Bio-inspired Vision in France and in Europe

Simon J. Thorpe

Centre de Recherche Cerveau & Cognition UMR 5549, 133, route de Narbonne, 31062, Toulouse, France, and SpikeNet Technology S.A.R.L., Labège, France

simon.thorpe@cerco.ups-tlse.fr

Abstract. Can biology be used as a source of inspiration for artificial vision systems? And can we harness the phenomenal processing power available from the latest computer technology to reproduce some of the capacities of biological vision? In this presentation I will report on some of the work in France and more generally within Europe that has attempted to meet this challenge. A number of different programs have provided finance for this sort of work. I will discuss in particular detail our own work that has investigated the neural underpinnings of Ultra-Rapid Visual Categorisation, our ability to determine rapidly whether a previously unseen natural image contains a target category such as an animal.

1. Introduction

There are many visual tasks where biological visual systems easily outperform even the most sophisticated machine vision systems. In insects such as the fly, behavioural responses need to be initiated within a few tens of milliseconds of a changed in the pattern of optical flow, and the underlying computations can be performed using hardware that weighs only a few milligrams and is extremely energy efficient. In both humans and monkeys, it is known that briefly flashed complex natural scenes can be categorised on the basis of 100-150 ms of processing in both humans (Thorpe et al 1996, VanRullen & Thorpe 2001b) and monkeys (Fabre-Thorpe et al 1998), and neurones in high-order regions of the visual system respond selectively to stimuli such as faces even when the images are being changed 72 times per second (Keysers et al 2001), implying that the processing at each step in the system can be done in as little as 14 ms. Such levels of performance are truly phenomenal given the constraints of the underlying neuronal hardware. Neurones transmit information using pulses (spikes) emitted at rates that rarely exceed 100 or so per second, 30 million times slower than the transistors in a modern CPU functioning at 3.6 GHz!! And, within the cortex, those pulses are transmitted at only 1-2 m/s. How is it that the brain can perform tasks so quickly with components that are intrinsically so slow?

The traditional answer has been to say that the brain is fast because it processes information in parallel – unlike a conventional Von Neuman computer architecture where each operation has to be performed sequentially. But modern computer technology is catching up rapidly. For example, today's microprocessors and graphics chips make extensive use of parallelism. ATI's latest graphics chip, the Radeon X800, uses multiple memory channels to achieve a memory bandwidth of over 28 Gbits/sec and is capable of performing over 200 billion floating point operations per second. Could this sort of phenomenal processing power be harnessed to do brain style visual processing?

The belief that biology can be a source of inspiration for development of new technological solutions is one that has motivated a number of research efforts both in France and more generally within Europe. In France, two recent programs have been relevant. One is the Integrated and Computational Neuroscience Research Initiative, started in 2001 under the direction of Prof Alain Berthoz
(see http://www.recherche.gouv.fr/recherche/fns/neurosciences.htm). This program has provided roughly 9Meuros of funding to around 100 research projects on a range of topics. Themes particularly relevant to the bioengineering question were the call in 2002 that targeted projects related to Time and the Brain. In 2004 there was another specific call for work on brain-machine interfaces. An example of the sort of work that has been funded by this action is the work by Thierry Bal and Gwendal le Masson who have been working on interfacing real and artificial neurons.

A second relevant program has been the ROBEA program (Robotics and Artificial Entities) which includes in its aims interactions between Robotics and Neuroscience, particularly with a view to studying and modelling sensory-motor and cognitive functions (for further information see http://www.laas.fr/robea/eng.html).
However, both the French Neuroscience and Robotics initiatives are relatively small scale relative to the ones at the European level where the Information Society Technologies program (IST) has invested considerable resources into this area. In 2001 the Future Emerging Technologies program launched a proactive initiative on "Neuroinformatics for living" artefacts together with a second call on "Life Like Perception Systems", which together formed the Neuro-IT cluster. A list of the 30 Neuro-IT related projects that have been funded by Europe can be found on the web page at http://www.cordis.lu/ist/fet/ni-sy.htm. Ones particularly related to vision include

- INSIGHT2+,
- ECOVISION http://www.pspc.dibe.unige.it/ecovision/
- BIBA http://www-biba.inrialpes.fr/
- ALAVLSI http://www.ini.unizh.ch/alavlsi/
- CAVIAR http://www.imse.cnm.es/~bernabe/CAVIAR/
- LOCUST http://www.imse.cnm.es/locust/

In 2004 a third initiative was launched called "Bio-inspired Intelligent Information Systems (Bio-I^3) that specifically targets Reverse Engineering of the brain. Further information can be found from the initiatives web site at http://www.cordis.lu/ist/fet/bioit.htm, Projects are currently being evaluated for selection and funding in 2005.

