Pergamon 0042-6989(94)E0086-Z

*Vision Res.* Vol. 34, No. 21, pp. 2917 2938, 1994
Elsevier Science Ltd. Printed in Great Britain

# A Model of Self-motion Estimation Within Primate Extrastriate Visual Cortex

JOHN A. PERRONE,*†‡ LELAND S. STONE§

*Received 22 September 1993; in revised form 28 February 1994*

**Perrone [(1992) *Journal of the Optical Society of America A*, *9*, 177–194] recently proposed a template-based model of self-motion estimation which uses direction- and speed-tuned input sensors similar to neurons in area MT of primate visual cortex. Such an approach would generally require an unrealistically large number of templates (five continuous dimensions). However, because primates, including humans, have a number of oculomotor mechanisms which stabilize gaze during locomotion, we can greatly reduce the number of templates required (two continuous dimensions and one compressed and bounded dimension). We therefore refined the model to deal with the gaze-stabilization case and extended it to extract heading and relative depth simultaneously. The new model is consistent with previous human psychophysics and has the emergent property that its output detectors have similar response properties to neurons in area MST.**

Optic flow Heading perception Depth perception Stability of gaze MT MST Vestibular-ocular reflex

## I. INTRODUCTION

The task of navigating through the environment involves a complex, coordinated sensorimotor process that uses visual, vestibular, proprioceptive, motor-corollary, and cognitive input. Determining one's own movement (self-motion or egomotion estimation) is a critical part of that task. The visual component of self-motion estimation has received particular attention (for reviews see Cutting, 1986; Heeger & Jepson, 1992; Warren, Morris & Kalish, 1988) because human self-motion perception, both rotational and translational, appears dominated by vision (for a review see Henn, Cohen & Young, 1980) as manifested by its ability to generate a compelling sense of self-motion without actual movement in space (vection).

The problem of visual self-motion perception was clearly outlined by Gibson (1950, 1966) who examined those two-dimensional (2-D) visual stimulus properties that could provide information about self-motion (see also Calvert, 1954; Llewellyn, 1971; Johnston, White & Cumming, 1973; Lee, 1974; Nakayama & Loomis, 1974; Koenderink & van Doorn, 1975; Warren, 1976; Regan & Beverly, 1979; Longuet-Higgins & Prazdny, 1980; Zacharias, Caglayan & Sinacori, 1985; Cutting, 1986). The strictest form of this problem limits the input to a single instant of 2-D image motion referred to as the "flow field" and no allowance is made for integration over time or for other three-dimensional (3-D) sources of information such as disparity, vergence, accommodation, shading, or perspective, despite the fact that such sources most likely play an important role in navigation. Solving the self-motion problem then reduces to using the flow field to recover one's instantaneous rotation and translation while moving through an unrestricted environment. After Gibson's initial investigations (Gibson, Olum & Rosenblatt, 1955), many attempts were made to show mathematically how 3-D self-motion parameters could be recovered from the 2-D flow field. This is a difficult non-linear problem in the general case of all possible observer translations and rotations with no constraints on the environmental layout (i.e. non-planar and discontinuous surfaces allowed) using only a single flow field for input (i.e. no extended sequences). Although algorithms were eventually developed to solve the general problem of extracting unrestricted 3-D self-motion from 2-D flow (e.g. Longuet-Higgins, 1981; Rieger & Lawton, 1985; Heeger & Jepson, 1990), their appropriateness as models of human self-motion processing remains questionable (see however, Lappe & Rauschecker, 1993).

Recent psychophysical experiments (Cutting, Springer, Braren & Johnson, 1992; De Bruyn & Orban,

*Aerospace Human Factors, NASA Ames Research Center, Moffet Field, CA 94035, U.S.A.
†Psychology Department, Stanford University, Stanford, CA 94305, U.S.A.
‡To whom all correspondence should be addressed at: Psychology Department, University of Waikato, Private Bag 3105, Hamilton, New Zealand.
§Life Science Divisions, NASA Ames Research Center, Moffet Field, CA 94035, U.S.A.

1990; Perrone & Stone, 1991; Stone & Perrone, 1991, 1993; Warren & Hannon, 1990) have systematically ruled out most algorithms previously proposed for human visual self-motion estimation. Furthermore, most previously proposed algorithms assume the existence of a precise 2-D velocity vector field (or, equivalently, a displacement field) as the input to the 3-D stage. Vector operations are then performed to decompose the flow field into its translational and rotational components. Motion-sensitive neurons within primate visual cortex could be used to derive the flow field, but there is no clear evidence if or how this is done. Alternatively, given that some neurons within primate extrastriate cortex appear broadly tuned to local image speed and direction (for a review see Maunsell & Newsome, 1987), Perrone (1992) proposed a mechanism by which the output of such neurons could be used directly as the input to a network designed to estimate self-motion. This obviates the need for an additional stage in which the flow field is explicitly extracted.

In this paper, we present, test, and discuss a new, neurally-based model of human self-motion estimation. In Section II, we summarize the template approach for heading estimation originally described elsewhere (Perrone, 1992). In Section III, we discuss the different conditions under which rotation can be introduced into the flow field including gaze stabilization. In Section IV, we motivate and justify our reasons for focusing on the gaze-stabilization case. In Sections V and VI, we discuss sampling issues. In Section VII, we present the new model in detail, including a new depth-extraction feature. In Section VIII, we show examples of the model's performance in both heading (direction of translation) and range estimation (relative depths of points within the environment). In Section IX, we compare the response properties of the output components of the model (detectors) with the known physiological properties of neurons within primate extrastriate visual cortex. In Section X, we compare the model's performance in heading estimation with previous human psychophysical results. In Section XI, we discuss the significance of the new model and outline areas appropriate for future research. Preliminary presentations of the model have appeared elsewhere (Perrone & Stone, 1992a, b).

## II. TEMPLATE APPROACH

A model which is not based on the assumption that the human visual system has access to an explicit 2-D flow field has recently been proposed (Perrone, 1992). This model (Fig. 1) uses direction- and speed-tuned sensors similar to neurons found in the Middle Temporal (MT or V5) area of the primate visual cortex (Zeki, 1980; Maunsell & Van Essen, 1983b; Albright, 1984) to solve the general self-motion problem by setting up maps of detectors or templates that mimic neurons in the Middle Superior Temporal area (MST) (Kawano, Sasaki & Yamashita, 1984; Kawano & Sasaki, 1984; Saito, Yukie, Tanaka, Hikosaka, Fukada & Iwai, 1986; Tanaka, Hikosaka, Saito, Yukie, Fukada & Iwai, 1986; Tanaka & Saito, 1989; Tanaka, Fukada & Saito, 1989; Duffy & Wurtz, 1991a, b; Orban, Lagae, Verri, Raiguel, Xiao, Maes & Torre, 1992). Briefly, image motion is first processed by MT-like sensors which respond to the local image motion according to the

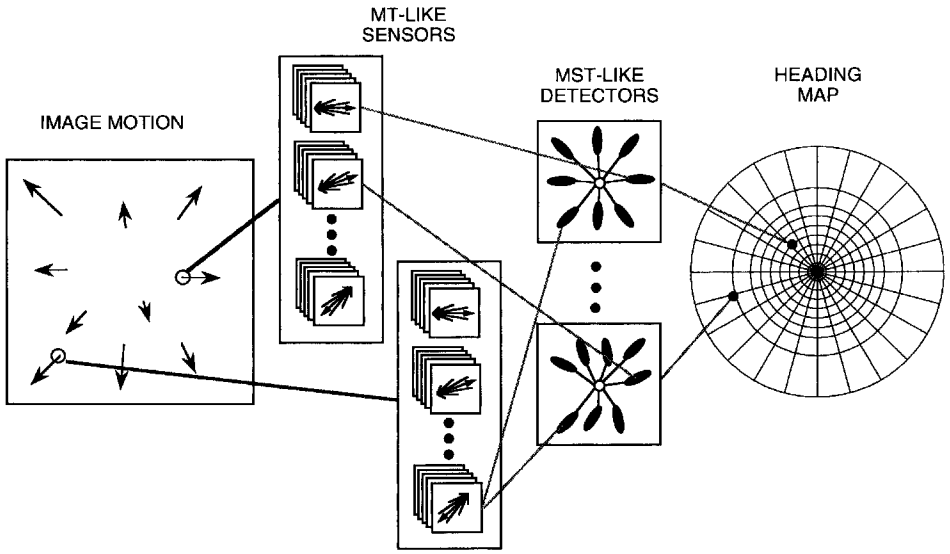


FIGURE 1. Overall structure of the template model. Image motion is analyzed using sets of speed- and direction-tuned MT-like motion sensors tiling the entire visual field. The output of specific sets of these sensors are then summed over a wide portion of the field by an MST-like detector. Because of the specificity of its MT inputs, the detectors are each "tuned" for a particular heading. Heading maps containing arrays of detectors are used to sample heading space. The detector with the largest output within all of the maps identifies heading. For clarity, only a small subset of the connections are shown.

