A Neural Network for Motion Detection of Drift-Balanced Stimuli

Hilary Tunley*
School of Cognitive and Computer Sciences
Sussex University
Brighton, England.

Abstract

This paper briefly describes an artificial neural network for preattentive visual processing. The network is capable of determining image motion in a type of stimulus which defeats most popular methods of motion detection – a subset of second-order visual motion stimuli known as drift-balanced stimuli(DBS). The processing stages of the network described in this paper are integratable into a model capable of simultaneous motion extraction, edge detection, and the determination of occlusion.

1 INTRODUCTION

Previous methods of motion detection have generally been based on one of two underlying approaches: correlation; and gradient-filter. Probably the best known example of the correlation approach is the Reichardt movement detector [Reichardt 1961]. The gradient-filter (GF) approach underlies the work of Adelson and Bergen [Adelson 1985], and Heeger [Heeger 1988], amongst others.

These motion-detecting methods cannot track DBS, because DBS lack essential components of information needed by such methods. Both the correlation and GF approaches impose constraints on the input stimuli. Throughout the image sequence, correlation methods require information that is spatiotemporally correlatable; and GF motion detectors assume temporally constant spatial gradients.

*Current address: Experimental Psychology, School of Biological Sciences, Sussex University.

The network discussed here does not impose such constraints. Instead, it extracts motion *energy* and exploits the spatial coherence of movement (defined more formally in the Gestalt theory of *common fate* [Koffka 1935]) to achieve tracking.

The remainder of this paper discusses DBS image sequences, then correlation methods, then GF methods in more detail, followed by a qualitative description of this network which *can* process DBS.

2 SECOND-ORDER AND DRIFT-BALANCED STIMULI

There has been a lot of recent interest in second-order visual stimuli, and DBS in particular ([Chubb 1989, Landy 1991]). DBS are stimuli which give a clear percept of directional motion, yet Fourier analysis reveals a lack of coherent motion energy, or energy present in a direction opposing that of the displacement (hence the term 'drift-balanced'). Examples of DBS include image sequences in which the contrast polarity of edges present reverses between frames.

A subset of DBS, which are also processed by the network, are known as micro-balanced stimuli (MBS). MBS contain no correlatable features and are drift-balanced at all scales. The MBS image sequences used for this work were created from a random-dot image in which an area is successively shifted by a constant displacement between each frame and *simultaneously* re-randomised.

3 EXISTING METHODS OF MOTION DETECTION

3.1 CORRELATION METHODS

Correlation methods perform a local cross-correlation in image space: the matching of features in local neighbourhoods (depending upon displacement/speed) between image frames underlies the motion detection. Examples of this method include [Van Santen 1985]. Most correlation models suffer from noise degradation in that any noise features extracted by the edge detection are available for spurious correlation.

There has been much recent debate questioning the validity of correlation methods for modelling human motion detection abilities. In addition to DBS, there is also increasing psychophysical evidence ([Landy 1991, Mather 1991]) which correlation methods cannot account for.

These factors suggest that correlation techniques are not suitable for low-level motion processing where no information is available concerning *what* is moving (as with MBS). However, correlation is a more plausible method when working with higher level constructs such as tracking in model-based vision (e.g. [Bray 1990]).

3.2 GRADIENT-FILTER (GF) METHODS

GF methods use a combination of spatial filtering to determine edge positions and temporal filtering to determine whether such edges are moving. A common assumption used by GF methods is that *spatial gradients are constant.* A recent method by Verri [Verri 1990], for example, argues that flow detection is based upon the notion

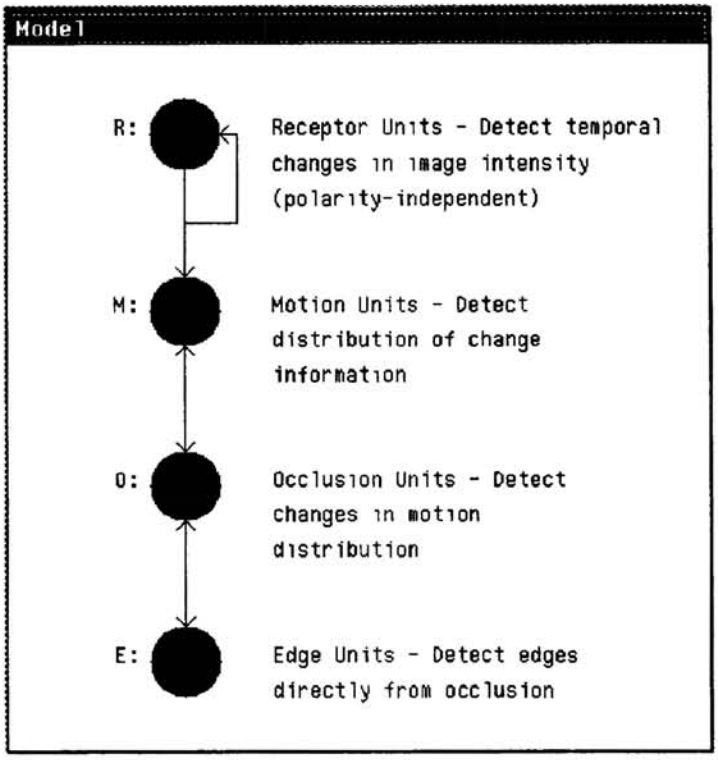


Figure 1: The Network (Schematic)

of tracking spatial gradient magnitude and/or direction, and that any variation in the spatial gradient is due to some form of motion deformation – i.e. rotation, expansion or shear. Whilst for scenes containing smooth surfaces this is a valid approximation, it is *not* the case for second-order stimuli such as DBS.

