Neural Networks, Vol. 4, pp. 303–317, 1991
Printed in the USA.

0893-6080/91 $3.00 + .00

ORIGINAL CONTRIBUTION

Visual Navigation With a Neural Network

NICHOLAS G. HATSOPOULOS AND WILLIAM H. WARREN, JR.

Brown University

(*Received* 14 *December* 1989; *revised and accepted* 30 *October* 1990)

Abstract—*A simple linear neural network modelled on areas MT and MST of primate visual cortex can determine the direction of self-motion of an observer by using the optical flow field induced by observer translation relative to a rigid planar environment. The model's input layer consists of a set of motion detectors covering a* 20° × 20° *portion of the visual field with a subset of eight detectors selective to four primary directions and two speeds representing the optical motion within a single receptive field. Heading is represented distributively on the output layer in terms of azimuth and elevation. The network's heading accuracy under ideal conditions is on the order of* 1° *of visual angle, which is in agreement with perceptual studies of heading accuracy in human observers. The network's performance under noisy optical flow conditions matches remarkably well that of human subjects. Moreover, the network's tolerance of noise makes it potentially useful in robotic vision. A subsequent problem is to extend the model to combined observer translation and rotation.*

Keywords—Optical flow, Direction of heading, J. J. Gibson, Noise tolerance.

1. INTRODUCTION

Understanding the control of locomotion is a major problem in psychology, neuroscience, and robotics. In this article, we focus on the information contained in the changing structure of light at an eye due to the movement of an organism relative to a rigid environment, known as optical flow. It is of great practical and theoretical interest whether an organism can rely on optical flow to determine its direction of self-motion, or heading. We present here a simple linear network, based on the known architecture of areas MT and MST in primate visual cortex, which models the low-level brain process by which an organism could determine its heading as it translates in a planar environment. We compare the performance of the model with that of human observers and present empirical results obtained under noisy optical flow conditions that test predictions made by the model.

J. J. Gibson (1950, 1954, 1968, 1979) argued that visual perception involves the active pick-up of information specific to the environment from the optic array, the changing ambient light surrounding the observer. The relevant information manifests itself in the form of invariant properties embedded in the varying pattern of ambient light. The important issue is, of course, whether such relevant invariants can be found in the changing optic array and, if so, whether the organism detects them. Gibson argued that there are two kinds of optical information: exterospecific information about the structure of the environment and propriospecific information about the organism's movement in the environment, such as the direction of heading and the time to contact with objects. Gibson was the first to point out the invariant radial pattern of optical flow emerging from translatory motion of an observer in a rigid environment (Gibson, 1947) (Figure 1).

If the temporally changing optic array is represented as a vector field of optical velocities, the vectors form a radial pattern emanating from a center point, called the focus of outflow or focus of expansion. This global radial pattern is a natural consequence of optical perspective and is invariant with respect to the 3D structure of the environment. Gibson proposed that this pattern is detected by the observer and used to control locomotion. Steering could be accomplished by placing the focus of radial outflow at the location in the visual field toward which the organism wishes to head.

Although the concept of a temporally changing optical array is rather abstract and makes no assumptions as to its specific mathematical representation, Gibson and others have represented it as a velocity vector field. The radial outflow invariant

We would like to thank James A. Anderson and Margaret E. Sereno for advice on the network simulations. Margaret Sereno's model inspired us to develop our own in the context of visual navigation. We would also like to thank Paul A. Kairis for his assistance in the graphics displays of the weight matrices.

Reprint requests should be sent to Nicholas G. Hatsopoulos, Department of Psychology, Brown University, Providence, RI 02912.

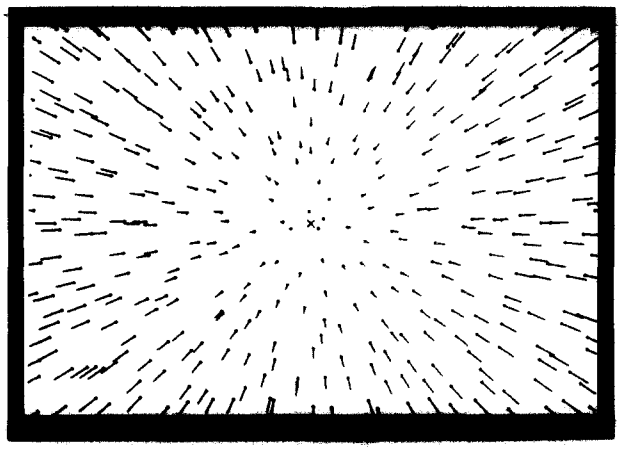


FIGURE 1. Radial velocity field induced by observer translation toward a rigid plane.

emerges under this particular mathematical description. However, the velocity field is a relatively high-level description which is not as trivial for the visual system to determine locally as it appears, given the problem of determining velocity from a changing luminance distribution (Horn & Schunck, 1981) and the so-called "aperture problem" (Hildreth, 1984a, b). The aperture problem refers to the indeterminacy of the true velocity of a luminance edge using a local motion detector. That is, if the diameter of the motion detector's receptive field is small compared to the length of the luminance edge, then it is only possible to pick up the velocity component perpendicular to the orientation of the edge. It is clear that velocity information from many local motion detectors must be pooled in order to determine the true velocity. There is considerable neurophysiological evidence that true velocity information is detected by the visual system (Maunsell & Van Essen, 1983a, b; Mikani, Newsome, & Wurtz, 1986; Saito et al., 1986; Rodman & Albright, 1987). We will assume that the visual system can represent the velocity field as a pattern of activity distributed over higher-level velocity-sensitive units.

In this article we are concerned with pure observer translation, that is, motion in a straight line. General rigid motion of an observer, however, consists of translation plus rotation about the three principal axes of the coordinate system centered at the nodal point of the eye. Rotations can be introduced through eye, head, and body movements. Unfortunately, the radial flow pattern on the retina is disrupted by observer rotation. The model presented here is limited to the case of pure translation, but we will return to consider extensions of the model to the general case in the discussion.

2. EXPERIMENTAL RESEARCH

The radial outflow hypothesis proposed by Gibson has been around for over 40 years. Until recently, there has been little empirical research to determine whether observers actually use this information. Cutting (1986) calculated that a heading accuracy of about 1° of visual angle is necessary for navigation during running and driving. However, research on heading judgments during translation both perpendicular (Llewellyn, 1971; Johnston, White, & Cumming, 1973) and parallel (Warren, 1976) to a random-dot plane found poor heading accuracies of 5° to 10°.

Riemersma (1981), Regan and Beverley (1982, 1984), and Cutting (1986) cited these disappointing results and concluded that observers do not use the radial outflow pattern. In its stead, they proposed their own hypotheses and attempted to support them experimentally. Riemersma suggested various visual cues (e.g., changing edgelines) induced by changing heading angle and changing lateral position under straight-road driving. Regan and Beverley discussed the maximum rate of change of magnification as a source of information because it is invariant to eye rotations. Cutting stressed depth differences and attempted to support a differential motion parallax hypothesis. Unfortunately, all these proposals are not true invariants of the changing ambient optic array. Riemersma's hypothesis only applies to particular environments, such as roads with well-defined edges. Regan and Beverley's maximum rate of magnification assumes a smoothness constraint and corresponds to heading direction only during perpendicular approach to a plane, and Cutting's theory requires significant depth differences and successive fixations.

However, recent research has revitalized the radial outflow hypothesis for observer translation. Using an improved methodology, Warren, Morris, and Kalish (1988) found heading accuracies on the order of 1° with translation parallel and perpendicular to a random-dot plane. Accuracy remained high over a range of observer speeds, dot densities, and angles of approach to the plane, except with very low densities (2 dots). Similar results were obtained with translation through a 3D cloud of dots, which violates surface smoothness constraints. These results supporting the radial outflow hypothesis motivated the development of the neural model presented here.

3. COMPUTATIONAL MODELS

Most, if not all, of the computational models that have appeared in the literature have taken an engineering or computer science approach to the problem of extracting propriospecific information without regard to biological or psychological plausibility. The problem is usually posed as one involving the derivation of the six motion parameters characterizing general motion of a rigid body in three dimensions: three translational parameters (V_x, V_y, V_z) and three rotational parameters (W_x, W_y, W_z). Actually, the

relative translational velocities $V_x : V_y : V_z$ are computed because it is impossible in theory to derive the magnitudes of all three translational components from the optical flow; only the ratio of the observer speed to surface distance (time-to-contact) is uniquely specified. So, in effect, only five motion parameters can be derived, two specifying the direction of translation and three specifying the rotation. The computational models can be broadly categorized into two classes according to the input representation used to compute these parameters: intensity-based and velocity-based representations.