Clearly, space will not allow a detailed presentation of all these projects. Instead I will concentrate on one particular idea that has been actively pursued in Europe and particularly in France, namely the idea of using networks of spiking neurons.

2. Reverse Engineering the Visual System with Spiking Neurons

There has been a great deal of work over the past two decades on artificial neural networks, but conventional artificial neural network approaches ignore one of the most obvious features of real biological neurones – namely, the fact that they send information using fixed amplitude pulses called spikes. Over the past few years, there is increasing evidence that the use of spikes is far from being a mere detail of biology, but radically transforms the sorts of computations that can be performed (Maass & Bishop 1999, Rieke 1997, Thorpe et al 2001). For example, the use of spikes opens up a wide range of largely unexplored coding strategies that include using the relative timing of spikes across populations of neurones, or the order in which they fire (Gautrais & Thorpe 1998, Hopfield 1995, Singer 1999).

The power of rank-order coding can be seen in SpikeNet, an image processing system that we have developed in Toulouse and which can simulate the spiking activity of millions of neurons very efficiently on standard computer hardware. This efficiency stems from a number of reasons including the use of "event-driven" simulation techniques that only perform computations when a neurone fires a spike. Since neurones in the visual system typically only fire about once per second, this greatly reduces the amount of processing required. In addition, it uses a very simplified neural model that ignores many of the details of individual neurons, thus allowing large numbers of neurons to be simulated efficiently (Delorme & Thorpe 2003).

Although in its early stages, SpikeNet has already proved to be useful for real world problems and has been used successfully to develop commercial applications (see http://www.spikenet-technology.com). It is already capable of identifying and localising objects such as vehicles and faces in images in real-time, i.e. at 25 frames per second, using a standard off-the-shelf PC hardware (Thorpe et al 2004). Admittedly, this is currently only true for relatively small input images (up to about 240*192 pixels) and with a relatively small number of objects in the list of targets. However, there is effectively no limit to the number of objects that can be analysed, nor on the size of the images that can be analysed. The only problem is that currently, increasing either value prohibits real time processing. However, by using the latest technology and by adapting the code to take advantage of the phenomenal increases in computing power that are currently being offered, real-time analysis of full-sized video images will be a practical reality within the time-course of the current project.

One of the biggest advantages of SpikeNet is that it is a technology that has been derived directly from fundamental research on the neurophysiology of the primate visual system. Because of this, we can make direct use of insights from computational neuroscience to reverse engineer the system. There is now a huge amount of information concerning the performance of the visual system in a wide range of tasks, the properties of neurones at different levels of the system, the detailed anatomical connectivity of the different areas and the biophysical properties at the cellular level. In parallel with the results of this detailed experimental work, there is also a wide range of different ideas concerning the way in which these different components might function during visual processing (Edelman 1999, Grossberg 2003, Rolls & Deco 2002, Ullman 1996). However, with very few exceptions, these different theoretical ideas are rarely tested in the form of explicit models that can be implemented in computer simulations. Even rarer are any sorts of applications that specifically make use of the details of the computational architecture of the visual system. But the current project is based on the premise that we already possess a substantial proportion of the knowledge required to put together an artificial system that can reproduce at least some of the higher level functions of the human and primate visual system. Implementing such ideas will have two major benefits: first, we will be able to test our understanding of the brain mechanisms of visual processing; second, we will be able to provide new sorts of radically original technology with almost unlimited potential. What has been lacking is a computational framework within which these different ideas can be tested, but we believe that simulations using very large-scale networks of spiking neurones provide just that framework.

