A Self-Organizing Neural Network Architecture for Navigation Using Optic Flow
A SELF-ORGANIZING NEURAL NETWORK
ARCHITECTURE FOR NAVIGATION
USING OPTIC FLOW

Seth Cameron, Stephen Grossberg, and Frank H. Guenther

December 1995

Technical Report CAS/CNS-95-032

Permission to copy without fee all or part of this material is granted provided that: 1. the copies are not made or distributed for direct commercial advantage, 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and/or special permission.

Copyright © 1995

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummingston Street
Boston, MA 02215
A Self-organizing Neural Network Architecture for Navigation Using Optic Flow

Running title: Navigation Using Optic Flow

Seth Cameron
Stephen Grossberg
Frank H. Guenther

Boston University
Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA, 02215
Fax Number: (617) 353-7755

Boston University Center for Adaptive Systems
Technical Report CAS/CNS 95-032

Seth Cameron was supported in part by the Office of Naval Research (ONR N00014-92-J-4015, N00014-92-J-4100, N00014-92-J-1309, N00014-92-J-0657). Stephen Grossberg was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0499) and the Office of Naval Research (ONR N00014-95-1-0409). Frank H. Guenther was supported in part by the Alfred P. Sloan Foundation, the Office of Naval Research (ONR N00014-95-1-0409), and the Air Force Office of Scientific Research (AFOSR F49620-92-J-0499).
ABSTRACT

This paper describes a self-organizing neural network architecture that transforms optic flow information into representations of heading, scene depth, and moving object locations. These representations are used to reactively navigate in simulations involving obstacle avoidance and pursuit of a moving target. The network's weights are trained during an action-perception cycle in which self-generated eye and body movements produce optic flow information, thus allowing the network to tune itself without requiring explicit knowledge of sensor geometry. The confounding effect of eye movement during translation is suppressed by learning the relationship between eye movement outflow commands and the optic flow signals that they induce. The remaining optic flow field is due only to observer translation and independent motion of objects in the scene. A self-organizing feature map categorizes normalized translational flow patterns, thereby creating a map of cells that code heading directions. Heading information is then recombined with translational flow patterns in two different ways to form maps of scene depth and moving object locations. All learning processes take place concurrently and require no external "teachers." Simulations of the network verify its performance using both noise-free and noisy optic flow information.
1. Introduction: Optic Flow, Heading, and Visual Navigation

As we move through the world, we experience flowing patterns of light on our retinas. Scientists have studied the abilities of humans to use this optic flow for a variety of tasks, including determination of heading (e.g., Crowell and Banks, 1993; Van den Berg, 1992, 1993; Warren and Hannon, 1988, 1990; Warren and Kurtz, 1992), observer velocity (e.g., Monen and Brenner, 1994), three-dimensional structure (e.g., Cornilleau-Peres and Droulez, 1993; Hildreth, Grzywacz, Adelson, and Inada, 1990; Treue, Andersen, Ando, and Hildreth, 1995), the locations and speeds of self-moving objects (e.g., Brenner, 1991) and distances to objects (e.g., Prazdny, 1980; Simpson, 1988, 1993). With the exception of determining observer velocity (Monen and Brenner, 1994), humans have proved to be very proficient in using optic flow to perform these tasks.

This article describes a neural network model that uses optic flow information to determine the direction of heading of a moving observer. This heading network is embedded in a larger modeling architecture that includes neural networks for determining distances to visible objects and the locations of moving objects. The resulting representations of heading, scene depth, and moving object locations are then used to reactively navigate around visible obstacles and to pursue moving targets. The aim of this modeling project is to use insights from biology to design a robust system that can be used for navigation in autonomous mobile vehicles. To this end, the system is designed as an interconnected collection massively parallel, self-organizing neural networks. As described below, massive parallelism allows the system to effectively utilize redundant sensor information in order to produce a system that is tolerant to errors in sensor responses, and the self-organization process allows the system to learn to use arbitrary sensor geometries and to remain tuned despite slow drifts in sensor or motor properties (e.g., due to wear).

A central problem addressed by the current architecture is the determination of the heading direction of a moving observer from the optic flow field. Gibson (1950) remarked that the optical flow pattern experienced by an observer moving along a straight line through a static environment contains a singularity he termed the focus of expansion (FOE). Gibson further hypothesized that the visual system might use the global pattern of radial outflow originating from the singularity to determine the translational heading of the observer. However, psychophysical experiments rejected this hypothesis because the flow pattern on the retina is radically altered by eye movements (Regan and Beverly, 1982). If the observer’s eyes rotate during translational movement, the resulting flow pattern is a superposition of two vector fields such that the FOE no longer necessarily coincides with heading direction. Such eye rotations are routine because people tend to visually track stationary objects as they walk.

Since Gibson, vision researchers have proposed a wide variety of models that extract egomotion from optical flow information. Some popular approaches have been:
(i) Models that use differential invariants (e.g. Koenderink and van Doorn, 1975; Waxman and Ullman, 1985). This technique solves for the motion of planar surfaces. A rigidly moving planar surface is a special case that gives rise to flow fields in which the flow vectors vary as a quadratic function of image position. Waxman and Ullman (1985), for example, first subdivided the flow field into patches that are approximately planar. The three-dimensional structure was then recovered in closed form from the parameters of the second order flow field. An advantage of these models is that the motion in each patch is computed independently, so the methods can deal with multiple moving objects. One drawback is that they require dense optic flow fields to compute derivatives. Humans, however, can determine heading quite successfully using sparse flow fields. Models that use differential invariants are also very vulnerable to noise in the flow field (Hatsopoulos and Warren, 1991). Noisy sensor information is inevitable, not just in robotics applications but also in the human visual system due to difficulties such as the aperture problem (e.g., Perrone, 1992).

(ii) Methods that rely on local differential motion. (e.g. Cutting, 1986; Hildreth, 1992; Longuet-Higgins and Prazdny, 1981; Reiger and Lawton, 1985). If two points have the same image location but are at different depths, then the vector difference between the two flow vectors is oriented toward the FOE. This technique relies on locating adjacent image features separated in depth so that their flow vector difference is oriented approximately toward the FOE. It fails to operate when approaching a wall with no depth variation, but human heading discrimination is also poor under this condition. As with the differential invariant models, local differential motion techniques do not work well on sparse flow fields and in the presence of noise.

(iii) Least-square minimization algorithms that use redundant information from available flow vectors. Heeger and Jepson (1990) proposed a solution to the heading calculation problem that uses least-square minimization. Their technique, termed the “subspace algorithm”, is robust and comparatively insensitive to noise. It has the advantage of not requiring proprioceptive information from eye muscles to cancel the effects of eye rotations, but as described, is not self-organizing.

(iv) Neural network models. Several researchers have recently posited neural network models for determining optic flow. Because of the massive parallel processing capabilities embodied in these networks, neural network solutions tend to be significantly more tolerant to noise in the flow field. Hatsopoulos and Warren (1991) describe a neural network motivated by neurophysiological results from areas MT and MST in primate visual cortex. Although the model is tolerant to both speed and directional noise in the flow field, the authors point out that it has two major shortcomings (p. 315): (1) unlike humans, the model’s accuracy degrades drastically in the presence of eye rotations, and (2) the model relies on supervised learning; i.e., it requires an external teaching signal that would not be available to a developing animal. Perrone (1992) incorporated both translational and rotational detector cells in a model that can determine heading even during eye rotations. Lappe and Rauschecker (1993) describe a neural network that effectively implements the
algorithm of Heeger and Jepson (1990). Perrone (1992) and Lappe and Rauschecker (1993) also compared various layers of their network to anatomical areas in the brain, most notably areas MT and MST.

