Suggestions for a Biologically Inspired Spiking Retina using Order-based Coding

Simon J. Thorpe[1, 3], Adrien Brilhault[1, 3], José-Antonio Perez-Carrasco[2]
[1]CNRS Université Toulouse 3 , Centre de Recherche Cerveau & Cognition,Toulouse, France
[2] Instituto de Microelectrónica de Sevilla (IMSE-CNM-CSIC), 41092 Sevilla, Spain
[3]SpikeNet Technology S.A.R.L, Labège, France. Email : simon.thorpe@cerco.ups-tlse.fr

Abstract—**This paper discusses some new suggestions for designing hardware vision systems that take inspiration from spike-based biological image processing. The key idea is to modify already existing Address Event Representation (AER) designs so that there is a periodic reset signal that can be generated every time some predefined proportion of "neurons" has emitted a spike. Each "neuron" only emits at most one spike per processing cycle, but the most strongly activated neurons fire first, ensuring that information about the image is transmitted with maximum efficiency. Simulations demonstrate that this sort of design can allow image reconstruction under conditions where only a few percent of the units have emitted a spike. This means that the reset signal can be triggered at high rates allowing images to be processed at very high clock rates but in a way that could automatically adjust to variations in environmental conditions including scene contrast.**

I. Introduction

In recent years it has become increasingly common for engineers designing image-processing systems to take inspiration from tricks that are used by biological vision systems. There are strong arguments in favor of such an approach. Both artificial and biological vision systems face very similar challenges. For example, both types of systems are often faced with the challenge of being able to identify, categorize and locate objects and events in complex cluttered scenes that are changing constantly. For both types of systems there is a clear requirement that these functions are performed as rapidly and as reliably as possible, using the most energy efficient hardware and while keeping weight and space requirements to the strict minimum. In the case of biological vision systems, it is clear that any variant that is either faster, more reliable, more energy-efficient, smaller or lighter than the competition will gain a selective advantage and become predominant as a result of natural selection. The result of hundreds of millions of year of natural selection can be seen in the designs used in our own visual systems, as well as those of much smaller organisms such as flies. For example, humans can make saccades towards animals in complex natural scenes in as little as 120-130 ms [1], and saccades towards human faces are even faster. Such levels of performance are achieved despite hardware constraints that would probably lead any electronic engineer to despair – conduction velocities of nerve fibers within the brain are typically only 1-2 m.s^{-1}, and the "clock speed" at which neurons operate (i.e. the maximum speed at which they can emit pulses) is limited to below 1KHz. Relative to today's consumer electronics, such constraints seem almost ridiculous. For example, the latest GPUs can achieve 2.4 Teraflops and have memory bandwidths of as much at 230 Gbytes/sec. The fact that we do not yet know how to reproduce the processing sophistication of the primate visual system may have more to do with our lack of understanding of the underlying computational architecture than a lack of processing power per se. Recent anatomical studies have shown that the neocortex of an adult human contains roughly 16 billion neurons [2], and probably less than 25% of these are directly involved in visual processing. Reproducing such hardware in an artificial system may be feasible within the next decade or so, assuming that we can understand the nature of the underlying computations.

II. Image Compression in the Retina

One particularly clear situation where efficient design is critical can be seen in the retina. In humans, there are only about 1 million retinal ganglion cells in each eye. These are the cells whose axons project to the brain. Although 1 million fibers may seem generous, it would only corresponds to an image 1000 x 1000 pixels in size, even if each fiber corresponded to one "pixel". But, as is well known, the optic nerve has to encode not only luminance, but also color, since there are three different types of cones, meaning that without very efficient encoding, the effective resolution of the retina would be even lower. In fact, those 1 million fibers have to encode information from around 130 million photoreceptors. Clearly, we can expect that the way information is encoded in the retina must be very highly optimized by natural selection.

