

# A Biologically Inspired Object Tracking System

Roger DuBois

Australian National University, Canberra ACT 2601, Australia

**Abstract.** The anatomy of the insect brain provides insights to neural architectures and visual processing algorithms which serve as blueprints for neuromimetic silicon chip designs. Selective attention reduces the amount of computation required by a biological system navigating through an information rich environment. In the insect visual system we see an example of task-specific sensor optimization for the detection of a topological invariant, the focus of expansion of the optic flow. A description of those regions of the optic lobe concerned with flow-field analysis is presented and this is followed by a description of a simple neural subsystem capable of detecting such a focus. This information is used in a feedback control system involving the peripheral sensors to gate object tracking and orienting systems. This robust and simple system is an ideal candidate for implementation in evolving silicon based vision systems.

## 1 Evolution: The Great Optimizer

Biological organisms have evolved a myriad of sensor optimizations to exploit information found in the natural environment. From the lowest phyla, some existing virtually unchanged for 70 million years to modern man, nature has achieved an elegance of design based upon the economy of energy. She does this by exploiting invariants from the multiple sources of information that may specify a given stimulus. Technological innovations frequently turn out to be mechanistic re-creations of existing biological functions. The concept of *bio-inspired systems* follows on naturally from this.

A typical notion about information in a perceptual system is that this information is correlated with affairs in the external world. We can imagine a huge covariance matrix describing all possible co-occurrences between an internal 'image' and the external environment. If we restrict the entries in the covariance matrix to only nonzero, non-unity values, perceptual processing is inferential, where the premise for each inference is a mental construction rather than what is captured by the sensors and determined by the system architecture. This form of synthesis can be computationally expensive both for the biological organism and the designer of an autonomous robot. The challenge is to minimize the cost of this synthesis and perhaps, to eliminate it altogether. While this may seem a contemporary problem it should be remembered that all biological organisms evolve subject to similar demands for efficiency. If we consider evolution to be the optimization of information processing then, given the amount of time involved and great diversity of living systems, if an optimization for a particular

task exists we can be reasonably sure that some organism will have evolved to take advantage of it. What optimization strategies might be useful?

### 1.1 Selective Attention

If we allow unity entries in the covariance matrix to represent invariant sources of information that specify an object in the external world, a biological system needs only to be constrained by evolutionary pressures to detect and attend to them. In the design of an autonomous robot we need to explicitly specify the invariants and then match the sensory subsystem to the detection task. For every invariant that may exist, perception becomes easier to understand and easier to implement on a machine. With sufficient numbers a perceptual system that does not rely *a priori* on synthetic model structures of the external world to work properly can be designed and built. Very sophisticated behaviour can be generated by systems which exhibit almost no measurable cognition. Take for instance the common fly: a high performance, visually guided animal, with limited neural architecture is capable of high speed navigation through optical flow-fields, collision avoidance and the detection and pursuit of small targets, with which it eventually mates. These are all desirable attributes for a robot! Reverse engineering of the more simple biological systems can be a profitable avenue of research.

Insects, flies for example, see each other and engage in various forms of high speed aerial pursuits prior to mating. The salient releasing mechanism for these orienting behaviours is the presence of a small target object in the animal's visual field. But by what mechanism does it focus its attention on the object of its desires in the presence of other distractors such as wind blown leaves or other insects.