4 THE NETWORK

A simplified diagram illustrating the basic structure of the network (based upon earlier work ([Tunley 1990, Tunley 1991a, Tunley 1991b]) is shown in Figure 1 (the edge detection stage is discussed elsewhere ([Tunley 1990, Tunley 1991b, Tunley 1992]).

4.1 INPUT RECEPTOR UNITS

The units in the input layer respond to rectified local changes in image intensity over time. Each unit has a variable adaption rate, resulting in temporal sensitivity – a fast adaption rate gives a high temporal filtering rate. The main advantages for this temporal averaging processing are:

- Averaging removes the D.C. component of image intensity. This eliminates problematic gain for motion in high brightness areas of the image. [Heeger 1988].
- The random nature of DBS/MBS generation cannot guarantee that each pixel change is due to local image motion. Local temporal averaging smooths the

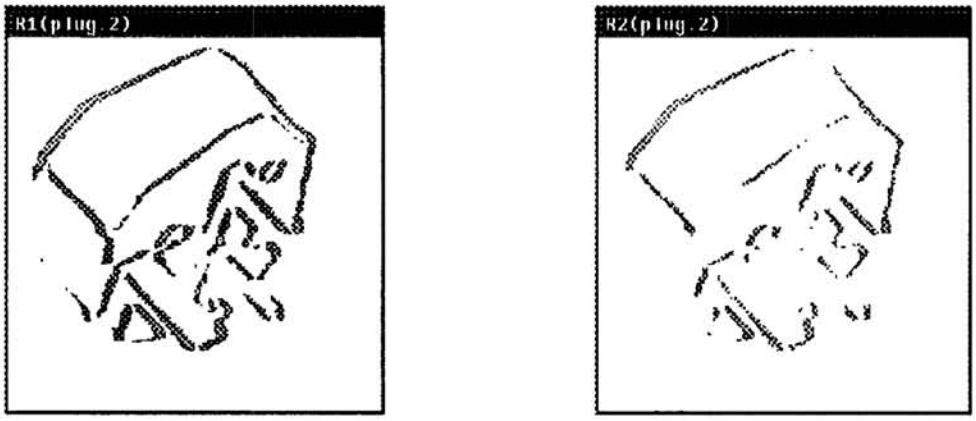
moving regions, thus creating a more coherently structured input for the motion units.

The input units have a pointwise rectifying response governed by an autoregressive filter of the following form:

$$R_n = (1 - \alpha).R_{n-1} + \alpha.|I_n - I_{n-1}| \tag{1}$$

where $\alpha \in [0, 1]$ is a variable which controls the degree of temporal filtering of the change in input intensity, n and $n - 1$ are successive image frames, and R_n and I_n are the filter output and input, respectively.

The receptor unit responses for two different α values are shown in Figure 2. α can thus be used to alter the amount of motion blur produced for a particular frame rate, effectively producing a unit with differing velocity sensitivity.



(a) (b)

Figure 2: Receptor Unit Response: (a) $\alpha = 0.3$; (b) $\alpha = 0.7$.

4.2 MOTION UNITS

These units determine the *coherence* of image changes indicated by corresponding receptor units. First-order motion produces highly-tuned motion activity – i.e. a strong response in a particular direction – whilst second-order motion results in less coherent output.

The operation of a basic motion detector can be described by:

$$M_{ijkdn} = |R_{ijn} - R_{i'j'n-1}| \tag{2}$$

where M is the detector, (i', j') is a point in frame n at a distance d from (i, j), a point in frame $n - 1$, in the direction k. Therefore, for coherent motion (i.e. first-order), in direction k at a speed of d units/frame, as $n \rightarrow \infty$:

$$M_{ijkdn} \rightarrow 0 \tag{3}$$

The convergence of motion activity can be seen using an example. The stimulus sequence used consists of a bar of re-randomising texture moving to the right in front of a leftward moving background with the same texture (i.e. random dots). The bar motion is second-order as it contains no correlatable features, whilst the background consists of a simple first-order shifting of dots between frames. Figures 3, 4 and 5 show two-dimensional images of the leftward motion activity for the stimulus after 3, 4 and 6 frames respectively. The background, which has coherent leftward movement (at speed d units/frame) is gradually reducing to zero whilst the microbalanced rightwards-moving bar, remains active. The fact that a non-zero response is obtained for second-order motion suggests, according to the definition of Chubb and Sperling [Chubb 1989], that first-order detectors produce no response to MBS, that this detector is second-order with regard to motion detection.