Although the scope of the current model is larger than the models of Hatsopoulos and Warren (1991), Perrone (1992), and Lappe and Rauschecker (1993) in that it determines distances to visible objects and the locations of moving objects in addition to observer heading, the model is similar to these models in that it uses massive parallelism to capitalize on the redundancy in the flow field in order to achieve noise tolerance. Unlike these models, however, the current model’s parameters are learned during an action-perception cycle rather than calculated off-line by the modeler or trained using an external teaching signal. The system learns appropriate parameters without a priori knowledge of the geometry and dynamics of the eye and head systems, and it can stay accurately tuned even when changes occur in the dynamics of these systems. Unlike the model of Hatsopoulos and Warren (1991), the model’s performance is not degraded by eye rotations. Furthermore, because the network’s components and learning laws are biologically based and the learning cycle uses only the kinds of information available to a biological organism, this work provides insight into the problems that biological systems may solve in order to use optic flow information to determine their heading, to identify the positions of objects and obstacles in the world around them, to determine the relative motions of these obstacles, and to navigate around them.

2. The Optic Flow Field

The optic flow field can be represented by the projection of object movement in the three-dimensional world onto a two-dimensional image plane. The motion of a point in three-dimensional space has six degrees of freedom: translational velocity $T = (T_x, T_y, T_z)'$ and rotation $\Omega = (\Omega_x, \Omega_y, \Omega_z)'$. If the position of an observer’s eye is $R = (X, Y, Z)'$, then motion may be described by $V_{observer} = T + (\Omega \times R)$. If the observer is moving through a static environment, a viewer centered coordinate frame may be established in which the observer appears to be standing still, and each point in the environment is moving with the opposite motion, $V = -(T + \Omega \times R)$.

Figure 1 shows how a point, $P$, in three-dimensional space is projected onto a point $r = (x, y)' = f(X/Z, Y/Z)'$ in the image plane, where $f$ denotes the focal length of the lens ($f = 1$ was used in the simulated model). Image velocity is defined as the derivative, with respect to time, of the $x$- and $y$- components of scene point projections, and can be written as follows (Heeger and Jepson, 1990):

$$\theta (x, y) = (v_x, v_y) = p (x, y) A (x, y) T + B (x, y) \Omega$$  

(1)
where \( p(x,y) = 1/Z \) is the inverse depth of points in the image plane, and \( A(x,y) \) and \( B(x,y) \) are matrices that depend only on image position and focal length, not on any unknowns:

\[
A(x,y) = \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix}
\]

\[
B(x,y) = \begin{bmatrix} \frac{xy}{f} & -(f + \frac{x^2}{f}) & y \\ \frac{f + y^2}{f} & -\frac{xy}{f} & -x \end{bmatrix}
\]

The various processing stages of the current model can be thought of as decomposing the right side of Equation 1 until we are left with a map representing the direction of the vector \( T \), which is the observer's heading direction.

In the following description, we will be interested in the speed and direction of the flow components at each retinal location since these components form the input to the model. The speed at a retinal position is:

\[
v = \| \mathbf{v} \| = \sqrt{v_x^2 + v_y^2}
\]

and the direction is:


\[ \phi = \tan^{-1}\left( \frac{v_x}{v_y} \right). \] 

(4)

3. Model Description

Figure 2 shows an overview of the neural network architecture. The following paragraphs describe the major components of this model.

3.1. Model Components

Motion detector field. The input layer is a population of cells sensitive to image flow on the retina. Several researchers have proposed physiologically plausible models that yield population encoding of optic flow vectors (e.g., Wang, Mathur, and Koch, 1989). A typical approach uses a separate floret of cells for each sampling point on the retina. Each cell of a floret is tuned to a preferred direction, \( e_k = [\cos(2\pi k/n), \sin(2\pi k/n)] \), where \( k = 1, ..., n \). Florets of this form constitute the first stage of the model, called the motion detector field. Motion at the \( i^{th} \) retinal location with speed \( v \) and direction \( \phi \) will generate a response \( S_{ki} \) according to the following equation:

\[ S_{ki} = \left( v \cos\left( \phi - \frac{2\pi k}{n} \right) \right), \quad 1 \leq k \leq n, \quad 1 \leq i \leq m \] 

(5)

For robotic applications, it suffices to represent the flow field with its x and y components

\[ S_{xi} = v \cos (\phi) \quad S_{yi} = v \sin (\phi) . \] 

(6)

The simulations in this report used a motion detector field that consisted of these two directionally tuned cells at each of 49 (7x7) retinal locations unless noted otherwise.

Eye velocity vector. The eye velocity vector consists of two pairs of opponent cells that represent pitch velocity \( (O_{\theta}, O_{\phi}) \) and yaw velocity \( (O_{\theta}, O_{\psi}) \) of the eye when it is rotating in its orbit. For positive rotations about the x axis, \( O_{\theta} \) is linearly related to the rotational speed and \( O_{\phi} \) is zero. For negative rotations, \( O_{\theta} \) is zero and \( O_{\phi} \) is linearly related to rotational speed. \( O_{\psi} \) and \( O_{\phi} \) code rotations about the y axis in a similar manner. The eye velocity vector activities project through adaptive inhibitory weights to the translation field as described below. This inhibitory input learns to cancel the portion of the flow field corresponding to eye rotations.

An ongoing argument in the egomotion literature concerns whether or not extraretinal information is required to interpret optic flow signals. In most situations, people tend to fixate on a point as they move. This behavior generates a mixed optic flow signal resulting
Navigation Using Optic Flow

FIGURE 2. Overview of the model. See text for details.

from the simultaneous translation of the body and rotation of the eyes. Unless the point of
fixation coincides with the direction of heading, the FOE that results when fixating during egomotion does not correspond to the direction of body translation. There are at least three approaches to solving the heading problem during eye rotations. One is to use an estimate of eye rotational velocity (e.g., from eye muscle proprioception or motor command output) to internally negate the rotation part of the signal from the flow field. The current model uses this approach. Another is to sample a large part of the flow field and subtract off any component that is common at all depths without using explicit knowledge of eye rotations (e.g., Perrone, 1992). A third approach, like the Heeger and Jepson (1990) algorithm, uses more sophisticated heading calculations that work in the presence of rotational components, again without requiring explicit knowledge of eye rotations.