3. Basic Mechanisms – From Temporal Order to Selectivity

SpikeNet is based on the idea that the order of firing within a population of cells can be used to encode information. As illustrated in Figure 1, an activation profile, produced for example by an image flashed on the retina will lead to a wave of spikes in which the earliest firing neurons will generally correspond to the most strongly activated neurons. The idea that order of firing can be used to encode information follows naturally from the fact that even the simplest integrate-and-fire neuron model will reach threshold more rapidly when the pattern of inputs matches its selectivity. Evidence in support of such a view came from a theoretical study by VanRullen and Thorpe (VanRullen & Thorpe 2001a) who used a simple model of the receptive field properties of ganglion cells in the retina to demonstrate that when the order of firing of retinal ganglion cells is taken into account, the input image can be reconstructed with sufficient accuracy to allow identification of many stimuli when a few as 0.5 to 1.0% of the cells have fired.

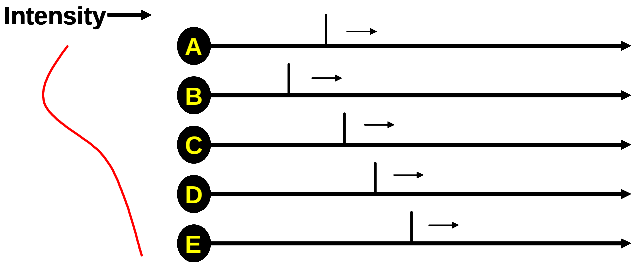


Figure 1. In response to an intensity profile, the neurons A-E will generate a wave of spikes in which the earliest firing units correspond to the most strongly activated cells.

Figure 2 shows how the addition of a feed-forward shunting inhibition circuit can be used to produce neurons that respond selectively as a function of the order of firing. The inhibitory unit S receives equally effective excitation from all the input units, and progressively desensitizes the two target neurons as a function of the number of inputs that have fired. Thus, while the first inputs to fire are fully effective, inputs that fire later produce less and less activation. Under such conditions, the total amount of excitation produced in units F and G will depend on how well the order of firing matches the pattern of weights from the input units. Maximal activation is produced when the order of firing of the inputs matches the set of synaptic weights. For example, since unit F has relatively strong weights from inputs C, D and E, it will respond well in response to the current input pattern.

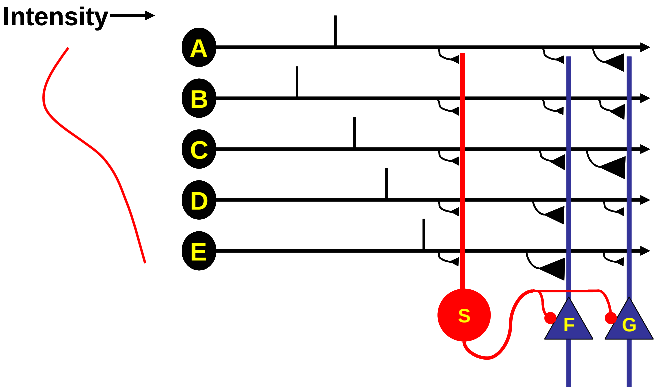


Figure 2. Units F and G receive excitatory synapses from the input units with variable weights. Unit S receives fixed excitatory synapses from all the inputs and generates shunting inhibition that progressively desensitizes units F and G as more and more of the inputs has fired.

Interestingly, recent experimental work on the properties of fast spiking interneurons in the cortex has provided evidence that supports the existence of such a mechanism. For example, in the somatosensory cortex, Swadlow has shown the fast spiking inhibitory interneurons have very small somas and can react very rapidly to the activation of their inputs. Their very short duration action potentials allow them to fire at very high rates up to 600 spikes per second, way above the values seen for the vast majority of cortical neurons. Furthermore, they receive strong but very non-selective inputs from the thalamus with the result that they show essentially no stimulus selectivity (Swadlow & Gusev 2002). Stimulus selectivity of the inhibitory units will also be reduced by the existence of electrical coupling between these cells (Galarreta & Hestrin 2001), which will tend to make the entire population fire together. Finally, there is now good evidence that the inhibitory units produce shunting inhibition in the target neurons. Indeed, the shunting inhibition develops very rapidly following the presentation of a visual stimulus, leading to a three fold increase in soma conductance (Borg-Graham et al 1998).