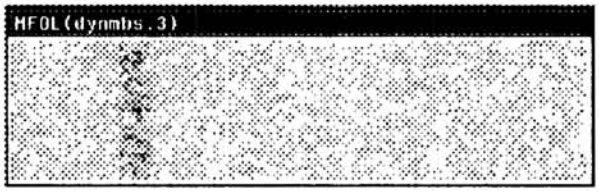


Figure 3: Leftward Motion Response to Third Frame in Sequence.

Figure 4: Leftward Motion Response to Fourth Frame.

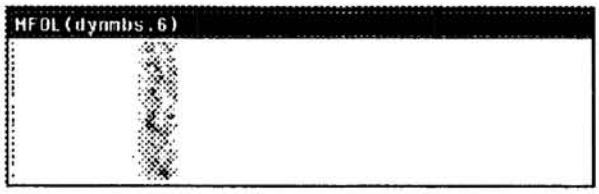


Figure 5: Leftward Motion Response to Sixth Frame.

The motion units in this model are arranged on a hexagonal grid. This grid is known as a flow web as it allows information to flow, both laterally between units of the same type, and between the different units in the model (motion, occlusion or edge). Each flow web unit is represented by three variables – a position (a, b) and a direction k, which is evenly spaced between 0 and 360 degrees. In this model each k is an integer between 1 and $k_{\max}$ – the value of $k_{\max}$ can be varied to vary the sensitivity of the units.

A way of using first-order techniques to discriminate between first and second-order motions is through the concept of coherence. At any point in the motion-processed images in Figures 3-5, a measure of the overall variation in motion activity can be used to distinguish between the motion of the micro-balanced bar and its background. The motion energy for a detector with displacement d, and orientation

k, at position (a, b), can be represented by E_{abkd}. For each motion unit, responding over distance d, in each cluster the energy present can be defined as:

$$E_{abkdn} = \frac{\min_k(M_{abkd})}{M_{abkd}} \tag{4}$$

where $\min_k(x_k)$ is the minimum value of x found searching over k values. If motion is coherent, and of approximately the correct speed for the detector M, then as $n \rightarrow \infty$:

$$E_{abk_m dn} \rightarrow 1 \tag{5}$$

where k_m is in the actual direction of the motion. In reality n need only approach around 5 for convergence to occur. Also, more importantly, under the same convergence conditions:

$$E_{abkdn} \rightarrow 0 \ \forall k \neq k_m \tag{6}$$

This is due to the fact that the minimum activation value in a group of first-order detectors at point (a, b) will be the same as the actual value in the direction, k_m. By similar reasoning, for non-coherent motion as $n \rightarrow \infty$:

$$E_{abkdn} \rightarrow 1 \ \forall k \tag{7}$$

in other words there is no peak of activity in a given direction. The motion energy is ambiguous at a large number of points in most images, except at discontinuities and on well-textured surfaces.

A measure of motion coherence used for the motion units can now be defined as:

$$M_c(abkd) = \frac{E_{abkd}}{\sum_{k=1}^{k_{max}} E_{abkd}} \tag{8}$$

For coherent motion in direction k_m as $n \rightarrow \infty$:

$$M_c(abk_m d) \rightarrow 1 \tag{9}$$

Whilst for second-order motion, also as $n \rightarrow \infty$:

$$M_c(abkd) \rightarrow 1/k_{max} \ \forall k \tag{10}$$

Using this approach the total M_c activity at each position – regardless of coherence, or lack of it – is unity. Motion energy is the same in all moving regions, the difference is in the distribution, or *tuning* of that energy.

Figures 6, 7 and 8 show how motion coherence allows the flow web structure to reveal the presence of motion in microbalanced areas whilst not affecting the easily detected background motion for the stimulus.

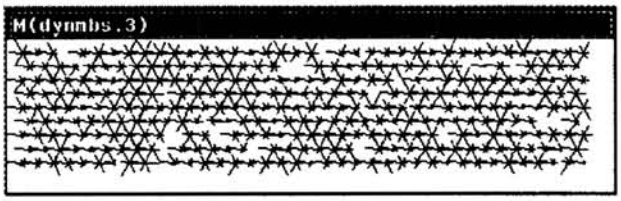


Figure 6: Motion Coherence Response to Third Frame

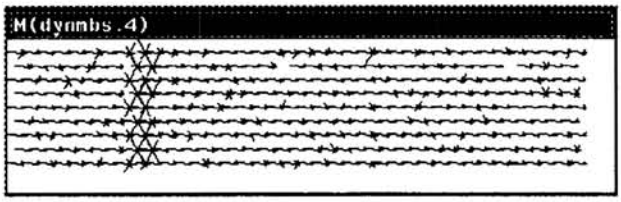


Figure 7: Motion Coherence Response to Fourth Frame

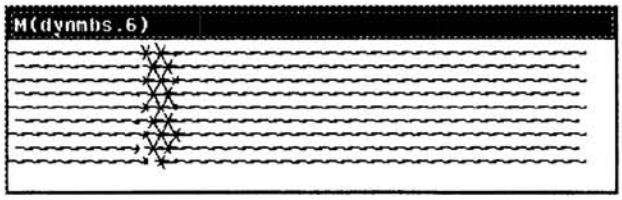


Figure 8: Motion Coherence Response to Sixth Frame

4.3 OCCLUSION UNITS

These units identify discontinuities in second-order motion which are vitally important when computing the direction of that motion. They determine spatial and temporal changes in motion coherence and can process single or multiple motions at each image point. Established and newly-activated occlusion units work, through a gating process, to enhance continuously-displacing surfaces, utilising the concept of visual inertia.