Warren and Hannon (1990) originally reported that no extraretinal information was necessary to determine heading from optic flow. They set up computer simulations of optic flow generated by both translational and rotational movements. In these experiments, the rotational component was artificially generated (rather than resulting from actual eye movements), thus insuring that subjects had no extraretinal information concerning the rotation. Subjects were required to decide, based on the flow information alone, if their perceived heading would take them to the left or right of a virtual object. Warren and Hannon reported that subjects were able to determine their heading accurately with or without an added rotational component. However, Royden, Banks, and Crowell (1992) performed a similar psychophysical study and found that people’s ability to determine heading was dramatically reduced when unknown rotational components were added to the translational flow field. Since this conflicted with Warren’s results, they recreated their experiment and determined that the discrepancy was due to unusually slow rotation rates used in the original experiment (0.2 - 1.2 deg/sec). Royden et al. simulated rotation speeds up to a more realistic 5 deg/sec. Their results suggest that extraretinal information about eye rotations may indeed be necessary for accurate heading perception. Still another set of experiments performed by van den Berg (1992, 1993) indicates that if the artificially generated rotational component of the flow field corresponds to eye movements that would be used to maintain fixation on a stable point in the image, subjects can successfully compensate even for higher rotation rates such as those used by Royden et al. Lappe and Rauschecker (1995) suggested that the success of subjects may depend on whether the flow field contains an FOE within the image plane, which is always the case when an observer rotates his eyes to maintain fixation of a stationary object but may not be the case during other eye rotations.

In summary, it is still not clear whether humans use extraretinal information when processing the optic flow field. The eye velocity vector in the current model is most naturally interpreted as either an efference copy of eye velocity commands or a proprioceptive representation of eye velocity. Such an efference copy of eye velocity is computed within a neural model of trajectory formation (Bullock and Grossberg, 1988). In a mobile robot with an active vision system, this vector would correspond to motor velocity commands and/or velocity feedback from eye position sensors. Within this interpretation, the model uses an extraretinal signal for processing the optic flow field. Alternatively, it is possible to
form an eye velocity vector from visual information (c.f. rotation detector cells in the model of Perrone, 1992) that could play a role in filtering the rotational component of the flow field.

**Translation field.** As shown in Figure 2, each cell $O_j$ in the eye velocity vector has a set of inhibitory weights $w_{jki}$ that project to each directionally tuned cell $k$ in the floret at every position $i$ in the translation field. Each translation field cell also receives excitatory input from the corresponding cell in the motion detector field. After learning (described in Section 4), the net result of these inputs is a field of cells that represents only the component of the flow field which is due to translational motion. That is, the translation field activity contains only the $p(x,y)A(x,y)T$ component of Equation 1.

A variant of the vector associative map (VAM) neural architecture (Gaudiano and Grossberg, 1991) is used to adjust the weights projecting from the eye velocity vector in such a way as to cancel out the portion of flow corresponding to voluntary rotations of the eyes as registered at the eye velocity vector. The activity of a cell in the translation field is formed by subtracting the inhibitory input from the eye velocity vector cells (corresponding to the rotational component of the flow field) from the excitatory input projecting from the motion detector field:

$$W_{ki} = S_{ki} - \sum_j O_j w_{jki}.$$  \hspace{1cm} (7)

and the learning equation for the weights projecting from the eye velocity vector is:

$$\frac{dw_{jki}}{dt} = \alpha_1 W_{ki} O_j.$$  \hspace{1cm} (8)

where $\alpha_1$ is a learning rate parameter ranging from 0.9 for noise-free conditions to 0.01 for noisy conditions. Before learning, the weights are set equal zero.

During purely rotational movements of the eye, the VAM learning law of Equation 8 adjusts the weights so that the inhibitory input from the eye velocity vector exactly cancels the excitatory input from the motion detector field. When the two inputs are equal, $W_{ki}$ in Equation 8 goes to zero and learning stops. At this point the system is tuned such that the rotational component of the flow field is effectively "filtered out" by eye velocity vector projections at the translation field.

During movements with both a translational and rotational component, only the rotational component is reliably correlated with the presynaptic activity of the eye velocity vector cells. The translational flow component will vary depending on heading direction and will not be strongly correlated with eye velocity vector activity. This translational component thus amounts to a form of noise in the learning process. A slower learning rate can be used
to “average out” the noisy translational component, as well as other forms of noise such as noisy sensor information. The ability to learn correct parameters in the face of sensor noise or translational motion is demonstrated in the simulations of Section 4.

Normalization field. The next stage of cells in the model performs a normalization operation on the outputs of the translation field. This is done because the magnitude of the flow vectors at the translation field contains information pertinent to observer speed and object distances only, not observer heading direction. In other words, we are not interested in the magnitude of the $p(x, y)A(x, y)T$ term from Equation 1. Only the direction of vector $T$ is needed. Normalization removes the unpredictable changes in flow vector magnitudes due to varying object distances and translation speeds. Neural networks capable of performing such a normalization operation were described in Grossberg (1973). They are recurrent on-center off-surround networks that undergo shunting, or membrane equation, interactions. Such networks have been described in the neocortex (Douglas et al., 1995). If the suprathreshold activity from the translation field is passed to another layer with recurrent linear feedback, the input pattern will become normalized in a manner that is approximated by Equation 9:

$$N_{ki} = \frac{[W_{ki}]^+}{\sum_m [W_{mi}]^+}$$

(9)

where $[\cdot]^+$ denotes a rectification function such that $[x]^+ = x$ for $x > 0$, $[x]^+ = 0$ otherwise.

Equation 9 describes an $L_1$ normalization. An $L_2$ normalization is given by

$$N_{ki} = \frac{[W_{ki}]^+}{\sqrt{\sum_m ([W_{mi}]^+)^2}}$$

(10)

Simulations indicated that either $L_1$ or $L_2$ normalization could be used but that $L_2$ was superior for matching flow fields with headings. The simulation results reported below were performed using $L_2$ normalization. All $N_{ki}$ are set to zero for retinal locations where the total flow signal is zero or negligible. These locations typically include points with no visible objects, very distant objects, or points near the FOE for a given heading.

Heading map. The next stage of cells is a self-organizing feature map (SOFM) that encodes heading. Cells in the heading map receive weighted excitatory input projections from the normalized flow field. Heading can be determined by classifying the pattern across these inputs. Kohonen (1982) described a learning algorithm, based on earlier work of von der Malsburg (1973) and Grossberg (1976), that can be used to self-organize a
topographical map whose cells classify input patterns. In the present application, neighboring cells in the map code similar heading directions. A neural mechanism that was proposed to create this kind of topographically organized map is a shunting neural network that utilizes distant-dependent competitive interactions with appropriate feedback signals (e.g., Grossberg, 1976). During learning, neighborhood interactions cause adjacent cells in the heading map to code similar headings. The interactions also ensure that the map uses all of its cells to encode headings. Otherwise, a single cell might learn large weights early on and continue to monopolize learning even though its match with the input pattern is poor. In addition, the neighborhood interactions cause heading map cells to automatically distribute themselves according to the frequency statistics of sampled headings. Greater discrimination among the more common headings is possible because these headings attract a higher density of map cells.

