Communicated by Richard Andersen

A Neural Network for the Processing of Optic Flow from Ego-Motion in Man and Higher Mammals

Markus Lappe
Josef P. Rauschecker
Laboratory of Neurophysiology, NIMH, Poolesville, MD 20837, USA and Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Interest in the processing of optic flow has increased recently in both the neurophysiological and the psychophysical communities. We have designed a neural network model of the visual motion pathway in higher mammals that detects the direction of heading from optic flow. The model is a neural implementation of the subspace algorithm introduced by Heeger and Jepson (1990). We have tested the network in simulations that are closely related to psychophysical and neurophysiological experiments and show that our results are consistent with recent data from both fields. The network reproduces some key properties of human ego-motion perception. At the same time, it produces neurons that are selective for different components of ego-motion flow fields, such as expansions and rotations. These properties are reminiscent of a subclass of neurons in cortical area MSTd, the triple-component neurons. We propose that the output of such neurons could be used to generate a computational map of heading directions in or beyond MST.

1 Introduction

The concept that optic flow is important for visual navigation dates from the work of Gibson in the 1950s. Gibson (1950) showed that the optic flow pattern experienced by an observer moving along a straight line through a static environment contains a singularity that he termed the focus of expansion. He hypothesized that the visual system might use the global pattern of radial outflow originating from this singularity to determine the translational heading of the observer.

A host of studies in human psychophysics have followed up Gibson's ideas (Regan and Beverly 1982; Rieger and Toet 1985; Warren *et al.* 1988; Warren and Hannon 1988, 1990). Regan and Beverly (1982) rejected his hypothesis on the basis that the optic flow pattern that arrives on the

Neural Computation **5, 374–391** (1993)

retina is radically altered by eye movements of the observer. Then the flow field becomes a superposition of the radial outflow pattern with a circular flow field that is obtained when the eyes move in the orbita. Generally the resulting vector field may also have a singular point similar to a focus of expansion, but this point does not necessarily coincide with the heading direction. If, for instance, the eye rotation results from the fixation of a point in the environment, the singularity will be at the fixation point instead of the destination point.

Nevertheless, Warren and Hannon (1990) found humans capable of judging their heading with great accuracy from optic flow patterns that simulated translation plus eye rotation. Their subjects were able to perceive their heading with a mean error between one and two degrees solely from the optic flow. No nonvisual information such as oculomotor signals was necessary. This ability persisted over a natural range of speeds and over a variation of the number of visible moving points between 10 and several hundred. The performance of the subjects was at chance, however, when no depth information in the form of motion parallax was available.

In the visual system there are at least two (maybe three) main streams of information flow (Mishkin *et al.* 1983; Livingstone and Hubel 1988; Zeki and Shipp 1988). In the simplest depiction, there is an inferotemporal system that is mainly responsible for the processing of form, and a parietal system that processes motion (Ungerleider and Mishkin 1982). Within the cortical motion system, one of the prominent and most investigated areas in primates is the middle temporal area or area MT (Allman and Kaas 1971). In cats the probable homologue for MT is the Clare–Bishop area (Clare and Bishop 1954), also called area PMLS (Palmer *et al.* 1978). Evidence from both areas suggests that they participate in the processing of flow field information. Both areas contain neurons that are highly direction selective and respond well to moving stimuli. It has first been found in cat area PMLS that a majority of neurons prefer movement away from the area centralis, that is, centrifugal motion (Rauschecker *et al.* 1987a,b; Brenner and Rauschecker 1990). The same has been found in monkey area MT (Albright 1989), thus strengthening the likelihood of a homology between these two areas. Other studies have revealed single neurons in PMLS that respond well to approaching or receding objects (Toyama *et al.* 1990).

More recently, a number of studies have described neurons in the dorsal part of monkey area MST (MSTd) that respond best to large expanding/contracting, rotating, or shifting patterns (Tanaka and Saito 1989a,b; Andersen *et al.* 1990; Duffy and Wurtz 1991a,b). The response of these neurons often shows a substantial invariance to the position of the stimulus. Duffy and Wurtz (1991a,b) found that a majority of the neurons in MSTd responded not only to one component of motion of the stimulus pattern (e.g., expansion or contraction), but rather to two or all three of them separately. About one-third of MSTd cells displayed selectivity to expansions or contractions *and* clockwise or counterclockwise rotations *and* showed broad directional tuning for shifting dot patterns when tested with these stimuli one after another. It is these "triple component cells" that our model is mainly concerned with. Furthermore, cells in MSTd are unselective for the overall speed of a stimulus and for the amount of depth information available in the stimulus.

There have been a number of computational approaches to extract navigational information from optic flow focusing on different mathematical properties of the flow field. The difficulty of the task is that in the mapping of three-dimensional movements onto a two-dimensional retina some information is lost that cannot be fully recovered. Models that use differential invariants (Koenderink and van Doorn 1981; Longuet-Higgins and Prazdny 1980; Waxman and Ullman 1985) require dense optic flow to compute derivatives. By contrast, humans are quite successful with sparse fields (Warren and Hannon 1990). Models based on algorithms that solve a set of equations for only a small number of vectors (Prazdny 1980; Tsai and Huang 1984), on the other hand, require precise measurements and are very sensitive to noise. Methods that rely on motion parallax or local differential motion (Longuet-Higgins and Prazdny 1980; Rieger and Lawton 1985) are in agreement with the psychophysical data in that they fail in the absence of depth in the environment. However, they require accurate measurements at points that are close to each other in the image but are separated in depth, which is an especially difficult task to accomplish. Furthermore, recent psychophysical studies (Stone and Perrone 1991) have shown that local depth variations are not necessary. Least-square minimization algorithms (Bruss and Horn 1983; Heeger and Jepson 1990) that use redundant information from as many flow vectors as are available are robust and comparatively insensitive to noise.

None of the above-mentioned algorithms is clearly specified in terms of a neural model. Given the current advances in visual neurophysiology, it seems desirable to construct a neural network for ego-motion perception that is consistent with the neurophysiological and psychophysical data. Recently a network model of heading perception in the simpler case without eye movements has been described (Hatsopoulos and Warren 1991), which accounts for some psychophysical findings. A neural model we presented in brief form earlier together with first results from the model described in this paper (Lappe and Rauschecker 1991) is also concerned with pure translations. It uses a centrifugal bias similar to the one found in PMLS and MT to achieve precise heading judgments with neuronal elements that are as broadly directionally tuned as the cells found in these areas.