The implementation details of the occlusion stage of this model are discussed elsewhere [Tunley 1992], but some output from the occlusion units to the above second-order stimulus are shown in Figures 9 and 10. The figures show how the edges of the bar can be determined.

References

[Adelson 1985] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.* 2, 1985.

[Bray 1990] A.J. Bray. Tracking objects using image disparities. *Image and Vision Computing*, 8, 1990.

[Chubb 1989] C. Chubb and G. Sperling. Second-order motion perception: Space/time separable mechanisms. In *Proc. Workshop on Visual Motion, Irvine, CA, USA*, 1989.

Figure 9: Occluding Motion Information: Occlusion activity produced by an increase in motion coherence activity.

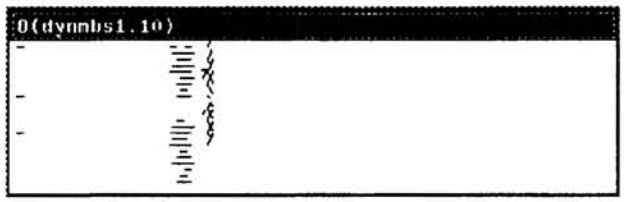


Figure 10: Occluding Motion Information: Occlusion activity produced by a decrease in motion activity at a point. Some spurious activity is produced due to the random nature of the second-order motion information.

[Heeger 1988] D.J. Heeger. Optical Flow using spatiotemporal filters. *Int. J. Comp. Vision*, 1, 1988.

[Koffka 1935] K. Koffka. *Principles of Gestalt Psychology.* Harcourt Brace, 1935.

[Landy 1991] M.S. Landy, B.A. Dosher, G. Sperling and M.E. Perkins. The kinetic depth effect and optic flow II: First- and second-order motion. *Vis. Res.* 31, 1991.

[Mather 1991] G. Mather. Personal Communication.

[Reichardt 1961] W. Reichardt. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. Rosenblith, editor, *Sensory Communications.* Wiley NY, 1961.

[Van Santen 1985] J.P.H. Van Santen and G. Sperling. Elaborated Reichardt detectors. *J. Opt. Soc. Am.* 2, 1985.

[Tunley 1990] H. Tunley. Segmenting Moving Images. In *Proc. Int. Neural Network Conf. (INNC90), Paris, France*, 1990.

[Tunley 1991a] H. Tunley. Distributed dynamic processing for edge detection. In *Proc. British Machine Vision Conf. (BMVC91), Glasgow, Scotland*, 1991.

[Tunley 1991b] H. Tunley. Dynamic segmentation and optic flow extraction. In. *Proc. Int. Joint. Conf. Neural Networks (IJCNN91), Seattle, USA*, 1991.

[Tunley 1992] H. Tunley. Sceond-order motion processing: A distributed approach. CSRP 211, School of Cognitive and Computing Sciences, University of Sussex (forthcoming).

[Verri 1990] A. Verri, F. Girosi and V. Torre. Differential techniques for optic flow. *J. Opt. Soc. Am.* 7, 1990.

Recurrent Eye Tracking Network Using a Distributed Representation of Image Motion

P. A. Viola
Artificial Intelligence Laboratory
Massachusetts Institute of Technology

S. G. Lisberger
Department of Physiology
W.M. Keck Foundation Center for Integrative Neuroscience
Neuroscience Graduate Program
University of California, San Francisco

T. J. Sejnowski
Salk Institute, Howard Hughes Medical Institute
Department of Biology
University of California, San Diego

Abstract

We have constructed a recurrent network that stabilizes images of a moving object on the retina of a simulated eye. The structure of the network was motivated by the organization of the primate visual target tracking system. The basic components of a complete target tracking system were simulated, including visual processing, sensory-motor interface, and motor control. Our model is simpler in structure, function and performance than the primate system, but many of the complexities inherent in a complete system are present.