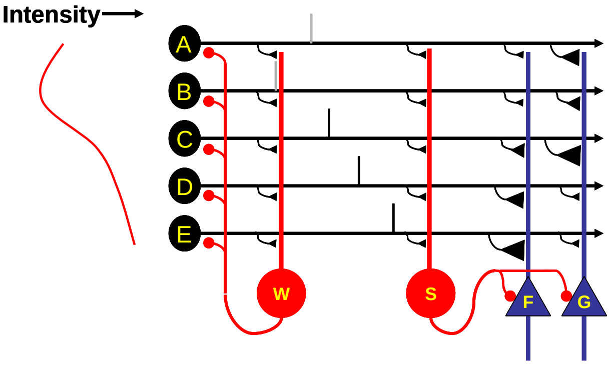


Figure 3. Addition of a feed-back inhibitory circuit (W) allows the implementation of a Winner-Take All operation that prevents more than a limited number of cells in the input layer from firing. Here, the threshold for triggering the feed-back inhibition has been fixed at 3 so that only the first 3 spikes are able to pass.

Feed-forward inhibition is not the only mechanism that may be involved. Figure 3 shows how the addition of a feedback inhibitory circuit can also be computationally useful. In this case, by adjusting the threshold for activation of the inhibition, one can prevent more than a certain number of cells in the input layer from generating a spike. If the inhibition is very strong, such a circuit will effectively perform a Winner-Take-All operation on the inputs, allowing only the most strongly activated input to fire. But with a higher threshold, the operation is more like a k-Winner-Take-All operation.

Together, rapidly acting feed-forward and feedback inhibitory circuits provide mechanisms that can (i) allow the percentage of active cells in the sensory pathways to be controlled, and (ii) make neurons in the next layer sensitive to the order in which the inputs fire. How might such mechanisms allow neurons at later stages to respond selectively to particular inputs?

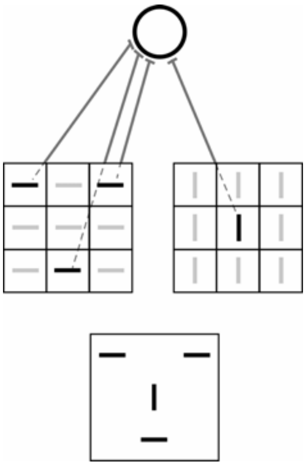
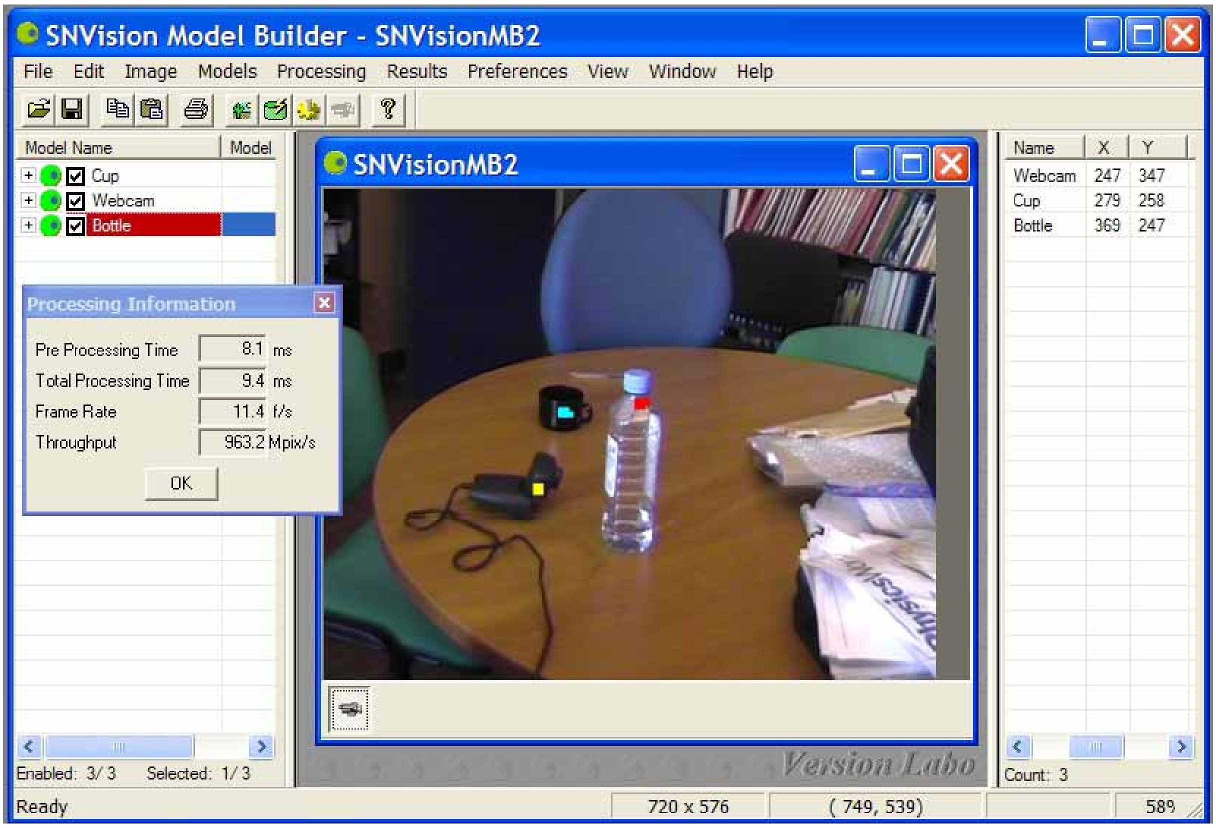