Bruss and Horn (1983) define a useful taxonomy of algorithms that use a velocity-based representation: discrete, differential, and continuous. Discrete algorithms rely on the precise velocities of a small sample of image elements. Emphasis is often placed on determining the minimum number of image points required to derive the observer's motion parameters. Prazdny (1981) determined the five parameters using a minimum of five image-point velocities by solving a set of three nonlinear equations. Tests of this technique yielded heading errors of less than 0.5°. Despite Prazdny's positive results, discrete models are incapacitated by even small amounts of noise in the optical flow field. The reason is that there is no redundant information which can be used to overcome internal and external noise.

Differential models depend on spatial derivatives of the velocity field to determine the five motion parameters. Longuet-Higgins and Prazdny (1980) used the first and second spatial derivatives at a single image element corresponding to a point on a smooth surface to solve for the motion parameters. Waxman and Ullman (1985) performed a differential analysis on a single image point and computed quantities inspired from fluid mechanics. The motion parameters were found by solving a set of 12, coupled, nonlinear equations. In the case of a planar environment there are generally two solutions.

Koenderink and van Dorn (1975, 1976, 1981) took a geometric approach rather than the algebraic approach described above. They discuss two first-order differential operators that can be applied to the optical velocity field. The first is called the divergence ($\nabla \cdot v$ where v is the velocity field). The divergence of a vector field produces a scalar field that measures the net vector flow outward from an infinitesimally small region at each location in the field. The second is called curl ($\nabla \times v$) and produces a vector field representing the magnitude and direction of rotation within an infinitesimally small region at each location in the field. They prove that the location of maximum divergence falls on a line bisecting the angle between the direction of heading and the normal to the surface. This raises doubts about Regan and Beverley's (1984) rate of change in magnification hypothesis, which is mathematically equivalent to divergence, because the divergence maximum does not generally lie in the direction of heading. However, Koenderink and van Doorn show that the two points of singularity corresponding to the maximum and minimum of divergence taken together do determine the axis of heading in a planar environment. Importantly, they are invariant under rotation, which is not true for the retinal focus of outflow. However, the differential approach, in general, is problematic for three reasons. First, it requires relatively smooth environmental surfaces in order that the spatial derivatives be computed. Second, the models are again highly vulnerable to noise, and differentiation tends to magnify noise already existing in the flow field. Third, there is a question as to how biologically plausible such differentiation is (Nakayama, 1985), and recent research indicates that human observers do not rely on differential invariants (Warren et al., 1988; Warren & Hannon, 1988, 1990; Lappin & Norman, 1988; Gillam, 1988).

The continuous approach takes advantage of the redundancy in the global flow field by sampling from many points in the image. Thus, it is relatively insensitive to noise in the flow field. Most of these models use a least-squares technique in which the space of possible flow fields is searched to minimize the difference between the observed and possible fields. Prazdny (1981), Bruss and Horn (1983), Lawton (1983), and Maybank (1986) use such a technique to solve for direction of heading under various conditions of observer motion. Rodford (1986) and Ballard and Kimball (1983) use a Hough transform of the flow field after which a least-squares technique is performed. In the case of pure translation, the transformed field best matches a sinusoid whose phase shift represents the focus of expansion. These models require redundant measures from a very large region of the flow field to produce a unique solution and often make assumptions about the structure of the environment. A major problem with these models is that they appear to be neither biologically nor psychologically plausible. Techniques such as Taylor series expansions (Maybank, 1986), and Hough transforms are engineering tools but shed little light on the problem of understanding control of locomotion in living organisms.

4. DERIVATION OF OPTICAL FLOW

Most mathematical analyses of optic flow assume a projection surface or model retina, although its exact shape is immaterial (Longuet-Higgins & Prazdny, 1980). A more appropriate optic array description represents optical flow in terms of visual angle (θ, ϕ) in spherical coordinates, making no assumptions about a projection surface. Our derivation of the optical flow produced by movement relative to an

environmental element is based on that of Rieger (1983) and begins with a Cartesian frame of reference with the observer at the origin (Figure 2). The observer can, in general, translate (V_x, V_y, V_z) along the axes and rotate (W_x, W_y, W_z) around those axes. A point **P** in the rigid environment can be specified in Cartesian coordinates as (x, y, z) and in spherical coordinates as (θ, ϕ, ρ) where θ is the elevation, ϕ is the azimuth, ρ is the radius. The point **P** moves relative to the observer's frame of reference as $\dot{x}\hat{\imath} + \dot{y}\hat{\jmath} + \dot{z}\hat{k}$ or $\dot{\theta}\hat{\theta} + \dot{\phi}\hat{\phi} + \dot{\rho}\hat{\rho}$ where $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$ and $\hat{\theta}$, $\hat{\phi}$, $\hat{\rho}$ are the basis vectors of their respective coordinate systems. From the point of view of the observer, the motion of point **P** in the $\hat{\rho}$ direction is not visible. Thus, the optical flow due to motion of point **P** becomes:

$$v_p = \dot{\theta}\hat{\theta} + \dot{\phi}\hat{\phi}. \tag{1}$$

Using simple coordinate transformations, this optical velocity can be expressed in terms of the Cartesian velocities of point **P** as:

$$v_p = (1/\rho)(\dot{x}\cos\theta\cos\phi + \dot{y}\cos\theta\sin\phi - \dot{z}\sin\theta)\hat{\theta} + (1/(\rho\sin\theta))(\dot{y}\cos\phi - \dot{x}\sin\phi)\hat{\phi}. \tag{2}$$

The Cartesian velocity of point **P** is related to the movement of the observer according to this vector equation.

$$\dot{x}\hat{\imath} + \dot{y}\hat{\jmath} + \dot{z}\hat{k} = V - \Omega \times r, \tag{3}$$

where $V \equiv (V_x, V_y, V_z)$, $\Omega \equiv (W_x, W_y, W_z)$, and $r \equiv \rho\hat{\rho}$. Thus, the optical flow of a point **P** can be expressed in terms of the translational and rotational movement of the observer:

$$v_p = [1/\rho(-V_x\cos\theta\cos\phi - V_y\cos\theta\sin\phi + V_z\sin\theta) + (W_x\sin\phi - W_y\cos\phi)]\hat{\theta} + [(1/(\rho\sin\theta)) \times (V_x\sin\phi - V_y\cos\phi) + \cot\theta(W_x\cos\phi + W_y\sin\phi) - W_z)]\hat{\phi}. \tag{4}$$

The variable ρ can be treated as a function of θ and ϕ and specifies the environmental structure. We are modelling a planar environment whose orientation is arbitrarily set parallel to the x–z plane. This constrains the function ρ as follows: $\rho\sin\theta\sin\phi = k$ where k is the distance of the observer to the plane. Finally, the components of observer translation (V_x, V_y, V_z) can be expressed in terms of elevation, α, and azimuth, β:

$$V_x = \|V\|\sin\alpha\cos\beta, \tag{5}$$

$$V_y = \|V\|\sin\alpha\sin\beta, \tag{6}$$

$$V_z = \|V\|\cos\alpha, \tag{7}$$

where $\|V\|$ is the speed of translation.