The heading map consists of a group of cells with activities $H_i$ that perform an instar classification (Grossberg, 1976; Kohonen, 1982) using weights $h_{kil}$ projecting from cells with activities $N_{ki}$ in the normalized flow field. For each presentation of a normalized flow pattern, one cell in the heading map will receive the largest total input. Early in development, this maximally excited cell and its $N$ nearest neighbors are set to activity level $1/(N+1)$, and all other map cell activities are set to zero. Weights projecting to each of the active cells are adjusted towards the flow pattern. Initially, large neighborhoods help the map cells, which begin with small random weights chosen from a uniform distribution between $0$ and $0.1$, to ‘unfold’ properly and thereby cover the entire range of experienced headings. Over time, the neighborhood size $N$ shrinks to zero so that flow patterns are categorized more precisely. The following equations were simulated:

\[
H_i^* = \sum_{ki} N_{ki} h_{kil}
\]

$H_{max} =$ cell with maximum activity $H_i^*$

\[
H_i = \frac{1}{N+1} \quad \text{for } H_{max} \text{ and } N \text{ neighbors}
\]

\[
H_i = 0 \quad \text{otherwise}
\]

\[
\frac{dh_{kil}}{dt} = \alpha_2 (N_{ki} - h_{kil}) H_i,
\]

where $H_i^*$ is the input to the $i^{th}$ heading map cell, and $\alpha_2$ is a learning rate parameter that starts at $0.1$ and shrinks with a linear decay rate to $0.001$ over $2000$ learning trials. The neighborhood $N$ starts as a $15 \times 15$ square centered at the maximally active cell. The width of this square shrinks by one after every $100$ heading samples until the neighborhood consists of only the maximally active map cell.
After training, a cell in the heading map will respond maximally to flow generated by a particular heading. However, most headings will fall between those encoded by cells in the heading map. During performance, winner-take-all competition among the heading cells is relaxed so that several candidate heading cells survive the competition based on the size of their total input. Allowing distributed activation across heading cells creates an interpolated output. For example, if an input heading falls halfway between headings encoded by two cells, then each of those cells will probably survive the competition and contribute approximately one-half of their information to the rest of the network. Activity in the surviving heading cells is also normalized so that the heading representation is a weighted average of active heading cells. Heading cell activity is thus calculated during performance as follows:

$$H_i = \sum_{kl} N_{kl} h_{kl}$$

\[ H_i = 0 \text{ if } H_i < (\max_m \{H_m\} - \Gamma) \]

\[ H_i = H_i \text{ otherwise} \]

$$H_i = \frac{H_i}{\sum_m H_m}$$

(13)

where \( \Gamma \) is a threshold parameter determining how many cells survive the competition. For the simulations, \( \Gamma \) was set to \( \max_m \{H_m\} / 15 \). In words, the heading map cell activities \( H_i \) are a contrast-enhanced version of their total inputs \( H_i \). Note that the sizes of the total inputs to the map cells determine the "neighborhood" during performance, whereas the active neighborhood during training is determined based on proximity to the maximally active cell. This was done because it provided the best map unfolding during training and noise tolerance during performance. Since the trained map is topographically organized, however, the maximally active cells during performance will still typically be neighboring cells in the map.

It is important to note that, as described thus far, the heading map can only discriminate between headings, not identify heading in terms of absolute azimuth and elevation angles. This is sufficient for the proper formation of the depth and independent motion maps in the model. However, testing the network's ability to accurately determine heading and using the heading estimate to perform navigation require that retinotopic labels be attached to each cell in the heading map. A simple technique for assigning a retinotopic label to a heading cell with activity \( H_i \) is to find the index \( i \) of the smallest weight \( d_{il} \) projecting from that cell to the retinotopically organized depth map. The smallest weight will
be located near the FOE on the retina, which is equivalent to the retinotopic heading. Alternatively, the following technique could be used to assign heading labels in a mobile robot:

1. Move towards a point of light at eye level and adjust the wheels until the image of the light is stationary on the optic sensor. When the image is stationary, the navigator is heading directly towards the point of light in body-centered coordinates.

2. As the robot approaches the light, rotate the eye so that the light image falls on many different regions of the optic sensor. Since the navigator is heading towards the light in body-centered coordinates, the projected point of light should still be motionless even though it is projected to a different retinal location. However, the retinotopic heading will have changed. The floor and other objects in the scene will create a flow pattern that excites a different cell in the heading map.

3. For each eye position, map the maximally excited heading cell to the current retinotopic location of the imaged point of light.

This technique was approximated in the current simulations by systematically sampling heading directions and labeling each cell with the heading that maximally excited it. The heading estimate is calculated using the following equation:

\[ H_{MAP} = \sum_i H_i \Phi_i \]  

(14)

where \( \Phi_i \) is the preferred heading of the \( i^{th} \) heading map cell, and \( H_i \) is determined using Equation 13.

**Depth map**. In addition to providing heading information, optic flow can be used to determine the distance to the object at each retinal location (if one is present) in a scene. For visual navigation, it suffices to form a distance measure that is scaled by the speed of the observer; e.g., an inverse time-to-collision measure \( p \|T\| \) for each retinal location, where \( p \) is the inverse depth, or proximity, of an object, as in Equation 1, and \( \|T\| \) is the translation speed of the observer. The general problem of determining scene depth from a sequence of images is a difficult one, largely due to the unknown rotational component of the flow field and the unknown heading direction. Solutions have been proposed (e.g., Koenderink and van Doorn, 1987; Longuet-Higgins, 1981; Prazdny, 1980; Waxman and Ullman, 1985; Waxman and Wohn, 1988; see Simpson, 1993 for a review), but these solutions typically require rather complex calculations due to the difficulty of the problem. In contrast, once the rotational component of the flow field is removed and heading direction is known, as in the current network, calculating the time-to-collision at each retinal position is relatively straightforward (see also Perrone, 1992). By removing the rotational flow...
component from Equation 1, the optic flow generated by translational motion alone reduces to:

\[
\begin{align*}
    v_{Tx} &= p\|T\|(-fT'_x + xT'_z) \\
    v_{Ty} &= p\|T\|(-fT'_y + yT'_z)
\end{align*}
\]

(15)

where \(v_{Tx}\) and \(v_{Ty}\) are the x and y components of the translational flow field, \(f\) is the focal length of the imaging system, \((T'_x, T'_y, T'_z)\) are the components of a unit vector that defines the direction of translation, and \((x, y)\) specifies retinal location. From Equation 15, one can derive the relationship between the magnitude of the flow vector at a retinal location, \(v_T\), and time-to-collision:

\[
v_T = p\|T\|\sqrt{(-fT'_x + xT'_z)^2 + (-fT'_y + yT'_z)^2}.
\]

(16)

Note that for a specific heading \((T'_x, T'_y, T'_z)\) and motion sampling position \((x, y)\) on the retina, the square root term in Equation 16 is a constant, which we can rename \(k_{Txy}\). In the current simulations, which involve only two directional cells at each retinal location of the translation field, the function \(g\) that determines the magnitude of the translational flow velocity at the \(i^{th}\) retinal location is:

\[
g(W_i) = \sqrt{W^2_{1i} + W^2_{2i}} = v_T.
\]

(17)

In order to produce a depth measure that is invariant across all headings and retinal locations, \(k_{Txy}\) must be removed from the flow speed measure \(v_T\). This is accomplished in the model by logarithmically compressing the flow speed represented at the translation field before passing it to the cell representing the corresponding retinal location in the depth map, then subtracting off an average value of this compressed flow speed. Specifically, the retinotopically organized depth map cell activities are governed by the following equation:

\[
D_i = \log[ g(W_i) ] - \sum_j H_{ij}d_{ji},
\]

(18)

where \(d_{ji}\) is the weight projecting from the \(j^{th}\) heading map cell to the \(i^{th}\) depth map cell. Tal and Schwartz (1994) have demonstrated that a logarithmic relationship between neuron firing rate and input activity is a property of integrate-and-fire neurons. This suggests that the logarithmic processing in Equation 18 could also be easily implemented in a biological system.
Navigation Using Optic Flow

Using a VAM learning mechanism, the weight \( d_{li} \) learns to represent the average of the compressed flow velocity signal at the \( i \)th retinal location when heading in the direction coded by the \( i \)th heading map cell. The VAM learning equation governing the weights \( d_{li} \) is:

\[
\frac{dd_{li}}{dt} = \alpha_3 D_i H_i,
\]

(19)

where \( \alpha_3 \) is a learning rate parameter that was set equal to \( \alpha_2 \) for the simulations.

