Spike-Based Image Processing: Can We Reproduce Biological Vision in Hardware?

Simon J. Thorpe

Centre de Recherche Cerveau & Cognition, Toulouse, France
simon.thorpe@cerco.ups-tlse.fr
http://www.cerco.ups-tlse.fr

Abstract. Over the past 15 years, we have developed software image processing systems that attempt to reproduce the sorts of spike-based processing strategies used in biological vision. The basic idea is that sophisticated visual processing can be achieved with a single wave of spikes by using the relative timing of spikes in different neurons as an efficient code. While software simulations are certainly an option, it is now becoming clear that it may well be possible to reproduce the same sorts of ideas in specific hardware. Firstly, several groups have now developed spiking retina chips in which the pixel elements send the equivalent of spikes in response to particular events such as increases or a decreases in local luminance. Importantly, such chips are fully asynchronous, allowing image processing to break free of the standard frame based approach. We have recently shown how simple neural network architectures can use the output of such dynamic spiking retinas to perform sophisticated tasks by using a biologically inspired learning rule based on Spike-Time Dependent Plasticity (STDP). Such systems can learn to detect meaningful patterns that repeat in a purely unsupervised way. For example, after just a few minutes of training, a network composed of a first layer of 60 neurons and a second layer of 10 neurons was able to form neurons that could effectively count the number of cars going by on the different lanes of a freeway. For the moment, this work has just used simulations. However, there is a real possibility that the same processing strategies could be implemented in memristor-based hardware devices. If so, it will become possible to build intelligent image processing systems capable of learning to recognize significant events without the need for conventional computational hardware.

1 Introduction

Biological vision systems can make very fast decisions despite hardware constraints that would lead most hardware engineers to despair. For example, biological neurons only emit electrical impulses, or spikes, a few hundred times a second at best, whereas electronic components can switch states billions of times a second. Likewise, the conduction velocity for transmitting spikes from one place in the cortex to another are typically only 1-2 $m.s^{-1}$ whereas signals can be transmitted within electronic devices

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part I, LNCS 7583, pp. 516–521, 2012.

at speeds approaching the speed of light. And yet, despite these limitations, biological systems work remarkably well. For example, in flies, the input-output delay necessary for a change in flight direction in response to a modification in the visual input can be as short as 20 ms - a value that includes not just the visual processing, but the initiation and modification of the motor response. And in humans, my own group has demonstrated that accurate saccades to small faces only 1-2° across and at eccentricities of 7-10° can be initiated after only 100-110 ms [2]. Again, this number includes the entire pathway from photoreceptor, to visual processing and the triggering of the motor response.

How can brains perform challenging tasks like face detection and localization in such so little time? Clearly, the selection pressure during the evolution of the visual system will have put a very high priority on fast processing. If our ancestors were not able to detect the presence of a predator in peripheral vision rapidly, then their survival hopes would be very impaired. So we can assume that biological vision systems will have made optimal use of the available hardware, and this means that if an electronically implemented system was designed that used the equivalent algorithms, it is plausible that it could achieve the same tasks with speeds that are orders of magnitude faster than their biological equivalents. Is there any chance of this happening in the foreseeable future? In my view, the answer to this question is yes. And in this presentation I will sketch how I think technology may move in the coming years.

2 Spike-Based Processing: Software Implementations

At the end of the 1980s it was already clear that the speed of biological vision posed a major problem for conventional views on how information gets processed [11]. The fact that neurons at the highest levels of the primate vision system can respond selectively to complex visual stimuli such as faces just 100 ms after stimulus onset means that each of the roughly 10 processing stages between the retinal photoreceptors and the top end of the visual system only has about 10 ms to reach a decision. And since the firing rates of cortical neurons rarely fire at above about 100 Hz, it means that the processing can presumably be done in a system where each neuron only has time to emit a single spike. This effectively rules out the conventional view that neurons transmit information using a firing rate code in which spike frequency is used to encode analogue values. As an alternative, I proposed that, even when only one spike per neuron is available, information can be encoded in the relatively timing of spikes across a population of neurons [12]. It is an idea that has been demonstrated experimentally using recordings from the salamander retina [4].