The flow equation becomes:

$$v_p = [((\sin\theta\sin\phi)/k)(-\|V\|\sin\alpha\cos\beta\cos\theta\cos\phi - \|V\|\sin\alpha\sin\beta\cos\theta\sin\phi + \|V\|\cos\alpha\sin\theta) + W_x\sin\phi - W_y\cos\phi]\hat{\theta} + [(\sin\phi/k) \times (\|V\|\sin\alpha\cos\beta\sin\phi - \|V\|\sin\alpha\sin\beta\cos\phi) + \cot\theta(W_x\cos\phi + W_y\sin\phi) - W_z]\hat{\phi} \tag{8}$$

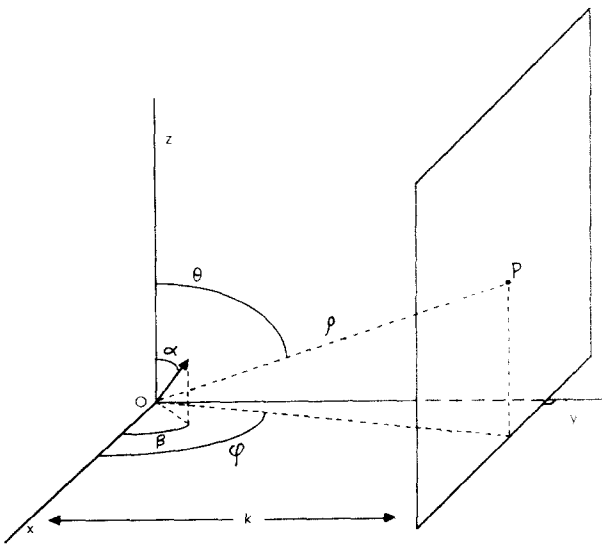


FIGURE 2. Cartesian frame of reference (x, y, z) moving with the observer who is at the origin O. The observer's direction of heading can be specified by two angles: an elevation, α, and an azimuth angle, β. The point P represents a point on the planar environment. k denotes the perpendicular distance of the observer to the plane and ρ denotes the distance of the observer to the point P.

5. JUSTIFICATION FOR A LINEAR TWO-LAYER NETWORK

The above mathematical analysis takes as given the observer's movement parameters and derives the flow produced by a point in the environment. The problem to be solved by the neural network is the inverse problem: given the flow at certain points in the ambient optic array, find the movement parameters of the observer. There are three reasons for choosing a two-layer linear neural network to solve this inverse problem. In the case of pure translation, the problem involves solving a set of linear equations. A minimum of two image points is required to solve for the direction of heading. This can be shown both geometrically and algebraically. According to the radial outflow hypothesis, the problem requires finding the focus of the radial optical flow pattern. Geometrically, the problem is solved by finding the point of intersection of any two velocity vectors in the field. Likewise, the algebraic solution requires the optical flow at two image points. Algebraically, the problem can be posed as follows:

$$v_p = aM\hat{V}, \tag{9}$$

where a is a scalar representing the ratio from which time to contact can be computed, M is the nonsquare matrix of trigonometric quantities:

$$\begin{pmatrix} -\cos\theta\cos\phi & -\cos\theta\sin\phi & \sin\phi \\ \sin\phi & -\cos\phi & 0 \end{pmatrix}$$

$\hat{V}$ is a unit vector pointing in the direction of translation, and v_p is the optical flow of point **P**. The problem can be solved by taking the inverse of M:

$$\hat{V} = M^{-1}v_p/a \tag{10}$$

However, M is not square and, therefore, does not have a unique inverse. A second image point must be sampled to solve for these three unknowns. It should be emphasized that the minimum solution requires only two velocity vectors. Our model, on the other hand, relies on redundant information from many velocity vectors.

A second reason that we chose a two-layer linear network architecture is that the problem of finding one's heading direction does not require high-order statistical relationships between flow vectors in different parts of the field, such as correlations. It is well known that a linear matrix model fails when feature correlations map to responses different than those from features treated independently. The optic flow at different points in the visual field can be treated independently, which makes the problem suitable for a simple matrix model.

The final reason is based on neurophysiological evidence in extra-striate cortex of certain primates. This evidence justifies the use of a two-layer network but does not really address the issue of linearity. Single unit recordings conducted in the middle temporal region (MT) of the superior temporal sulcus of the rhesus, owl, and macaque monkeys (Maunsell & van Essen, 1983a, 1983b; Albright et al., 1984; Mikami et al., 1986; Rodman & Albright, 1987) show that MT cells are primarily selective to visual motion and relatively insensitive to form and color. In particular, they are selective to both direction and speed. Cells in striate cortex have also been found which are responsive to motion (Schiller, Finlay, & Volman, 1976), but two properties of MT make it unique in regard to motion analysis. First, MT cells have larger receptive fields, ranging from about 4° × 4° to 10° × 10°. This implies spatial summation from motion-sensitive cells in striate cortex because it is known that major afferents to MT come from primary visual cortex. Second, MT cells seem to respond to two-dimensional pattern motion, suggesting an integration of multiple, one-dimensional image velocity detectors (Movshon et al., 1985).

MT cells project primarily to another region in the superior temporal sulcus called the medial superior temporal area (MST) (Saito et al., 1986; Weller et al., 1984). MST cells have even larger receptive fields (some as large as 40° × 40°) which in turn suggests spatial summation from MT. Saito and his colleagues categorized cells in MST into three classes according to their functional properties: D, R, and S type cells. D cells (50% of cells tested) responded to motion in a particular direction much as MT cells do. R cells (about 14% of cells tested) were sensitive to overall rotational motion in various orientations including rotation in depth. S cells (about 16% of cells tested) responded to expansion and contraction. Some S cells were selective to the radial expansion and contraction of a small object like a circle or square (figure-type S cell) while others responded best to a large global expansion or contraction of the visual scene (field-type S cells). It is possible that these field-type S cells are used to pick up radial optic flow patterns induced by observer translation in the environment. Interestingly, there were three times as many S cells found selective to expansion as there were those selective to contraction. Moreover, these S cells are very broadly tuned to the speed of radial motion, which is to be expected if they are responding to heading direction. It is this projection from MT to MST which inspired the architecture of the neural network.

6. ARCHITECTURE OF THE NEURAL NETWORK

The basic two-layer feedforward architecture of the network was constant throughout the research described in this article (Anderson, 1983) (Figure 3) and incorporated features of recent work by Sereno

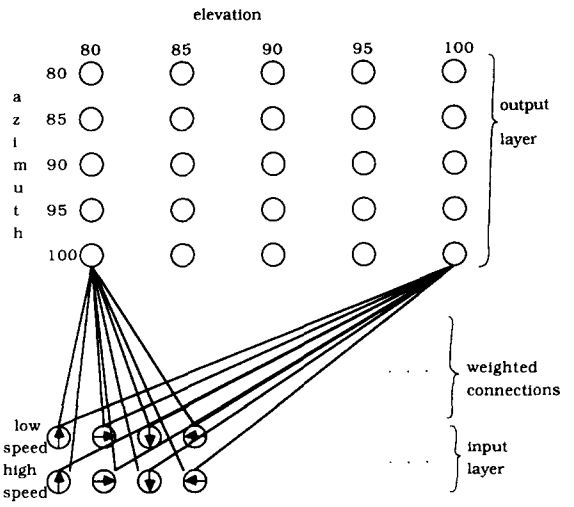


FIGURE 3. Diagram of the linear two-layer feedforward network used in the simulations. The eight input cells for one location in the visual field and their connections to two output cells are indicated. Each output cell is fully connected to all input cells.

(1987). A set of neuron-like input nodes or cells representing the velocity field of a sample of the ambient optic array fed signals through a fully connected matrix of weighted connections to a layer of output cells representing the direction of heading. The tuning characteristics of the input cells, the input sampling size, and output representation evolved quite a bit in the course of the research, and we will focus here on the final structure and properties of the network.

As stated above, the input layer of nodes are modelled after the cells in the middle temporal area of the superior temporal sulcus. The tuning characteristics of MT cells have been extensively examined, making the design of the input layer relatively easy. Rodman and Albright (1987) discuss three tuning characteristics of MT cells which are particularly relevant to the design of the input layers: the shape of the speed and direction tuning curves, tuning bandwidth, and the interaction between speed and direction tuning. For simplicity, we used a triangular tuning curve for both speed and direction coding in the input layer instead of the Gaussian-like curves found by Rodman and Albright (Figure 4). They found direction tuning bandwidths ranging from 25° to 151° with a mean of 88° (bandwidth is measured as the full width between half-maximum response). Direction bandwidths in the neural model were set to 60°. Speed tuning bandwidths ranged between 1.35 and 4.60 octaves where an octave is defined as one log unit. In the later simulations, we used logrithmic speed tuning with a bandwidth of four octaves. The most interesting finding reported by Rodman and Albright is that no interaction was discovered between speed and direction tuning. That is, the sharpness of the direction tuning was not affected by different speeds and vice versa. We implemented a product rule (cell firing rate = speed activation × direction activation) in the network which resulted in no interaction. The single cell recordings reported by Rodman and Albright are consistent with data by others with results of the same order of magnitude (Maunsell & van Essen, 1983a; Mikani et al., 1986). For computational reasons, we used cells with only four different primary direction selectivities and two primary speed selectivities for most of the simulations, making a total of eight input cells per visual field location. Also, we limited the visual field in our simulations to 20° × 20° consisting of 4° × 4° nonoverlapping input cell receptive fields. Therefore, the input layer consisted of 200 cells, 8 cells per receptive field region times 25 receptive field regions.