Training the network according to Equations 18 and 19 during random movements leads to depth map cell activities that each code the following depth measure for objects at the corresponding retinal location:

\[
D_i = \log v_T - \log v_f \\
= \log (p\|T\|k_{Txy}) - \log (p\|T\|k_{Txy}) \\
= (\log p\|T\| + \log k_{Txy}) - (\log p\|T\| + \log k_{Txy}) \\
= (\log p\|T\| + \log k_{Txy}) - (\log p\|T\| + \log k_{Txy}) \\
= \log p\|T\| - \log p\|T\|,
\]

(20)

where the term \(|\log p\|T\|\) is a constant determined by the environmental experience of the network during learning. A large positive cell activity in the depth map corresponds to a short time-to-collision; these large cell activities identify the retinal locations of nearby obstacles for use by the navigation module. This depth measure has several advantages. First, it is easily learned and calculated in the network using the VAM mechanism described above. Second, it is invariant with respect to retinal position. In particular, the same value of the depth measure corresponds to the same time-to-collision regardless of the retinal location, even though objects at the same depth generate different velocity signals at different places on the retina. Third, logarithmic processing leads to a depth measure that is compressed such that nearby objects garner a larger percentage of the cell’s dynamic range. This allows the depth map to represent proximity of nearby objects more accurately than distant objects, which is a useful property for avoiding collisions. Finally, because the depth map encodes an inverse time-to-collision measure \( p\|T\| \), a large signal means a short time-to-collision. Therefore, the most salient signals in the depth map are those that present the most danger, another useful property for obstacle avoidance.

Although faithfulness to biology is not the top priority of this model, it is interesting to note that with a larger number of directional cells at each retinal location (as would be expected in motion processing pathways in vivo), a more biologically plausible function \( g \)
Navigation Using Optic Flow

than the one defined by Equation 17 can be used to determine the magnitude of translational flow velocity; namely,

\[ g(W_i) = \sum_k [W_{ki}]^+ \]  

(21)

where \([x]^+\) denotes a rectification function such that \(x^+ = x \) for \(x > 0\), \(x^+ = 0\) otherwise. For a small number of directionally tuned cells at each retinal location, this measure will vary for the same translational flow speed depending on the direction of the flow vector. However, as the number of directionally tuned cells increases, this undesirable dependency on the angle of the retinal flow vector decreases, as indicated in Figure 3.

**FIGURE 3.** Speed measure produced by Equation 21 as a function of translational flow angle and number of cells in the floret. The speed measure is scaled so that the maximum output for each collection of cells is 1.0. More cells reduce the angular dependence of this flow speed measure.

**Independent motion map.** While sitting still, it is a common experience to have one's attention drawn to a moving object. This is not surprising since the isolated motion signal uniquely identifies the location of the object. During locomotion, however, large optic flow signals fill the entire retina, yet independently moving objects may still retain their perceptual saliency. For example, while driving down a street, the optic flow velocities in peripheral vision can be quite large, yet the location of a moving vehicle is still salient because the optic flow signal it creates differs from the pattern of flow generated by surrounding stationary objects.

The independent motion map in Figure 2 is formed by a VAM mechanism, similar to the one used for estimating depth, that detects independently moving objects. Whereas the depth VAM compares the magnitude of a flow signal to a learned average, the motion VAM compares the direction of flow with an expected pattern of flow. The net input to a cell in the independent motion map is:

\[ I_{ki} = N_{ki} - \sum_l H_l m_{lki} \]  

(22)
where $N_{ki}$ is the excitation from the corresponding cell in the normalization field and the $m_{ki}$ are adaptive weights tuned to represent the expected flow field when heading in the direction represented by heading map cell output $H_p$. The cell outputs $M_{ki}$ in the independent motion map are calculated from this input as follows:

$$M_{ki} = [I_{ki} - \Lambda]^+, \quad (23)$$

where $\Lambda$ is a constant threshold designed to keep slight mismatches from cluttering the motion detection field. The weights $m_{ki}$ are adjusted as follows:

$$\frac{dm_{ki}}{dt} = \alpha_4 I_{ki} H_I, \quad (24)$$

where $\alpha_4$ is a learning rate parameter that was set equal to $\alpha_2$ for the simulations. This learning law leads to weights $m_{ki}$ that represent the average directional flow pattern seen at retinal location $i$ when moving in the heading direction coded by $H_p$. After training, if the pattern of flow received from the normalization field corresponds to the expected pattern for heading $H_p$, then weighted inhibition from $H_I$ will cancel the input from the normalization field, and all activity in the independent motion map will be quenched. However, independently moving objects will likely generate flow directions that differ from the directions expected for the perceived heading and retinal locations. A non-zero output $M_{ki}$ indicates that an independently moving object is located at the retinal location indexed by $i$ since the flow in direction $k$ at this retinal location is significantly larger than the expected flow for the current heading as encoded by the weights $m_{ki}$. The larger the value of $\Lambda$, the larger the angular difference between the normalized flow signal and the expected flow signal must be before it produces a positive activation in the independent motion map.

**Navigation module.** Together, the heading, depth, and independent motion maps provide a representation of the visual world that can be used to navigate around obstacles toward stationary or moving targets. Goal-based navigation in the presence of obstacles is carried out in the simulations reported here using a simple approach-avoidance algorithm. For simplicity, the navigator was limited to fixed-speed motion on the ground plane ($\|T\| = 1$), thus requiring the generation of only an azimuthal translation command, $T_x$.

Navigation toward a target involves keeping the output of the heading map $H_{MAP}$, defined in Equation 14, equal to the direction of the goal $G$ in retinal coordinates. If the target and heading do not match, then a non-zero difference vector generates motor commands that adjust the navigator’s translation $T_x$ towards the target.
\[ T_x(t+1) = T_x(t) + \epsilon (G - H_{MAP}) \]  \hspace{1cm} (25)

where \( \epsilon \) is a gain factor set to 0.02. Translational commands were limited to \(-45^\circ < T_x < 45^\circ\).

