection table, which allows any arbitrary connectivity.

During a saccade, classifier excitatory neurons do not receive output spikes from the retina. The neurons are held instead at their reset potential by stimulating the global inhibitory neuron with a high frequency. As soon as the suppression is released, the inhibition stops, and a constant input current starts charging up the excitatory neurons (Fig. 3). If a neuron receives more excitation than inhibition, it will spike earlier than neurons receiving no retina input. The output of the classifier is determined by a time-to-first-spike code after the end of the saccade since the neuron receiving the most excitation will spike first.

It is easy to combine this time-to-first-spike code with the hard winner-take-all circuit discussed in [4]: as soon as the winning first neuron spikes, it excites the global inhibitory neuron, which discharges all neurons to their resting potential $V_0$. The winner itself is reset to its reset potential $V_R$. Since we set $V_R > V_0$, and all neurons receive the same constant input current, the winner will continue to spike and will be the only spiking neuron in the classifier array.

The time-to-first-spike code allows the use of additional signals for setting up the classifier. For example, one of the neurons can be used as a *reference neuron* in that it is only charged by a constant current and receives no retina spikes. It will then spike at a fixed, but adjustable delay after the reset. We can use the output of this reference neuron to tune the classifier neurons: they should be biased so that no classifier neuron spikes before the end of the saccade. We then implement the following biasing rule: if the winning neuron spikes before the reference neuron whose delay is timed to the end of the saccade, the total amount of the excitation is decreased. If the winning neuron spikes only slightly after the reference neuron, the total amount of excitation is increased.

## IV. EXPERIMENTS

We tested whether the spike-based saccadic system can recognize patches on the 'Lenna' image (Fig. 4). All experiments were done under normal office lighting conditions.

We manually selected nine target points of saccades (crosses in Fig. 4(a)) in the image which correspond to high-salient points for an observer. From another nine randomly selected start points (circles in Fig. 4(a)), we performed saccades from
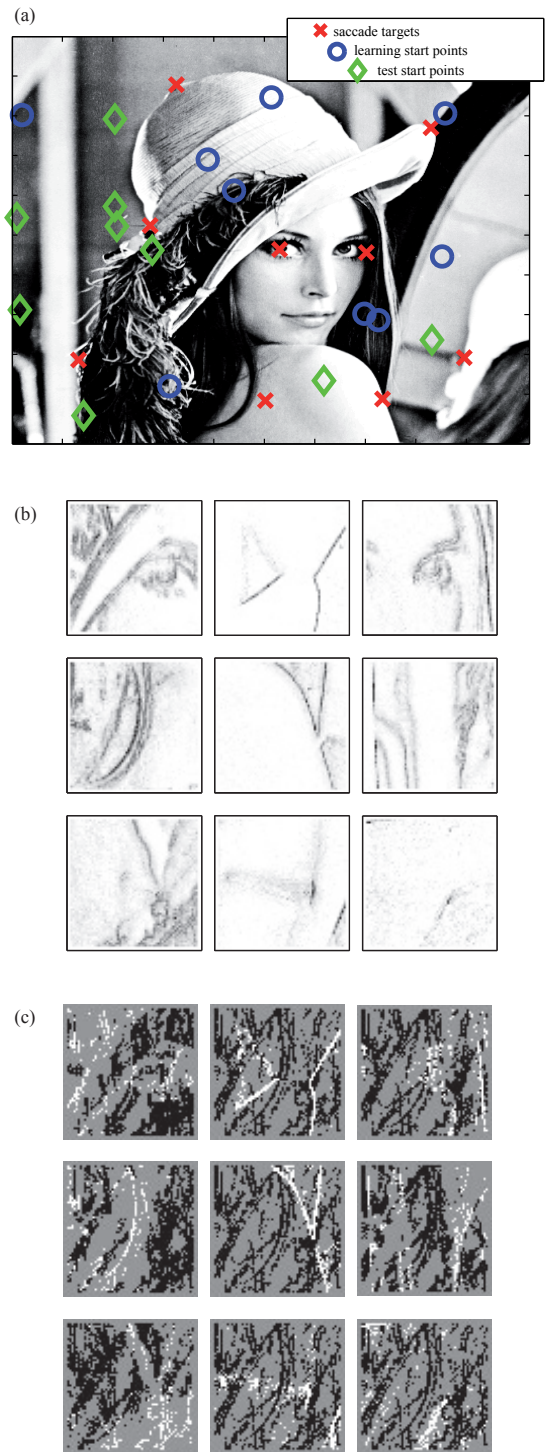