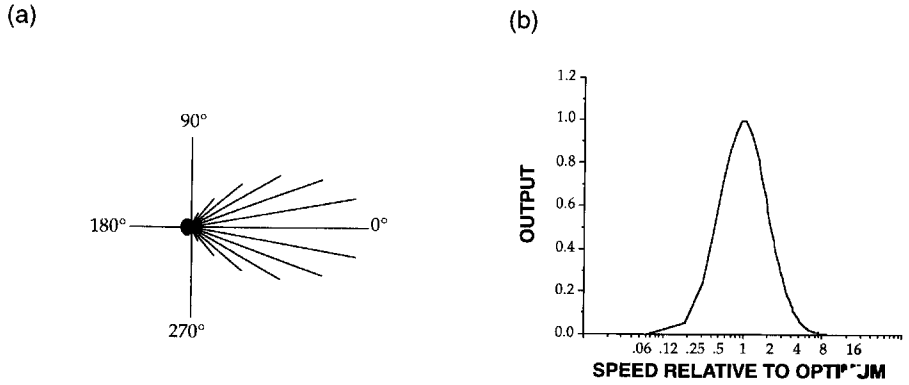


FIGURE 2. Idealized MT neuron responses. (a) Direction tuning curve in polar plot form. The curve is based on a Gaussian function with SD = 30 deg [see Perrone, 1992, Fig. 2(a)]. This can be compared to the typical direction tuning curves found from single-unit recordings (e.g. Albright, 1984, Fig. 9.). (b) Speed-tuning curve. This is also based on a Gaussian with SD = 1 octave. The horizontal axis is plotted on a $\log_2$ scale and represents the ratio of the image speed to the optimum speed for the sensor. Compare to Fig. 6(b) of Maunsell and Van Essen (1983b).

product of their direction and speed responses.* It is assumed that multiple speeds and directions are represented at each location in the visual field and that direction- and speed-tuning can be approximated by Gaussian curves (Fig. 2). In Fig. 1, each MST-like detector is connected to a number of MT-like sensors at each location, each tuned to a specific direction and speed. For clarity, only one MT input from each of two locations is shown for the two different detectors. Note that different detectors can share the same MT input and that MT-sensors at all locations process information simultaneously for all of the detectors that they feed. This shared parallel architecture provides for extremely efficient information processing. Each detector integrates motion information across a large portion of the visual field by summing the activities of the most active sensor-input at each location. A biologically plausible way to simulate this winner-take-all strategy would be to have the inputs from each sensor sum linearly but also mutually inhibit each other (standard lateral inhibition). (The flower-like icons symbolizing detectors are explained in detail in Fig. 6.) The detector acts as a template for a specific global pattern of image motion and responds optimally to (i.e. detects) a particular instance of observer motion (translation plus rotation). Orderly arrays or maps of such detectors are then set up because we need different detectors within each map for each of the possible headings and a different map for each of a set of possible rotations (not shown in Fig. 1). Heading is estimated by identifying the most active detector within all of the maps.

This approach is both biologically plausible (i.e. input and output neurons consistent with the general receptive field properties of MT and MST neurons, and the maps are consistent with the general idea of cortical maps) and psychophysically plausible (i.e. consistent with previous data on human performance in heading estimation). While many researchers have proposed "looming-detectors" of various forms which respond to the radial expansion patterns that often occur during self-motion (Koenderink & van Doorn, 1975; Regan & Beverley, 1978; Saito, *et al.*, 1986; Perrone, 1987, 1990; Albright, 1989; Tanaka *et al.*, 1989; Glünder, 1990; Hatsopoulos & Warren, 1991; Verri, Straforini & Torre, 1992), such translation detectors cannot solve the self-motion estimation problem alone and do not work in the presence of rotation. More complex detectors must be used and the processing of rotation must be considered as it is regularly encountered during normal locomotion.

Although Perrone (1992) demonstrated the feasibility of the template approach, one main objection to applying this approach to human self-motion estimation could be raised: an enormous number of detectors are necessary to encode every possible situation. In order to make this approach more plausible, the number of templates must be minimized. A straightforward way would be to handle only those types of image motion that can reasonably be expected to occur during human locomotion. We have therefore refined the original model to deal only with the most common situation, forward translation with image rotation limited to that produced by fixation of a stationary point in the scene (the gaze-stabilization case). This refinement restricts the possible rotations that can be experienced and therefore greatly reduces the number of maps needed to support accurate heading estimation in most realistic circumstances.

## III. DIFFERENT TYPES OF ROTATION

Figure 3 shows two different types of flow fields in which rotation has been added to a pure translation field. In each case, the translation field is radial,

*It is assumed that, unlike V1 neurons, MT neurons are truly speed- and direction-tuned largely indedendent of spatial frequency. There is some evidence that this is true, at least for a subset of MT neurons (Movshon, Newsome, Gizzi & Levitt, 1988).

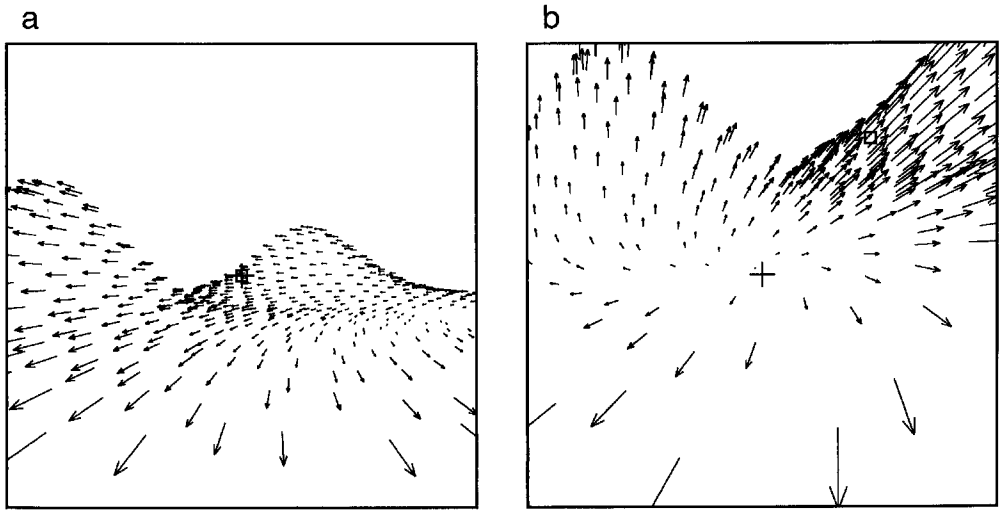


FIGURE 3. Two types of image motion patterns (flow fields) produced by different translation/rotation situations. (a) An example of the type of motion that results when the observer rotates to the right about a vertical axis during an instant of translation towards the small square. (b) This flow field simulates the situation in which an observer is tracking a fixed point on the ground (cross) during the translation towards the square.

expanding out from the position of the square. If rotation were absent the square would indicate the location of the focus of expansion (FOE) and therefore heading (Gibson, 1950). However the rotation results in a complex pattern of image motion from which the heading must be discerned. Note that these patterns occur retinally and are not necessarily perceived in this form. The situation shown in Fig. 3(a) can occur if the observer tracks a moving point in the environment during translation (e.g. a bird flying past) or moves along a curved path and fixates an object moving with him (e.g. the car in front). In both of these cases, any arbitrary combination of translation and rotation can occur as they are produced independently. These cases are however not the most common. In most cases, gaze will be stabilized on a stationary point in the environment. A typical flow field produced by forward translation with gaze stabilization is shown in Fig. 3(b).

Primates possess a number of eye-movement mechanisms that serve to stabilize gaze during self-motion. In addition to the classical rotational vestibulo-ocular reflex (VOR) (for reviews see Wilson & Mellville-Jones, 1979; Robinson, 1981; Miles & Lisberger, 1981; Leigh & Brandt, 1993), a linear VOR provides stabilization of gaze during translation (Buizza, Leger, Droulez, Berthoz & Schmid, 1980; Smith, 1985; Baloh, Beykirch, Honrubia & Yee, 1988; Paige, 1989; Paige & Tomko, 1991a, b; Israël & Berthoz, 1989; Schwarz, Busettini & Miles, 1989; Schwarz & Miles, 1991). Visually-driven reflexive eye movements (ocular following) that would serve to stabilize gaze during translation have also been described (Miles, Kawano & Optican, 1986; Gellman, Carl & Miles, 1990; Busettini, Miles & Schwarz, 1991). Finally, in addition to these reflexes, voluntary smooth-pursuit eye-movements can also be used to assist fixation of a stationary object during locomotion (for reviews see Lisberger, Morris & Tyschen, 1987; Stone & Lisberger, 1989; Keller & Heinen, 1991).

The combination of these three oculomotor pathways would presumably be quite effective in keeping the image of a stationary point stabilized on the fovea [the cross in Fig. 3(b)] particularly since postural strategies appear to minimize head motion during locomotion, at least to within the working range of the vestibular-ocular reflex (VOR) (Grossman, Leigh, Abel, Lanska & Thurston, 1988; Pozzo, Berthoz & Lefort, 1990) thereby providing an additional tier of control for gaze stabilization. The saccadic eye-movement system (for reviews see Robinson, 1981; Fuchs, Kaneko & Scudder, 1985; Sparks & Mays, 1990) provides a mechanism for jumping from one object of interest to another in a ballistic manner but, during the intervening fixations, foveal image motion is kept low presumably to allow for accurate visual acuity (Westheimer & McKee, 1975; Murphy, 1978; Barnes & Smith, 1981). Finally, the ocular following response is greatly enhanced immediately following saccades. This should serve to expedite the re-establishment of gaze stabilization as soon as possible after a saccade (Kawano & Miles, 1986; Gellman *et al.*, 1990).

Unless these gaze-stabilization mechanisms are consciously overridden, humans and monkeys will therefore generally stabilize their gaze during locomotion (Collewijn, 1977; Grossman, Leigh, Bruce, Huebner & Lanska, 1989; Solomon & Cohen, 1992; Leigh & Brandt, 1993). In fact, deficits in gaze stabilization are associated with impaired vision and oscillopsia during locomotion (Takahashi, Hoshikawa, Tsujita & Akiyama, 1988; Grossman & Leigh, 1990). In addition

to bringing the increased processing power of the fovea onto an object of interest and keeping the foveal image as stable as possible, such tracking can be shown to simplify self-motion estimation. Longuet-Higgins and Prazdny (1980) pointed out how such a strategy offers computational advantages and others (Bandopadhay, Chandra & Ballard, 1986; Sandini, Tagliasco & Tistarelli, 1986; Sandini & Tistarelli, 1990) have explicitly shown how the dimensionality of the problem is reduced.

The model originally proposed by Perrone (1992) was designed to deal with any possible combination of rotation and translation and did not differentiate between the two cases shown in Fig. 3(a, b). Because the type of flow field shown in Fig. 3(a) is experienced less often, while that shown in Fig. 3(b) is a common experience, we suggest that a special mechanism may have evolved to deal with the gaze-stabilization case. The original model can be significantly simplified when only flow fields induced during gaze stabilization need be processed. The next section will present these simplifications.