Since the pioneering work of Lord Adrian in the 1920s it has almost universally been assumed that the fibers in the optic nerve transmit information in the form of a rate code – since more strongly activated cells fire at higher rates. Furthermore, the center-surround receptive field organization of retinal ganglion cells means that, to a first approximation, the retina can be thought of as performing a sort of local convolution on the image. Another key point is that retinal ganglion cells are divided into "On-center" cells (that are maximally excited by a bright point on a dark background) and "Off-center" cells that respond best to dark points on a bright background.

978-1-4244-5309-2/10/$26.00 ©2010 IEEE

While the idea of coding the image using a rate code may seem plausible, recent experimental work has actually ruled it out in the mouse retina, since the amount of information available by counting spikes within a given amount of time was insufficient to explain the animal's behavioral performance[3]. Other coding schemes in which information is encoded in the details of spike timing seem inevitable. In one such scheme, originally proposed by Thorpe [4], it is the relative order of firing across the population of neurons that is used (see also [5]). The idea follows naturally from the fact that an integrate-and-fire neuron can be thought of as a capacitance with a threshold. In response to a visual stimulus, retinal ganglion cells will charge up progressively until they reach a threshold for generating a spike, and the time taken to reach threshold will depend on how well the stimulus matches the cell's receptive field. For example, one would expect an on-center receptive field to fire very quickly when the image contains an appropriately positioned bright spot. Simulation studies have demonstrated that by using the order in which the cells fire, it is possible to reconstruct an image sufficiently well to allow the key objects to be identified even when less than 1% of the cells in the retina have had time to emit a spike [6]. Furthermore, the idea that relative spike timing can be used as an efficient code has recently been demonstrated experimentally in the salamander retina [7].

In this paper, we discuss ways in which this sort of order based coding could be integrated into a next generation of spike-based hardware.

III. Current Systems for Spike-based Processing

The possibility of using spike-based processing in hardware systems has become increasingly popular in recent years. Starting with early work at Caltech, the notion of Address Event Representation (AER) has been used in a number of systems. The basic idea is that communications between devices can be thought of in terms of sending sequences of spikes, where each spike is encoded in terms of the identity of the neuron that has spiked. For example, in the CAVIAR project, this sort of system has been used to develop Convolutional Networks that can implement a number of interesting processing architectures [8]. Other related work from the group at ETH in Zurich has resulted in the development of a 128 by 128 pixel temporal-contrast retina device that uses an asynchronous mode of processing [9]. Each pixel is effectively associated with two spiking channels. One generates a spike when there is an increase in the luminance (positive time-derivative) of a pixel that exceeds a certain threshold value. The other channel generates a spike when there is a decrease in luminance (negative time-derivative). When a spike is generated, the pixel is effectively reset and a new spike is only generated when the luminance has changed again. In a static world, spiking is virtually absent, but as soon as there is motion, large numbers of spikes can be generated. Importantly, since there is no notion of a clock in the system, there is nothing like the fixed frame rate that characterizes virtually all conventional imaging technology.

However, it is important to realize that virtually all these spike-based coding schemes effectively assume that the underlying coding is rate based. For example, in the CAVIAR project, the temporal-contrast retina chip generates spikes that depend on the luminance time-derivative value at each pixel. High luminance derivative at a given pixel will result in a high density of spikes from the corresponding neuron in the chip's output. This spike sequence can then be fed to a convolution chip that adds the weights of the convolution into a new array of neurons, and the correct result can be obtained by adding the convolution values more frequently for the highly active pixels that have high illumination derivative levels. The strategy certainly works, but given the previous theoretical and modeling work, it seems likely that the computations could be made more efficient by using a coding scheme in which there is a periodic reset, and information encoded not in the firing rate of the neurons, but rather in their order of firing. This sort of temporally encoded spiking has already been used in some hardware systems (see for example, [10]).

In the remainder of this paper we will outline a design for a retinal chip that uses order of firing to encode local spatial contrast, and show how such a system is potentially capable of transmitting image information extremely efficiently, using a very small number of spikes. The proposed system builds on already existing technologies that have been developed for performing spike based processing of image contrast [11].