Fig. 4. Experimental procedure. (a) We manually selected 9 points as target points for the saccades (red crosses). For training, we recorded the retina spikes during the saccades that started from each of 9 randomly selected start points (blue circles) to each of the target points. (b) The figures show the spike averages for the 9 patches. (c) The most relevant pixels (see text) were connected either to excitatory or inhibitory synapses of the classifier neurons (white-excitatory, black-inhibitory). Classification performance was determined by response of classifier neurons during the saccades from 9 random starting points (green diamonds in (a)).
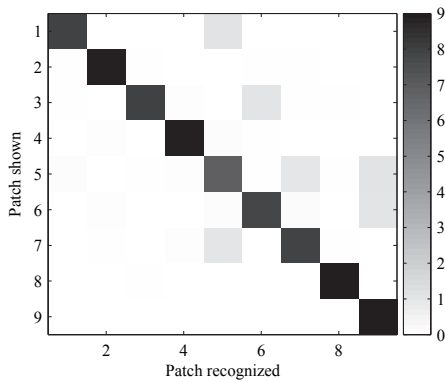
Fig. 5. Hit matrix quantifying the classification performance. There were nine trials from randomly selected start points to each patch. The classifier recognized 92.6% of the patches correctly (entries on the diagonal).

each of these start points to every target point. The recorded spikes from the retina during the saccades served as a training set.

The training samples were averaged to obtain a class representation of every target point (Fig. 4(b)). For each class, the pixels were categorized as excitatory or inhibitory, depending on whether the spike count of the pixel was larger or smaller than the average pixel count over all classes. We then sorted the pixels by their variance across the classes. Since only binary synaptic efficacies were available, we had to normalize the input to the neurons by assigning the same number of excitatory and inhibitory connections to each neuron. We used the same number of connections that the class with the lowest number of connections received, which turned out to be 1365 (out of the 4096 retinal pixels) in these experiments. These excitatory and inhibitory connections were programmed into the connection table (shown in Fig. 4(c)).

To test the performance of the system, we selected 9 additional random start points (diamonds in Fig. 4) and recorded the response of the classifier for saccades from every start point to every target point. From the hit matrix shown in Fig. 5, we computed the performance of the classifier. In 92.6% of the 81 trials, the patch was correctly classified (chance level for 9 classes: 11.1%), which is equivalent to a mutual information of 2.83 bit (stimulus entropy for the 9 patches: 3.17 bit).

## V. Discussion

We presented a spike-based saccadic recognition system that uses the information contained in the spikes of a temporal contrast spike-based retina at the end of a saccade for recognition of patches on a test image. Recognition takes place in real-time; the system uses only as many neurons as classes (or patches); and the classifier output is determined in one spike. We also discussed a new form of temporal stimulus representation, consisting of retina spikes in the last 150 ms of the saccade. The classifier based on a time-to-first-spike code is robust and provides a recognition rate of 92.6% for the 9 patches in this experiment.

## VI. Acknowledgments

## References

[1] P. Lichtsteiner and T. Delbrück, "64x64 event-driven logarithmic temporal derivative silicon retina," in *Proceedings of the 2005 IEEE Workshop on Charge-Coupled Devices and Advanced Imager Sensors*, June 2005, Nagao Prefecture, Japan, 9–11 June.

[2] V. Dante, P. Del Guidice, and A. M. Whatley, "Hardware and software for interfacing to address-event based neuromorphic systems," in *The Neuromorphic Engineer*, vol. 2, no. 1, 2005, pp. 5–6.

[3] S.-C. Liu and M. Oster, "Feature competition in a spike-based winner-take-all VLSI network," in *Proceedings of the 2006 IEEE International Symposium on Circuits and Systems*, May 2006, pp. 3634–3637.

[4] M. Oster and S.-C. Liu, "Spiking inputs to a winner-take-all network," in *Advances in Neural Information Processing Systems*, Y. Weiss, B. Scholköpf, and J. Platt, Eds.,Cambridge, MA: MIT Press, 2006, vol. 18, pp. 1051–1058.

[5] M. Oster, A. M. Whatley, S.-C. Liu, and R. J. Douglas, "A hardware/software framework for real-time spiking systems," in *International Conference on Artificial Neural Networks (ICANN)*, W. Duch, J. Kacprzyk, and E. Oja, Eds. Warsaw, Poland: Springer Lecture Notes, Sep 2005, vol. 3696, pp. 161–166.

[6] S. Thorpe, R. Guyonneau, N. Guilbaud, J. Allegraud, and R. VanRullen, "Spikenet: real-time visual processing with one spike per neuron," *Neurocomputing*, vol. 58-60, pp. 857–864, 2004.

[7] E. R. Kandel, J. Schwartz, and T. M. Jessell, *Principles of Neural Science*, Mc Graw Hill, 2000, pp. 510, 784–786.

[8] J. Ross, D. Burr, and M. C., "Suppression of the magnocellular pathway during saccades," *Behavioural Brain Research*, vol. 80(1-2), pp. 1–8, 1996.