## IV. SIMPLIFICATIONS DUE TO GAZE STABILIZATION

General unconstrained self-motion has six degrees of freedom: three for observer speed and heading direction (azimuth and elevation) plus three for rotation (yaw, pitch, and roll). However, observer speed cannot be recovered from image flow alone (it trades off with the absolute depth of points in the environment), so the general visual self-motion estimation problem is actually five-dimensional.

Roll body motion (sway) is generally small (less than about 4 deg/sec peak) during human locomotion (Waters, Morris & Perry, 1973; Cappozzo, 1981) and is at least partially compensated for by ocular counter-rolling driven by both vestibular and visual inputs (Henn *et al.*, 1980). Although ocular counterrolling in response to static head tilt has a relatively low gain, recent studies have shown that in the frequency range of standard walking (fundamental around 1–3 Hz, see Waters *et al.*, 1973; Cappozzo, 1981; Grossman *et al.*, 1988), ocular counterrolling can have a gain as high as 0.7 (Vieville & Masse, 1987; Ferman, Collewijn, Jansen & Van den Berg, 1987; Peterka, 1992). Finally, head counterrolling may be used to augment the range of roll stabilization (Gresty & Bronstein, 1992). If roll is indeed neglected, the problem reduces to four dimensions.

During gaze stabilization, the direction of rotation is fixed because the rotation axis is constrained to be perpendicular to the plane defined by the fixation and heading directions. This fact provides the main advantage of restricting the problem to the gaze-stabilization case, as it reduces the problem to only three dimensions by fixing the yaw/pitch ratio for each possible heading.

Unfortunately, gaze stabilization does not constrain the rotation rate which is a function of the unknown fixation distance and observer speed. To illustrate this, for convenience but without loss of generality, we use an exocentric heading-centered coordinate frame to define the fixation point in 3-D space ($x, y, z$ with $z$ pointing in the direction of translation, i.e. $\dot{x} = \dot{y} = 0$). If we make the simplifying assumption that the observer is fixating parallel to the ground plane (i.e. $y = 0$), the vertical component of the motion is therefore zero and horizontal gaze speed is then given by:

$$\omega = \frac{\dot{z}x}{x^2 + z^2} \tag{1}$$

or, in polar coordinates,

$$\omega = \frac{V}{F} \sin \alpha. \tag{2}$$

Equation (1) gives the rate of gaze-rotation that will occur if a particular point $(x, 0, z)$ is tracked while the observer moves along the $z$-axis. Equation (2) emphasizes that gaze speed is a function of $\alpha$, the fixation (or gaze) angle with respect to heading, $V$, the speed of forward translation, and $F$, the fixation distance. Figure 4 illustrates this relationship graphically. A particular $(x, z)$ position in this space defines the point in the world that is being fixated. The solid lines represent the loci of fixation points for which the eye rotates at 0.5, 1, 2, and 4 deg/sec, respectively. These are equivalent to the "isoangular displacement contours" derived by Cutting (1986). This figure assumes a forward speed of 1 m/sec. If the observer moves at a different speed, the rotation rates associated with each contour scale linearly [see equation (2)].

In the general case ($y \neq 0$), the gaze-rotation rate is given by:

$$\omega = \frac{\dot{z}\sqrt{x^2 + y^2}}{x^2 + y^2 + z^2} \tag{3}$$

in Cartesian coordinates or by:

$$\omega = \frac{V}{F}\sqrt{\sin^2 \beta + \sin^2 \alpha \cos^2 \beta} \tag{4}$$

in spherical coordinates ($\alpha$ and $\beta$, azimuth and elevation of gaze with respect to heading). Note that equation (3)

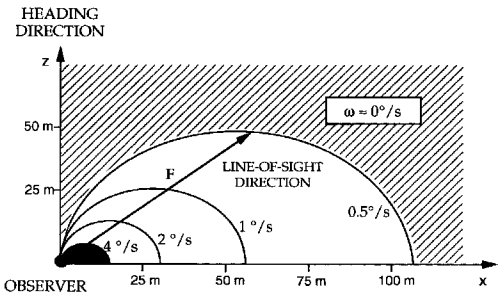


FIGURE 4. Iso-eye-rotation rate contours for points in the field at various fixation distances from the observer. This is a top down view showing only the field to the right of the observer and points lying in a horizontal plane located at eye-level.

reverts to equation (1) if $y = 0$ and equation (4) reverts to equation (2) if $\beta = 0$.

All other parameters being the same, each possible fixation distance results in a different gaze-rotation rate, which then generates a different pattern of image motion. At first glance, the space of possible fixation distances and, therefore, of flow fields seems infinite. However, a large part of the space consists of regions where rotation is negligible. For points in the striped region of Fig. 4, gaze stabilization produces $<0.5$ deg/sec of rotation and the observer would experience nearly pure translational flow. Conversely, rotation rates above 4 deg/sec are only found in the very compressed solid area as they require both close fixation and eccentric gaze, a situation not likely to be conducive to accurate navigation.

Given that rotation speed falls off inversely with fixation distance [equation (4)] it is reasonable to sample the possible range of fixation distances using logarithmic steps and a relatively small number of rotation speeds. For most of our simulations, we arbitrarily decided to generate just four maps, each corresponding to four possible gaze-rotation speeds (0, 1, 2 and 4 deg/sec). Although this constraint does not reduce the dimensionality of the problem, we capitalized on the fact that the range of possible rotation speeds is logarithmically compressed and effectively bounded.

## V. SAMPLING HEADING SPACE

We can also reduce the number of templates by sampling heading space judiciously. First, we can constrain the model to process points in the forward hemisphere only. Humans rarely experience backward locomotion and anomalous perceptual effects occur when they do (Perrone, 1986). Second, we need only sparsely sample the periphery. Psychophysics has shown that human heading estimation deteriorates rapidly as heading moves into the periphery (Crowell & Banks, 1993; Warren, 1976; Warren & Kurtz, 1992). Therefore, for most simulations, we used step sizes which produce a 5% difference in the peak activity of adjacent pure translation detectors (see below) in response to a given heading. We adopted a polar map for sampling the heading space: the 5% sensitivity criterion resulted in sampling the radial direction at 0, 3, 6, 9, 12, 15, 18, 21, 26, 36, 56, and 89 deg and the axial direction was arbitrarily sampled in equal steps of 15 deg. Although this arrangement creates a convenient circularly symmetric map of headings, the radial and axial values do not directly correspond to azimuth and elevation.

## VI. SAMPLING THE ENVIRONMENT

To select the appropriate sensors needed to provide input for the detectors tuned to particular patterns of image motion, one needs to specify quantitatively the image motion that can occur at each location in the visual field as a result of a specified observer motion through the environment. To do so, we start with the following standard equation (Longuet-Higgins & Prazdny, 1980):

$$\begin{pmatrix}\dot{X}\\ \dot{Y}\end{pmatrix}=\begin{pmatrix}-1/z & 0 & X/z\\ 0 & -1/z & Y/z\end{pmatrix}\begin{pmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{pmatrix} + \begin{pmatrix}XY & -(X^2+1) & Y\\ Y^2+1 & -XY & -X\end{pmatrix}\begin{pmatrix}\omega_x\\ \omega_y\\ \omega_z\end{pmatrix} \quad (5)$$

where $(\dot{X}, \dot{Y})$ is the expected retinal velocity caused by the combined rotation $(\omega_x, \omega_y, \omega_z)$ and translation $(\dot{x}, \dot{y}, \dot{z})$ of the observer with a stationary point at $(x, y, z)$ in the environment which projects onto the retina at position $(X, Y) = (x/z, y/z)$ (given a focal length of 1). In the simulations of the model, the image motion is sampled by assuming that a set of MT-like sensors exists at each image location occupied by an input flow vector. No attempt was made to simulate the topographic sampling of MT (Van Essen, Maunsell & Bixby, 1981). Unlike the previous equations, the coordinate system for equation (5) is retinotopic, i.e. egocentric with the $z$-axis aligned along the line of sight, because the preferred direction and speed of the sensors must be appropriate for their retinotopic locations.

Notice for equation (5) that the first component of the motion which arises from translation is an explicit function of the depth of the points ($z$) whereas the second component which arises from rotation is not. The translational component alone can be rewritten as:

$$\begin{pmatrix}\dot{X}_T\\ \dot{Y}_T\end{pmatrix}=\frac{V}{D_p} \times\begin{bmatrix}\frac{-1}{\cos\alpha_p\cos\beta_p} & 0 & \frac{\sin\alpha_p}{\cos^2\alpha_p\cos\beta_p}\\ 0 & \frac{-1}{\cos\alpha_p\cos\beta_p} & \frac{\sin\beta_p}{\cos^2\alpha_p\cos^2\beta_p}\end{bmatrix} \times\begin{pmatrix}\sin\alpha_H\cos\beta_H\\ \sin\beta_H\\ \cos\alpha_H\cos\beta_H\end{pmatrix}. \quad (6)$$

The matrix projects the 3-D motion onto the 2-D image plane. The last vector is simply a unit vector pointing in the heading direction. Figure 5 illustrates our nomenclature within the spherical egocentric coordinate system that we use to derive equation (6). Equation (6) explicitly shows that, in addition to the azimuth and elevation of heading $(\alpha_H, \beta_H)$ and of the stationary environmental point $(\alpha_p, \beta_p)$, the translational component of the flow vector is a function of the distance of the point $(D_p)$ and observer speed $(V)$. We will consider the effect of the latter two parameters more closely.

