

Printed in Eighth Scandinavian Conference on Artificial Intelligence, SCAI'03, eds. Bjørnar Tessem, Pekka Ala-Siuru, Patrick Doherty, and Brian Mayoh, pp. 183-88, IOS Press, Amsterdam, 2003.

Artificial Vision: What can we learn from biology?

G.T. EINEVOLL and H.E. PLESSER

*Department of Mathematical Sciences and Technology
Agricultural University of Norway, 1432 Ås, Norway*

1. Biological and Artificial Vision

The computational approach to perception, advocated in David Marr's influential book from 1982 [1], led to significant optimism in the machine vision community. However, despite the enormous growth in computer power over the last decades, machine perceptual systems are still extremely impoverished versions of their biological counterparts. Object recognition is, for example, one of the most important aspects of visual perception. Babies can recognize immediately and effortlessly a large variety of objects, a task still beyond the capabilities of artificial systems. A natural strategy is thus to look closer at the organization and functioning of biological vision, hoping that this could suggest more effective algorithms to use in machine vision.

An interesting and encouraging point is that biological "hardware" is intrinsically very slow compared to modern electronics. In nerve cells (neurons) the signal is essentially carried by ions across the cell membrane. The flow of information between neurons is predominantly carried by so called *action potentials*, standardized pulse-like disturbances in the electrical potential across the cell membrane. The duration of an action potential is about 1 ms. Thus 1 kHz seems to be a maximal "clock frequency", more than six orders of magnitude slower than the processor in a modern silicon-based computer where the signal is carried by the extremely light electron. Thus, if we were able to understand the algorithms used in biological visual perception and implement them in electronic hardware, then

2. Biological Vision: Early Visual Pathway

When light hits the eye, neurons in the retina on the back side of the eye get excited. Several types of neurons are involved in the signal processing in the retina, but the output action potentials which are transmitted towards cortex, leave from *retinal ganglion cells*. The ganglion cell axons (long, thin neuronal out-growths which propagate the signal to other neurons) constitute the optical nerve. This nerve transmits visual information to a part

of the brain called *dorsal lateral geniculate nucleus (LGN)* which is a part of thalamus. The main function of thalamus is to transmit sensory stimuli from the outer sensory systems to cortex. The *relay cells* in LGN receive visual signals from retinal ganglion cells and transmit processed information to the primary visual cortex. From primary visual cortex the signals are then fed to other parts of cortex (as well as other brain structures), and this eventually results in a visual perception of the surrounding world.

An important notion in studies of the visual system is the *receptive field* of a neuron. This term refers to the limited area of the visual field where stimulation with light (or darkness) influences the firing of action potentials in the neuron. For retinal ganglion cells and cells in the LGN the receptive fields are small, roughly circular, areas, and they exhibit *centre-surround antagonism*. This means that the cells have highest response when stimulated by a circular spot of light (on a dark background) exactly covering the *receptive-field centre*. Illumination of an area outside this receptive-field centre, on the other hand, will contribute to reducing the activity. Therefore, the receptive field can be described as a circular excitatory area surrounded by a ring-shaped inhibitory area. Neurons which responds to light spots in this way are called *on-cells*. *Off-cells* have opposite response, i.e., they have largest activity when a dark spot covers the receptive field centre. The antagonistic centre-surround organization makes the system more suited to detect changes in the light intensity than the absolute magnitude of the intensity. In primary visual cortex the receptive fields are more complex, and one finds cells that respond best to elongated (e.g., edges, bars) and/or moving stimuli.

At the levels of retina and LGN the on- and off-cells form two roughly independent networks of cells both covering the entire visual field. In addition, these cells also group into three distinct classes, labelled *P, M* and *K* in primates, distinguished by their physiological properties such as receptive-field size and signal propagation velocity [2]. Also these cell populations seem to be essentially uncoupled before reaching the visual cortex. Thus, one take-home message from the biological visual system is that there exist six independent pathways (on/off for each of P, M, and K). These pathways carry different types of information regarding the visual input; information which somehow are merged at a higher processing level to build our mental representation of the visual world.

For many decades vision has largely been viewed as a *bottom-up process*, in which images are converted to electric signals in the retina, and then propagated through increasingly specialised filter systems within the brain, from simple spot detectors in the retina itself, via edge detectors in primary visual cortex, to "grandmother cells" in inferiotemporal cortex. This feedforward paradigm unfortunately gives rise to nearly as many new questions as it answers. The most famous of these is no doubt the binding problem: each leg of a chair will excite a different edge-detector cell in primary visual cortex. So if there are two chairs in the room, how does the brain know which leg belongs to which chair? Another issue is why, if the visual system is a sequence of feedforward filters, one finds such an enormous number of feedback connections from higher visual areas to lower ones? For example, a mere 10% of all incoming connections to LGN arise from the retinal ganglion cells, while some 90% of all input to LGN is cortical feedback [2].