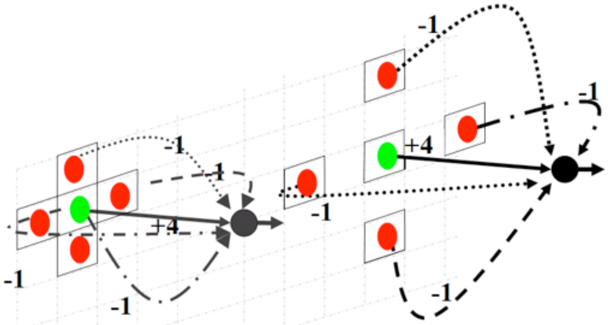


Figure 1. Basic design for the connections between a photoreceptor array and the summation units. At the finest scale (left), each "neuron" receives a strong positive current from the central photoreceptor and negative current inputs from the four neighbours. The centre has a weight such that the circuit effectively performs a local convolution. The panel on the right shows how a coarser convolution can be obtained by using more widely spaced inputs.

IV. A Proposal for Spike-Order based Processing

Consider a device containing an array of photoreceptors that each generates a current that depends purely on local luminance. Suppose that each photoreceptor injects current into a local summing point, but also (with negative polarity, and via a resistive circuit that produces a four-fold reduction in current) into the neighboring points. At a fixed point in time, all the summation points are reset to zero, and then they start accumulating either negative or positive current, effectively performing a local convolution on the image as illustrated in figure 1. This particular choice of coefficients is obviously only one of a wide range of convolutions that could potentially performed, although the design would be more complicated with larger convolutions. Larger range on-center/off-center kernels can be implemented via diffusive networks (in practice, up to about 20 pixels wide [12]). Let us then suppose that for each summing point, there are two voltage thresholds that determine if and when spikes are initiated. As soon as the voltage reaches the positive threshold, a spike is added to the output stream with the appropriate *x,y* location and with a polarity corresponding to an On-center response (see Figure 2). If the negative threshold is exceeded, an "Off-center" spike is added to the list. In both cases the summing point is then disabled so that only one spike per point can be generated.

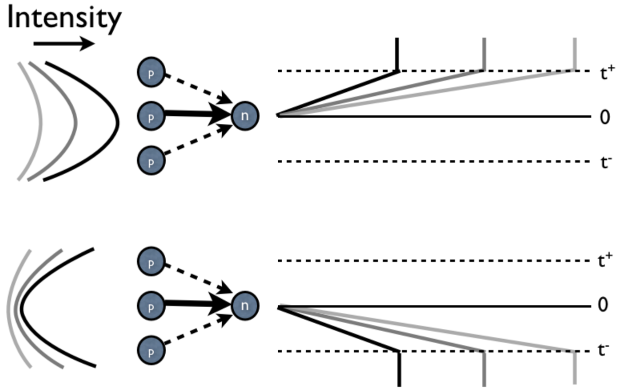
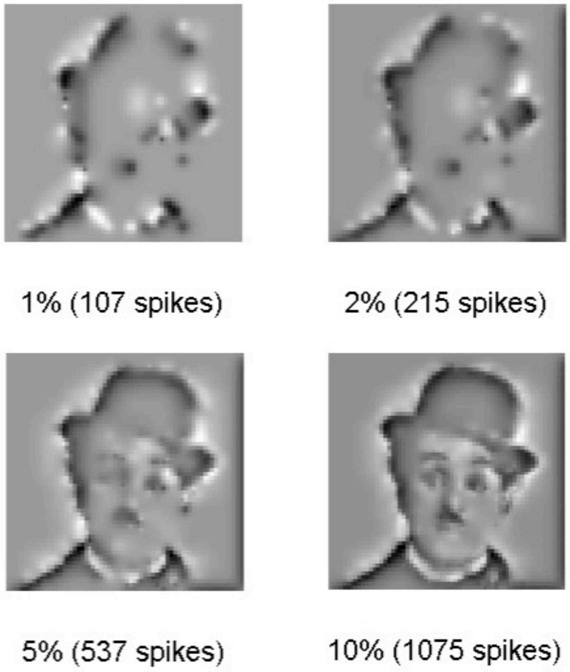


Figure 2. The principle of contrast to latency convertion. Photoreceptors (p) feeding into a "neuron" (n) with weights that perform a convolution on the image. Starting from an initial value at reset, the neuron charges until it reaches either the positive threshold (t+) or the negative one (t-). The top panel shows how spike latency in the positive channel will vary with contrast (high values producing shorter latencies), while the lower panel shows spikes in the negative (Off-center) channel when the image contrast is inversed.