### *Distance*

Humans can navigate successfully through a wide range of environments and therefore we have chosen not to put constraints on the layout. It would seem that an

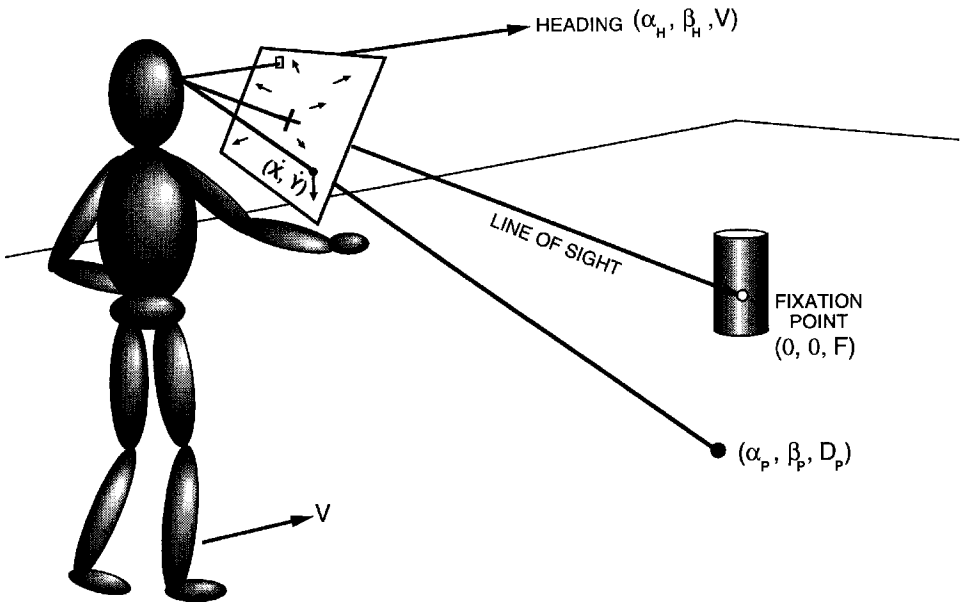


FIGURE 5. Our egocentric coordinate system used to derive equations (5) (8). The $z$-axis of the coordinate frame is aligned with the line of sight. Observer moves through environment at speed $V$ and heading direction $(\alpha_H, \beta_H)$ while fixating point $(0, 0, F)$. This results in rotation of the line of sight $(\omega_x, \omega_y, \omega_z)$ with $\omega_z$ (roll) $\approx 0$ deg/sec. The combined translation and rotation generate image velocity $(\dot{X}, \dot{Y})$ for a projected point $(\alpha_p, \beta_p, D_p)$.

impossibly large number of motion sensors are therefore required to sample the wide range of possible object distances that can occur at each location in the visual field. However, the translational component of the image speed (the only component of the flow vector sensitive to changes in object distance) falls off as an inverse function of distance [equation (6)]. We can therefore once again capitalize on logarithmic compression to sample the wide range of possible point distances using a small set of speed-tuned sensors at each image location. For convenience, we associate the set of sensors with a set of reference planes, orthogonal to the line of sight, located 2, 4, 8, 16 and 32 m away. This arbitrary, but judicious, choice of distances covers a wide range of possible layouts and the broad speed tuning of each sensor (Fig. 2) allows the system to respond to interpolated values. For an observer moving at 1 m/sec (walking pace) and looking in the direction of translation, image motion is only 0.6 deg/sec at 20 deg in the periphery (much less near the fovea) at the maximum range (32 m) so sampling distances further out will not yield much additional information. Conversely, at a pace of 1 m/sec, the observer will hit the closest reference plane (2 m) in only 2 sec so it is not likely that the observer will need to consider objects much closer than this because it is already too late to stop or avoid them (see Cutting, 1986; Cutting *et al.*, 1992). Therefore, for each detector, a set of five sensors at each retinotopic location based on reference planes located at 2, 4, 8, 16, 32 m along the line of sight captures much of the usable information. This arrangement provides a convenient egocentrically centered space that easily converts into a time-to-impact space when the line of sight is directed along the heading vector. It should be emphasized that this sampling of depth does not increase the number of detectors required in the heading map, it merely determines the number of sensor inputs to a given detector.

Figure 6 shows an example of a detector and how it is constructed. Each detector samples each image location using five motion sensors with preferred speeds and directions corresponding to the expected image velocity for each reference plane. Consider first the set of labelled vectors in the right-hand region of Fig. 6. For points on the nearest reference plane (a), the image motion resulting from just translation is high ($T_a$) and so the preferred direction of the motion sensors does not deviate by much from the radial direction out from the FOE. For points on the far reference plane (e), the translational image motion is low ($T_e$) and rotation, if present, will perturb the preferred direction of the motion sensors by a large amount from the radial direction out from the FOE. For the in-between planes (b, c, d), the deviation will be intermediate. It should be remembered that image speed is also a function of location on the image plane [$X$, $Y$ in equation (5)]. Sensors at different locations can be tuned to quite different speeds, yet still correspond to points on the same reference plane (cf. locations 1 and 2 in Fig. 6).

*Observer speed*

We now consider what happens if observer speed is faster or slower than that (1 m/sec) used to set the

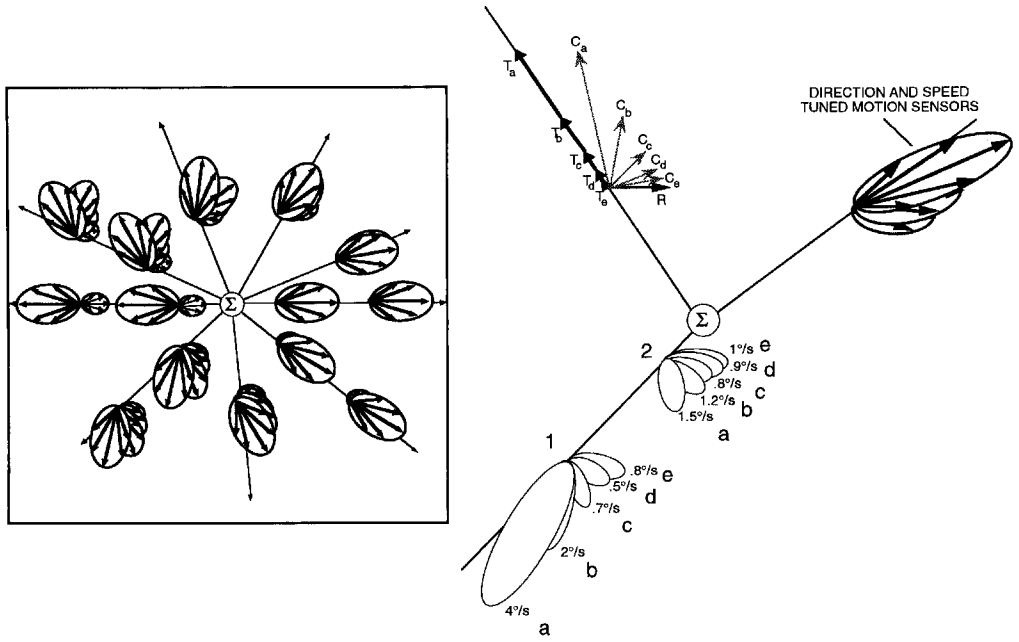


FIGURE 6. Representation of MST-like heading detector (left) and construction details. Sets of speed- and direction-tuned local motion sensors are connected and the activities of the most active sensor at each location are summed across the visual field to yield the response of the detector. At each location, the five translational components, corresponding to the five reference depth planes (a e), are given by $T_a$, $T_b$, etc. The fixed rotation rate (0, 1, 2, or 4 deg/sec) and candidate heading direction determine the rotational component ($R$) which depends weakly on the retinotopic position but is independent of depth. The rotation and translation components add vectorially to give the final sensor preferred velocity (speed and direction) used in the template ($C_a$, $C_b$, etc).

sensors. It should first be pointed out that we specified the absolute observer speed and the absolute reference plane depths merely for didactic purposes in order to make the examples specific. In reality, one can only specify a ratio of speed and distance [because $V$ and $D$ trade off in equation (6)]. So, for example, if the observer actually moves at 2 m/sec, the translational components of flow double and the pre-calculated translational component of the sensors' preferred velocities are therefore half of what they should be for their assigned reference plane, but exactly correct for a reference plane twice as far. The net effect of doubling $V$ is therefore that the model now works as if sampling reference planes at 4, 8, 16, 32 and 64 m. Note that the time-to-impact to the closest plane remains the same (2 sec). Observer speed different from that used to derive the model could therefore potentially produce less than optimal sampling of depth as near points carrying important translational information may be moving too fast to be processed properly. However, in reality, this is unlikely to be a problem because of the natural co-variation of point distance and observer speed (e.g. when driving most objects of interest are further away than when walking). Furthermore, if desired, the robustness of the model could be increased arbitrarily by having a larger number of reference planes (e.g. at . . . 0.5, 1, 2, 4, 8, 16, 32, 64, 128 m, . . .).

Since the observer is tracking an environmental point, the gaze-rotation rate will increase proportionally with increases in observer speed if fixation distance is held constant [equation (4)]. However, as long as there exists a detector map tuned to this higher rotation rate, model performance will be unchanged. Again, the model could be made arbitrarily robust by including additional detector maps corresponding to higher eye-rotation rates but this seems unnecessary given the natural covariation of observer speed and fixation distance [$V$ and $F$ trade-off exactly in equation (4)]. Faster moving observers will naturally tend to fixate further away (e.g. when driving, fixation points will tend to be further away than when walking).

## VII. THE NEW MODEL

### *Overall flow during gaze stabilization*

In order to construct detectors tuned for specific instances of translation with gaze-stabilization, we need an equation which describes overall retinal flow under these conditions. In the previous sections, we have highlighted particular factors that need to be considered in this endeavour. We now are in a position to generate the equation describing the overall flow. Again, our retinotopic coordinate system is illustrated in Fig. 5. First, equation (6) is substituted into equation (5).