Obstacles are ignored until one becomes a threat by registering a short time-to-collision as indicated by cell activity in the depth map, defined by Equation 18, greater than a threshold value \( \beta \). A value of \( \beta = 3.5 \) was used in the simulations reported here. The center of mass \( C_x \) of these suprathreshold cells in the depth map is calculated, and the heading is altered to move away from this center according to the following algorithm:

\[
\text{if } (H_{MAP} > C_x) \quad \text{then } T_x(t+1) = T_x(t) + \delta \\
\text{if } (H_{MAP} < C_x) \quad \text{then } T_x(t+1) = T_x(t) - \delta
\]  \hspace{1cm} (26)

where \( \delta \) is a small positive constant set to 0.2 in these simulations. The navigator veers away from the looming obstacle until it is no longer a threat. Once clear, the approach signal regains dominance, and the navigator resumes progress towards the goal.

4. Simulations

4.1. Training the model

A major advantage of this model over previous models for heading perception in mobile robots is its ability to self-organize through an action-perception cycle rather than requiring teaching signals and supervised learning. The model is trained by randomly generating rotational and/or translational movements, then utilizing the combination of eye velocity information and the flow pattern resulting from the movements to tune the parameters in the translation field and the heading map. These learning processes are detailed in the following paragraphs. Throughout training and testing, optic flow information was corrupted by varying amounts of three types of noise:

**directional noise.** Each flow vector is perturbed by a uniform randomly distributed angle between +/- the amount of directional noise.

**speed noise.** Each flow vector is multiplied by a uniform randomly distributed number between 0 and 2.

**aperture noise.** Each flow vector is perturbed by a uniform randomly distributed angle between +/- the amount of aperture noise, then the magnitude of the flow signal is reduced by the cosine of the angular difference between the original and perturbed vectors. Aper-
ture noise attempts to model the uncertainty of using local flow detectors to measure the motion of a luminance edge.

4.2. Translation field

The weights projecting to the translation field from the eye velocity cells learn to cancel the rotational optic flow generated by eye movements. The easiest way for this to happen is simply to generate random eye motions without any translational movement (imagine an infant scanning around a room before it can crawl). Under these circumstances, a fast learning rate may be used and the system is completely trained after only 20-30 random eye movements. Figure 4(A) shows the weights projecting from each eye velocity vector cell to the floret of cells at one retinal location after training with purely rotational movements of the eyes. To more clearly illustrate the pattern of weights projecting from each eye velocity vector cell, 12 directional cells were used in each floret of the translation field in this simulation.

If necessary, the weights may also be trained in the presence of translational movements. Random translations have an effect similar to noise on the desired training signal. In an actual implementation, noise may also result from limitations of sensor arrays such as those due to the aperture problem (e.g., Wallach, 1976; Perrone, 1992), although this problem may be reduced in vivo by motion preprocessing stages that convert aperture ambiguities into unambiguous and coherent segmentations of moving objects (Chey et al., 1995). A slower learning rate is required for stable learning with noise. Still, the system can learn to cancel the effects of eye rotations on the flow field in the presence of noise with a relatively small number of movements, requiring on the order of 500 randomly generated eye movements to tune the parameters. Such a simulation is summarized in Figure 4(B), which shows how the error decreases as training proceeds in noiseless case and with +/- 45 degrees of random (uniformly distributed) directional noise added to each cell's input at the motion detector field.

4.3. Heading map

Like learning at the translation field, learning at the heading map is carried out during an action-perception cycle. Here, however, the goal is to train the model to use the translational component of the flow field to determine heading direction. This is done by randomly generating translational movements of the eye (as if it were mounted on a moving body), then using the resulting translational flow field to self-organize a map representation of heading direction.

As noted in Section 3.1, the heading map is a variant of a SOFM. During learning, the cells in the heading map spread out to code different heading directions. The angular separation of neighboring heading cells will depend upon the number of cells in the map and the statistical distribution of heading directions sampled during learning (Kohonen, 1982). This provides the map with the desirable feature of efficiently allocating its resources toward more commonly experienced input patterns. For example, when the model is
FIGURE 4. Learning at the translation field. (A) Inhibitory weights $w_{jki}$ in Equation 8 projecting from eye velocity signals after training with purely rotational eye movements. Each of the four curves on the plot indicate the weight values projecting from one of the four eye velocity vector cells to all of the floret cells at one retinal location. An eye rotation corresponding to the activation of a single eye velocity vector cell results in a flow pattern that takes a cosine shape across each floret of cells (see Equation 5). The cosine shapes of these curves thus indicate that the weights have successfully learned to cancel the flow field components due to eye rotations as reflected at the eye velocity vector. (B) Noise tolerance while learning eye velocity parameters without noise (left) and with $\pm 45^\circ$ degrees of random directional noise (uniformly distributed) added to the motion detector field input. Percent error is measured as the amount of residual activity at the translation field during a rotational movement divided by the amount of activity that would occur without rotational nulling.
trained on a body that spends most of its time moving forward (as would be expected for most animals), it develops an accuracy bias as illustrated in Figure 5(A). The heading map on the left is the result of training on a set of 2000 movements with headings distributed uniformly between +/- 45° in both azimuth and elevation. The map on the right was trained on a set of headings biased toward small deviations from straight ahead. The tightly grouped heading cells in the center of the map (corresponding to straight ahead movement) allows the map to code these directions more accurately than more peripheral directions. In this regard, Crowell and Banks (1993) noted that people are more accurate at judging headings with small eccentricities (forward) than those with large eccentricities (sideways), and similar degradation of heading detection as the FOE moves away from the fovea was noted by Warren and Kurtz (1992). When trained with a distribution of headings as shown in the right side of Figure 5(A), the model develops a similar accuracy bias. Other researchers have attributed this accuracy bias to over-representation of outward flow vectors in visual cortex area MT in monkey and corresponding areas in the cat cortex (Lappe and Rauschecker, 1995).

As discussed above for learning in the translation field, it is important for the heading map to be tolerant of noise in the direction and speed of local optic flow signals. Figure 5(B) shows that the heading map is still able to topographically organize when trained with noisy optic flow information. In these “aperture noise” simulations, the effects of the aperture problem were simulated by randomly perturbing each perceived flow vector by an angle uniformly distributed between +/- 0°, 40°, and 90° and then reducing the magnitude of each flow vector by a factor equal to the cosine of the perturbed angle minus the actual angle. 2000 randomly generated movements with headings between +/- 45° for azimuth and elevation were used to train the network.

Figure 6(A) shows heading estimate accuracy under various kinds and amounts of noise in the optic flow signal. To allow comparisons with the simulations of Hatsopoulos and Warren (1990), training was carried out on headings between +/- 25° in azimuth and elevation, and error was averaged over test headings between +/- 20°. The network’s performance is comparable to the model of Hatsopoulos and Warren. This can be seen from the directional noise simulations, which indicate that the network performs with about 0.75° average error with no directional noise and 3.9° average error with +/- 90° directional noise; Hatsopoulos and Warren (1990) reported average errors of approximately 0.8° for no noise and 3.4° for +/- 90° noise. As these authors point out, these directional noise results are consistent with psychophysical results showing that humans perform slightly worse than the network with noisy flow fields. The current model’s performance under aperture noise (bottom of Figure 6(A)) is similar to the directional noise results. Finally, a simulation using “speed noise” (i.e., an increase or decrease of the magnitude of the flow vector) is reported in Figure 6(A). Like the network of Hatsopoulos and Warren (1990), speed noise has little effect on the network’s performance, again in keeping with psychophysical results showing that speed noise has little effect on the ability of
FIGURE 5. Unfolding of the heading map. (A) Example heading maps with uniform heading sampling (left) and sampling biased toward straight ahead (right). When the distribution of training samples is biased toward straight ahead, the distribution of map cells is more concentrated for movements near straight ahead, resulting in more accurate heading estimates for these movement directions. (B) Heading map after 2000 training steps for three levels of simulated aperture error in the motion detection field.

humans to determine heading. Thus, the current network shows the high degree of noise tolerance achieved by the model of Hatsopoulos and Warren (1990).