The output of the chip will generate a sequence of spikes in which the ordering corresponds to the relative convolution values at the various points in the image. The interesting point of such a design is that since the highest value convolutions are sent first, the information transfer per spike is optimized. Indeed, as we will show, a downstream mechanism can then be used to reconstruct the image by inserting the receptive field of the active neurons in the appropriate position and with the appropriate polarity and scale. By progressively reducing the weighting depending on the order, it is possible to produce a high quality reconstructing of the original image using a surprisingly small number of spikes. However, this desensitization mechanism may not be strictly necessary when the percentage of cells that are allowed to fire is kept low.

Of course, the time required to for the first image point to reach one of the two thresholds will depend on a number of factors including the contrast of the image – low contrast images will take longer to generate spikes. But, the other factor will be the threshold values that in principle can be arbitrarily small. With very low thresholds, the chip will generate large numbers of spikes very rapidly, but unless there was a lot of noise in the circuits, the ordering should remain relatively constant irrespective of the actual threshold values used.



1% (107 spikes)

2% (215 spikes)

5% (537 spikes)

10% (1075 spikes)

Figure 3. Recontruction of an image using rank-order based coding. We simulated a retina with a resolution of 64x64 pixels, and three spatial scales using the simple 5 point convolution illustrated in figure 1. Even when only a few hundred spikes have been propagated, it is clear that the image is clearly recognizable.

One interesting feature of such a circuit is that it would be possible to reset the chip once a certain fixed number of spikes has been generated. Consider the case of a hypothetical device with 64*64 pixels that uses convolutions shown in figure 1 at three scale, i.e. with separations of 1, 2 and 4 pixels. If each convolution has two different polarities corresponding to ON- and OFF-center responses, this means there would be a total of 10752 different spike identities. Figure 3 shows that even when only the first few hundred spikes have been propagated, it is already to reconstruct the input image with enough detail for recognition. To obtain this sort of reconstruction, we use a weighting function that gives the highest impact to the first inputs to fire, and then progressively decrease the weighting with increasing rank[6]. As a result, it would be possible to reset the chip when only a few percent of the "neurons" have fired. Using reasonable values for the spike generation process and typical values derived from the designs used in the CAVIAR chip sets, we estimate that effective frame rates as high as 10

KHz should be feasible, allowing the chip to operate under conditions impractical with conventional designs.

Note that in many ways this sort of chip design provides a compromise between conventional frame-based imaging designs in which the image is transmitted with a fixed frame rate, and the sorts of fully asynchronous and frame-free approaches used by Delbruck and others. Effectively, the frame rate can be dynamically modified according to the quality of the images being processing, slowing the effective frame rate in the case of low contrast images, and allowing extremely high rates in the case of high contrast images.

V. Practical Considerations

A number of practical circuit oriented considerations are required to map the ideas outlined above to functional hardware. First, photocurrents vary from a few femto amperes to fractions of microamperes. Consequently, integrating directly photocurrents yields systems where timings depend directly on illumination level, resulting in delays that may vary up to 9 decades. This is absolutely impractical, unless illumination conditions are restricted. Therefore, in general, a first step would be to scale photocurrents to a fixed level of "*pixel operation current*", preserving scene contrast.

Second, manipulating photocurrents (such as a simple scaling) with present day CMOS technology results in excessive inter-pixel mismatch (spatial fixed pattern noise). Unless direct manipulation of photocurrents is avoided [9, 11], some calibration means will be required [12]. Although present day reported calibration circuits are bulky and offer rather low precision, the combination of CMOS technology with new nanoscale memory devices could be a promising solution [13].