Second, eye rotation direction is constrained by the choice of heading and is fully defined in terms of the heading direction $(\alpha_H, \beta_H)$ and the fixation distance $(F)$. It can be derived by setting $\dot{X} = \dot{Y} = 0$ for $\alpha_p = \beta_p = 0$ and $D_p = F$ (the gaze-stabilization constraint) and $w_z = 0$ (the no-roll assumption) and solving for $(\omega_x, \omega_y, \omega_z)$. The flow vector at image position $(X, Y)$ is thus completely defined by observer translation $(\alpha_H, \beta_H, V)$, the position of the point in space $(\alpha_p, \beta_p, D_p)$, and the fixation point $(0, 0, F)$ all given in spherical coordinates (azimuth, elevation, and magnitude)

$$\begin{aligned}\begin{pmatrix}\dot{X}\\ \dot{Y}\end{pmatrix} = \frac{V}{D_p} &\times \begin{bmatrix} \dfrac{-1}{\cos\alpha_p \cos\beta_p} & 0 & \dfrac{\sin\alpha_p}{\cos^2\alpha_p \cos\beta_p} \\ 0 & \dfrac{-1}{\cos\alpha_p \cos\beta_p} & \dfrac{\sin\beta_p}{\cos^2\alpha_p \cos^2\beta_p} \end{bmatrix} \\ &\times \begin{pmatrix}\sin\alpha_H \cos\beta_H \\ \sin\beta_H \\ \cos\alpha_H \cos\beta_H\end{pmatrix} + \frac{V}{F\cos^2\alpha_p} \\ &\times \begin{pmatrix}\sin\alpha_p \tan\beta_p & -1 & 0 \\ \tan^2\beta_p + \cos^2\alpha_p & -\sin\alpha_p \tan\beta_p & 0\end{pmatrix} \\ &\times \begin{pmatrix}\sin\beta_H \\ -\sin\alpha_H \cos\beta_H \\ 0\end{pmatrix}. \end{aligned} \tag{7}$$

Rather than sampling various fixation distances $(F)$, for the reasons described in Section III, we chose to sample across a compressed and bounded set of possible eye-rotation rates $(\omega_0)$ corresponding to the fixation loci described in Fig. 4. Rather than sampling point distances $(D_p)$, for the reasons described in Section V, we chose to sample points on references planes perpendicular to the line of sight across a compressed and bounded range of possible depths $(z)$. The flow equation can be modified to reflect these choices, yielding:

$$\begin{aligned}\begin{pmatrix}\dot{X}\\ \dot{Y}\end{pmatrix} &= \frac{V}{z}\begin{pmatrix}-1 & 0 & \tan\alpha_p \\ 0 & -1 & \tan\beta_p/\cos\alpha_p\end{pmatrix}\begin{pmatrix}\sin\alpha_H \cos\beta_H \\ \sin\beta_H \\ \cos\alpha_H \cos\beta_H\end{pmatrix} \\ &+ \frac{\omega_0}{\cos^2\alpha_p\sqrt{\sin^2\beta_H + \sin^2\alpha_H \cos^2\beta_H}} \\ &\times \begin{pmatrix}\sin\alpha_p \tan\beta_p & -1 & 0 \\ \tan^2\beta_p + \cos^2\alpha_p & -\sin\alpha_p \tan\beta_p & 0\end{pmatrix} \\ &\times \begin{pmatrix}\sin\beta_H \\ -\sin\alpha_H \cos\beta_H \\ 0\end{pmatrix}. \end{aligned} \tag{8}$$

*Heading detector maps*

We can now generate detectors for the two degrees of freedom of possible headings $(\alpha_H, \beta_H)$ and the additional compressed and bounded degree of freedom defined by the four preset rotation rates $(\omega_0)$ after fixing $V$ at walking pace (1 m/sec). For each detector, we determine the appropriate sensors for an array of image locations $(\alpha_p, \beta_p)$ at the five reference depths $(z)$ using equation (8). For convenience, in the present simulations, we assume that we have, at each retinotopic location, an unlimited set of motion sensors from which to select the five required for each detector. In previous simulations (Perrone, 1992), a fixed set of 12 directions in 30 deg steps and a fixed range of speed preferences was used. The sensors required were drawn from this limited set yet the model still performed well.

We set up $[(11 \times 24) + 1] = 265$ detectors in each of four detector maps for a total of 1060 detectors. This is a manageable number that could easily be implemented within MST or another extrastriate cortical area, yet this set can handle a large number of common self-motion scenarios. Several examples of detectors are shown in Fig. 7. Note that although all four examples are tuned to the same heading, they are quite different due to their rotational tuning. In particular, at higher rotation rates, a false FOE appears on the side of gaze opposite to true heading.

*Relative depth maps*

Once heading has been determined, the relative ranges of the points in the scene can then be derived. The velocity tuning of each motion sensor within each detector was set to correspond to a particular reference plane, therefore in the winning heading detector, the distribution of activity across the set of motion sensors at each location provides a readout of the relative depth of the point. If the motion sensor corresponding to the nearest reference plane is the most active at one location and the sensor corresponding to the farthest reference plane is most active at another location, we know that the second point has 16 times the depth of the first point. If observer speed is known then we would know the absolute depths. Therefore, simply, by identifying the most active sensor at each location within the winning detector, we can derive a complete relative depth map.

## VIII. TESTING THE MODEL

*Heading estimation*

Figure 8 shows a flow field simulating observer motion 10 deg to the left and 5 deg up from fixation at 1 m/sec towards a field of random points while fixating a point 10 m away. The overall flow field is shown on the right. The detector responses to this stimulus for each of the four rotation-rate maps is shown in Fig. 9. Detector output is plotted along the vertical axis and has been normalized relative to the maximum. The peak response is for the template tuned to an actual heading of $(-7.8$ deg, 4.5 deg) in the map tuned to 1 deg/sec of eye rotation. This is the detector whose tuning is closest to the actual heading and, therefore, the model's performance was optimal given our course sampling of heading space. Interpolation could be used to improve performance further.

Our model was developed under the assumption that gaze is perfectly stabilized and roll is completely

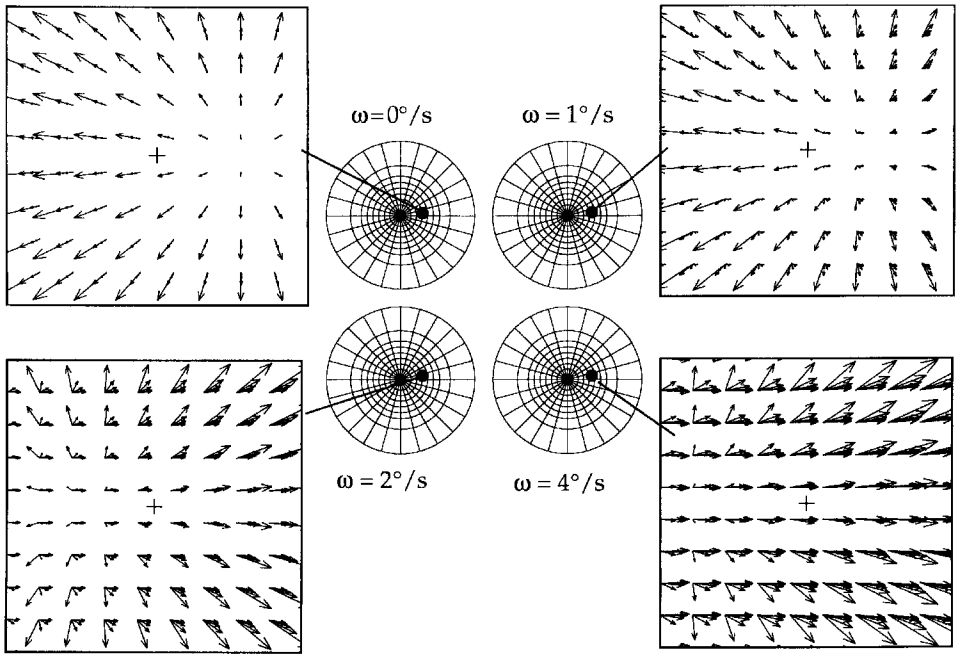


FIGURE 7. Four examples of detectors. Each detector is tuned to a heading direction of $\alpha = 12$ deg and $\beta = 0$ deg. This heading direction implies an eye rotation to the left about the yaw axis. The five vectors at each location represent the optimum velocity tuning for the five motion sensors corresponding to the five reference planes. The four different maps correspond to the four different eye-rotation rates used in the model.

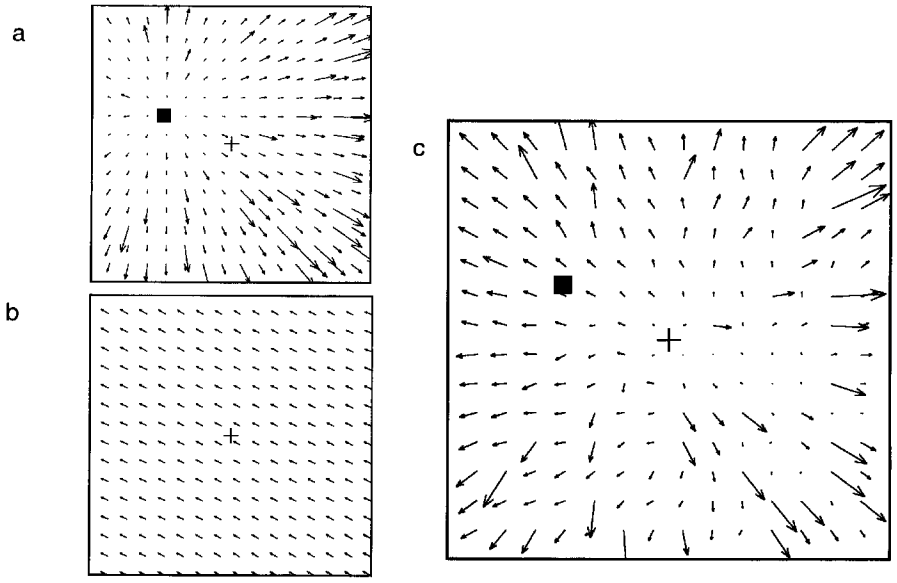


FIGURE 8. Input flow field used to test the model. The instantaneous flow field was approximated by a displacement field over 1 sec of motion. The environmental points were located such that they projected as a uniform array (14 × 14) on the image plane with a 40 × 40 deg field of view. The radial distance of the points out from the eye was randomly determined and could be from 5 to 20 m away. (a) The translational component of the flow field with heading at $\alpha = -10$ deg, $\beta = 5$ deg relative to fixation. Because all detectors are constructed retinotopically, the fixation point is always represented in the center of the field (cross). (b) Rotational component caused by gaze stabilization. (c). The overall input flow field is simply the vector sum of (a) and (b) The solid square indicates the heading direction that must be extracted.