In this article we present a new neural network that succeeds when the radial flow pattern is disturbed by eye movements. The network is capable of reproducing many of the psychophysical findings, and the single units exhibit great similarity to the triple component cells of Duffy and Wurtz (1991a,b) in area MSTd.

2 The Model

Our network is built in two layers. The first layer is designed after monkey area MT and represents the input to the network. The second layer is constructed to yield a representation of the heading direction as the output of the net and thus could form a model of MSTd. In each network layer we employ a population encoding of the relevant variables, namely the speed and direction of local movements in layer one and the heading direction of the individual in layer two. The computation of the direction of translation is based on the subspace algorithm by Heeger and Jepson (1990). Its main course of action is to eliminate the dependencies on depth and rotation first and thereby gain an equation that depends only on the translational velocity. Therefore it bears some similarity to Gibson's original claim that the visual system can decompose the optic flow into its translational and rotational components. We will restrict the scope of our model to such eye movements as occur when the observer keeps his eyes fixed on a point in the environment while he is moving. This is a natural and frequently occurring behavior, and we believe that using assumptions that are a reflection of the behavior of an animal or a human being makes it more likely to gain results that can be compared with experimental data. Although it is mathematically possible to include any type of eye movements, it is not very likely that the eyes would rotate around their long axis to a significant amount during locomotion. Note that our assumption includes the case of no eye movements at all, since it can be described as gazing at a point infinitely far away.

2.1 Optic Flow and the Subspace Algorithm.

Optic flow is the projection of the motion of objects in the three-dimensional world onto a two-dimensional image plane. In three dimensions, every moving point has six degrees of freedom: The translational velocity $\mathbf{T} = (T_x, T_y, T_z)^t$ and the rotation $\mathbf{\Omega} = (\Omega_x, \Omega_y, \Omega_z)^t$. When an observer moves through a static environment all points in space share the same six motion parameters. The motion of a point $\mathbf{R} = (X, Y, Z)^t$ in a viewer-centered coordinate system is $\mathbf{V} = -(\mathbf{\Omega} \times \mathbf{R} + \mathbf{T})$. This motion is projected onto an image plane. Writing two-dimensional image vectors in small letters, the perspective projection of a point is $\mathbf{r} = (x, y)^t = f\ (X/Z, Y/Z)^t$, where f denotes the focal length. Following Heeger and Jepson (1990) the image velocity can be written as the sum of a translational and a rotational component:

$$\boldsymbol{\theta}(x,y) = (dx/dt, dy/dt) = p(x,y)\mathbf{A}(x,y)\mathbf{T} + \mathbf{B}(x,y)\mathbf{\Omega} \tag{2.1}$$

where $p(x,y) = 1/Z$ is the inverse depth, and

$$\mathbf{A}(x,y) = \begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}$$

$$\mathbf{B}(x,y) = \begin{pmatrix} (xy)/f & -(f + x^2/f) & y \\ f + y^2/f & -(xy)/f & -x \end{pmatrix}$$

The unknown depth and translational velocity are multiplied together and can thus only be determined up to a scale factor. Regarding therefore the translation $\mathbf{T}$ as a unit vector, one is left with six unknowns: p, the two remaining components of $\mathbf{T}$, and the three components of $\boldsymbol{\Omega}$, but only two known quantities, θ_x, θ_y. The subspace algorithm uses flow vectors at five distinct image points to yield an overdetermined system of equations that is solved with a minimization method in the following way: The five separate equations are combined into one matrix equation $\boldsymbol{\Theta} = \mathbf{C}(\mathbf{T})\mathbf{q}$ where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_5)^t$ is now a 10-dimensional vector consisting of the components of the five image velocities, $\mathbf{q} = [p(x_1,y_1), \ldots, p(x_5,y_5), \Omega_x, \Omega_y, \Omega_z]^t$ an eight-dimensional vector, and $\mathbf{C}(\mathbf{T})$ a 10×8 matrix composed of the $\mathbf{A}(x_i,y_i)\mathbf{T}$ and $\mathbf{B}(x_i,y_i)$ matrices:

$$\mathbf{C}(\mathbf{T}) = \begin{pmatrix} \mathbf{A}(x_1,y_1)\mathbf{T} & & & \mathbf{B}(x_1,y_1) \\ & \ddots & & \vdots \\ & & \mathbf{A}(x_5,y_5)\mathbf{T} & \mathbf{B}(x_5,y_5) \end{pmatrix}$$

Heeger and Jepson (1990) then show that the heading direction can be recovered by minimizing the residual function

$$R(\mathbf{T}) = \|\boldsymbol{\Theta}^t \mathbf{C}^{\perp}(\mathbf{T})\|^2,$$

where $\mathbf{C}^{\perp}(\mathbf{T})$ is a matrix that spans the two-dimensional orthogonal complement of $\mathbf{C}(\mathbf{T})$.

2.2 Restriction to Fixations during Locomotion. We now restrict ourselves to only those eye movements that arise through the fixation of a point $\mathbf{F} = (0, 0, 1/p_{\mathrm{F}})^t$ in the center of the visual field while the observer moves along a straight line. The rotation that is necessary to fixate this point can be derived from the condition that the flow at this point has to be zero:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = p_{\mathrm{F}} \begin{pmatrix} -f & 0 & 0 \\ 0 & -f & 0 \end{pmatrix} \mathbf{T} + \begin{pmatrix} 0 & -f & 0 \\ f & 0 & 0 \end{pmatrix} \boldsymbol{\Omega}$$

Choosing $\Omega_z = 0$ we find $\boldsymbol{\Omega} = p_{\mathrm{F}}(T_y, -T_x, 0)^t$. The optic flow then is:

$$\boldsymbol{\theta}(x,y) = \left[p(x,y) \begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix} + p_{\mathrm{F}} \begin{pmatrix} f + x^2/f & (xy)/f & 0 \\ (xy)/f & f + y^2/f & 0 \end{pmatrix} \right] \mathbf{T}$$

The case of a straight translation without any eye movements can easily be described within this framework by considering a fixation point that is infinitely far away. Then p_F and the rotational velocity $\mathbf{\Omega}$ are zero, resulting in a purely translational flow.