VI. Conclusions

Although retina chips based on the ideas presented here have not yet been implemented, there is every reason to believe that they could have very interesting properties. None of the basic features require technologies that have not already been implemented in previous systems. Simulation work has already shown that the principle of order based coding can be remarkably powerful [14]. Furthermore, this sort of coding can be coupled with other biologically inspired mechanisms such as Spike Time Dependent Plasticity (STDP) to produce systems capable of learning to detect frequently encountered stimuli using purely unsupervised learning techniques [15, 16]. Such principles can also be included in electronic devices.

[1] H. Kirchner and S. J. Thorpe, Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Res, 2006, vol 46, pp. 1762-76.

[2] F. A. Azevedo, et al., Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J Comp Neurol, 2009, vol 513, pp. 532-541.

[3] A. L. Jacobs, et al., Ruling out and ruling in neural codes. Proc Natl Acad Sci U S A, 2009, vol 106, pp. 5936-41.

[4] S. J. Thorpe, *Spike arrival times: A highly efficient coding scheme for neural networks*, in *Parallel processing in neural systems and computers*, R. Eckmiller, G. Hartmann, and G. Hauske, Editors. 1990, Elsevier: North-Holland. p. 91-94.

[5] R. VanRullen, R. Guyonneau, and S. J. Thorpe, Spike times make sense. Trends Neurosci, 2005, vol 28, pp. 1-4.

[6] R. VanRullen and S. J. Thorpe, Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. Neural Comput, 2001, vol 13, pp. 1255-83.

[7] T. Gollisch and M. Meister, Rapid neural coding in the retina with relative spike latencies. Science, 2008, vol 319, pp. 1108-11.

[8] R. Serrano-Gotarredona, et al., CAVIAR: A 45k Neuron, 5M Synapse, 12G Connects/s AER Hardware Sensory-Processing-Learning-Actuating System for High-Speed Visual Object Recognition and Tracking. IEEE Transactions on Neural Networks, 2009, vol 20, pp. 1417-1438.

[9] P. Lichsteiner, C. Posch, and T. Delbruck, A 128x128 120 dB 15 µs Latency Asynchronous Temporal Contrast Vision Sensor. IEEE Journal of Solid State Circuits, 2008, vol 43, pp. 566-576.

[10] S. S. Chen and A. Bermak, Arbitrated time-to-first spike CMOS image sensor with on-chip histogram equalization. Ieee Transactions on Very Large Scale Integration (Vlsi) Systems, 2007, vol 15, pp. 346-357.

[11] P. F. Ruedi, et al., A 128x128, pixel 120-dB dynamic-range vision-sensor chip for image contrast and orientation extraction. Ieee Journal of Solid-State Circuits, 2003, vol 38, pp. 2325-2333.

[12] J. Costas-Santos, T. Serrano-Gotarredona, R. Serrano-Gotarredona, and B. Linares-Barranco, A spatial contrast retina with on-chip calibration for neuromorphic spike-based AER vision systems. Ieee Transactions on Circuits and Systems I-Regular Papers, 2007, vol 54, pp. 1444-1458.

[13] J. Borghetti, et al., A hybrid nanomemristor/transistor logic circuit capable of self-programming. Proceedings of the National Academy of Sciences of the United States of America, 2009, vol 106, pp. 1699-1703.

[14] S. J. Thorpe, R. Guyonneau, N. Guilbaud, J. M. Allegraud, and R. Vanrullen, SpikeNet: Real-time visual processing with one spike per neuron. Neurocomputing, 2004, vol 58-60, pp. 857-864.

[15] T. Masquelier, R. Guyonneau, and S. J. Thorpe, Competitive STDP-Based Spike Pattern Learning. Neural Comput, 2009, vol 21, pp. 1259-1276.

[16] T. Masquelier and S. J. Thorpe, Unsupervised Learning of Visual Features through Spike Timing Dependent Plasticity. PLoS Comput Biol, 2007, vol 3, pp. e31.