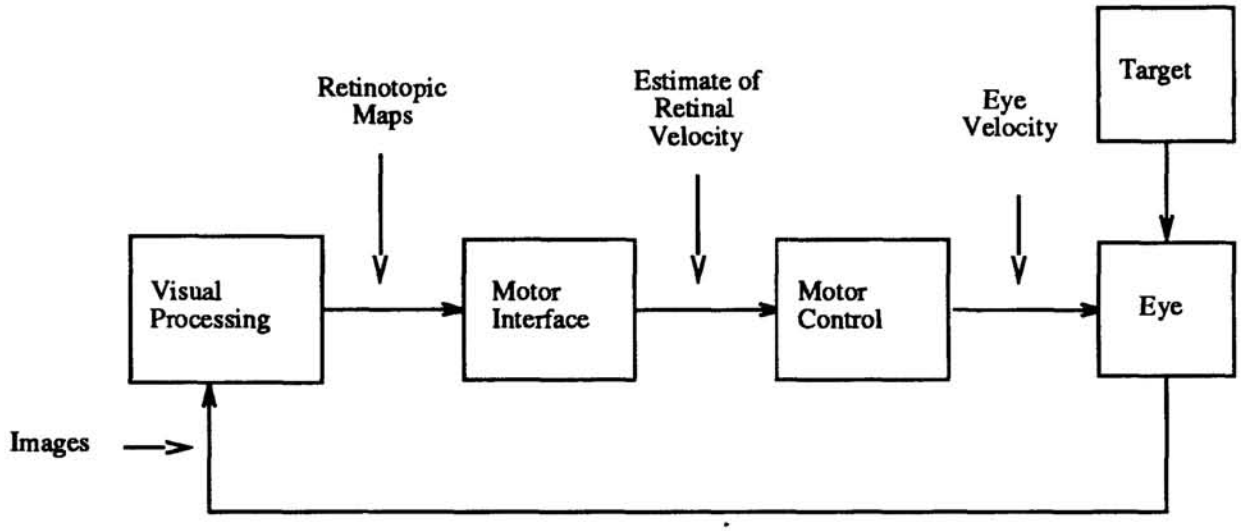


Figure 1: The overall structure of the visual tracking model.

1 Introduction

The fovea of the primate eye has a high density of photoreceptors. Images that fall within the fovea are perceived with high resolution. Perception of moving objects poses a particular problem for the visual system. If the eyes are fixed a moving image will be blurred. When the image moves out the of the fovea, resolution decreases. By moving their eyes to foveate and stabilize targets, primates ensure maximum perceptual resolution. In addition, active target tracking simplifies other tasks, such as spatial localization and spatial coordinate transformations (Ballard, 1991).

Visual tracking is a feedback process, in which the eyes are moved to stabilize and foveate the image of a target. Good visual tracking performance depends on accurate estimates of target velocity and a stable feedback controller. Although many visual tracking systems have been designed by engineers, the primate visual tracking system has yet to be matched in its ability to perform in complicated environments, with unrestricted targets, and over a wide variety of target trajectories. The study of the primate oculomotor system is an important step toward building a system that can attain primate levels of performance. The model presented here can accurately and stably track a variety of targets over a wide range of trajectories and is a first step toward achieving this goal.

Our model has four primary components: a model eye, a visual processing network, a motor interface network, and a motor control network (see Figure 1). The model eye receives a sequence of images from a changing visual world, synthetically rendered, and generates a time-varying output signal. The retinal signal is sent to the visual processing network which is similar in function to the motion processing areas of the visual cortex. The visual processing network constructs a distributed representation of image velocity. This representation is then used to estimate the velocity of the target on the retina. The retinal velocity of the target forms the input to the motor control network that drives the eye. The eye responds by rotating,