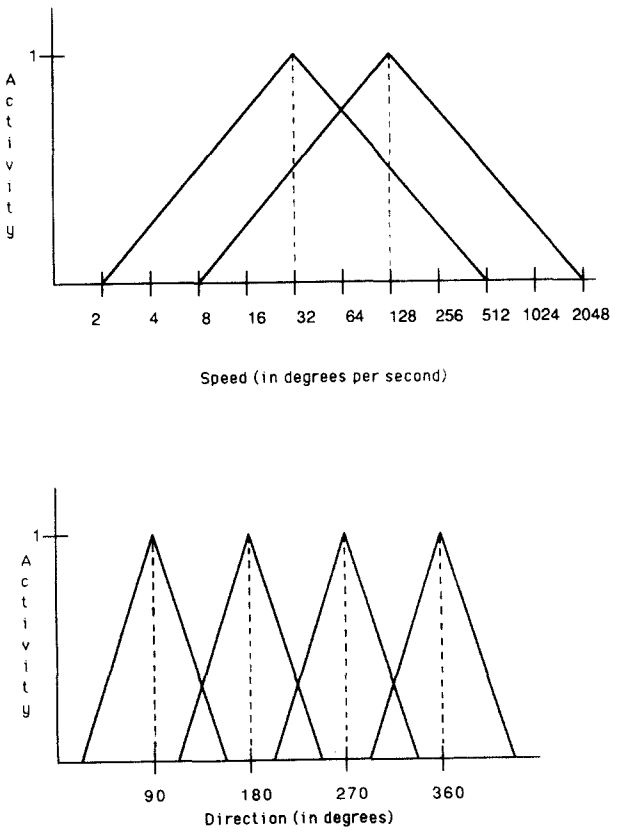


FIGURE 4. Speed and direction triangular tuning curves used for the input cells.

The output layer of nodes correspond to MST cells and represent direction of heading. We decided on a topographic coding of heading in which azimuth and elevation are coded conjunctively. After initial failures with a local coding scheme, we switched to a distributed coding of heading with overlapping receptive fields, which has some physiological support (Saito et al., 1986). To be consistent with the input layer, we used 25 cells to form a 5 × 5 grid representing a 20° × 20° portion of the visual field. Each cell responded maximally to a particular heading direction. Response dropped off linearly as a function of Euclidian distance between the heading of maximal response and the heading direction to be coded:

$$cr_i = 1 - (1/r)((\alpha_{m_i} - \alpha)^2 + (\beta_{m_i} - \beta)^2), \quad (1)$$

where cr_i is the coding response of cell i, β_{m_i} is the azimuth and α_{m_i} is the elevation of the heading of maximal response, β is the azimuth and α is the elevation of the direction of heading to be coded, and r is the radius of response. For most of the simulations, the maximum radius of response for each cell was 10°.

The above equation produced the output layer coding necessary for the training phase of the simulations. However, the inverse mapping from distributed coding of heading to azimuth and elevation necessary during the testing phase required an interpretive rule. The azimuth and elevation were calculated by taking the average of the azimuths and elevations of maximal response for all cells weighted

by each cell's level of activity:

$$\alpha = \sum_{i=1}^{25} \alpha_{m_i} a_i \tag{12}$$

$$\beta = \sum_{i=1}^{25} \beta_{m_i} a_i \tag{13}$$

where β_{m_i} and α_{m_i} are the azimuth and elevation of maximal response for cell i and a_i is the activity of cell i. It should be stressed that this nonlocal interpretive rule is only necessary as an aid to us, the investigators, because a real number representation of heading is most convenient. In an actual nervous system, a distributed representation is more tolerant to internal noise and may be more natural for the visual cortex and subsequent processes.

7. DYNAMICS OF THE NETWORK

7.1. Activation Rule

The activation rule of each output cell determined the activity of the cell due to the activities of the input cells. It was a simple linear summation function of the input cell activities weighted by the connections feeding the input layer to the output cell in question:

$$a_i = \sum_{j=1}^{200} x_j w_{ij} \tag{14}$$

where x_j is the activity of input cell j, a_i is the activity of output cell i, and w_{ij} is the weight of the connection between them.

7.2. Learning Rule

We implemented the standard Widrow–Hoff error correction learning rule (Widrow & Hoff, 1960):

$$\Delta w_{ij} = \eta x_j (cr_i - a_i), \tag{15}$$

where Δw_{ij} is the incremental change in the weight connecting input cell i and output cell j. η is a learning rate parameter. x_i and a_j are the activities of input cell j and output cell i. cr_i is the coding response of cell i which acts as a teaching signal.

We should stress that we are not attempting to model the course of synaptic modification that takes place during visual learning or development in biological systems. There is some evidence that environmental input can affect the tuning characteristics of motion-sensitive cells (Kennedy & Orban, 1983). However, we are using a supervised learning rule simply to establish correct associations between flow fields and their corresponding headings, so that we can model the known architecture of the motion system and compare the pattern of weights resulting in the network with the organization of connections between MT and MST.

8. SIMULATIONS

The simulations performed with the neural model consisted of two phases: learning and testing. The learning phase involved associating a training set of optical flow patterns represented on the input layer with their correct heading direction coded on the output layer. The distributed activities on the input and output layers can be treated as vectors in a multidimensional state space. Learning performance was measured in two distinct ways. During the learning process, the absolute value of the error signal used to modify the weights was averaged over the training set for each learning cycle: $\langle |cr_i - a_i| \rangle$. This measure allowed us to observe progress between cycles and to stop the learning phase when the measure appeared to asymptote.

Performance was also assessed during the test phase, which consisted of presenting optical flow patterns to the input layer and comparing output layer activity with that corresponding to the correct heading. The cosine of the angle between the actual output state vector and the state vector corresponding to the correct heading was calculated. In addition, heading accuracy was measured in terms of the average heading error ($\varepsilon_{heading}$) between the correct azimuth and elevation and the azimuth and elevation found using the interpretation algorithm:

$$\varepsilon_{heading} = 1/2(|\alpha_c - \alpha_a| + |\beta_c - \beta_a|) \tag{16}$$

where the subscript c refers to the correct angle and a, the algorithm.

A sum-squared deviation measure of heading error could have also been used. In fact, any monotonic function of deviation would be appropriate. The network was tested with both the original optical flow patterns and novel patterns with which it had not been trained. The model should be able to generalize its performance to new heading directions, new random dot configurations, and new time-to-contact values (yielding different velocity magnitudes).

The network's sensitivity to noise was also examined. This is an important issue because flow fields in the real world are often perturbed by noise from the environment (blowing snow, leaves, trees, etc.) and from the receptor system, and previous discrete and differential models are highly vulnerable to noise (Bruss & Horn, 1983). The network should be relatively insensitive to noise, and, more importantly, its performance under noisy conditions should be qualitatively similar to that of human subjects.

Twenty-five simulations were performed in all, each defined with respect to the stimulus set used to train the network. We present the results of three simu-

lations which are particularly informative, summarized in Table 1. For all three simulations, each input stimulus represented the flow field produced by the relative motion between the observer and a random dot plane composed of 50 random dots to make it highly probable that each receptive field region received a single random dot. Also, the ratio of distance to observer speed (time-to contact) was randomly chosen from a predetermined interval in order to cover the speed tuning range of the input cells. The training stimulus set was cycled randomly through the network 10 times with either a constant learning rate equal to 1 over the number of training stimuli or an exponentially decaying rate with a starting value of 0.05 and a decay coefficient of .25. We will also briefly discuss results of simulations in which various relevant parameters were varied to find out how these parameters affected the network's performance.