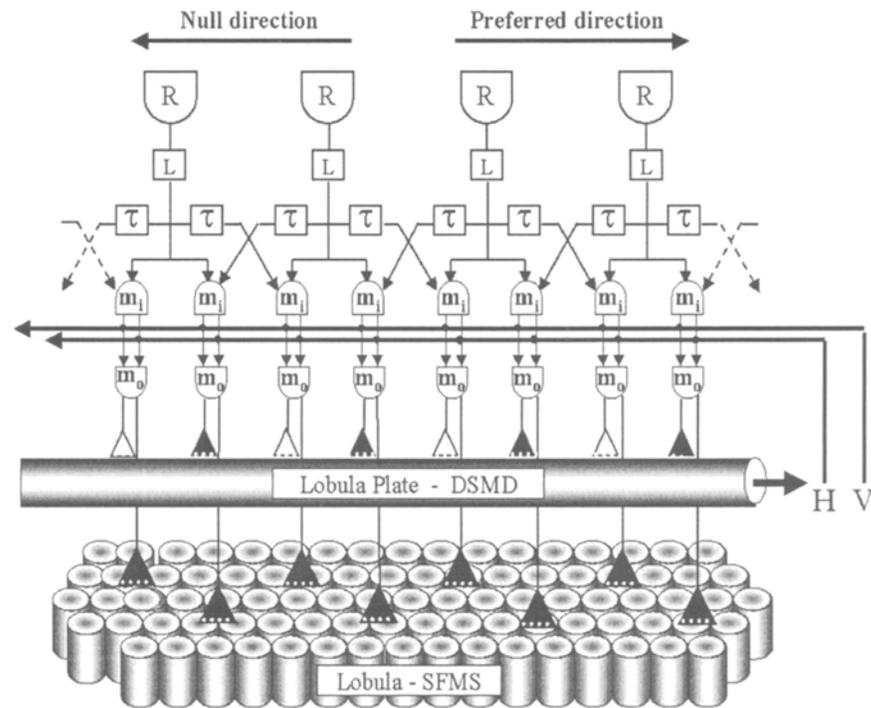
### 1.2 Neuroanatomy — VLSI in anima

The primary visual system of the insect brain is highly structured, dominated by a retinotopic organization. It comprises three visual ganglia, namely the *lamina*, the *medulla* and the *lobula complex*. Each neuropil is strictly organized into columns and strata. Both the lamina and medulla are composed of parallel synaptic compartments, or columns, which exactly match in number the ommatidia in the retina [14][6][4][9].

Each column in the lamina (L in Fig. 1) known as an *optic cartridge*, receives inputs from a group of photoreceptors (R1-R6) that share the same visual axis as the overlying ommatidium. For a detailed description of the lamina synaptic pathways see [11]. The principal neurons of the lamina are concerned with encoding luminance contrast in environments having large intensity variations and exhibit high pass filter transfer characteristics. The lamina contains no motion-detection neurons and projects its filtered signals to the input (distal) layers of the medulla.

The medulla is a bistratified neuropil characterized by an extensive network of lateral interconnections and is the most peripheral structure in which movement

detection takes place. The input layers of the medulla ( $m_i$  in Fig.1) are connected to the output layers ( $m_o$ ) by arrays of short transmedullary interneurons [6]. These interneurons play a crucial role in gating signals through this structure.



**Fig. 1.** Schematic representation of the insect motion detection system. The R units represent the photoreceptor array, the L units are lamina cells. The medulla is shown as a bi-stratified structure consisting of an input ( $m_i$ ) and an output ( $m_o$ ) side. DSMD is a wide-field, direction-selective motion-detecting neuron and SFMS is the small-field motion system and  $\tau$  represents a delay line. Shown also is the feedback of horizontal (H) and vertical (V) flow-field information from the output region of the DSMDs to the peripheral medulla.

The initial stages of movement detection rely on sequence discrimination between spatially separate, retinotopically organized signals having an asymmetric non-linear interaction between them. The computational elements are known as elementary movement detectors or EMD's [3] and it is this array of flickering elementary, motion signals that constitute the common currency for all visually guided behaviour in subsequent processing. These retinotopically organized signals are located in the proximal output layers of the medulla ( $m_o$ ) but from here on the spatial organization of the signal becomes increasingly degraded as higher

order properties of the visual world are computed by the third optic neuropil, the lobula complex.

The lobula complex consists of two complementary neuropils the *lobula* and the *lobula plate* each of which receive the parallel retinotopically organized flicker signals arriving from the output layers of the medulla.

In the lobula these spatially localized flickering signals are used to drive a position detection system whereby a position-sensitive torque signal is generated to orient the animal towards a particular target. These elemental small-field signals evoke very strong orienting responses, which increase in a dramatic manner as the spatial extent of the stimulating target decreases and on their own are perfectly adequate to direct visual flight. Observe for example the particularly impressive performance of a dragonfly in aerial pursuit of prey against an open sky.

However it must be stressed at this point that this sort of control, based purely on positional flicker information works *only in the absence of any other stimuli* or a moving textured background. The very high gain of the small-field system is inversely proportional to spatial extent of the stimulus and comes at the cost of a lateral inhibitory surround which rapidly shuts down the system in the presence of too many detractors. As I propose below, this operating limitation is overcome using the output from the complementary wide-field lobula plate neurons.

The lobula plate contains large field *directionally-selective movement-detecting* (DSMD) neurons [4][5], which share a common network of presynaptic elements derived from the medulla. The DSMD neurons (Fig. 1) comprise several classes of fan-shaped, tangential cells, which integrate the EMD inputs and transmit the resulting signal, coding either global horizontal (H) or vertical (V) image motion via a centrally projecting axon to the optomotor systems controlling the flight musculature. Thus the average, global motion of the visual flow field is directly involved in flight control.

## 2 Engineering *in vivo* — Neurophysiology

Flies were immobilized with wax and presented with computer generated optic flow-fields. The optic lobe was explored with microelectrodes until neurons responsive to image movement were located. These neurons exhibited rapid ( $\approx 1$  ms.) electrical depolarizations known as action potentials or ‘spikes’ whose occurrence times were recorded as a sequence of pulses. The resulting point process appeared stochastic with Poisson statistics in the unstimulated animal. The statistical properties of this process changed with stimulation. Cells, with roughly circular receptive fields ( $\approx 30^\circ$ ), were recorded which exhibited classical bipolar optomotor responses, *i.e.* they exhibited continuous spontaneous spiking activity, modulated by movement of the visual field. Image movement in a preferred direction excited the response, while movement in the opposite, non-preferred, direction resulted in response inhibition.