Many of these problems can be resolved if one considers the human (mammalian) visual system not as a passive and neutral signal processing system, but as an active and biased information collection apparatus. As we go through life, the brain permanently forms and updates hypotheses about our environment. On the basis of these hypotheses, and our innate and acquired knowledge of the laws of nature, the brain will direct attention to particularly relevant parts of our environment, e.g., to especially puzzling aspects of a visual scene. The effectiveness of the algorithms underlying visual perception is likely due to the interplay between such *top-down* and the traditional bottom-up processes.

A solid understanding of this interplay is clearly necessary if one aims to build similar artificial systems. As always when “many things happen at the same time” such an understanding must come from mathematical modelling [3].

3. Mathematical Neuroscience

The list of success stories of the use of mathematics and mathematical simulation in physics, chemistry and engineering is endless. In the biological sciences the situation is different. The training of students in biology or medicine traditionally puts little emphasis on mathematics and physics, and scepticism towards any benefits of mathematics in describing living systems prevails in the biological and medical research communities.

Neuroscience is maybe the biological sub-discipline where the use of mathematical techniques is most established and recognized. An important reason for this is the success of Hodgkin and Huxley 50 years ago modelling action potentials in neurons using an equivalent electrical-circuit formalism [4]. From their model they could predict both the shape and velocity of action potentials down the axons of neurons. From their model they calculated the propagation velocity of the action potential along the squid giant axon to be 18.8 m/s (at 18.3°C) which was roughly 10% off their experimental value of 21.2 m/s. Such quantitatively accurate model predictions are rare in theoretical biology, and this success story has given modellers a relatively firm starting point for mathematical explorations of both single neurons and neural networks.

The mathematical models used in neuroscience can be categorized into three types: *Descriptive*, *mechanistic*, and *interpretive* models [5]. The goal of *descriptive*, or *statistical*, models is to summarize experimental data compactly yet accurately. In *mechanistic* modelling one attempts to account for nervous system activity on the basis of neuronal morphology, physiology and circuitry. This type of modelling follows the traditional physics approach to mathematical modelling of natural systems. In *interpretive* modelling the goal is to model the functional roles of neural systems, i.e., relating neuronal responses to the task of processing useful information for the animal. This type of modelling is unique to biological systems having developed under evolutionary pressure. While it makes sense to ask, e.g., *why* the receptive field of retinal ganglion cells exhibits centre-surround antagonism, the question of, e.g., *why* an apple falls to the ground is not fruitful.

4. Mathematical modelling of the early visual pathway

Descriptive modelling of the receptive field of cells in the early visual pathway has a long tradition. Cells in the retina and LGN have traditionally been modelled as the difference between two spatial Gaussian functions in combination with suitable temporal filters [6]. For the more complex receptive fields in primary visual cortex Gabor functions have commonly been used (see Ch. 2 in [5]).

An example of *interpretive* modelling in the early visual pathway is the success of information theory in explaining important features of receptive fields for neurons in retina, LGN and primary visual cortex (Ch. 4 in [5]). The basic idea is that the form of these receptive fields is optimised to convey the maximum amount of information about the natural world. Based on measured statistics on visual scenes, and taking the presence of noise into account, one can with the use of information theory calculate the optimal form of the receptive field. For example, under low-noise conditions the optimal spatial organization of the receptive field was found to be the circular centre-surround organization seen in retinal ganglion cells and LGN relay cells. With a higher noise level the

theoretically estimated optimal form retains only the circular excitatory centre and loses the inhibitory surround. In the retina one can expect the signal-to-noise ratio to be controlled by the level of ambient light, and low levels of illumination should correspond to the high-noise case. If this is so, the predicted change in the spatial form of the receptive field at low illumination (high noise) corresponds to what is seen experimentally in the retina.

Since 1996 a research program in computational neuroscience has been established at the Agricultural University of Norway (NLH) in connection with the establishment of M.Sc. programs in physics, mathematics and informatics. Specifically we have focussed on *mechanistic* mathematical modelling, including simulation, of the early visual pathway with a particular emphasis on LGN. This work has been done in close collaboration with the experimental group of Paul Heggelund at the Department of Physiology at the University of Oslo. Mechanistic modelling requires detailed knowledge about the circuit neurons and their connections. Fortunately, LGN is (compared to other brain parts) modestly complex, and following significant experimental efforts during the last years, the LGN circuit is now mapped out relatively well [2]. A schematic overview is given in Fig. 1.