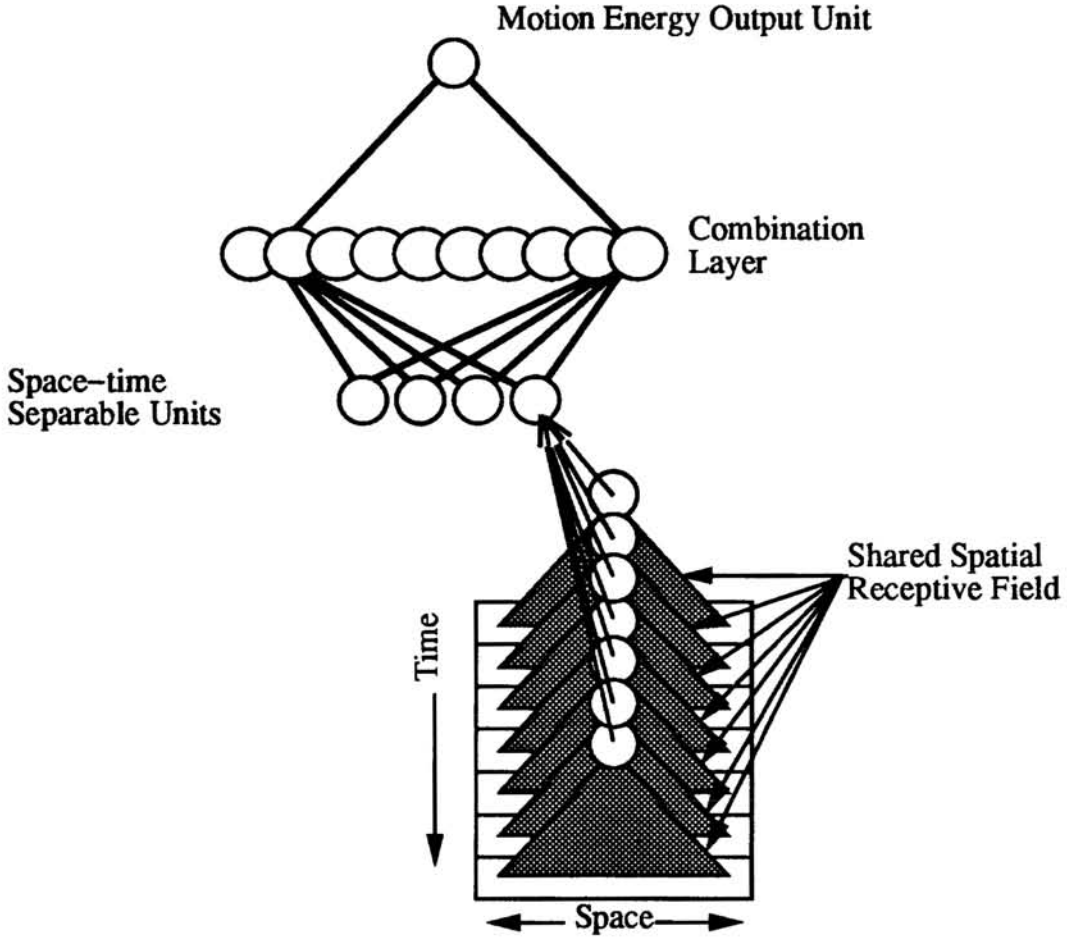


Figure 2: The structure of a motion energy unit. Each space-time separable unit has a receptive field that covers 16 pixels in space and 16 steps in time (for a total of 256 inputs). The shaded triangles denote complete projections.

which in turn affects incoming retinal signals.

If these networks function perfectly, eye velocity will match target velocity. Our model generates smooth eye motions to stabilize smoothly moving targets. It makes no attempt to foveate the image of a target. In primates, eye motions that foveate targets are called saccades. Saccadic mechanisms are largely separate from the smooth eye motion system (Lisberger et. al. 1987). We do not address them here.

In contrast with most engineered systems, our model is adaptive. The networks used in the model were trained using gradient descent[1]. This training process circumvented the need for a separate calibration of the visual tracking system.

2 Visual Processing

[1]Network simulations were carried out with the SN2 neural network simulator.

The middle temporal cortex (area MT) contains cells that are selective for the direction of visual motion. The neurons in MT are organized into a retinotopic map and small lesions in this area lead to selective impairment of visual tracking in the corresponding regions of the visual field (Newsome and Pare, 1988). The visual processing networks in our model contain directionally-selective processing units that are arranged in a retinotopic map. The spatio-temporal motion energy filter of Adelson and Bergen (Adelson and Bergen, 1985) has many of the properties of directionally-selective cortical neurons; it is used as the basis for our visual processing network. We constructed a four layer time-delay neural network that implements a motion energy calculation.

A single motion-energy unit can be constructed from four intermediate units having separable spatial and temporal filters. Adelson and Bergen demonstrate that two spatial filters (of even and odd symmetry) and two temporal filters (temporal derivatives for fast and slow speeds) are sufficient to detect motion. The filters are combined to construct 4 intermediate units which project to a single motion energy unit. Because the spatial and temporal properties of the receptive field are separable, they can be computed separately and convolved together to produce the final output. The temporal response is therefore the same throughout the extent of the spatial receptive field.

In our model, motion energy units are implemented as backpropagation networks. These units have a receptive field 16 pixels wide over a 16 time step window. Because the input weights are shared, only 32 parameters were needed for each space-time separable unit. Four space-time separable units project through a 16 unit combination layer to the output unit (see Figure 2). The entire network can be trained to approximate a variety of motion-energy filters.

We trained the motion energy network in two different ways: as a single multilayered network and in stages. Staged training proceeded first by training intermediate units, then, with the intermediate units fixed, by training the three layer network that combines the intermediate units to produce a single motion energy output. The output unit is active when a pattern in the appropriate range of spatial frequencies moves through the receptive field with appropriate velocity. Many such units are required for a range of velocities, spatial frequencies, and spatial locations. We use six different types of motion energy units – each tuned to a different temporal frequency – at each of the central 48 positions of a 64 pixel linear retina. The 6 populations form a distributed, velocity-tuned representation of image motion for a total of 288 motion energy units.

In addition to the motion energy filters, static spatial frequency filters are also computed and used in the interface network, one for each band and each position for a total of 288 units.

We chose an adaptive network rather than a direct motion energy calculation because it allows us to model the dynamic nature of the visual signal with greater flexibility. However, this raises complications regarding the set of training images. Assuming 5 bits of information at each retinal position, there are well over 10 to the 100th possible input patterns. We explored sine waves, random spots and a variety of spatial pre-filters, and found low-pass filtered images of moving random spots worked best. Typically we began the training process from a plausible set of

weights, rather than from random values, to prevent the network from settling into an initial local minima. Training proceeded for days until good performance was obtained on a testing set.

Krauzlis and Lisberger (1989) have predicted that the visual stimulus to the visual tracking system in the brain contains information about the acceleration and impulse of the target as well as the velocity. Our motion energy networks are sensitive to target acceleration, producing transients for accelerating stimuli.