Figure 4. A very simple illustration of how selectivity for fairly high level representations (here a face-like pattern) could be generated by using a limited number of connections and low levels of activation. The output neuron receives connections from two 3*3 arrays of feature detectors, tuned to vertical and horizontal orientation. If, as a result of learning, only four of the input neurons have strong connections, and if when a new input image is presented only the first four input units are allowed to fire, the output neuron will only receive maximal activation when all four features (and no others) are present.

Figure 4 illustrates the very simple case of a neuron receiving from two 3 by 3 unit arrays of orientation selective feature detectors. Suppose that initially, the unit receives weak connections from all 18 input units, but that as a result of learning, all the weights are concentrated on just four of the inputs – 3 horizontally tuned units, and one vertically tuned unit. If we now use the sort of inhibitory circuits described in the preceding paragraphs to limit the number of active units in the input arrays to 4, it should be clear that the chances of all four units matching the input pattern of the cell will be very low. If we fix the threshold for firing in the output neuron at 4 active inputs, only one of the 3060 possible ways of activating four units in the input array will generate a spike in the output neuron. This illustrates that even a very simple mechanism can be selective for something that can be quite high level because in the case illustrated here, the features correspond to something quite specific, namely, a face.

4. Simulations with SpikeNet

Could this sort of simple mechanism be used in the brain? And could it go some way to explaining the extraordinary processing power of the visual system? One approach is to use computer simulations to see whether a system based on such principles is able to perform tasks requiring visual recognition. SpikeNet is a simulation package that aims to do just this. There are actually two different versions of SpikeNet. The original version was developed by Arnaud Delorme during his doctoral thesis (Delorme et al 1999), whereas a more recent version has been developed for image processing. The source code of Delorme's original version can now be downloaded under GNU licence from his website and is described in detail in a recent article (Delorme & Thorpe 2003). The more recent version is a commercial product developed by SpikeNet Technology SARL under licence from the CNRS, but a demonstration version can be downloaded from the company's website at http://www.spikenet-technology.com. While for the commercial version, priority has been given to providing a software package capable of reliable real-time image processing in real-world situations, both versions share the same underlying computational principles. In particular, both versions test the idea that high level visual processing tasks can be performed under conditions where each neuron in the system only gets to fire at most one spike, and where the percentage of neurons that actually emit spikes is kept to a strict minimum. Obviously, in the real nervous system, it is (at least for the foreseeable future) impossible to prevent neurons from firing multiple spikes. As a result, it is unlikely that it will be possible to *prove* that the nervous system can perform high level visual tasks with only one spike per neuron. However, by building a synthetic system such as SpikeNet in which multiple spiking is prevented we can ask just how much visual processing can be achieved under such conditions.



Model Name	Model
☑ Cup	
☑ Webcam	
☑ Bottle	

Enabled: 3/ 3   Selected: 1/ 3

Pre Processing Time	8.1	ms
Total Processing Time	9.4	ms
Frame Rate	11.4	f/s
Throughput	963.2	Mpix/s

Name	X	Y
Webcam	247	347
Cup	279	258
Bottle	369	247

Count: 3

Ready   720 x 576   (749, 539)   589

Figure 5. Interface of SpikeNet. The image to be processed is shown in the central panel. To the left is a list of the different models to be recognized. On the right, the panel shows the locations of the three objects. The Processing Information window shows that the total processing time for analyzing the image was 9.4ms on a 2GHz Pentium 4 based machine and that throughout with a video input was 11.4 frames per second.