The optic flow equation above has only four unknowns: $p(x, y)$, p_F, T_x, and T_y. Combining the equations for two different flow vectors into one matrix equation in the same way as before yields $\mathbf{\Theta} = \mathbf{C}\ (\mathbf{T}) \cdot [p(x_1, y_1), p(x_2, y_2), p_F]^t$, where $\mathbf{C}(\mathbf{T})$ is now only a 4×3 matrix, the orthogonal complement of which is a line given by the vector $\mathbf{C}^{\perp}(\mathbf{T})$. The residual function becomes the scalar product between this vector and the observed flow:

$$R(\mathbf{T}) = |\mathbf{\Theta}^t \mathbf{C}^{\perp}(\mathbf{T})|^2 \tag{2.2}$$

Since the optic flow is a linear function of the translational direction, $R(\mathbf{T})$ does not have a single minimum but is equal to zero along a line in the (T_x, T_y) plane. Therefore one such minimization alone cannot give the translational velocity, rather several pairs of flow vectors with different $R(\mathbf{T})$ functions have to be used in conjunction.

2.3 The Network. In the first layer of the network, which constitutes the flow field input, 300 random locations within 50° of eccentricity are represented. We assume a population encoding of the optical flow vectors at each location by small sets of neurons that share the same receptive field position but are tuned to different directions of motion. Each such group consists of n' neurons with preferred directions $\mathbf{e}_k$, $k = 1, \ldots, n'$. The flow vector $\boldsymbol{\theta}$ is represented by the sum over the neuronal activities s_k in the following way:

$$\boldsymbol{\theta} = \sum_{k=1}^{n'} s_k \mathbf{e}_k \tag{2.3}$$

We do not concern ourselves with how the optic flow is derived from the luminance changes in the retina or how the aperture problem is solved. Neural algorithms that deal with these questions have already been developed (Bülthoff *et al.* 1989; Hildreth 1984; Yuille and Grzywacz 1988). A physiologically plausible network model that yields as its output a population encoding like the one we use here has been proposed by Wang *et al.* (1989). It can be thought of as a preprocessing stage to our network, modeling the pathway from the retina to area MT or PMLS.

Since we start out with a layer in which the optic flow is already present, we have to guarantee that the tuning curves of the neurons and the distributions of the preferred directions match the requirement of equation 2.3. As the simplest choice for our model, we use a rectified cosine function with $n' = 4$. It preserves the most prominent feature of the observed directional tuning curves in MT/PMLS, namely broad unidirectional tuning with no response in the null direction. The preferred directions are equally spaced, $\mathbf{e}_k = [\cos(\pi k/2), \sin(\pi k/2)]$, and for the unit's response to a movement with speed θ_0 and direction ϕ, the tuning curve is

$$s_k = \begin{cases} \theta_0 \cos(\phi - \pi k/2) & \text{if } \cos(\phi - \pi k/2) > 0 \\ 0 & \text{otherwise} \end{cases}$$

The second layer represents a population encoding of the translational direction of the movement of the observer, which is represented by the intersection point of the 3D-movement vector $\mathbf{T}$ with the image plane. There are populations of n neurons at possible intersection points whose combined activities u_l give the perceived direction. But here the sum of the activities $U = \sum_{l=1}^{n} u_l$ at each position yields a measure of how likely this position is to be the correct direction of movement. The perceived direction is chosen to be the one that has the highest total activity.

The output of a second layer neuron is a sigmoid function $g(x)$ of the sum of the activities of its m input neurons weighted by synaptic strengths J_{jkl} and compared to a threshold μ:

$$u_l = g\left(\sum_{i=1}^{m}\sum_{k=1}^{n'} J_{ikl} s_{ik} - \mu\right) \tag{2.4}$$

Here J_{ikl} denotes the strength of the connection between the lth output neuron and the kth input neuron in the population that represents image location i. The sigmoid function is symmetric such that $g(-x) = 1 - g(x)$.

The connections and their strengths are set once before the network is presented with any stimuli, and are fixed afterward. First a number of image locations are randomly assigned to a second layer neuron. Then, values for the synaptic strengths are calculated so that the population of neurons encoding a specific $\mathbf{T}$ is maximally excited when $R(\mathbf{T})$ equals zero. Although the neuron may receive input from a large number of image locations we start the calculation of the connections with only two in order to keep it simple. We want the sum in equation 2.4 to equal the scalar product on the right side of equation 2.2:

$$\sum_{i=1}^{2}\sum_{k=1}^{n'} J_{ikl} s_{ik} = \mathbf{\Theta}^t \mathbf{C}^{\perp}(\mathbf{T})$$

For every single image location i we have

$$\sum_{k=1}^{n'} J_{ikl} s_{ik} = \boldsymbol{\theta}_i^t \begin{pmatrix} C^{\perp}_{2i-1}(\mathbf{T}) \\ C^{\perp}_{2i}(\mathbf{T}) \end{pmatrix}, \qquad i = 1, 2$$

Substituting equation 2.3 we find

$$\sum_{k=1}^{n'} J_{ikl} s_{ik} = \sum_{k=1}^{n'} s_{ik} \mathbf{e}_{ik} \begin{pmatrix} C^{\perp}_{2i-1}(\mathbf{T}) \\ C^{\perp}_{2i}(\mathbf{T}) \end{pmatrix}, \qquad i = 1, 2$$

Therefore we set the synaptic strengths to

$$J_{ikl} = \mathbf{e}_{ik} \begin{pmatrix} C^{\perp}_{2i-1}(\mathbf{T}) \\ C^{\perp}_{2i}(\mathbf{T}) \end{pmatrix}$$

If the neuron is connected to more than two image locations the input connections are divided into pairs and the connections are calculated separately for each pair.