Some comments should be made regarding these comparisons. First, the current network achieves comparable results despite using a self-organizing learning scheme as compared to the supervised learning scheme of Hatsopoulos and Warren. Second, the current network works in the presence of eye rotations due to the removal of rotational flow components at the translation field; the Hatsopoulos and Warren does not work in the presence of eye rotations. Although the results reported here use a slightly larger retina (7x7 vs. 5x5)
Navigation Using Optic Flow

and more cells for heading detection (49 vs. 25), the number of input patterns required to adequately train the network is less in the current model than in the Hatsopoulos and Warren (1990) network (2000 vs. 4000). The fundamental difference is that the current network can learn from randomly generated headings whereas the Hatsopoulos and Warren model requires a teacher to explicitly train the network using optic flows that correspond to known headings.

Figure 6(B) shows a more graphical representation of heading estimation performance under two different amounts of aperture noise, this time for a network trained with headings between +/- 45°. The ‘o’ marks denote an arbitrarily selected grid of sample headings. The ‘+’ marks plot the responses of the heading map.

4.4. Depth map

Figure 7 shows the effect of directional flow noise on relative depth estimation. The optic flow field used in these simulations had 49 cells arranged in a 7x7 grid. A random depth was selected for objects at each of the 49 retinal locations. Then, movements were made towards those objects along 36 randomly chosen headings between +/- 37° in both azimuth and elevation. The first three panels of Figure 7 compare the depth measure calculated by the model to the actual depth of the object at each retinal location. (For purposes of comparison, actual depth was processed according to Equation 20 and scaled to compensate for translational speed.) For each of these simulations, noise was present during both training and performance. It should be noted that the effects of noise during training are quite small relative to the effects of noise during performance; i.e., most of the error in the plots is attributable to noise during performance rather than incorrect values of model parameters learned during training. The final plot in Figure 7 shows error vs. the amount of directional noise. The model was trained in the absence of noise for this plot. Error was calculated as the difference between the actual object depth and the depth predicted by the network divided by the actual object depth. This was multiplied by 100 to obtain percent error, and the result was averaged over all retinal locations and 36 different headings. As indicated in Figure 7, the depth map is unusually robust to noise because the optic flow signals were perturbed in direction but not magnitude. Recall that for a given heading, object depth depends only on the magnitude of the optic flow signal. Therefore, as long as the heading estimate is reasonably correct (see Figure 6), the depth estimates will also be correct regardless of directional noise.

If any speed noise is present in the flow field, depth estimation will unavoidably degrade. This is evident from Figure 7, which shows the effect of aperture noise on relative depth estimation. Since aperture noise contains a speed component in addition to the directional component, depth map accuracy degrades significantly more than with directional noise. Nonetheless, the depth map still performs well with 40° aperture noise, and later simulations will show that the 100% average error in depth estimates under 90° aperture noise is still adequate for navigation because multiple snapshots of the scene, taken as the navigator moves, tend to “average out” the unwanted effects of noise over time. Performance could be further improved if predictions about the continuity of surfaces in the field of view were made.
FIGURE 6. Heading estimation under various noise conditions. (A) Error in heading judgment plotted as a function of the amount of directional noise, speed noise, and aperture noise in the flow field for a network trained with headings between +/- 25°. (B) Example of heading error for 0° noise and 40° aperture noise for a network trained with headings between +/- 45° (top). The circles mark the actual heading direction and the + signs mark the estimated heading direction.

view were used to perform neighborhood smoothing operations on the output of the depth
FIGURE 7. Effect of different levels of directional noise on relative depth estimates. Each of the 49 cells on the 7x7 retina estimates the relative depth of the object in its receptive field. The first three plots show network depth estimates (solid lines) compared to actual relative depths (dotted lines) for three different levels of directional noise presented during both training and performance. The final plot shows the average percent error of depth estimates as a function of directional noise for a network trained in the absence of noise.

map, but this topic is beyond the scope of the present article. It should also be noted that the worst depth estimation errors occur for objects near the FOE, where optic flow signals tend to be very small. This is another inherent problem with using optic flow for depth estimation. One technique that can help alleviate this problem for real-world scenes is to ignore retinal locations with very small depth weights and fill in depth estimates by averaging estimates from neighboring cells.

4.5. Independent motion map

Figures 9 and 10 illustrate the performance of the independent motion map. The scene consists of a tiled ground plane, a distant wall, and a box floating above the ground plane between the observer and the wall. Each panel shows the visual field as seen by the network's retinal array. Objects in the visual field are indicated by dotted lines. Arrows in the
FIGURE 8. Effect of different levels of aperture noise on relative depth estimates. Each of the 49 cells on the 7x7 retina estimates the relative depth of the object in its receptive field. The first three plots show network depth estimates (solid lines) compared to actual relative depths (dotted lines) for three different levels of aperture noise presented during both training and performance. The final plot shows the average percent error of depth estimates as a function of aperture noise for a network trained in the absence of noise.

left panels indicate flow components in the translation field, and arrows in the right panels indicate flow components at the independent motion map after thresholding according to Equation 23.

Figure 9 indicates performance at the independent motion map in the absence of noise. In Figure 9(A), the navigator moves forward while the box remains stationary. Since optic flow in the scene is commensurate with the flow expected by the activated heading map, all activity in the independent motion map is quenched. In Figure 9(B), the navigator again moves forward, but this time the box is moving independently to the left. The flow generated by the box does not fit with the flow pattern established by the rest of the scene, so it is singled out in the independent motion map. Figure 9(C) shows an example of detecting the same box motion while the navigator moves at 37° to the right instead of
straight ahead. One might note that the direction coded by the active cells in the independent motion map only roughly corresponds to the actual motion direction of the object with respect to the navigator. This highlights the fact that this map is primarily useful for identifying the retinal locations of moving objects, rather than their direction of movement relative to the navigator. The directional error arises because the motion VAM detects mismatches between expected direction and the incoming flow direction. The difference of these two directions will point approximately in the actual direction of the moving object. However, the exact calculation of direction requires knowledge of both the depth of the moving object and the navigator’s speed so that the component of optic flow due to object motion may be separated from the component due to self motion.

Figure 10 illustrates independent motion map performance under various noise conditions. Figures 10(A) and (B) illustrate the performance of the independent motion detector in the presence of +/- 40° and +/- 90° aperture noise, respectively. Noise was present during both training and performance. The network successfully detects the moving box in both conditions, but the network erroneously detects motion at several retinal locations in the +/- 90° condition. The independent motion detector is the most sensitive part of the network to directional noise since it relies on the accuracy of optic flow directions at each retinal location and cannot average over the entire retina. Directional noise can be countered to some degree by choosing a higher value of the threshold A in Equation 23, but raising the threshold also increases the chances of missing a moving object. Figure 10(C) shows the performance of the independent motion map in the presence of 100% speed noise. Because both the heading and independent motion maps are essentially unaffected by perturbations in the magnitude of optic flow signals, the moving box is easily detected in the speed noise condition. It should also be noted that faster moving objects perturb the optic flow signals more than slower moving objects and are therefore easier to detect.