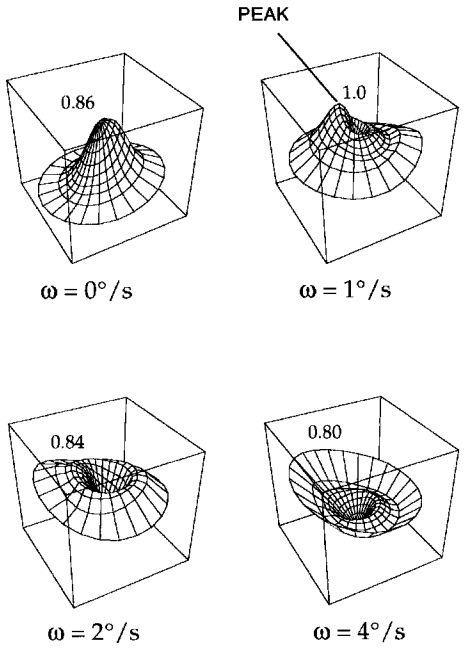


FIGURE 9. Response of the model to the stimulus shown in Fig. 8. The activity of each detector is shown in the vertical dimension of each map.

suppressed. In order to test the robustness of the model to violations of these strict assumptions, we performed three series of simulations (Fig. 10). In all three series, we simulated forward translation in the (−7.8 deg, 4.5 deg) direction with the fixation distance set to 9 m to generate a 1 m/sec eye movement. Given that this exact scenario is represented by one of the detectors, it is not surprising that the nominal performance of the model was perfect. First, we examined the effect of various stabilization gains (defined as the fraction of the speed needed to stabilize gaze perfectly). In each case, the direction of eye rotation was correct (150 deg) but the amplitude was variable [Fig. 10(a)]. Note that as long as gain was 0.6 or higher, performance was good. Second, we simulated stabilization such that the eye movement had the correct speed (1 deg/sec) but its direction was variable [Fig. 10(b)]. Note that performance is good as long as the direction is within 15 deg of correct and degrades slowly for higher direction errors. Third, we added various levels of extraneous roll around the line of sight [Fig. 10(c)]. In this case, simulated gaze stabilization was perfect and roll did not seriously degrade performance until it surpassed 2 deg/sec. We conclude from these data that our model is robust to violations of our initial assumptions.

Royden, Banks and Crowell (1992) found evidence that humans cannot recover heading during simulated gaze stabilization at higher rotation rates without the benefit of eye-movement information. Because this could be construed to rule out all self-motion algorithms that attempt to handle rotation on a purely visual basis, we address their results specifically. In order to replicate their simulated stabilization condition (5 deg/sec average rotation), we created flow-field inputs that corresponded to 0.5 m/sec of translation in the (5 deg, 0 deg) direction with 2.2 m/sec of rotation (start of the trial) and in the (11.2 deg, 0 deg) direction with 11.1 deg/sec of rotation (end of the trial). The scene consisted of points randomly distributed in depth between 0 and 37.3 m and a field of view of 30 × 30 deg. In simulations of the beginning of their trial, the model produced a small bias (~5 deg) in the direction of rotation while, in simulations of the end of the trial, the model produced heading estimates completely dominated by the rotation (~90 deg bias). This is consistent with their empirical findings of biases between 15 and 20 deg. However, two questions arise. First, why doesn't the model perform perfectly at the beginning of the trial given the fact that the rotation rate is only 2.2 deg/sec? Second, would the model fare better at the end of the trial if it had a map tuned to a higher rotation rate?

To address these questions, we performed additional simulations under slightly modified conditions in which the range of point distances was reduced by a factor of 4 (0–9.3 m) and observer speed was increased to 1 m/sec in order to provide a better balance of translational and rotational flow. Simulations of the beginning of the trial then yielded optimal estimation; i.e. the closest detector, the one tuned to (6 deg, 0 deg), had the maximal response although simulations of the end of the trial still yielded high biases (~60 deg).) We then extended the model to incorporate a fifth heading map tuned to 8 deg/sec. While an additional 8 deg/sec map did not change the performance of the model under the original experimental conditions used by Royden *et al.* (for the simulations described in the preceding paragraph), it reduced the bias corresponding to the end of the trial to ~5 deg under the modified conditions.

These results show that our model's performance is consistent with the results of Royden *et al.* (1992), even if the model has heading maps tuned to higher rotation rates. However, they also suggest that not just rotation rate but also layout and observer speed may affect how accurately heading can be extracted from the flow field. Further psychophysics will be required to determine to what extent translational flow can compensate for high rotation rates and allow humans to estimate heading accurately even when extra-retinal information is unavailable.

*Range extraction*

In this simulation, the environment consisted of an array of points lying in a vertical plane located at 40 m from the observer and a second superimposed circular patch of points 5 m from the observer. The simulated observer motion was 9 deg to the right at a speed of 0.5 m/sec. Fixation as always is set on the center of the image and in this case generates a pure yaw eye rotation of −0.89 deg/sec. The resulting simulated image motion is depicted in Fig. 11(a). When this flow field was run through the model, the most active

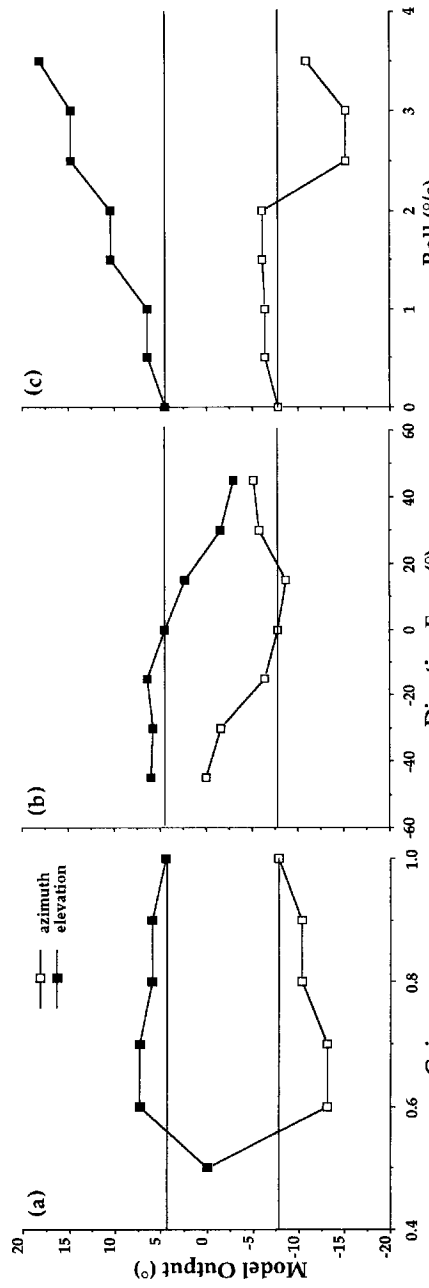


FIGURE 10. Testing robustness of model. Test stimulii were similar to that used in Fig. 8 but the point distances ranged from 5 to 45 m. Horizontal lines indicate true azimuth and elevation of heading direction (−7.8 deg, 4.5 deg). (a) Effect of stabilization gain errors. (b) Effect of stabilization direction errors. (c) Effect of extraneous roll (clockwise about line of sight).

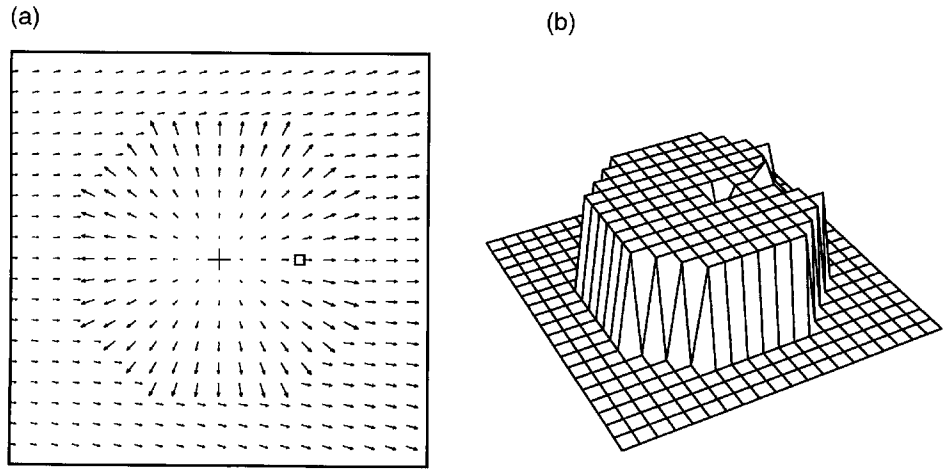


FIGURE 11. Testing the depth estimation function of the model. (a) The input flow field. The layout consisted of points at two possible depths. A far vertical background plane of points was 40 m from observer. A nearer circular region was 5 m away. Heading direction was (9 deg, 0 deg) represented by square. The motion was sampled using a 20 × 20 array of sensors over the 40 × 40 deg field of view. (b) A perspective plot showing the relative depths assigned to each point by the model.

detector was that tuned to (9 deg, 0 deg) in the 1 deg/sec map.