Now the question of when $R(\mathbf{T})$ is minimal comes down to the question of when all the neurons' inputs balance each other to give a net input of zero. Consider two output neurons u_l and $u_{l'}$ receiving input from the same set of first layer neurons but with inverse connections such that $J_{ikl'} = -J_{ikl}$. Then, if the threshold μ equals zero, the sum of both neurons' activities is equal to 1 regardless of their inputs, since the sigmoid input/output function is symmetric. If, however, μ has a slightly negative value, both sigmoid functions will overlap and the sum will have a single peak at an input value of zero. Such a matched pair of neurons generates its maximal activity when $R(\mathbf{T}) = 0$.

MSTd neurons have very large receptive fields and do certainly receive input from more than 2 image locations. Also MSTd neurons show the same response in the case of as little as 25 visible moving dots as they do in the case of 300 (Duffy and Wurtz 1991a). We chose each of our model neurons to receive input from 30 image locations. We restrict the space for the encoded heading directions to the innermost $20 \times 20^\circ$ of the visual field, since this approximates the range over which the psychophysical experiments have been carried out. Nevertheless, each layer-two neuron may receive input from a much larger part of the visual field. The layer-two neurons form a three-dimensional grid with 20×20 populations encoding one degree of translation-space each, and 20 pairs of neurons in each population.

3 Results

3.1 Comparison of the Network's Performance with Human Psychophysical Data. The network was tested with simulated flow fields with different motion parameters. We used a cloud-like pattern that consisted of a number of dots, the depths of which were randomly distributed within a given range. To test the behavior without eye movements a translational direction was randomly chosen within the innermost $20 \times 20^\circ$ and the rotation was set to zero. To test cases with eye rotation the translational direction was again chosen randomly and the fixation point was set in the center of the image plane and assigned a specific depth. The rotational component was then calculated from the condition that the flow at the fixation point must be zero. Each simulation run consisted of 100 presentations of different flow fields, after which we calculated the mean error as the mean angular difference between the network's computed direction and the correct direction.

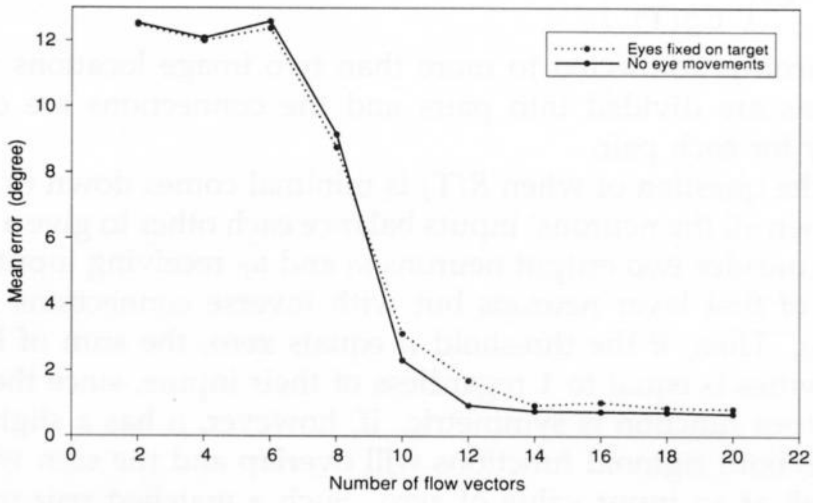


Figure 1: Performance with sparse flow fields. The heading error becomes small with as little as 10 vectors. The number of dots necessary is about the same with or without eye movements.

We found the network's performance to be well within the range of human performers (Warren *et al.* 1988). For pure translation as well as with eye movements the mean error settled between 0.5 and 1.5°, showing that the network always has its activity maximum at a position close to the translational direction. Consistent with the experiments of Warren *et al.* (1988) we found very little influence of speed on the performance of the network.

Humans are able to detect their heading with very sparse flow fields consisting of only ten dots (Warren and Hannon 1988, 1990). In order to test how many flow vectors are needed in our model under otherwise optimal conditions we made an additional assumption: We assumed that a given pair of vectors in the flow field serves as input to at least one pair of neurons in each population of the output layer. If this were not the case, some populations would receive more information than others and the number of dots neccessary for correct heading estimation would depend on the heading direction. Our assumption ensures that all heading directions are represented equally. Considering the large number of cortical neurons this assumption is biologically reasonable since it would be approximately fulfilled if the number of neurons in the output layer were large. For the simulations, we distributed the connections between input and output neurons in such a way as to fulfill the assumption. The results of the simulations are shown in Figure 1. The cloud of dots extended in depth from 11 to 31 m with a fixation point at 21 m. The translational speed was 2 m/sec. In both the pure translation and in the eye rotation case the network started to detect the heading with the desired accuracy at approximately 10 points, although with eye rotation the error did not quite reach the optimum and continued to decrease as more flow vectors were provided. Mathematically two vectors are sufficient to compute the heading of a purely translational movement (Prazdny 1980), but humans fail to detect their heading with only two visible dots (Warren *et al.* 1988). Our network does not know a priori if the flow field is generated by a translation alone. It therefore has to rely on the flow pattern and needs about the same number of vectors as with the eye movements.

Humans also fail when eye rotations are paired with a perpendicular approach to a solid wall, where all points are at the same depth (Rieger and Toet 1985; Warren and Hannon 1990). In this case the subjects' performances are at chance and they often report themselves as heading toward the fixation point. Because of a well-known ambiguity in planar flow fields (Tsai and Huang 1984), we were not able to test the depth dependence of the network with approaches to a plane at different angles. We therefore varied the depth range of the cloud. Doing this revealed that with decreasing depth the peak in the second layer grows broader and covers the fixation point as well as the heading direction. This can be seen in Figure 2 where the summed population activities in the output layer are shown on a grayscale map, together with reduced pictures of the input flow fields. Input and output are compared for situations that differ in the amount of depth in the image. In Figure 2a a flow field is shown in which the depth range of the cloud of dots is large, extending from 7 to 30 m. The observer moves toward the cross while he is keeping his eyes fixed on an object (×) in the center. There is no apparent focus of expansion. The network output (Fig. 2b) shows an easily localizable brightness peak in the upper left that corresponds to the correct heading direction as indicated by the cross. Figure 2c shows the same movement as Figure 2a, but here the depth range of the cloud is much smaller, ranging from 19 to 21 m. In this case the flow field looks very much like an expansion centered at the fixation point. In the corresponding network output (Fig. 2d), the peak is very broad and includes the fixation point in the center. A maximum nevertheless still exists, although much less pronounced, and in the simulations the network was still able to compute the right heading. However, the solution is unstable and very sensitive to noise. To illustrate this, we randomly varied the amplitudes of the flow vectors by stretching them by a factor distributed uniformly between 0.9 and 1.1, thus adding 10% noise. The results for all conditions are shown in Figure 3 for different depth ranges. This small amount of noise increases the error for the rotational movement to around 7°, whereas in the purely translational case the network performance is unaffected. With growing depth differences this separation becomes less pronounced and the error values for the rotational case decrease.