4.6. Navigation

The simulations shown in Figure 12 demonstrate the utility of the self-organizing heading and depth maps for visual navigation. Figure 12(A) shows an overhead view of a field of rectangular obstacles and the path that the navigator takes to reach the goal indicated by the ‘+’ sign. The task for the navigator is to start at the bottom center and reach the ‘+’ sign at the top of the scene without hitting any obstacles. No high level path planning was used, only the simple approach-avoidance algorithm described in Section 3. The ‘o’ symbols mark places along the path where the navigator was in approach mode, and the ‘x’ symbols mark the places where the navigator was in avoidance mode.

Because the components of the network are robust to noisy optic flow fields, the approach-avoidance calculation based on the network’s output is also robust to noise. Figure 12(B) shows two example paths taken by the navigator in the presence of 90° aperture noise. The main difference between performance in 0° noise and 90° noise is that at higher noise levels, the navigator occasionally misjudges steps and clips the corners of obstacles. Also, random depth map errors occasionally cause the navigator to veer unnecessarily. Figure 12(C) shows two example paths taken by the navigator in the presence of 100%
FIGURE 9. Independent motion detection with no noise. Each panel shows the visual field as seen by the network's retinal array. Arrows in the left panels indicate flow components in the translation field, and arrows in the right panels indicate flow components at the independent motion map after thresholding. (A) The navigator is moving straight ahead over a tiled ground plane. The suspended box is stationary with respect to the ground plane, so no activity remains in the independent motion map. (B) The navigator is moving straight ahead over a tiled ground plane. The suspended box is moving to the left. The flow field at the retinal location that corresponds to the box does not match the expected field, so activity at that location pops out in the independent motion field. (C) Here the navigator is moving at 37° to the right while the box still moves to the left, indicating that independent motion may be detected for a wide range of navigator headings.
FIGURE 10. Independent motion detection under various noise conditions. (A) Example of independent motion map performance when the incoming flow field is randomly perturbed by +/- 40° aperture noise. (B) At +/- 90° aperture noise, some errant vectors survive the threshold along with the moving box. (C) Performance of the independent motion map in the presence of 100% speed noise. Independent motion detection is essentially unaffected by perturbations in the magnitude of the flow signals.
FIGURE 11. (A) Two example paths taken by the approach-avoidance algorithm, viewed from above. The circles represent times when the approach signal was dominant. The x's represent times when the avoidance signal was dominant. The navigator starts at the bottom and attempts to reach the + sign while avoiding the blocks. (B) Two example paths taken by the approach-avoidance algorithm in the presence of 90° aperture noise. (C) Two example paths taken by the approach-avoidance algorithm in the presence of 100% speed noise.

speed noise. 100% speed noise is guaranteed to degrade the performance of the depth map by at least that amount, and one can see by the x's that the navigator is sometimes avoid-
ing ghosts. However, on average the ghosts tend to cancel each other, and the network extracts useful information about the true location of the obstacles.

Finally, Figure 12 shows examples of motion pursuit simulations that require all elements of the visual navigation network. One navigator, designated by the ' + ' symbol, attempts to reach a goal indicated by a '-' at the top of the picture. A second navigator, designated by the ' o ' symbol, is slightly faster and uses the first navigator's changing location as its goal. The ' o ' navigator uses its independent motion map to determine the location of the '+' navigator. As shown in the two simulations of Figure 12, the ' o ' navigator is typically successful in its attempt to overtake the slower navigator.

![Figure 12. Pursuit behavior viewed from above. The navigator designated by the ' o ' symbols is pursuing the navigator designated by the ' + ' symbols. The '+' navigator is slightly slower but gets a head start in its attempt to reach the goal. The 'o' navigator is faster and attempts to capture the '+' navigator before it reaches its goal. The 'o' navigator detects the location of the '+' navigator using its independent motion map. In both simulations, the 'o' navigator is able to successfully pursue and overtake the slower '+' navigator before it reaches its goal.](image)

5. Concluding Remarks

The network described in this article was developed primarily as a module for autonomous robot navigation. Its features include self-organization, fast learning, noise tolerance, operation in the presence of eye movements, and reasonable memory and computational demands. The structure and learning principles were inspired by several previous computational and neural models. Some of these models were at least partly designed to explain single cell recording data from areas in the middle temporal (MT) and medial superior temporal (MST) regions of primate visual cortex (e.g., Hatsopoulos and Warren, 1990; Lappe and Rauschecker, 1993, 1995; Perrone, 1992). Although this was not the primary goal of the current model, many of its cell types also show similarities to MT and MST cell properties. Cells in MT have been shown to be sensitive to the orientation and velocity of visual stimuli (Rodman and Albright 1987), as are cells in the motion...
Navigation Using Optic Flow

detection, translation, and normalization fields in the model. Many cells in MST respond maximally to radially expanding patterns and patterns that include expansion and full field translation, suggesting a role in heading detection and/or depth estimation (e.g., Lagae et al., 1994); cells in the heading map are similarly tuned. Finally, MST receives strong fiber projections from MT (e.g., Maunsell and Van Essen, 1983; Ungerleider and Desimone, 1986). Heading map cells in the model similarly receive projections from MT-like cells in the normalization field. It should be noted, however, that cells in MT and MST have complex properties that are by no means completely explained by the current model or other models of heading detection, including sensitivity to rotational patterns, spiraling patterns, and other complex stimuli (e.g., Lagae et al., 1994; Graziano, Andersen, and Snowden, 1994).

The translation and normalization fields of the model also share properties with MT cells but the accuracy bias shown in Figure 5(A) differs from this interpretation in two respects. First, the current model's bias results from tuning anisotropies in cells related more closely to area MST cells than to MT cells. Second, the bias in the current model is the result of visual experience. Visual experience does not appear to play a role in the anisotropies seen in neurophysiological data, at least in the cat equivalent of area MT (Brenner and Rauschecker, 1990). The role of visual experience in shaping MST cell properties has not been fully investigated. For the main purposes of this article, however, the important thing to note is that a potentially useful accuracy bias toward commonly used headings arises during the learning process.

To the extent that the present model does capture brain heuristics, the use of eye velocity information to nullify the rotational component of the flow field suggests that MST cells may change their flow field sensitivities in the presence of eye movements. Specifically, a cell that fires maximally to a radially expanding flow field in the absence of eye movements should not respond well to the same flow field if it is generated by an eye movement combined with an appropriately designed flow stimulus. The model predicts this because the eye movement in the latter case would result in removal of the rotational component of the flow field, resulting in a flow field at the translation and normalization fields that is no longer radially expanding. Another way to think of this is that the same flow pattern at the retina corresponds to two different heading directions in the two cases, thus maximally activating two different cells in the heading map. An experiment investigating this might help clarify the role of MST cells and eye velocity information in primate heading detection.

6. References


Navigation Using Optic Flow


December 22, 1995
Navigation Using Optic Flow


Navigation Using Optic Flow