Figure 5 is a screen-shot of an application based on the SpikeNet kernel in which a three different objects have been localised in 9.4 ms. In fact, there are no limits on the number of different "models" that can be implemented. The basic architecture used in these simulations is illustrated in Figure 6. A pre-processing stage corresponding roughly to the retina and V1 and the resulting representation contains a set of orientation-selective "maps" each roughly corresponding to "simple" type orientation selective neurons. These neurons reach threshold and fire at a latency that depends on the strength of the input. Thus, if a high contrast vertical edge is present at a particular point in the input image, the neuron with a vertically tuned receptive field at the appropriate location will be one of the first to reach threshold and fire. In this way, the order of firing within V1 contains information about the contours present in the image. When a new visual form (or model) is learnt, an array of neurons with the same dimensions as the input image is created, with one unit for each pixel. All of these neurons share the same pattern of weights which is determined by a learning algorithm which fixes high weights with the earliest firing inputs and low or zero weights for the others. This allows us to recognize (and locate) the same visual form anywhere within the image which effectively provides translation invariance, although the cost in terms of the number of neurons required is extremely high.

Clearly such an arrangement is very different to the one used by the visual system in which several layers of processing are interleaved between V1 and the equivalent of the "recognition layer" which is presumably located in something like inferotemporal cortex. As one progresses through extrastriate areas such as V2 and V4, receptive field sizes increase until inferotemporal cortex where receptive fields can include much of the visual field. It is likely that this is a way of obtaining position invariance which reduces the total number of neurons required. In SpikeNet, we can effectively recognize hundreds or even thousands of different visual forms by creating a new array of recognition units for each new object. But the cost in terms of the number of neurons involved would be astronomical because one would need one unit for each point in the image. With images containing up to a million different pixels, it would be totally prohibitive to adopt such an approach in the human visual system. On the other hand, SpikeNet does not suffer from the binding problem since object identity and position are explicitly coded. Indeed, the fact that the system can report the number and location of each object in the scene is a distinct advantage.

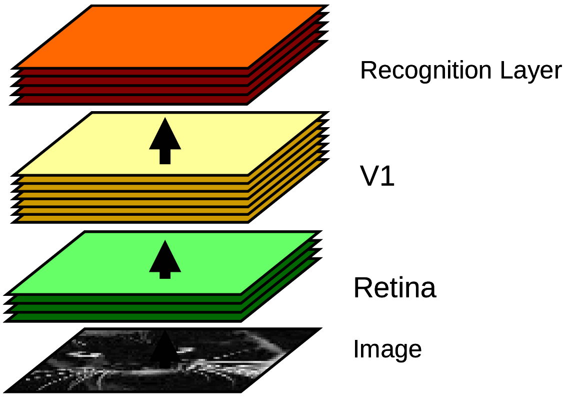


Figure 6. The current version of SpikeNet generally uses a very simple architecture in which arrays of units in the recognition layer receive direct inputs from the equivalent of V1.

5. Conclusions and Perspectives

We make no claim that the current architecture used in SpikeNet is realistic. However, the main point that we wish to make concerns the efficiency of the underlying algorithm. By restricting the number of neurons that fire and using only the first 1-2% of cells that fire, highly selective visual responses can be produced very rapidly. These results allow us to draw the following tentative conclusions.

Single spike coding is viable. The first point is that sophisticated visual processing with just one spike per neuron is clearly possible, despite that fact that with only one spike, traditional coding schemes based on determining the firing rates of individual neurons are ruled out.

Pure feed-forward mechanisms are computationally powerful. Although SpikeNet can include both horizontal and feedback connection patterns, the current version does not use them. Despite this, accurate identification is possible even with noisy images and at low contrasts**.** Clearly, the initial feed-forward wave of processing is capable of considerably more than is conventionally assumed.

Sparse coding is very efficient. One of the main reasons for the speed of SpikeNet lies in its very sparse coding scheme. Typically, we have found that only 1-2% of neurons in any given processing stage need to fire in order to allow identification. The key is to use a coding scheme in which the most strongly activated neurons fire first (Rank Order Coding) since this guarantees that decisions are made as quickly as possible.