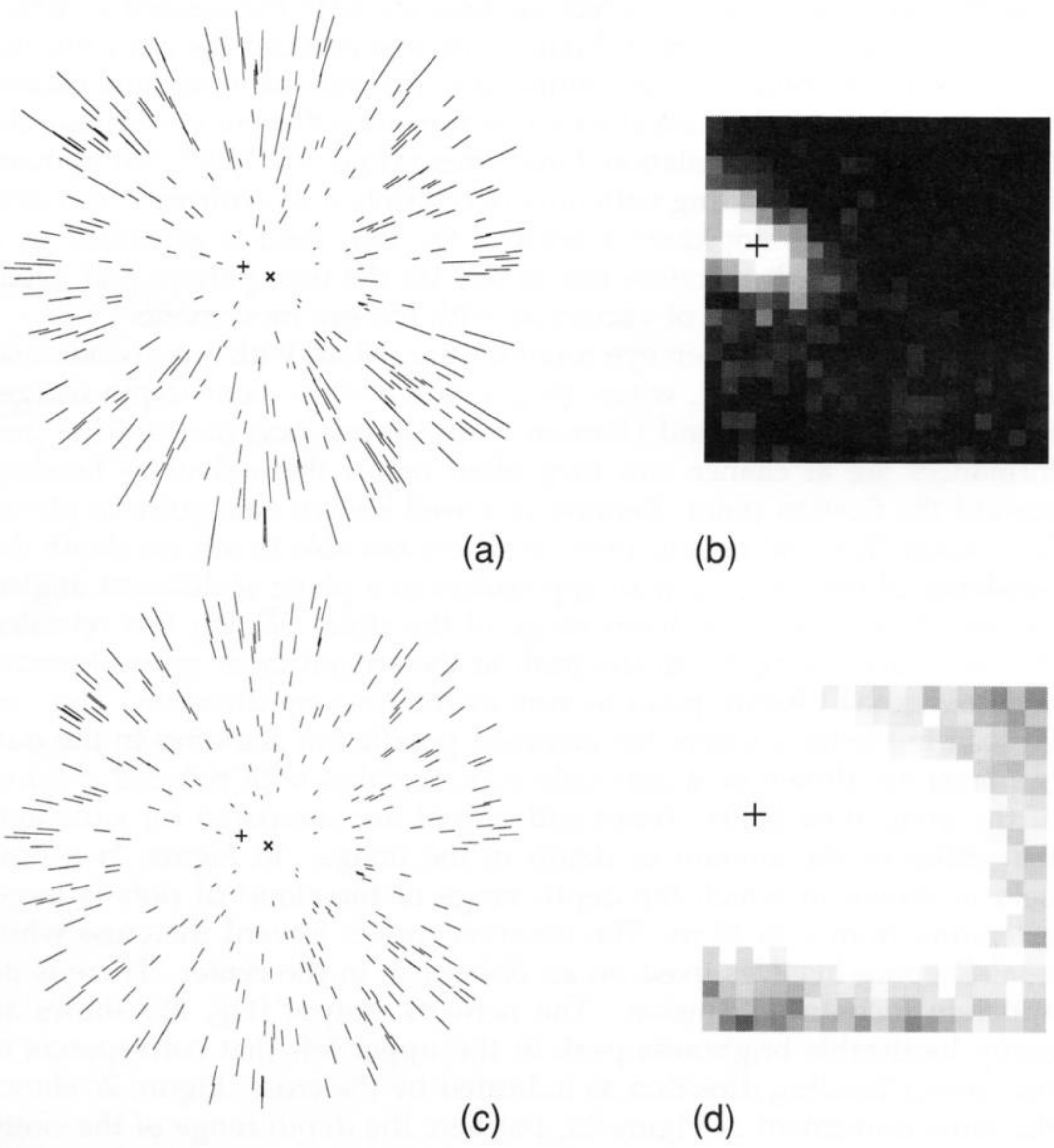


Figure 2: Influence of image depth on the heading judgment of the network. (a) Depth-rich flow field. Movement is toward the cross (+) while the × in the center is fixated. (b) Output of the network. The response peak gives the correct heading. (c) Same movement with only little depth differences. (d) Brightness maximum in the output of the network is very broad and includes the fixation point.

3.2 Comparison with Single Cell Properties in MSTd. The output layer cells of our model network exhibit a remarkable resemblance to some triple component neurons in MSTd. Figure 4 shows the response of *one* output layer cell to presentations of each of the components (e.g., expansions, rotations) at different places in the visual field. The neuron