If a cell's receptive field was centered on the focus of expansion of the flow-field such that the preferred and non-preferred motion vectors balanced, the excitation and inhibition cancelled and the cell's firing rate was no different from the normal unstimulated spontaneous rate *i.e.* a NULL signal. Thus, the combined influence of two cells of orthogonal directional preference and similar receptive fields would constitute a detector for the focus of expansion, a *topological invariant*.

## 2.1 Linear analysis

The recorded cells have an exceptional ability to resolve sudden changes in contrast frequency ( $0.625^\circ / 5.32 \text{ ms}$ ) and this behaviour shows rapid adaptation to the mean background image motion. The response to an arbitrary stimulus is reliably predicted by convolution of the stimulus history with the image velocity-impulse-response. These velocity kernels regarded as a linear transforming process, map the stimulus to the observed response. The transfer function  $G$ , estimated for each level of adaptation used was, in each case, 3rd-order:

$$G(s) = K \frac{(s - a)}{(s - b_1)(s - b_2)(s - b_3)}$$

As  $s \rightarrow b_n$  the roots of the denominator polynomial,  $G(s) \rightarrow \infty$  where each value  $b_n$ , is known as a pole of the transfer function and represents a characteristic natural frequency, or rate of change for the system. The poles completely describe a linear system. If the poles are plotted in the complex frequency plane (s-plane) we observe a remarkable feature: the root loci migrate along radial lines of constant damping ratio ( $\xi \approx 0.5$ ) with increasing levels of adaptation. In control system theory a damping factor of 0.5 indicates a system is optimized for the best possible transient performance. Thus, the system not only measures the steady state velocity of the flow-field but also, by maintaining a constant damping ratio, encodes the velocity contrast as well.

## 3 Closing the Loop

The cells recorded from during these experiments had identical response characteristics to the classic lobula plate DSMD cells which convey flow-field information to the motor systems [8][10] but dual electrode studies and subsequent anatomical verification revealed them to be something entirely different.

Rather than being orthograde, centrally projecting afferent neurons they were discovered to be efferent *i.e.* they conveyed the horizontal or vertical flowfield information (H & V in Fig. 1) peripherally outwards from the midbrain to the input region of the medulla ( $m_i$ ). These cells, known as the *medullary tangentials* (MTH & MTV) receive their inputs from the output regions of the wide-field lobula plate cells coding global image motion, the DSMDs. In their turn these medullary tangentials relay these signals to the precise location where the transmedullary cells receive input from the lamina.

The transmedullary cells maintain the retinotopic organisation of the flicker signals arriving from the lamina as they relay information to the motion processing array in the medulla. The most likely candidates to receive this re-afferent signal from the lobula plate are the SUB, TM1 and TM5 transmedullary cells [14]. Interestingly these cells are the only transmedullary cells to have dendrites restricted to one group of terminals arising from a single optic cartridge and as such are ideal candidates for inclusion in an small object detection system.

Thus we have a recurrent signal path, inward through the medulla on to the lobula complex and back to the medulla. This feedback from the lobula plate to the distal medulla gates signal transmission through the transmedullary cells which in turn feed into the position detecting system, the SFMS array. It is proposed that this arrangement constitutes a mechanism for static orientation and dynamic tracking of an object in both free-field and textured environments.

When the animal is stationary the (presumed inhibitory) NULL signal derived from the integrated output of the DSMD lobula plate neurons, is projected over the sensory array and every constituent high gain small-field element of the SFMS array spanning the entire visual field is gated open to receive input and provoke an orienting response.

However, when the animal moves through the environment its ego-motion creates visual flow-fields. When the receptive field of a particular MT cell is centered on a focus of expansion the feedback signal to that cell's projective field is locally NULL unlocking the high performance positional tracking system only along a receptive 'beam' aimed at those sensors directed at this focus. The two systems work in harmony since the lobula plate neurons are hardwired to the flight motor muscles specifically to balance the visual flow across each eye to enable steady flight.

This orienting and tracking behaviour is observed in flying insects on a hot day. When they are stationary they are receptive to any small object passing overhead and will chase it. If they are moving it is only when the object is caught in the glare of these perceptual headlights that it will be 'captured'.

### 3.1 Anima in silico

The insect visual system provides us with a neural architecture with which to construct a silicon based machine vision system capable of pursuit and obstacle avoidance. This 'smart-sensor' technology to be used in various autonomous devices or as adjuncts to piloted vehicles for line sensing, collision avoidance, missile tracking and so on.