3 The Interface Network

The function of the interface is to take the distributed representation of the image motion and extract a single velocity estimate for the moving object. We use a relatively simple method that was adequate for tracking single objects without other moving distractors. The activity level of a single motion energy unit is ambiguous. First, it is necessary for the object to have a feature that is matched to the spatial frequency bandpass of the motion energy unit. Second, there is an array of units for each spatial frequency and the object will stimulate only a few of these at any given time. For instance, a large white object will have no features in its interior; a unit with its receptive field located in the interior can detect no motion. Conversely, detectors with receptive fields on the border between the object and the background will be strongly stimulated.

We use two stages of processing to extract a velocity. In the first stage, the motion energy in each spatial frequency band is estimated by summing the outputs of the motion energy filters across the retina weighted by the spatial frequency filter at each location. The six populations of spatial frequency units each yield one value. Next, a 6-6-1 feedforward network, trained using backpropagation, predicts target velocity from these values.

4 The Motor Control Network

In comparison with the visual processing network, the motor control network is quite small (see Figure 3). The goal of the network is to move the eye to stabilize the image of the object. The visual processing and interface networks convert images of the moving target into an estimate for the retinal velocity of the target. This retinal velocity can be considered a motor error. One approach to reducing this error is a simple proportional feedback controller, which drives the eye at a velocity proportional to the error. There is a large, 50-100 ms delay that occurs during visual processing in the primate visual system. In the presence of a large delay a proportional controller will either be inaccurate or unstable. For this reason simple proportional feedback is not sufficient to control tracking in the primate. Tracking can be made stable and accurate by including an internal positive feedback pathway to prevent instability while preserving accuracy (Robinson, 1971).

The motor control network was based on a model of the primate visual tracking motor control system by Lisberger and Sejnowski (1992). This recurrent artificial neural network includes both the smooth visual tracking system and the vestibulo-ocular system, which is important for compensating head movements. We use a

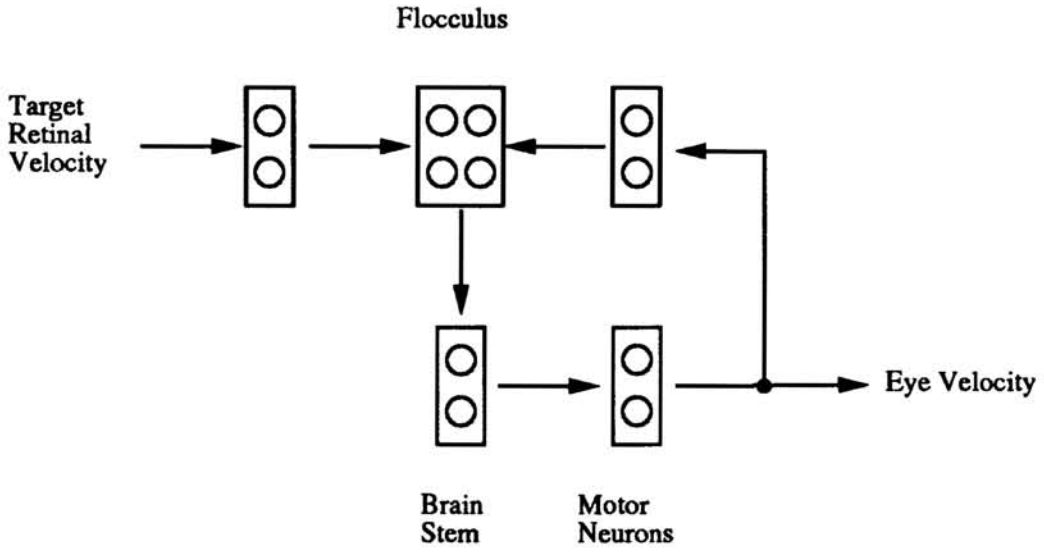


Figure 3: The structure of the recurrent network. Each circle is a unit. Units within a box are not interconnected and all units between boxes were fully interconnected as indicated by the arrows.

simpler version of that model that does not have vestibular inputs. The network is constructed from units with continuous smooth temporal responses. The state of a unit is a function of previous inputs and previous state:

$$s_j(t + \Delta t) = (1 - \tau \Delta t) s_j(t) + I \tau \Delta t$$

where $s_j(t)$ is the state of unit j at time t, τ is a time constant and I is the sigmoided sum of the weighted pre-synaptic activities. The resulting network is capable of smooth responses to inputs.

The motor control network has 12 units, each with a time constant of 5 ms (except for a few units with longer delay). There is a time delay of 50 ms between the interface network and control network. (see Figure 3). The input to the network is retinal target velocity, the output is eye velocity. The motor control network is trained to track a target in the presence of the visual delay.

The motor control network contains a positive feedback loop that is necessary to maintain accurate tracking even when the error signal falls to zero. The overall control network also contains a negative feedback loop since the output of the network affects subsequent inputs. The gradient descent optimization procedure uses the relationship between the output and the input during training—this relationship can be considered a model of the plant. It should be possible to use the same approach with more complex plants.