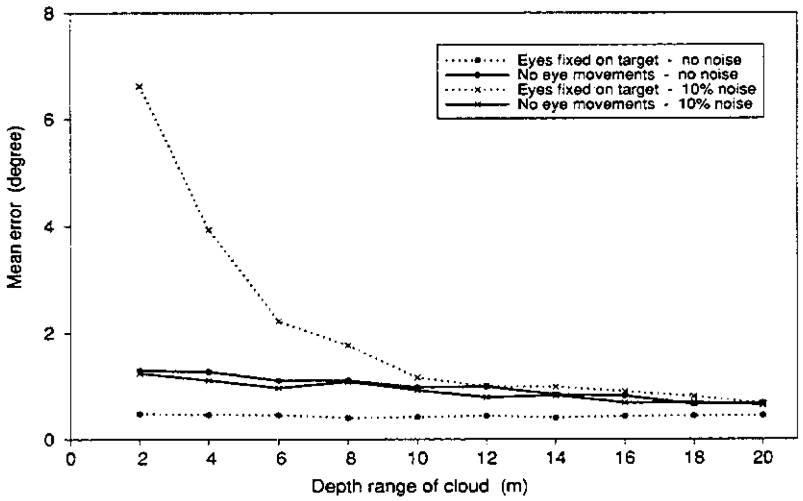
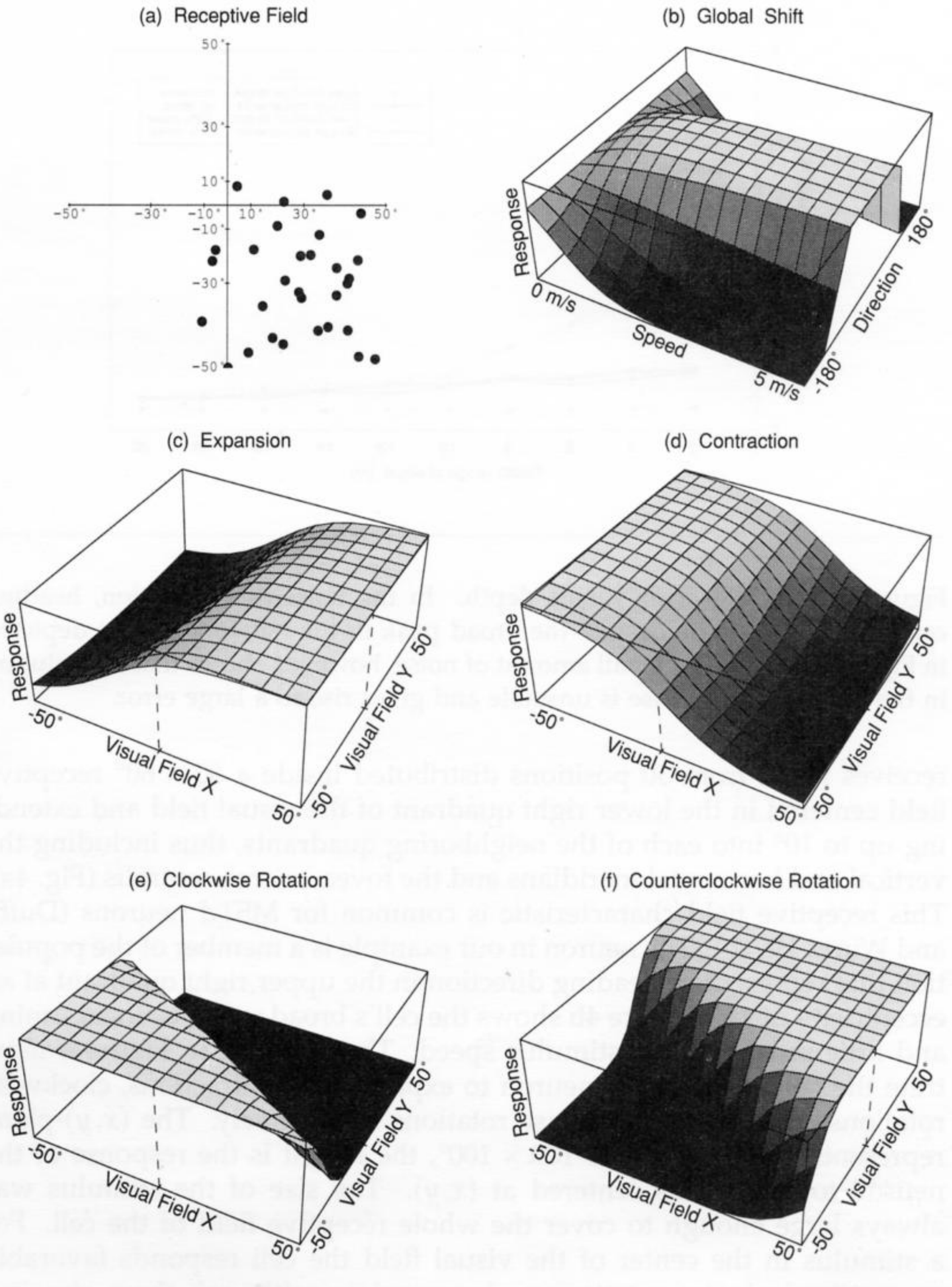


Figure 3: Heading error versus depth. In the noise-free condition, heading calculation is accurate despite the broad peak in the network output depicted in Figure 2d. Adding a small amount of noise, however, shows that the solution in the eye movement case is unstable and gives rise to a large error.

receives input from 30 positions distributed inside a 60 × 60° receptive field centered in the lower right quadrant of the visual field and extending up to 10° into each of the neighboring quadrants, thus including the vertical and horizontal meridians and the fovea or area centralis (Fig. 4a). This receptive field characteristic is common for MSTd neurons (Duffy and Wurtz 1991b). The neuron in our example is a member of the population that represents a heading direction in the upper right quadrant at an eccentricity of 11°. Figure 4b shows the cell's broad unidirectional tuning and little selectivity for stimulus speed. The plots c–f in Figure 4 illustrate the responses of the neuron to expansions, contractions, clockwise rotations, and counterclockwise rotations, respectively. The (x, y)-plane represents a visual field of 100 × 100°, the height is the response of the neuron to a stimulus centered at (x, y). The size of the stimulus was always large enough to cover the whole receptive field of the cell. For a stimulus in the center of the visual field the cell responds favorably to counterclockwise rotations and expansions, although there also is a smaller response to contractions. There are very large areas of position invariance covering almost half of the visual field for a given stimulus movement. The response to counterclockwise rotations, for instance, is constant in most of the upper two quadrants.

The cell also shows the reversals in selectivity observed in 40% of triple-component neurons in MSTd (Duffy and Wurtz 1991b). In our example, moving the center of the stimuli to the right causes the response



(a) Receptive Field

(b) Global Shift

(c) Expansion

(d) Contraction

(e) Clockwise Rotation

(f) Counterclockwise Rotation

Figure 4: Responses of *one* output layer cell. (a) Receptive field of the cell as defined by its input connections. (b) Broad unidirectional response to global shifts of a dot pattern. No tuning to a particular stimulus speed. (c–f) Responses to expanding, contracting, and rotating patterns centered at different positions within the visual field reveal large areas of position invariance and sudden reversals of selectivity.

to contractions to disappear. Moving the center of the stimuli to the lower left causes the cell's selectivity to change to favor contractions and clockwise rotations. There are intermediate positions where the cell responds to both modes of one component. For example, in plots b and c, there is a vertical strip in the center where the cell responds to expansions as well as to contractions.