Chip design is intimately linked with the choice of algorithm and several groups have made some progress with single-chip vision systems, some with more success than others. The innovative work of John Tanner and Carver Mead at Caltech on the *Correlating Optical Motion Detector* [15] led to several VLSI designs which were flawed by computationally expensive algorithms, based on the gradient scheme, [16][12] and the Marr and Hildreth zero-crossing detector [2]. Both of these schemes are excessively sensitive to noise since they rely on the

calculation of spatial and temporal derivatives and at low ambient light intensity velocity estimates become problematic.

An approach taken by the Centre for GaAs VLSI technology and the Departments of Electrical & Electronic Engineering at the University of Adelaide, South Australia [1] based on the *Template Model* [7] is more promising. They have constructed a device whose input is an array of photodetectors interfaced with an array of graded refractive index (GRIN) lenses with parallel contrast differentiation. Changes in contrast are multiplexed for adjacent channels in space and in  $t_{i-1}$  sample time to form a 'Horridge template', an index into a lookup table providing information on object motion, direction, orientation, angular velocity and bearing. Several iterations of this design have already proved the feasibility of the approach and the advantages of using gallium arsenide for low power, high speed operation, in low ambient light levels. It is proposed that the architecture/algorithm outlined above would be added easily to their design. Access to a spectrum of different image motion parameters also opens up the possibility of tuning the system to more subtle and salient features in the environment, *e.g.* objects moving with a particular angular velocity and so on.

## References

1. Abbott, D., Yakoleff, A., Moini, A., Nguyen, X.T., Blanksby, A., Beare, R., Beaumont-SSmith, A. Kim, G., Bouzardoum, A., Bogner, R.E., Eshraghi, K. (1995) Technology inspired obstacle avoidance - a technology independent program, Proc. SPIE, Int. Soc. Opt. Eng., 2591.
2. Bair, W. and Koch, C. (1989) Real-time motion detection using an analog VLSI zero-crossing chip. Proc. SPIE Visual Information Processing: From Neurons to Chips. Proc. SPIE, Vol 1473, pp. 59-65.
3. Buchner, E. (1976) Elementary movement detectors in an insect visual system, Biol. Cybern., 24: 85-101.
4. Hausen, K. (1982a) Motion sensitive interneurons in the optomotor system of the fly I. The horizontal cells: structure and signals, Biol. Cybern., 45: 67-79.
5. Hausen, K. (1982b) Motion sensitive interneurons in the optomotor system of the fly II. receptive field organization and response characteristics, Biol. Cybern., 46: 143-156.
6. Hausen, K. (1984) The lobula complex of the fly: structure, function and significance in visual behaviour. In M.A. Ali (Ed.), Photoreception and vision in invertebrates, Plenum, New York and London, pp. 523-560.
7. Horridge, G.A. (1991) Ratios of template responses as the basis of semivision. Phil. Trans. Roy. Soc. London B, Vol. 331, pp 189-197.
8. Ibbotson, M.R., Maddess, T. and DuBois, R.A. (1991) A system of insect neurons sensitive to horizontal and vertical image motion connects medulla to midbrain, J. Comp Physiol A,
9. Kirschfeld K (1972) The visual system of *Musca*: studies on structure and function. In: Wehner R. (ed) Information processing in the visual system of arthropods. Springer, Berlin, pp 61-74.
10. Maddess, T., DuBois, R.A. and Ibbotson, M.R.(1991) Response properties and adaptation of neurons sensitive to image motion in the butterfly *Papilio aegus* , J. Comp Physiol A,

11. Meinertzhagen, L.A. and O'Neil S.D. (1991) Synaptic organization of columnar elements in the lamina of the wild type in *Drosophila melanogaster*, J. Comp. Neurol., pp. 232-236.
12. Moorea, A. and Koch, C. (1991) A multiplication based motion detection chip. In: Visual Information Processing: From Neurons to Chips. Proc. SPIE, Vol 1473, pp. 66-75.
13. Strausfeld, N.J. (1982) Functional neuroanatomy of the blowfly's visual system. In M. A. Ali (Ed.), Photoreception and Vision in Invertebrates, London New York: Plenum Press, pp. 483-522.
14. Strausfeld, N.J. (1989) Beneath the compound eye: neuroanatomical analysis and psychophysical correlates in the study of insect vision. In D.G. Stavenga and R.C Hardie (Eds.), Facets of Vision, Springer-Verlag, Berlin Heidelberg, pp. 360-390.
15. Tanner, J. and Mead, C. (1984) A correlating optical motion detector. Proc. Conferences on Advanced Research in VLSI, M.I.T., pp57-64.
16. Tanner, J. and Mead, C. (1986) An integrated analog optical motion sensor. In: VLSI Signal Processing II, pp.59-76.