The control network was trained with the visual processing network frozen. A training example consists of an object trajectory and the goal trajectory for the eye. A standard recurrent network training paradigm is used to adjust the weights to minimize the error between actual outputs and desired outputs for step changes in target velocity.

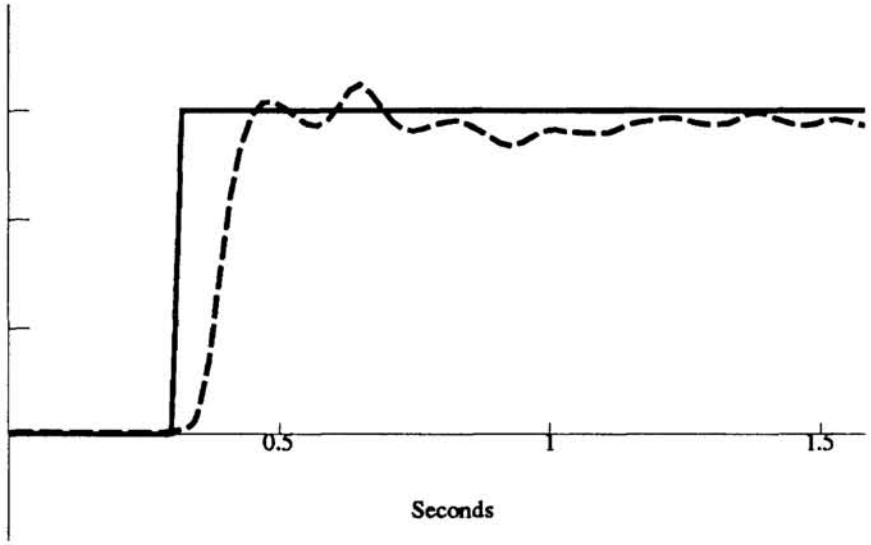


Figure 4: Response of the eye to a step in target velocity of 30 degrees per second. The solid line is target velocity, the dashed line is eye velocity. This experiment was performed with a target that did not appear in the training set.

5 Performance

After training the network on a set of trajectories for a single target, the tracking performance was equally good on new targets. Tracking is accurate and stable - with little tendency to ring (see Figure 4). This good performance is surprising in the presence of a 50 millisecond delay in the visual feedback signal[2]. Stable tracking is not possible without the positive internal feedback loop in the model (eye velocity signal to the flocculus in Figure 3).

6 Limitations

The system that we have designed is a relatively small one having a one-dimensional retina only 64 pixels wide. The eye and the target can only move in one dimension—along the length of the retina. The visual analysis that is performed is not, however, limited to one dimension. Motion energy filters are easily generalized to a two-dimensional retina. Our approach should be extendable to the two-dimensional tracking problem.

The backgrounds of images that we used for tracking were featureless. The current system cannot distinguish target features from background features. Also, the interface network was designed to track a single object in the absence of moving distractors. The next step is to expand this interface to model the attentional phenomena observed in primate tracking, especially the process of initial target

[2]We selected time constants, delays, and sampling rates throughout the model to roughly approximate the time course of the primate visual tracking response. The model runs on a workstation taking approximately thirty times real-time to complete a processing step.

acquisition.

7 Conclusion

In simulations, our eye tracking model performed well. Many additional difficulties must be addressed, but we feel this system can perform well under real-world real-time constraints. Previous work by Lisberger and Sejnowski (1992) demonstrates that this visual tracking model can be integrated with inertial eye stabilization—the vestibulo-ocular reflex. Ultimately, it should be possible to build a physical system using these design principles.

Every component of the system was designed using network learning techniques. The visual processing, for example, had a variety of components that were trained separately and in combinations. The architecture of the networks were based on the anatomy and physiology of the visual and oculomotor systems. This approach to reverse engineering is based on the existing knowledge of the flow of information through the relevant brain pathways.

It should also be possible to use the model to develop and test theories about the nature of biological visual tracking. This is just a first step toward developing a realistic model of the primate oculomotor system, but it has already provided useful predictions for the possible sites of plasticity during gain changes of the vestibulo-ocular reflex (Lisberger and Sejnowski, 1992).

References

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models of the perception of motion. *Journal of the Optical Society of America*, 2(2):284–299, 1985.

[2] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.

[3] R.J. Krauzlis and S. G. Lisberger. A control systems model of smooth pursuit eye movements with realistic emergent properties. *Neural Computation*, 1:116–122, 1992.

[4] S. G. Lisberger, E. J. Morris, and L. Tychsen. *Ann. Rev. Neurosci.*, 10:97–129, 1987.

[5] S.G. Lisberger and T.J. Sejnowski. Computational analysis suggests a new hypothesis for motor learning in the vestibulo-ocular reflex. Submitted for publication., 1992.

[6] W.T. Newsome and E. B. Pare. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neuroscience*, 8:2201–2211, 1988.

[7] D. A. Robinson. Models of oculomotor neural organization. In P. Bach y Rita and C. C. Collins, editors, *The Control of Eye Movements*, page 519. Academic, New York, 1971.