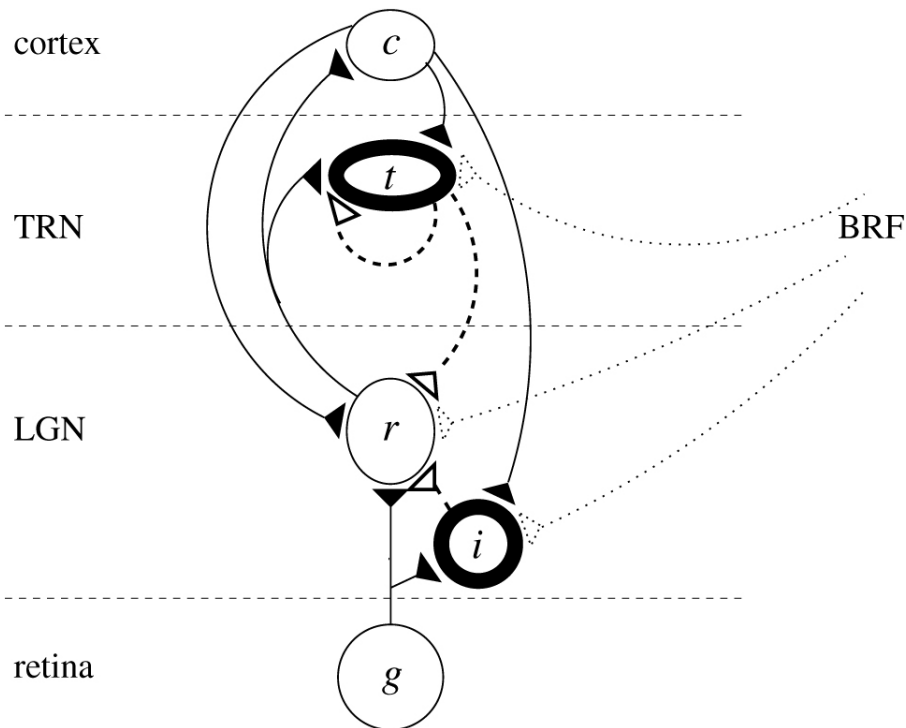


Figure 1: Schematic view of LGN circuit. The neurons involved are retinal ganglion cells (g), LGN relay cells (r), LGN interneurons (i), thalamic reticular (TRN) cells (t), and cortical cells (c). The excitatory connections are shown as solid lines while the inhibitory connections are shown with dashed lines. In addition the LGN and thalamic reticular cells receive modulatory inputs (dotted lines) from the brainstem reticular formation (BRF). Note that each of the circles represents cell *populations*. Taken from [9].

In our initial studies we have focussed on the extra feedforward inhibition on LGN relay cells due to the LGN interneurons using either a *firing-rate* formalism [7] or a more realistic (and computationally more demanding) simulation of networks of neurons where individual action potentials are included explicitly [8]. Lately, we have focussed more on the more interesting question regarding the effect of the massive cortical feedback on the LGN response properties [9]. Here one of the goals has been to develop mechanistic, rate-based receptive-field formulas for LGN cells to elucidate how this feedback contributes in the visual perception process.

In collaboration with Honda Research Institute Europe in Germany we aim to construct an extensive model of the entire early visual pathway including retina, LGN, and the visual cortex. Since a goal here is to study in detail the effect of temporal interactions between the different pathways, action potentials must be incorporated explicitly. This requires computer simulations of possibly several hundreds of thousands of neurons. For this we will use the NEST simulator (<http://www.synod.uni-freiburg.de/>) which is tailor-made for such simulations.

However, despite the progress that has been made in understanding biological vision, it seems safe to conclude that the present understanding is too limited to allow for clear and specific prescriptions for new machine-vision algorithms.

References

- [1] D. Marr, *Vision*, W.H. Freeman, New York, 1982.
- [2] S.M. Sherman and R.W. Guillery, *Exploring the thalamus*, Academic, New York, 2001.
- [3] G.T. Einevoll: *Mathematical modeling in the early visual system: Why and how?* in *Modulation of Neuronal Signaling: Implications for Active Vision*, Vol. 334 in NATO Science Series: Life and Behavioural Sciences, eds. G.T. Buracas, O. Ruksenas, T.D. Albright and G.M. Boynton (IOS Press, Amsterdam, 2003). Available from http://arken.nlh.no/~itfgev/forskning/einevoll_NATO_2003.pdf .
- [4] A.L. Hodgkin and A.F. Huxley, *Journal of Physiology* **117**, 500-544 (1952)
- [5] P. Dayan and L.F. Abbott, *Theoretical Neuroscience*, MIT Press, Cambridge, 2001.
- [6] R.W. Rodieck, *Vision Research* **5**, 583-601 (1965).
- [7] G.T. Einevoll and P. Heggelund, *Visual Neuroscience* **17**, 871-886 (2000).
- [8] H.E. Plesser, G.T. Einevoll and P. Heggelund, *Neurocomputing* **44-46**, 973-978 (2002).
- [9] G.T. Einevoll and H. E. Plesser, *Network: Comput. Neural Systems* **13**, 503-530 (2002).