Image segmentation is not required for high level identification. One of the most striking features of SpikeNet is that there is nothing even remotely like image segmentation going on. Everything is done by using large numbers of neurons tuned to diagnostic combinations of features that will fire as soon as there is enough evidence to allow activation. It could be that the traditional view that the first step in processing requires scene segmentation is a major error, and that intelligent segmentation involves feedback that occurs only one the initial feed-forward pass has been completed.

The processing architectures used by SpikeNet are still a long way from those used by biological vision and future work will be aimed at reducing the gap. For example, SpikeNet does not have the equivalent of separate ventral and dorsal pathways specialised for object identification and localisation. Instead, there is a retino-topically organised map of neurons for each object or feature constellation that needs to be identified. For applications, this is actually quite useful, because the system automatically provides about the *xy* coordinates of each identified object (unlike object selective neurons in inferotemporal cortex that have only limited spatial selectivity). However, there is a very high cost in terms of the number of neurons required. Future versions will try are use a more biologically realistic strategy that almost certainly will allow a major reduction in the number of neurons required. Nevertheless, the biological reverse engineering approach used in SpikeNet has already proved remarkably successful and a number of important computational issues have already been addressed using this sort of approach.

References

1. Borg-Graham LJ, Monier C, Fregnac Y. 1998. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* 393: 369-73
2. Delorme A, Gautrais J, van Rullen R, Thorpe S. 1999. SpikeNET: A simulator for modeling large networks of integrate and fire neurons. *Neurocomputing* 26-7: 989-96
3. Delorme A, Thorpe SJ. 2003. SpikeNET: an event-driven simulation package for modelling large networks of spiking neurons. *Network* 14: 613-27
4. Edelman S. 1999. *Representation and recognition in vision*. Cambridge, Mass.: MIT Press.

5. Fabre-Thorpe M, Richard G, Thorpe SJ. 1998. Rapid categorization of natural images by rhesus monkeys. *NeuroReport* 9: 303-8
6. Galarreta M, Hestrin S. 2001. Electrical synapses between gaba-releasing interneurons. *Nat Rev Neurosci* 2: 425-33.
7. Gautrais J, Thorpe S. 1998. Rate coding versus temporal order coding: a theoretical approach. *Biosystems* 48: 57-65
8. Grossberg S. 2003. How Does the Cerebral Cortex Work? Development, Learning, Attention, and 3-D Vision by Laminar Circuits of Visual Cortex. *Behav Cogn Neurosci Rev* 2: 47-76
9. Hopfield JJ. 1995. Pattern recognition computation using action potential timing for stimulus representation [see comments]. *Nature* 376: 33-6
10. Keysers C, Xiao DK, Foldiak P, Perrett DI. 2001. The speed of sight. *J Cogn Neurosci* 13: 90-101.
11. Maass W, Bishop CM. 1999. *Pulsed neural networks*. Cambridge, Mass.: MIT Press.
12. Rieke F. 1997. *Spikes : exploring the neural code*. Cambridge, Mass.: MIT Press.
13. Rolls ET, Deco G. 2002. *Computational Neuroscience of Vision*. Oxford: Oxford University Press
14. Singer W. 1999. Time as coding space? *Curr Opin Neurobiol* 9: 189-94.
15. Swadlow HA, Gusev AG. 2002. Receptive-field construction in cortical inhibitory interneurons. *Nat Neurosci* 5: 403-4.
16. Thorpe S, Delorme A, Van Rullen R. 2001. Spike-based strategies for rapid processing. *Neural Networks* 14: 715-25.
17. Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. *Nature* 381: 520-2
18. Thorpe SJ, Guyonneau R, Guilbaud N, Allegraud JM, Vanrullen R. 2004. SpikeNet: Real-time visual processing with one spike per neuron. *Neurocomputing* 58-60: 857-64
19. Ullman S. 1996. *High-level vision : object recognition and visual cognition*. Cambridge, Mass.: MIT Press
20. VanRullen R, Thorpe SJ. 2001a. Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput* 13: 1255-83.
21. VanRullen R, Thorpe SJ. 2001b. The time course of visual processing: from early perception to decision- making. *J Cogn Neurosci* 13: 454-61.