The distribution of activity in the five motion sensors at each location feeding into the winning detector was examined and the maximum value found. If the winning sensor corresponded to the nearest reference plane, this point was assigned a depth of $D$. This was repeated at the other locations and each point was assigned a depth value of $D$, $2D$, $4D$, $8D$ or $16D$ [Fig. 11(b)]. The majority of the points in the circular region were found to lie on the $4D$ plane and all of the "background" points were found to lie on the $16D$ plane. Some range errors are apparent for points close to the actual heading direction (square), where the template sensors tend to be tuned to a narrow band of slow speeds. Small errors in the gaze-rotation rate estimate contribute to the errors in the range estimates at such locations.

This example demonstrates the simplicity of the depth extraction part of the model. It does not require any extra mechanisms and the depth information is simply coded by the relative activity of the sensors within the structure of the detectors set up to extract heading. Furthermore, the depth estimation process generalizes to a wide range of observer/scene combinations. For instance, the common "motion parallax" stimuli used to study structure from motion, in which the scene moves parallel to the image plane (e.g. Rogers & Graham, 1979), are analyzable using the templates in the model tuned to the 89 deg radial direction.

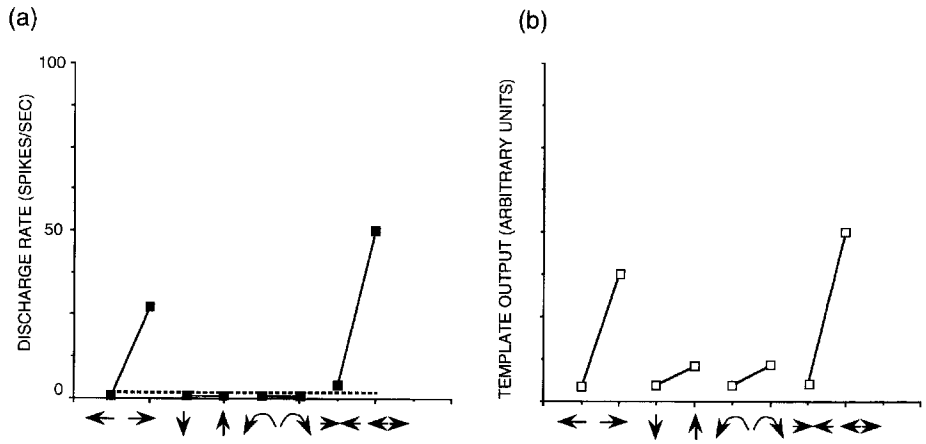


FIGURE 12. Simulating MST neuronal responses. Icons along horizontal axis represent type of image motion used to test the neuron and the model. (a) Replotted data from Fig. 6(e) of Duffy and Wurtz (1991a). Planoradial neuron 53XL68. Dashed line shows activity for their control condition. (b) Normalized output of a template tuned to (10 deg, 0 deg) heading and 1 deg/sec eye rotation.

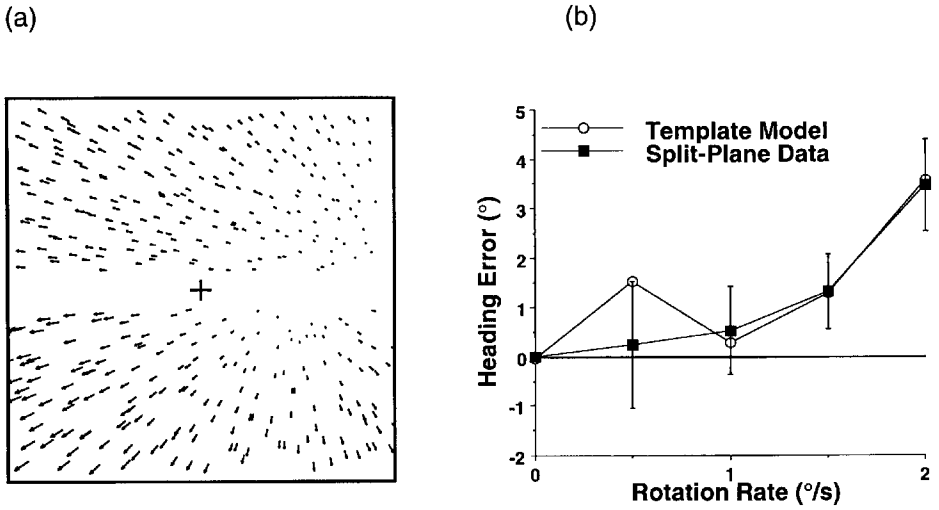


FIGURE 13. Comparing simulated and actual human performance. (a) "Split-plane" stimulus for (0 deg, 0 deg) heading and 1.5 deg/sec yaw rotation to the right. Observer speed was 2 m/sec and the distance to near plane (bottom) was 10.2 m and the distance to the far plane (top) was 22.7 m. (b) Graph of heading error vs rotation rate (positive values indicating rightward) for human observers (solid squares) and for the model (open circles). Some parametric changes were made for this simulation; (see text for other details): the rotation rates for the detector maps were lowered to 0, 0.45, 0.9 and 1.8 deg/sec and noise was added to the output of each of the templates in each of the four maps. The added noise was normally distributed with a SD equal to 5% of the original activity of the template. Psychophysical data is the mean of three observers and the error bars represent ±1 SD. Model estimates are based on 12 replications. The model input consisted of only two frames while observers viewed a sequence of 35 frames over 2.3 sec.

## IX. RELATIONSHIP TO THE NEUROPHYSIOLOGY OF EXTRASTRIATE CORTEX

### *Heading detectors act like MST neurons*

The model is based on a network of idealized MT-neurons whose outputs are integrated across the visual field in order to construct putative MST neurons. Therefore, by design, the motion sensors behave like MT-neurons. In this section, we show that the resulting heading detectors, constructed to solve the self-motion problem have the emergent property that they do indeed resemble neurons within MST.

Several groups have recorded from MST neurons using large fields of random dots to simulate optic flow patterns (Duffy & Wurtz, 1991a, b; Orban *et al.*, 1992; Saito *et al.*, 1986; Tanaka *et al.*, 1986, 1989; Tanaka & Saito, 1989): Figure 12(a) shows replotted data from Duffy and Wurtz's (1991a) study using their convention for the stimuli presented. The icons along the bottom of the graph correspond to leftward, rightward, downward and upward planar motion, anti-clockwise and clockwise roll motion, radial in and radial out, respectively from left to right. The points plotted represent the outputs of a neuron (53XL78) in response to 100 × 100 deg planes of moving random dots undergoing the type of motion specified by the icon. Duffy and Wurtz described this neuron as "planoradial" because it was selective to both planar motion and radial expansion. Figure 12(b) shows the output of a template tuned to (10 deg, 0 deg) and 1 deg/sec eye-rotation. The "responses" of this detector to the same large-field stimuli are comparable to the experimental data.

We also tested this same detector with small (33 × 33 deg) patches of flow-field components (pure radial, roll, and planar flow within a single depth plane) presented at nine different locations within the detector's receptive field as Duffy and Wurtz (1991b) did with MST neurons. At all positions, this detector's pure radial responses were similar (ranging from around 10–50% of the response to the optimal stimulus) and maintained their selectivity to expansion over contraction. Furthermore, its pure roll responses were of similar magnitude to the expansion responses, largely independent of location, but not directionally selective. Finally, its response to planar motion was highly dependent on stimulus location. Such behavior is not uncommon in MST neurons (see Figs 7 and 8 in Duffy & Wurtz, 1991b). In particular, this detector shows, at least for radial and roll motion, the kind of position invariance reported to be common among MST neurons (Tanaka *et al.*, 1989; Duffy & Wurtz, 1991b; Andersen, Graziano & Snowden, 1991; Orban *et al.*, 1992).

### *Multiple detector maps*

By examining the case of gaze stabilization, we were able to reduce the self-motion estimation problem to only three dimensions. However, maps within extra-striate visual cortex have traditionally been viewed as 2-D maps. In order to make the new model more biologically plausible, it is important to consider how it could be implemented within a 2-D cortical structure.

We now present two ways by which this could be achieved.

First, the multiple heading maps for different gaze-rotation rates (or fixation distances) could be embedded within a columnar structure. Just as orientation or ocular dominance in striate cortex (Hubel & Wiesel, 1968, 1974; Hubel, Wiesel & Stryker, 1978; Weisel, Hubel & Lam, 1974) and direction in MT (Albright, Desimone & Gross, 1984) allow for the coding of additional stimulus dimensions as well as the two spatial dimensions, so too could the third dimension of gaze-rotation rate be coded within a 2-D map.

Second, the detectors could be dynamic templates whose sensor inputs change as a function of rotation. Specialized detectors could extract the rotation component independently of translation (Perrone, 1992). This visual information could then be used to retune the detectors dynamically. Recent physiological data (Orban *et al.*, 1992) suggest, however, that this is unlikely. The dynamic changes could, however, be triggered by extra-retinal inputs. In fact, only a single 2-D heading map would be necessary if the sensor to detector (MT to MST) inputs were dynamically altered according to vergence. Furthermore, there is recent evidence that non-retinal inputs related to fixation distance can be used to modify visual cortical responses even at the level of V1 (Trotter, Celebrini, Stricanne, Thorpe & Imbert, 1992).

## X. RELATIONSHIP TO PSYCHOPHYSICS OF HEADING PERCEPTION

The test example shown in Fig. 8(c) is similar to the stimuli used by Warren and colleagues (Warren & Hannon, 1988, 1990; Warren *et al.*, 1988) to test humans' ability to extract heading information using only visual inputs. It is not surprising that the model was also able to extract heading in this case because it was designed specifically for that purpose. However, removal of the depth information in this type of stimulus makes it difficult for the templates to discriminate between rotation and non-rotation scenarios. In the extreme, if there is only a single plane of points, the model fails to extract true heading and falsely indicates that heading is biased in the direction of rotation towards the false FOE. This is both consistent with previous psychophysical results showing that, when single planes of points are used, observers consistently misperceive their heading as towards the false FOE (Rieger & Toet, 1985; Warren & Hannon, 1990; Perrone & Stone, 1991; Stone & Perrone, 1991) and with the fact that heading is not uniquely defined by such single-plane visual stimuli.