8.1. Simulation 1

The training set consisted of 400 flow fields corresponding to 400 heading directions sampled from a uniform distribution. We restricted the range of heading directions so that the focus of expansion would fall within the 20° by 20° visual field. The network was first tested on the training set after learning had stopped. Average heading error (averaged over elevation and azimuth) was about 0.83°. It is crucial that the network also generalize to flow fields produced by new random dot planes and new contact times. The network's accuracy on 100 new flow fields corresponding to 100 randomly chosen headings was nearly identical, about 0.89°.[1]

The network was also tested on optical flow patterns representing lower random dot densities. As stated above, a mathematical solution to the problem of finding one's heading direction given the optical flow induced by pure observer translation requires the visual motion of only two image points. The optical flow of more than two image points simply introduces redundant information. By testing the network on lower dot density fields, it could be determined what role redundancy played in the network's solution to the problem. Test stimuli containing 5, 10, 15, 20, 30, 40, and 50 dots were presented to the network and average heading error is plotted as a function of number of dots in Figure 5. Comparable human data from Warren et al. (1988) are plotted on the same graph, representing the heading threshold at which observers were 75% correct in judging whether they were heading to the left or right of a visible target, during translation parallel to a ground plane. The flattening of the simulation curve beyond 25 dots was a floor effect because there are 25 receptive field regions represented on the input layer, a maximum of 25 random dots can be registered by the network (one dot per region). If more

TABLE 1
Summary of Results of the Three Simulations

	No. of Stimuli	Average Heading Error
Simulation 1		
Learning set: 400 random headings	400	0.83°
Test set: 100 random headings	100	0.89°
Simulation 2		
Learning set: 25 random headings	400	0.81°
Test set: 100 random headings	100	0.86°
Simulation 3		
Learning set: 25 evenly spaced headings	400	1.16°
Test set 1: 100 random headings	100	0.77°
Test set 2: 16° evenly spaced headings, 2.5 out of phase with learning set	16	0.49°
Test set 3: 50 random headings with added uniform speed noise with a range of 16° per second	50	0.70°
with a range of 64° per second	50	0.73°
Test set 4: 50 random headings and all vector magnitudes set to 32° per second	50	0.98°
to 128° per second	50	0.92°

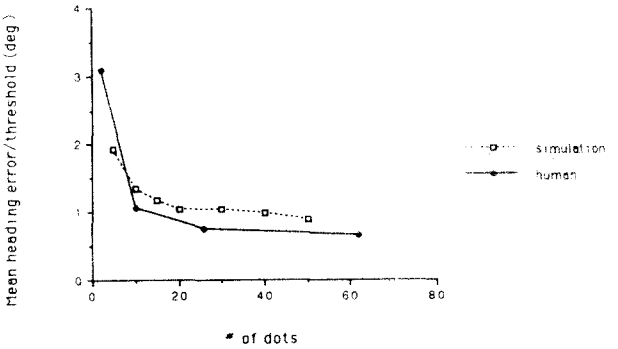


FIGURE 5. Mean heading error as a function of number of random dots in Simulation 1. Also, mean heading threshold for human subjects.

[1] These new flow fields may correspond to novel heading directions although the chances are very slim that many of them were not contained in the training set. On the other hand, these new flow fields are the result of newly generated random dot planes and new contact times.

than one dot happens to fall in the same receptive field, only the most recently generated dot is kept. The monotonic decline in heading accuracy with decreasing dot density (under 25 dots) is similar to the data from human subjects. We will show that they both may be using the redundant information provided by a larger number of random dots in a similar way.

8.2. Simulations 2 and 3

These two simulations were performed to determine how well the network could generalize to optical flow fields for new heading directions as well as new random dot planes. In Simulation 2, the training set consisted of 400 flow fields corresponding to only 25 unique random headings. As expected, average heading accuracy on the learning set was quite good (0.81°). However, it was surprising how well the network generalized to flow fields for 100 randomly chosen heading directions (0.86°). This is a case of good generalization because the 100 randomly chosen heading directions could include any of 400 unique heading directions whereas the training set consisted of only 25 unique headings.

In Simulation 3, the 400 training optical flow fields corresponded to 25 heading directions which were evenly spaced throughout the 20° × 20° visual field. The results of this simulation were quite remarkable. The network performed better on 100 novel flow fields corresponding to 100 randomly selected heading directions (0.77°) than it did on the training set (1.16°), and more than twice as accurately on optical flow fields for heading directions that were 2.5° out of phase with respect to those in the training set (0.49°). One possible reason is that we only cycled through the training stimuli 10 times. We chose 10 cycles as a standard because the average errors for most of our simulations asymptoted at 10 cycles. However, average error could drop substantially if we had included more training cycles in Simulation 3. On the other hand, we are most interested in performance to novel stimuli. This result suggests that it may not always be necessary to excessively train the network.

Tests of noise sensitivity were conducted with Simulation 3. Two kinds of additive, independent noise were introduced: speed and directional. Additive noise was introduced into the optical flow field by taking each velocity vector independently and perturbing either its magnitude (speed) or its direction according to a probability distribution with a mean of zero and a specified standard deviation. Speed noise was introduced using a uniform sampling distribution with standard deviations of 4.62°/s and 18.48°/s, producing ranges of 16° and 64°/s. This created a pattern of vector magnitudes similar to that produced by translation through a 3D cloud of random dots rather than a plane. Test stimuli were novel flow fields corresponding to 50 random heading directions with 50 trials in each noise condition. Heading accuracies were almost identical to the previous test with proper vectors (0.77°), yielding mean errors of 0.70° and 0.73°, respectively.

We also introduced another form of speed noise in which the magnitudes of all flow vectors were set to the same arbitrary value, 32°/s for one test and 128°/s for a second test. These values were chosen because they correspond to the primary speed selectivities of the input cells. This is a major distortion of the field, akin to heading toward a convex surface, yet heading accuracies remained high, 0.98° and 0.92°, respectively.

These findings are very similar to those of Warren et al. (1990), who found that human observers perform as well with flow fields whose magnitudes are randomly scrambled (heading thresholds of 0.9°) as they do with proper vector fields (0.7°). Their displays contained only velocity field information, as each dot existed for only two frames and was replaced at a random position elsewhere in the field. These findings are also consistent with neurophysiological results from area MST (Tanaka, Fukada, & Saito, 1989). Thus, both the network and the visual system appear to have abstracted the fact that it is the radial pattern of vector directions that provides information for heading, generalizing over variations in vector magnitudes.

To examine the network's resistance to noise in this radial pattern, additive directional noise was introduced by randomly perturbing the direction of each vector from its proper orientation within a specified envelope (see Figure 6).

A uniform sampling distribution was used to test the network. Standard deviations of 13°, 26°, 39°, and 52° corresponding to uniform distributional ranges of 45°, 90°, 135°, and, 180° were used. A monotonic increase in heading error was observed as a function of directional noise (Figure 7). Nevertheless, heading accuracy was quite good even with a 180° range of uniform noise, yielding a heading error of only 3°. The network's tolerance of noise again

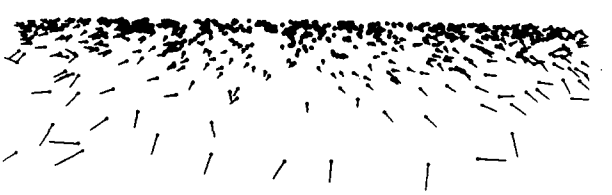


FIGURE 6. An optical flow field perturbed with uniformly distributed, additive, directional noise. The flow field is the result of observer translation over parallel to a ground plane of random dots.