By ignoring the firing rate of neurons, and just concentrating on which cells fire first, we showed that it was perfectly possible to implement neural networks capable of detecting faces even under conditions where no neuron gets to fire more than once [13]. And we went on to show that a simple feed-forward network based on a set of orientation selective feature maps (similar to neurons in cortical area V1) could drive a recognition layer capable of state of the art face identification – at least at the time of publication [3]. These ideas led to the creation of a high-tech spin-off company – SpikeNet Technology – in the summer of 1999. The company has been developing software systems based on these basic principles ever since, and currently employs a staff of 12 in its offices near Toulouse in south-west France (see www.spikenet-technology.com).

However, while software implementations based on spiking neural networks and rank-order coding can be very compact and fast, they still need a standard computer architecture to run. Admittedly, the options for very large scale computing systems are improving every day, and projects like Henry Markram's BlueBrain (http://bluebrain.epfl.ch/) demonstrate that it is indeed possible to simulate even very complex neural systems with current technology [7]. Other groups such as Steve Furber's group at Manchester University are developing specialized chips based on multiple ARM processor cores coupled with very speed interconnection hardware that will hopefully allow networks containing up to a billion spiking neurons to be simulated in real time [10].

But given the nature of the underlying computations, is seems possible that it would be possible to design specific hardware systems that implement the ideas directly, rather than relying on software simulation. It is this possibility that I would like to address in the remainder of this presentation.

3 Spike Based Coding of Sensory Information

In some early simulation studies, we showed that using a spike-based scheme could be a particularly efficient way to encode information about images. Specifically, we showed that a retina equipped with ON- and OFF-centre receptive fields at a range of sizes can be used to transmit an image efficiently using only the order of firing as a code. We found that by the time roughly 1% of the neurons in the retina had fired a single spike, it was possible to recognize most of the objects in a image [14].

In recent years, a number of groups have been working on retina-like chips in which information is transmitted off-chip in the form of a series of events that are functionally quite similar to spikes [6]. Tobi Delbruck's presentation in this symposium illustrates this approach very well. Rather than trying to send information about the image in the form of a series of gray-scale values, these new designs allow information to be transmitted continuously in a completely asynchronous manner.

4 Learning with Spike-Time Dependent Plasticity

How can this sort of spike-like information be used by subsequent processing stages? Over a number of years, we have been looking at the possibility that Spike-Time Dependent Plasticity (STDP) could provide a powerful way for a neuronal system to learn to recognize patterns of afferent activity. In an early study, we reported that when a neuron equipped with STDP is subjected to repeating patterns of spikes, high weights will invariably concentrate on those inputs that fire early during the pattern [5]. This tendency is remarkably robust, and can operate even when there is a considerable amount of added noise. More recently, we extended this work by showing that a single neuron equipped with STDP is capable of learning to respond selectivity to any repeating pattern of incoming spikes, even when the pattern involves no change in the underlying firing rate statistics [8]. Thus, even when the neuron is receiving totally random activity through 2000 different afferents, the existence of a repeating motif involving a subset of those inputs can be detected after just a few tens of repetitions. Furthermore, with a pattern lasting just 50 ms, we were able to show that while the receiving neuron may start to fire anywhere within the pattern, after a relatively short period of time, the neuron gradually fires earlier and earlier so that in the end, it fires only 5-10 ms after the beginning of the pattern.

When multiple neurons are receiving from the same set of afferents, and there are inhibitory connections between the neurons, the system will operate as a competitive learning mechanism [9]. The inhibition acts to prevent two neurons learning to fire to the same pattern, and as a consequence, for a given pattern lasting 50 ms, you may have three different neurons firing a specific times during the pattern. Alternatively, if multiple patterns are present in the inputs, different neurons will learn to respond to different patterns.

5 STDP Learning with a Dynamic Vision Sensor

In some recent work with Olivier Bichler and other colleague at the CEA in Saclay we have been looking at how an STDP-based learning mechanism would react to real spiking events provided by one of Tobi Delbruck's Dynamic Vision Sensor chips [1]. In this initial study, we used a dataset available on the web that was produced by recording the spikes generated by one of the chips in response to cars travelling along a six-lane freeway. The imaging chip has a resolution of 128*128 pixels, and each pixel can generate one of two types of spiking events – one corresponding to a local increase in luminance (roughly equivalent to an "ON" event), the other corresponding to a local decrease (an "OFF" event) [6]. This makes a total of 32,768 afferents that were each connected to a first layer containing 60 neurons, each of which implemented a modified STDP rule and had inhibitory connections to the other neurons in the same layer. Each of the 60 first layer neurons was then connected to a second layer of containing 10 neurons. Following the presentation of just a few minutes of data, it was noted that the neurons in the first layer had learned to respond selectively to cars moving a particular locations in the image – a result that simply reflects the fact that these were the patterns of activation that reoccurred over and over again in the incoming data. But even more remarkably, the neurons in the second layer had learned to respond to repeating sequences of activation in the first layer of units. Specifically, this meant that they effectively ended up counting cars passing on each of the six lanes of the freeway. It is important to realize that this learning process occurred with absolutely no supervision. There was no need to provide any labeling of the data – the system simply learned to encode the most "meaningful" reoccurring events in the incoming data.