The response reversals take place along edges running across the visual field, which is similar to the findings of Duffy and Wurtz (1991b). The reason for this is that the residual function, which is computed by the neuron, equals zero along a line in the (T_x, T_y) space, as mentioned before. The edge of the surface that marks the neuron's response to expansions follows this line. The neuron signals only that the heading direction lies somewhere along the edge. The edges of all neurons in one population overlap at the point that corresponds to the heading represented by that population. When the network is presented with a flow field, the population encoding the correct heading is maximally excited since all of their neurons will respond. In populations representing other directions, only part of the neurons will be active, so that the total activity will be smaller.

It is worth noting that the edges of reversal do not necessarily cross the receptive field of the cell. In the example of Figure 4, the reversal from selectivity for expansion to selectivity for contraction takes place in the left half of the visual field outside the cell's receptive field, which occupies the lower right quadrant. Likewise, it sometimes occurred in the simulations that the reversal for rotation was not even contained within the 100 × 100° visual field.

Another interesting observation is that the edges for rotation and expansion/contraction often cross each other approximately orthogonally. The position of the intersection point, on the other hand, can vary widely between cells.

4 Discussion

We have designed a neural network that detects the direction of egomotion from optic flow and is consistent with recent neurophysiological and psychophysical data. It solves the traditional problem of eye movements distorting the radial flow field by means of a biologically reasonable mechanism.

The model reproduces some key properties of human ego-motion perception, namely, the ability to function consistently over a range of speeds, the ability to work with sparse flow fields, and the difficulties in judging the heading when approaching a wall while moving the eyes.

The network also generates interesting neuronal properties in its output layer. Simple intuitive models for heading perception might expect a single neuron to show a peak of activity for an expansion at a certain preferred heading direction. Instead, our model uses a population encoding in which single cells do not carry all the information about the perceived heading, but rather the combined activity of a number of cells gives that information. At the level of a single neuron, the position information is contained in the edges of reversal of the cell's preferred direction of stimulus motion.

The resulting characteristics of the output neurons in our network show great similarity to the response properties of a particular cell class recently described in MSTd, the triple-component neurons (Duffy and Wurtz 1991a,b). These cells, which comprise about one-third of all neurons in MSTd, display selectivity not only for expansion or contraction, but also for one type of rotation and one direction of shifting patterns. Most of the neuronal outputs produced by our network have similar properties. It appears tempting to postulate, therefore, that the output of triple-component cells could be used to compute directional heading, either within MST or in another area.

A potential problem for using the output of MSTd neurons to compute heading direction concerns their apparent position invariance. In a neural network that is supposed to signal the directional heading, the response of the output layer cells has to depend on the position of the stimulus in some way. Most neurons in MSTd seem to be insensitive against changes of stimulus position, although the proportions of position invariant cells reported in different studies vary and obviously depend on the exact stimulus paradigm (Andersen *et al.* 1990; Duffy and Wurtz 1991b; Orban *et al.* 1992). In our network model many output neurons would appear position invariant when tested over a limited, wide range of stimulus positions. Interestingly, the proportion of position dependent responses seems to be highest among triple-component neurons (Duffy and Wurtz, 1991b): In about 40% of these cells component selectivity for a flow field stimulus is reversed along oriented edges, which conforms exactly with the behavior of our model neurons. It is conceivable, therefore, that it is this subtype of triple-component neurons that is involved in the computation of heading direction. More neurons of this type might be encountered in MSTd if one specifically looks for them. Their frequency of occurrence may depend on laminar position, or they might be found even more frequently at another processing stage.

A closer look at the experimental data reveals that the number of triple component cells in MSTd may indeed have been underestimated. The different cell types in MSTd do not fall in strictly separate classes but rather form a continuum changing smoothly from triple to single component cells (Duffy and Wurtz 1991a). Therefore, double and single component cells might be regarded as possessing some, albeit weak, responses to the other components. It is equally possible, however, that single and double component cells simply do not participate in the detection of heading direction, but serve some other purpose. Single component cells, for example, could be involved in the analysis of object motion.

The network can also generate cells that are selective to fewer components when the restriction is removed that rotations are due to the fixation of an object. Allowing arbitrary rotations, including ones around a sagittal axis through the eye, results in neurons that are unselective for rotations and respond only to translations and expansions/contractions. Under the different assumption that only frontoparallel rotations, including for instance pursuit eye movements, will occur, the neurons show strong, fully position invariant responses to rotational stimuli, which dominate over the selectivity for translation and expansion/contraction (Lappe and Rauschecker 1993).

We would like to emphasize that the neurons in our model do not decompose the flow field directly. At no point is the translational part of the optic flow actually computed. The neurons rather test the consistency of a measured optic flow with a certain heading direction. In this way, a response selectivity for rotations, for example, does not mean that the neuron is actually *tuned* to the detection of a rotation in the visual field, but this property rather has to be regarded as the result of a more complex selectivity.

The cells in the output layer of our model form a computational map of all possible heading directions. However, it would not be easy to find this map in an area of visual cortex, since the topography reveals itself only in the properties of cell populations. Simultaneous recording from an array of electrodes would perhaps be the only way to demonstrate this computational map experimentally. Our model suggests that one has to focus on the mapping of selectivity reversals and explore these more thoroughly, especially in triple component cells: Neurons in neighboring columns should show smooth shifts of their preferences. The concurrent activity of such cells in a hypercolumn would signal one particular heading direction in space, which is given by the intersection point of their reversal edges for expansion and contraction.

References

Albright, T. D. 1989. Centrifugal directionality bias in the middle temporal visual area (MT) of the macaque. *Visual Neurosci.* **2**, 177–188.

Allman, J. M., and Kaas, J. H. 1971. A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (Aotus trivirgatus). *Brain Res.* **31**, 85–105.