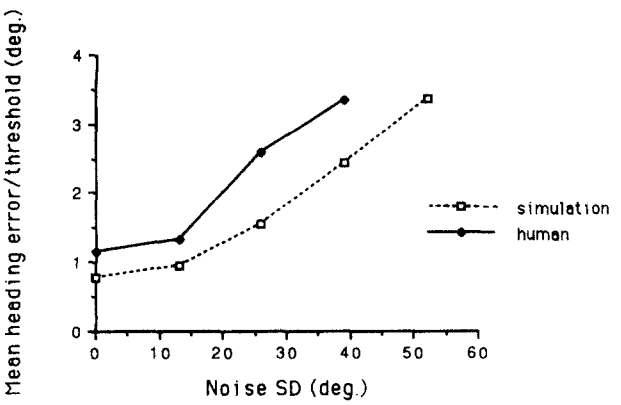


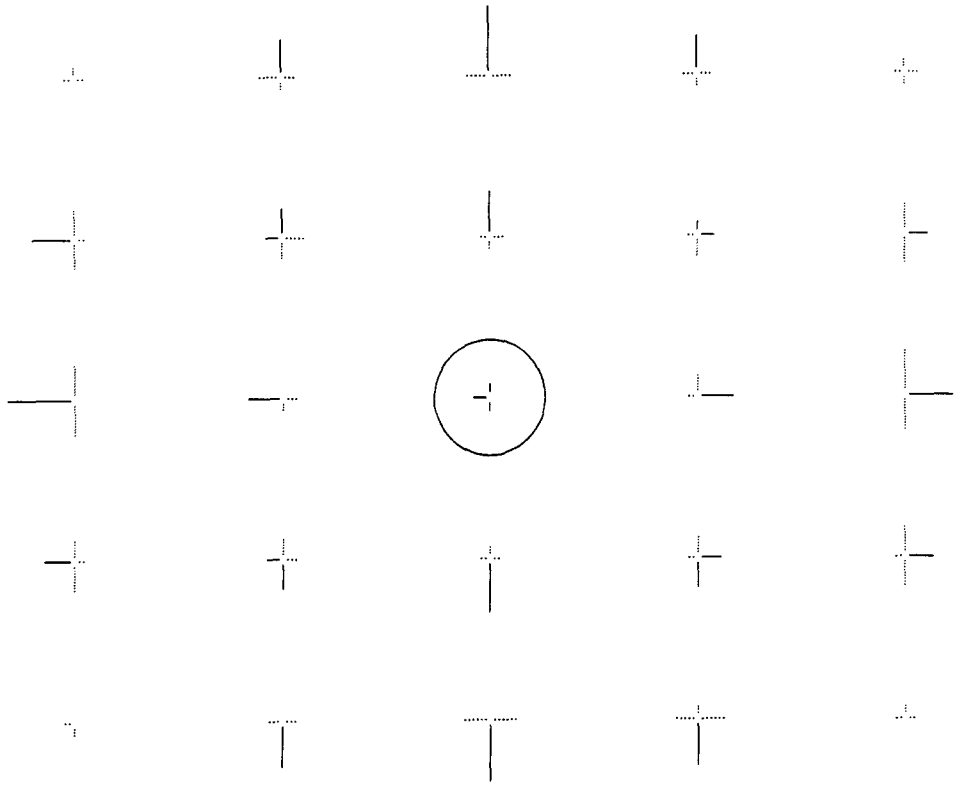
FIGURE 7. Mean heading error for uniform directional noise as a function of standard deviation of noise distribution in Simulation 3. Also heading threshold for human subjects as a function of uniform directional noise.

suggests that the redundancy in the global flow field is being used to solve the problem. These results provide a prediction from this simple model which can be compared with the performance of human subjects.

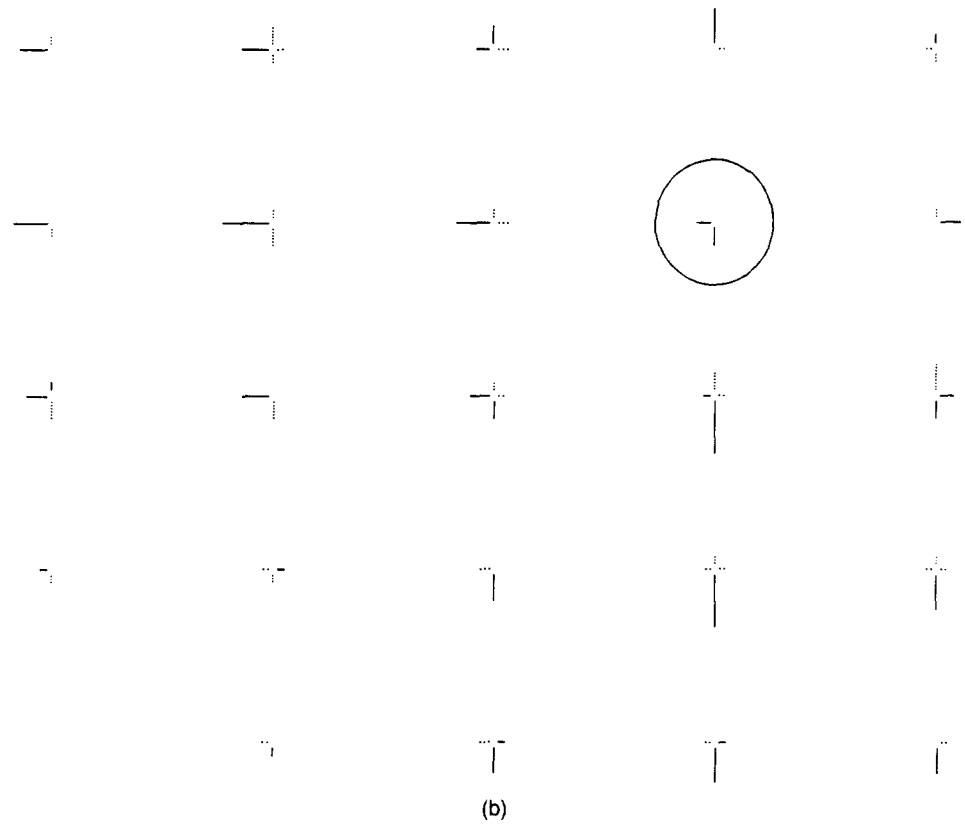
An experiment was conducted to test this prediction, reported in Warren et al. (1990). Human subjects were presented with velocity field displays with a two-frame dot life on a CRT screen, simulating observer translation parallel to a random dot ground plane. Directional noise was introduced by perturbing the direction of displacement of each dot between frames according to a uniform distribution. The standard deviation of the uniform noise could take on one of four values: 0°, 13°, 26°, 39°, corresponding to distributional ranges: 0°, 45°, 90°, and 135°. (The zero degree range corresponds to a no-noise control condition.) Thus, the three noise conditions in the experiment matched the first three in the network simulations. Because the human displays simulated movement parallel to a plane surface, a perfect stimulus match did not exist between the experiment and the simulations: Only half of the radial flow pattern was visible and only azimuth was varied in the experiment. Nevertheless, a qualitative comparison between the two is possible, and the results for human observers also appear in Figure 7. Despite a small absolute difference, the curves are nearly parallel. Thus, the performance of both systems appears to be affected by directional noise in a highly similar way.

In order to visualize the connectivity structure of the network, the weights connecting the input layer with the output layer in Simulation 1 are displayed in Figures 8a and 8b. Each diagram shows the weights connecting all the low speed input cells to a single output cell, where the circle indicates the direction of heading to which the output cell is most selective. The orientations of the lines represent the directions of primary selectivity of the input cells, and their lengths represent the magnitudes of the weights; solid lines indicate positive weights, dotted lines indicate negative weights. There are three things to notice in these diagrams. First, it is clear that the output cells have become selective to radial patterns of optical flow similar to the S cells that Saito et al. (1986) identified in MST. Each output cell is maximally responsive to a radial pattern with a particular focus. This result seems to be inconsistent with neurophysiological data (Tanaka et al., 1989) which indicates that expansion-sensitive MST cells respond to global expansion regardless of the position of the focus. We increased the output cell radius from 10° to 20° in Simulation 1 during training, thus broadening the selectivity to position of the focus of expansion. Heading accuracy dropped from 0.83° to 1.22° of visual angle. Despite the increase in heading error, performance was still on the order of 1°. Perhaps, expansion-sensitive MST cells are very broadly tuned to position of the focus. Second, the weights connecting input cells near the focus of outflow are generally very small. Thus, optical flow very near the focus of outflow does not play a major role in determining heading direction. One reason for this is the fact that flow velocities near the focus are always small and fall near the low end of the speed tuning bandwidth. In addition, flow vectors falling in the receptive field that contains the focus could possess any orientation depending on their location about the focus. Because the input cell is not sensitive to the relative location of the flow vector, flow near the focus does not uniquely determine the heading direction. This has yet to be tested in human observers. Third, it is apparent that the strengths of the weights connecting input cells far away from the focus of expansion are also relatively small. One possible explanation for this is that these flow vectors are so large that they fall either near the end or outside the bandwidth of speed tuning. Similar connectivity structures were observed for the other two simulations.