6 Towards Memristor Based Hardware Implementations

For the moment, this work has been done entirely using software simulations – specifically using a software simulation package developed by Olivier Bichler called XNet. However, the longer-term aim of the project is to try and find a way of implementing the same sorts of STDP-based learning algorithms in Memristor-based hardware systems. Memristors are a class of semiconductor devices capable of storing information by varying the effective resistance of a junction by applying a high voltage, and it has been demonstrated that such devices could potentially be used to implement an STDP-type learning rule [15]. Ultimately, it could be possible to build a complete imaging processing system using just two basic components – a Dynamic Vision Sensor device for generating the basic spiking data at the input, coupled with a network of spiking neurons arranged in a competitive network and which have memristor-based STDP synapses as inputs. Each neuron in the processing layer would have connections from the input device that could be modified whenever the postsynaptic neuron fires a spike. Specifically, the resistance of the connection would need to be reduced every time the activation of the input was followed by the initiation of a spike in the postsynaptic cell.

While there are still a number of technical issues to be dealt with, it already seems clear that such a system could potentially work. And if successful, it would open the way towards a radically different way of processing images – one that was directly inspired by the way in which biological vision systems operate. One particularly revolutionary aspect would be the complete absence of anything resembling a CPU. All processing would be done by local processing elements, none of which would be more complicated than a simple integrate and fire neuron.

Acknowledgements. Some of the work described here has been supported by the ANR (NAVIG and NEMESIS projects).

References

1. Bichler, O., et al.: Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity. Neural Networks 32, 339–348 (2012)
2. Crouzet, S.M., Kirchner, H., Thorpe, S.J.: Fast saccades towards faces: Face detection in just 100 ms. J. Vis. 10(4), 1–17 (2010)
3. Delorme, A., Thorpe, S.J.: Face identification using one spike per neuron: resistance to image degradations. Neural Netw. 14(6-7), 795–803 (2001)
4. Gollisch, T., Meister, M.: Rapid neural coding in the retina with relative spike latencies. Science 319(5866), 1108–1111 (2008)
5. Guyonneau, R., Vanrullen, R., Thorpe, S.J.: Neurons Tune to the Earliest Spikes Through STDP. Neural Comput. 17(4), 859–879 (2005)
6. Lichtsteiner, P., Posch, C., Delbruck, T.: A 128x128 120 dB 15 mu s latency asynchronous temporal contrast vision sensor. IEEE Journal of Solid-State Circuits 43(2), 566–576 (2008)
7. Markram, H.: The blue brain project. Nat. Rev. Neurosci. 7(2), 153–160 (2006)

8. Masquelier, T., Guyonneau, R., Thorpe, S.J.: Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. PLoS one 3(1), e1377 (2008)
9. Masquelier, T., Guyonneau, R., Thorpe, S.J.: Competitive STDP-based spike pattern learning. Neural computation 21(5), 1259–1276 (2009)
10. Plana, L.A., et al.: SpiNNaker: Design and Implementation of a GALS Multicore System-on-Chip. ACM Journal on Emerging Technologies in Computing Systems 7(4) (2011)
11. Thorpe, S., Imbert, M.: Biological constraints on connectionist modelling. In: Pfeifer, R., et al. (eds.) Connectionism in Perspective, pp. 63–92. Elsevier, Amsterdam (1989)
12. Thorpe, S.J.: Spike arrival times: A highly efficient coding scheme for neural networks. In: Eckmiller, R., Hartmann, G., Hauske, G. (eds.) Parallel Processing in Neural Systems and Computers, pp. 91–94. Elsevier, North-Holland (1990)
13. VanRullen, R., et al.: Face processing using one spike per neurone. Biosystems 48(1-3), 229–239 (1998)
14. VanRullen, R., Thorpe, S.J.: Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. Neural Comput. 13(6), 1255–1283 (2001)
15. Zamarreno-Ramos, C., et al.: On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. Frontiers in Neuroscience (May 2011)