Andersen, R., Graziano, M., and Snowden, R. 1990. Translational invariance and attentional modulation of MST cells. *Soc. Neurosci. Abstr.* **16**, 7.

Brenner, E., and Rauschecker, J. P. 1990. Centrifugal motion bias in the cat's lateral suprasylvian visual cortex is independent of early flow field exposure. *J. Physiol.* **423**, 641–660.

Bruss, A. R., and Horn, B. K. P. 1983. Passive navigation. *Computer Vision, Grahics, Image Process.* **21**, 3–20,

Bülthoff, H., Little, J., and Poggio, T. 1989. A parallel algorithm for real-time computation of optical flow. *Nature (London)* **337**, 549–553.

Clare, M. H., and Bishop, G. H. 1954. Responses from an association area secondarily activated from optic cortex. *J. Neurophysiol.* **17**, 271–277.

Duffy, C. J., and Wurtz, R. H. 1991a. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* **65**(6), 1329–1345.

Duffy, C. J., and Wurtz, R. H. 1991b. Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J. Neurophysiol.* **65**(6), 1346–1359.

Gibson, J. J. 1950. *The Perception of the Visual World.* Houghton Mifflin, Boston.

Hatsopoulos, N. G., and Warren, W. H., Jr. 1991. Visual navigation with a neural network. *Neural Networks* **4**(3), 303–318.

Heeger, D. J., and Jepson, A. 1990. Visual perception of three-dimensional motion. *Neural Comp.* **2**, 129–137.

Hildreth, E. C. 1984. *The Measurement of Visual Motion.* MIT, Cambridge, MA.

Koenderink, J. J., and van Doorn, A. J. 1981. Exterospecific component of the motion parallax field. *J. Opt. Soc. Am.* **71**(8), 953–957.

Lappe, M., and Rauschecker, J. P. 1991. A neural network for flow-field processing in the visual motion pathway of higher mammals. *Soc. Neurosci. Abstr.* **17**, 441.

Lappe, M., and Rauschecker, J. P. 1993. Computation of heading direction from optic flow in visual cortex. In *Advances in Neural Information Processing Systems*, Vol. 5, C. L. Giles, S. J. Hanson, and J. D. Cowan, eds. (in press). Morgan Kaufmann, San Mateo, CA.

Livingstone, M., and Hubel, D. 1988. Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science* **240**, 740–749.

Longuet-Higgins, H. C., and Prazdny, K. 1980. The interpretation of a moving retinal image. *Proc. R. Soc. London B* **208**, 385–397.

Mishkin, M., Ungerleider, L. G., and Macko, K. A. 1983. Object vision and spatial vision: Two cortical pathways. *Trends Neurosci.* **6**, 414–417.

Orban, G. A., Lagae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H., and Torre, V. 1992. First-order analysis of optical flow in monkey brain. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2595–2599.

Palmer, L. A., Rosenquist, A. C., and Tusa, R. J. 1978. The retinotopic organization of lateral suprasylvian visual areas in the cat. *J. Comp. Neurol.* **177**, 237–256.

Prazdny, K. 1980. Egomotion and relative depth map from optical flow. *Biol. Cybern.* **36**, 87–102.

Rauschecker, J. P., von Grünau, M. W., and Poulin, C. 1987a. Centrifugal organization of direction preferences in the cat's lateral suprasylvian visual cortex and its relation to flow field processing. *J. Neurosci.* **7**(4), 943–958.

Rauschecker, J. P., von Grünau, M. W., and Poulin, C. 1987b. Thalamocortical connections and their correlation with receptive field properties in the cat's lateral suprasylvian visual cortex. *Exp. Brain Res.* **67**, 100–112.

Regan, D., and Beverly, K. I. 1982. How do we avoid confounding the direction we are looking and the direction we are moving? *Science* **215**, 194–196.

Rieger, J. H., and Lawton, D. T. 1985. Processing differential image motion. *J. Opt. Soc. Am. A* **2**, 354–360.

Rieger, J. H., and Toet, L. 1985. Human visual navigation in the presence of 3-D rotations. *Biol. Cybern.* **52**, 377–381.

Stone, L. S., and Perrone, J. A. 1991. Human heading perception during combined translational and rotational self-motion. In *Soc. Neurosci. Abstr.* **17**, 857.

Tanaka, K., and Saito, H.-A. 1989a. Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62**(3), 626–641.

Tanaka, K., and Saito, H.-A. 1989b. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62**(3), 642–656.

Toyama, K., Fujii, K., and Umetani, K. 1990. Functional differentiation between the anterior and posterior Clare-Bishop cortex of the cat. *Exp. Brain Res.* **81**, 221–233.

Tsai, R. Y., and Huang, T. S. 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. Pattern Anal. Machine Intelligence* **6**, 13–27.

Ungerleider, L. G., and Mishkin, M. 1982. Two cortical visual systems. In *Analysis of Visual Behavior*, D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, eds., pp. 549–586. MIT Press, Cambridge, MA.

Wang, H. T., Mathur, B. P., and Koch, C. 1989. Computing optical flow in the primate visual system. *Neural Comp.* **1**(1), 92–103.

Warren, W. H., Jr., and Hannon, D. J. 1988. Direction of self-motion is perceived from optical flow. *Nature (London)* **336**, 162–163.

Warren, W. H., Jr., and Hannon, D. J. 1990. Eye movements and optical flow. *J. Opt. Soc. Am. A* **7**(1), 160–169.

Warren, W. H., Jr., Morris, M. W., and Kalish, M. 1988. Perception of translational heading from optical flow. *J. Exp. Psychol.: Human Percept. Perform.* **14**(4), 646–660.

Waxman, A. M., and Ullman, S. 1985. Surface structure and three-dimensional motion from image flow: A kinematic analysis. *Int. J. Robotics Res.* **4**, 72–94.

Yuille, A. L., and Grzywacz, N. M. 1988. A computational theory for the perception of coherent visual motion. *Nature (London)* **335**, 71–74.

Zeki, S., and Shipp, S. 1988. The functional logic of cortical connections. *Nature (London)* **335**, 311–317.

Received 7 January 1992; accepted 2 October 1992.