By observing the resulting weighted connections between the input and output layers, it becomes clear now why the network generalized so well to stimuli corresponding to new random dot configurations and to new observer speeds in Simulation 1. The network's output nodes are acting like a set of radial flow filters or templates. The output node whose template best matches the novel flow field will be most activated. A stimulus corresponding to a novel random dot configuration differs from a training stimulus in that the random dots fall in different locations in the receptive fields of the input layer so that the flow vectors registered by each receptive field vary somewhat for a given heading direction. Novel random dot configurations produce a sort of



(a)



(b)

FIGURE 8. Diagrams illustrating the postlearning weights connecting low speed input cells to one output cell whose primary heading selectivity lies at the center of the circle. Line orientation represents the direction of primary selectivity of the four input cells at each receptive field location, and line length represents the magnitude of the weights: solid lines are positive weights and dotted lines are negative weights. (a) Output cell selective to headings at the center of the visual field. (b) Output cell selective to headings at upper right of visual field.

directional noise in the flow field. However, it is a special kind of noise that decreases with visual angle from the focus of expansion. Because flow vectors near the focus of expansion are not weighted very much, and this directional noise decreases with distance from the focus, a stimulus corresponding to a novel random dot configurations will highly activate the output node corresponding to the correct heading. Moreover, the network generalizes to new observer speeds because they only affect the magnitudes of flow vectors and not the radial pattern of the flow field. Because the network's output nodes are tuned to the radial structure of the field, variability in vector magnitude will not greatly affect the network's performance. A similar connectivity structure is observed for high speed input cells so that performance will be similar across very different observer speeds.

The results of Simulation 1 using lower dot density stimuli can also be understood in light of the observations of the connectivity structure. As the dot density decreases, the stimulus becomes more ambiguous to the network because more radial templates will match the input flow field. Moreover, in this context, the output layer can be thought of as a statistical estimator of the true heading direction. The reason is that the finite size of the input layer's receptive fields introduces noise, as stated above, so that heading is no longer a deterministic function of the input flow field. Therefore, as the sample size decreases, the standard error of the estimator increases and so average heading error increases. If we hypothesize the heading estimator as a sort of average of the flow data, estimator theory tells us that the standard error of the estimator is proportional to one over the square root of the number of data points. We tested this hypothesis by fitting the results of the simulation and the human experiment with a power function with an exponent of -0.5. We did this by transforming the independent variable (i.e., the number of dots) (see Figure 9) and fitting least-squares regression lines to the data. Both fits were remarkably good (correlations of .98 and .99 for the simulation and human data, respectively). Despite a difference in the y intercepts (0.40 vs. -0.07) the slopes were quite similar (3.25 for the simulation data vs. 4.36 for the human data) and a difference in y intercept (.40 vs. $-.07$). This suggests that the network and human subjects are using the redundant flow information in a similar fashion.

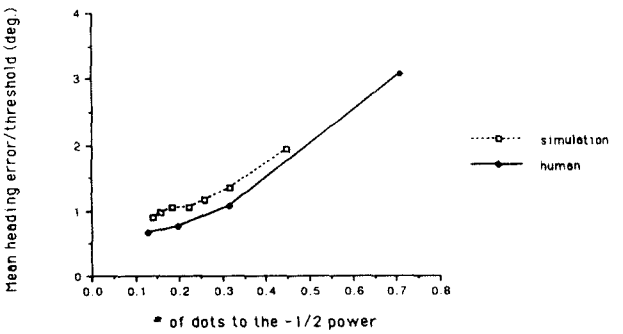


FIGURE 9. Mean heading error/threshold for Simulation 1 and human subjects as a function of transformed dot density.

The network's ability to generalize to new headings in Simulation 2 can be attributed to the distributed coding of heading on the output layer. The distributed coding results in a considerable overlap of activity for distinct but similar headings during training. This allows the network to interpolate to new headings during the test phase.

Finally, the network's noise tolerance can be explained as follows. The reason the network performed so well under speed noise is similar to that given for the network's ability to generalize to new observer speeds: vector magnitude is not diagnostic of direction of heading. The network's tolerance of moderate directional noise is explained by treating the output layer as a linear estimator of the true heading direction, as stated above. Given a large enough sample size, the estimate of the heading will have a small variance and so the average heading error will be small.

A small parametric study was performed to determine what factors significantly affect the network's performance using a training set of 400 random heading directions, as in Simulation 1. Two factors were systematically varied: (a) input and output quantization, and (b) number of input cells per receptive field. Input quantization refers to the size of the input layer's receptive fields. Because the receptive fields are not overlapping, the number of receptive fields increases as the receptive field size decreases. To our surprise, input quantizations from 4° to 2.5° had little effect on performance. Although performance on the training set did get better with higher quantization or resolution, average heading error was on the order of 1° with novel stimuli even for a receptive field size of 2.5°.[2] Output quantization refers to the number of output cells on the output layer. Likewise, an increase in output quantization to 11 × 11 (compared to 5 × 5) yielded little difference in average heading error except when it was accompanied with a decrease in the output response radius from 10° to 4°. In this case, average heading error was 0.68° for novel stimuli.

On the other hand, the number of input cells per receptive field had an effect on performance but only

[2] Average heading error was 1.19° compared to 1.12° with a 4.0° receptive field. The reason that the performance of the 4.0° r.f. simulation differs from that of Simulation 1 is that the range of observer speeds used during training differed greatly.

if the number of direction selectivities changed. Average heading error decreased to 0.62° with novel stimuli when the number of direction selectivities increased to 16. Increasing the number of speed selectivities per receptive field from 2 to 4 had no effect. Because the accuracy of representation of an attribute improves with an increase in the number of different selectivities coding the attribute, these results indicate that directional accuracy is more important than speed accuracy. This is, of course, not surprising because we have shown that directional noise affects performance much more than does speed noise. In a sense, an increase in the number of direction selectivities lowers the internal noise of directional coding.

9. CONCLUSION

We have presented a simple two-layer neural network that has been trained to form a linear mapping between the pattern of optical flow and the direction of heading. It is apparent from the analysis of its connections that the network is picking up the global radial flow pattern of optical flow, as does the human visual system (Warren et al., 1988). Many of the results found in perceptual experiments with humans have been replicated qualitatively with the model. First, heading accuracies on the order of 1° were found using high density, noiseless optical flow fields. Second, the network's behavior in response to additive noise in the flow fields was strikingly similar to that of humans. Heading accuracy was hardly affected by speed noise, which is not only consistent with the human data but agrees nicely with Gibson's radial outflow hypothesis. Likewise, the degradation of heading accuracy with increasing directional noise variance is comparable to the human data described above.

There are three appealing characteristics of the neural model besides the close match with the psychological data. First, unlike many computational models, it makes no assumptions about the structure of the environment given that the velocity vector field is properly detected. Differential models, in particular, require differentiable flow fields resulting from relatively smooth environmental surfaces. Since the network is picking up the radial invariant structure, it succeeds in more complex nonplanar environments, such as 3D clouds. Second, the model is, on a crude level, biologically plausible. The brain contains simple velocity-sensitive elements working in parallel. In particular, there is neurophysiological evidence that a mapping from velocity-sensitive units to radial pattern-sensitive units exists in primate visual cortex. Also various neurophysiological constraints were incorporated in the model such as tuning and receptive field properties of MT cells. Third, in contrast to discrete and differential models, the network is surprisingly tolerant of noise. This property makes it potentially useful for applied problems in computer vision and robotics.

A major shortcoming of the model is its inability to handle general observer movement including rotation as well as translation. Warren and Hannon (1988, 1990) showed that human observers can determine their heading to within one degree with flow fields that simulated translation plus eye rotation. A simulation was performed with training stimuli that added observer rotation about the z axis, simulating eye rotations. As expected, the heading error was quite large. Future research will extend the existing architecture. One possibility is to implement a neural analogue of the Rieger and Lawton (1986) difference-vector algorithm. An extra layer of cells would be added between the two layers of the existing network, the connections between it and the input layer wired so that the cells in the middle layer would have relative motion sensitive receptive fields, as found in the pigeon (Nakayama & Loomis, 1974; Frost & Nakayama, 1983) and monkey (Orban, Gulyan, & Vogels, 1987). These cells would, in effect, be determining difference vectors in local regions of the flow field and would, in turn, feed to radially selective output cells. Another option is to add a hidden layer and modify the connections using a learning rule such as back propagation to grow an optimal solution to the wiring problem (Zipser & Anderson, 1988).

Another shortcoming with the model is its reliance on supervised training. It is unclear from where such an external teaching signal would come. We address this issue in a follow-up article that discusses the problem of observer rotation and translation (Hatsopoulos & Warren, 1990), and suggest possible mechanisms such as Kohonen's feature-map unsupervised learning rule (Kohonen, 1989). Because the mapping from a difference vector field to heading direction is linear, similar difference vector fields should map to neighboring heading cells as the feature-map learning rule dictates.

Finally, although we have made no assumptions about the structure of the environment, we have postulated rigidity. Our model treats any violation of this assumption, such as blowing snow or sand, as noise. At first glance, it does seem reasonable to consider such violations of rigidity as noise. On the other hand, it may be more appropriate to be sensitive to such rigidity violations because they may provide valuable information about local object motions and deformations. However, such information could be detected by a separate subsystem, while the navigational system detects global properties of the flow. These issues have yet to be investigated experimentally.

In this research we have attempted to bring to-

gether the framework of Gibson's perceptual theory, the biological constraints provided by neurophysiology, and the power of a neural network model to offer a possible explanation of visual navigation in biological systems and provide robust methods for robotic control.

REFERENCES

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the Macaque. *Journal of Neurophysiology*, **52**, 1106–1130.

Anderson, J. A. (1983). Cognitive and psychological computation with neural models. *IEEE Transactions on Systems, Man and Cybernetics*, **13**, 799–815.

Ballard, D. H., & Kimball, O. A. (1983). Rigid body motion from depth and optical flow. *Computer Vision, Graphics, and Image Processing*, **22**, 95–115.

Bruss, A. R., & Horn, B. K. P. (1983). Passive navigation. *Computer Vision, Graphics, and Image Processing*, **21**, 3–20.

Cutting, J. E. (1986). *Perception with an eye for motion*. Cambridge, MA: MIT Press.

Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, **35**, 528–532.

Frost, B., & Nakayama, K. (1983). Single visual neurons code opposing motion independent of direction. *Science*, **220**, 744–745.

Gibson, J. J. (1947). *Motion picture testing and research* (AAF Aviation Psychology Research Report No. 7). Washington, DC: Government Printing Office.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Gibson, J. J. (1954). The visual perception of objective motion and subjective movement. *Psychological Review*, **61**(5), 304–314.

Gibson, J. J. (1968). What gives rise to the perception of motion? *Psychological Review*, **75**(4), 335–346.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gillam, B. (1988). *Is orientation disparity a stereoscopic primitive*? Psychonomic Society Annual Meeting, Chicago.

Hatsopoulos, N., & Warren, W. H. Jr. (1990). *Visual navigation during pursuit eye-movements*. Unpublished manuscript.

Hildreth, E. C. (1984a). The computation of the velocity field. *Procedures of the Royal Society of London*, **221**, 189–220.

Hildreth, E. C. (1984b). *The measurement of visual motion*. Cambridge, MA: MIT Press.

Horn, B. K. P., & Schunk, B. G. (1981). Determining optical flow. *Artificial Intelligence*, **17**(1–3), 185–203.

Johnston, I. R., White, G. R., & Cumming, R. W. (1973). The role of optical expansion patterns in locomotor control. *American Journal of Psychology*, **86**(2), 311–324.

Kennedy, H., & Orban, G. A. (1983). Response properties of visual cortical neurons in cats reared in stroboscopic illumination. *Journal of Neurophysiology*, **49**(3), 686–703.

Koenderink, J. J., & van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, **22**, 773–791.

Koenderink, J. J., & van Doorn, A. J. (1976). Local structure of movement parallax of the plane. *Journal of the Optical Society of America*, **66**, 717–723.

Koenderink, J. J., & van Doorn, A. J. (1981). Exterospecific component of the motion parallax field. *Journal of the Optical Society of America*, **71**, 953–957.

Kohonen, T. (1989). *Self-organization and associative memory, 3rd edition*. Berlin: Springer Verlag.

Lappin, J., & Norman, F. (1988). *Detectabilities of the components of optical flow*. Psychonomic Society Annual Meeting, Chicago.

Lawton, D. T. (1983). Processing translational motion sequences. *Computer Vision, Graphics, and Image Processing*, **22**, 116–144.

Lewellyn, K. R. (1971). Visual guidance of locomotion. *Journal of Experimental Psychology*, **91**, 245–261.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London, B*, **208**, 385–397.

Maunsell, J. H. R., & van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area (MT) of macaque monkey: I. Selectivity for stimulus direction, velocity, and orientation. *Journal of Neurophysiology*, **49**, 1127–1147.

Maunsell, J. H. R., & van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area (MT) of macaque monkey: II. Binocular interactions and the sensitivity to binocular disparity. *Journal of Neurophysiology*, **49**, 1148–1167.

Maybank, S. J. (1986). Algorithms for analyzing optical flow based on the least-squares method. *Image and Vision Computing*, **4**(1), 243–246.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *Journal of Neurophysiology*, **55**(6), 1308–1327.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. G. Gross (Eds.), *Study group on pattern recognition mechanisms* (pp. 117–151). Vatican City: Pontifica Academia Scientiarum.

Nakayama, K. (1985). Biological image motion processings: A review. *Vision Research*, **25**, 625–660

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity sensitive neurons, and space perception: A hypothesis. *Perception*, **3**, 63–80.

Orban, G. A., Gulyan, B., & Vogels, R. (1987). Influence of a moving textured background on direction selectivity of cat striate neurons. *Journal of Neurophysiology*, **57**, 1792–1812.

Prazdny, K. (1980). Egomotion and the relative depth map from optical flow. *Biological Cybernetics*, **36**, 87–102.

Prazdny, K. (1981). Determining the instantaneous direction of motion from optical flow generated by a curvilinear moving object. *Computer Graphics and Image Processing*, **17**, 238–248.

Regan, D., & Beverley, K. I. (1982). How do we avoid confounding the direction we are looking with the direction we are moving? *Science*, **215**, 194–195.

Regan, D., & Beverley, K. I. (1984). Psychophysics of visual flow patterns and motion in depth. In L. Spillman & B. R. Wooten (Eds.), *Sensory experience, adaptation, and perception: Festschrift for Ivo Kohler* (pp. 215–240). Hillsdale, NJ: Lawrence Erlbaum.

Rieger, J. H. (1983). Information in optical flows induced by curved paths of observation. *Journal of the Optical Society of America*, **73**(3), 339–344.

Rieger, J. H., & Lawton, D. T. (1986). Determining the instantaneous axis of translation from optic flow generated by arbitrary sensor motion. In N. I. Badler & J. K. Tsotsos (Eds.), *Motion: representation and perception* (pp. 33–41). New York: North-Holland.

Riemersma, J. B. J. (1981). Visual control during straight road driving. *Acta Psychologica*, **48**, 215–225.

Rodford, C. J. (1986). Optical flow fields in Hough transform space. *Pattern Recognition Letters*, **4**, 293–303.

Rodman, H. R., & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Visual Research*, **27**(12), 2035–2048.

Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., & Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience*, **6**(1), 145–157.

Sereno, M. (1987). *Implementing stages of motion analysis in neural networks*. Ninth Annual Conference of the Cognitive Science Society.

Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *Journal of Neurophysiology*, **39**, 1288–1319.

Tanaka, K., Fukada, Y., & Saito, H. (1989). 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. *Journal of Neurophysiology*, **62**, 642–656.

Warren, R. (1976). The perception of egomotion. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 448–456.

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

Warren, W. H., & Hannon, D. J. (1990). Eye movements and optical flow. *Journal of the Optical Society of America*, **7**, 160–169.

Warren, W. H. Jr., Griesar, W., Blackwell, A. W., Hatsopoulos, N. G., & Kalish, M. (1990). On the sufficiency of the velocity field for the perception of heading. Unpublished manuscript.

Warren, W. H. Jr., Morris, M. W., & Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 646–660.

Waxman, A. M., & Ullman, S. (1985). Surface structure and 3D motion from image flow: A kinematic analysis. *International Journal of Robotics Research*, **4**, 72–94.

Weller, R. E., Wall, J. T., & Kaas, J. H. (1984). Cortical connections of the middle temporal visual area (MT) and the superior temporal cortex in owl monkeys. *Journal of Comparative Neurology*, **228**, 81–104.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. 1960 *WESCON Convention, Record Part IV* New York: IRE, 96–104.

Zipser, D., & Anderson, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, **331**, 679–684.