A more challenging stimulus for the model [Fig. 13(a)] is the image motion obtained when an observer is translating towards two vertical half-planes of points at different depths separated by a 6 deg vertical gap across the horizontal meridian while the observer's line of sight rotates to the right. This flow field represents one frame from a sequence in which the observer is following a curvilinear path while tracking a point moving with him/her (i.e. cancelling his/her gaze stabilization reflexes). This stimulus is inadequate for algorithms based on local differential motion (Longuet-Higgins & Prazdny, 1980; Rieger & Lawton, 1985; Hildreth, 1992) because the only depth differences in the scene are for points on each side of the 6 deg gap. The fact that humans could in fact accurately estimate their heading under this condition argues that the heading estimation must be based on a global process similar to the template model presented above and not on a local differencing scheme (Perrone & Stone, 1991; Stone & Perrone, 1991, 1993).

The model was tested with this "split-plane" stimulus under a variety of heading/rotation rate conditions (at yaw rates of 0, 0.5, 1.0, 1.5 and 2.0 deg/sec and at heading angles of $-8$, $-4$, 0, 4 and 8 deg in the horizontal plane). Recall that the gaze-stabilization model does not have the complete set of arbitrary heading and rotation combinations necessary to process perfectly this type of curvilinear path flow field. Because we were only interested in estimates of heading azimuth, only those detectors along the horizontal meridian were sampled but more finely than in previous simulations. Heading estimates were plotted as a function of the true heading direction and linear interpolation was used to find the point at which the model estimated heading to be in the 0 deg direction (straight ahead) in order to match the human psychophysical data in which perceived straight ahead was determined using a staircase procedure. If heading were extracted perfectly, then the estimates would all lie on the solid line, independent of rotation rate. Figure 13(b) is a graph of the heading errors produced by the model and by human observers as a function of rotation rate. The human data represents the mean from three subjects and has been reported elsewhere (Perrone & Stone, 1991). The model provides an excellent fit to the psychophysical data and exhibits the same tendency to produce a bias in the direction of the rotation.*

## XI. DISCUSSION

The problem of extracting 3-D self-motion information from 2-D image motion has generally been treated from a purely analytical viewpoint (e.g. Lee, 1974; Koenderink & van Doorn, 1975; Longuet-Higgins & Prazdny, 1980). It has long been known that the image motion that occurs as a result of self-motion can be exactly specified mathematically [equation (5)]. Many models of how the reverse process might be done have been proposed (see Heeger & Jepson, 1992 for a review) but most have been shown to be incompatible with human psychophysical results (Warren & Hannon, 1988, 1990; Perrone & Stone, 1991; Stone & Perrone, 1991,

*The gaze-rotation rates chosen for the heading maps were selected to provide the best fit to the data. Using a different set of values produces different predictions. For other values, the heading error produced by the model could be in the direction opposite to that of the rotation direction (i.e. below the solid line). This "reversed" behavior was also sometimes seen in our psychophysical data.

1993; Cutting *et al.*, 1992). Motivated by a desire to explain both the physiology of primate extrastriate cortex and our own human psychophysical results, we have presented a model of visual heading and depth estimation which uses neural-like sensors and detectors with realistic MT- and MST-like receptive field properties, respectively, and whose overall performance reasonably predicts human perception in many heading estimation tasks. The usual problem with template models is that the number of templates needed to mimic behavior is usually too high to be of practical use. Using constraints provided by examining the consequences of the gaze-stabilization mechanisms that are active and presumably effective during locomotion, we have successfully lowered the number of templates to a manageable level while producing overall performance consistent with previous psychophysical data.

The template model does not rely on a decomposition of the translational and rotational components of the image motion in order to recover heading. Cutting *et al.* (1992) also question the concept of "decomposition" and suggest a means by which heading direction can be estimated from the differential motion parallax that occurs around the point of fixation. However their approach relies on an active scanning pattern and a set of rules by which the "observer" can infer heading. Our detectors pick up this information automatically, not just around the fixation point but over the whole field, without the need for higher-level decision rules.

*Alternate model of MST-based heading estimation*

Recently, Lappe and Rauschecker (1993) proposed a model which uses a two-layered network, ostensibly MT to MST, to estimate heading by performing a partial decomposition of the flow field using a modified version of the subspace algorithm (Heeger & Jepson, 1990, 1992). The original subspace algorithm is a mathematical procedure by which heading can be computed even in the presence of arbitrary rotation. This algorithm could be used as the basis for creating heading detectors which are totally immune to observer rotation. The problem with this approach as a model for human heading estimation has been two-fold: (1) that it finds the exact solution and therefore appears inconsistent with the psychophysical results of Fig. 13 and others (Royden *et al.*, 1992), and (2) that it would result in detectors that would be inconsistent with the preliminary finding that the responses of MST neurons to their preferred flow component are not immune to the presence of unpreferred flow components (Orban *et al.*, 1992).

Lappe and Rauschcker have modified the subspace algorithm to provide immunity only to that rotation caused by gaze stabilization and have implemented it using a neural network. Their model however requires that the flow field be derived before the modified subspace algorithm can be applied. Unfortunately, the method they use to derive the flow field explicitly assumes that the output of MT neurons is proportional to local image speed. This is inconsistent with the well-known responses of MT neurons (Maunsell & Van Essen, 1983b; Albright, 1984) but might however be remedied if they used a more biologically realistic pre-processing algorithm for deriving the flow field. In addition, the biological implementation of the output of their model is ill-defined. The heading maps they show in their Fig. 2 are misleading as they are not maps of MST neurons but rather "maps" of the sums of outputs of subpopulations of MST neurons. Unlike our output maps, this is not a place code unless the summing is explicitly done by a new set of neurons. Therefore, their model requires an additional stage of processing at the output to readout the sums of the activities of the MST subpopulations, each tuned to a particular heading.

Despite these shortfalls, the model is an interesting alternate view that deserves further examination. In particular, because the output neurons are only immune to the rotations in the appropriate direction for gaze stabilization, it no longer suffers from being "too good" as the original subspace algorithm appears to be. It would be interesting to see however whether the model is robust to errors in the direction of stabilization and to roll [see our Fig. 10(b, c)] and whether the model shows the systematic errors seen in heading estimation during curvilinear motion (see our Fig. 13) or under the conditions used by Royden *et al.* (1992).

*Human heading estimation*

Although numerous models exist that can solve the visual self-motion problem in the general case, there is some debate as to whether or not humans are able to extract heading using visual motion cues alone. The main evidence for purely-visual heading estimation came from psychophysical experiments using visual stimuli that simulated fixation of a stationary point in the environment during forward translation (Warren & Hannon, 1988, 1990; Cutting *et al.*, 1992). These authors showed that heading can be accurately perceived in this case although simulated eye rotation rates were low (below $\sim 2$ deg/sec). In addition, Rieger and Toet (1985) and Perrone and Stone (1991) showed that in the case of visually simulated curvilinear motion in which gaze is fixed with respect to heading and not stabilized on a stationary target, heading estimation is also possible although systematic biases were found [see Fig. 13(b)].

Recently, Royden *et al.* (1992) presented convincing evidence that the combination of a flow field and extra-retinal signals, presumably oculomotor in origin, can allow humans to make accurate heading judgements under conditions where the visual flow field alone cannot. However, they used visual flow fields that were impoverished either because of a high rotation/translation ratio or a lack of depth variation in the scene. They concluded that "humans require extra-retinal information about eye-position to perceive heading accurately in the presence of rotation rates $> 1$ deg/sec". This conclusion however should await further psychophysical testing under high rotation conditions in which the flow field is not impoverished. Nevertheless, this

caveat does not belittle the role of extra-retinal inputs which can allow humans to perceive heading accurately even under visually impoverished conditions (see below). Finally, regardless of the controversy surrounding heading estimation at higher rotation rates, it is clear that humans do indeed have the capacity to use visual-motion information alone to extract heading relatively accurately at the low rotation rates that occur most commonly during gaze stabilization.

Simulations of our model are consistent with previous psychophysical findings (Rieger & Toet, 1985; Warren & Hannon, 1988, 1990; Perrone & Stone, 1991; Cutting *et al.*, 1992; Royden *et al.*, 1992). Furthermore, the model is not only able to extract heading quite accurately in response to simulated forward translation during gaze stabilization (Figs 8 and 9—the stimulus it was designed to deal with), but also to mimic human performance during curvilinear motion (Fig. 13—a stimulus it was not expressly designed to deal with). The model shows the same type of systematic errors that humans can make under the curvilinear condition although heading estimation is much better than that which would be achieved if the rotation were simply ignored. These results suggest that the human visual system may possess specialized detectors tuned for gaze stabilized self-motion and that observed heading errors in response to curvilinear motion may reflect the operation of such detectors on non-optimal stimuli.

*Extra-retinal signals*

Our model does not preclude the possibility that information other than the flow field are used in heading estimation. Visual signals such as disparity and extra-retinal signals such as vergence, accommodation, or eye-movement motor corollary could all contribute potentially useful information. In fact, the model output-map format is aptly suited to incorporate such inputs to improve performance. There is often competing activity in all four maps. With noise both in the input flow field and at the level of the detectors, one can see that competing responses within the detector maps could on occasion result in incorrect self-motion estimates. In Fig. 9, the peak in the 1 deg/sec map provides a visual signal, indicating that the eye is rotating at 1 deg/sec in the 150 deg direction. An extra-retinal signal signalling eye-speed at 1 deg/sec could be used to facilitate all the responses in the 1 deg/sec map and/or to inhibit the responses in the other maps. If both eye-speed and eye-direction are taken into consideration, oculomotor facilitation could be constrained to the appropriate sector of the appropriate map. This is a mechanism by which extra-retinal information could greatly enhance performance in heading judgments as is observed empirically (Royden *et al.*, 1992). Furthermore, any information concerning the actual distance to the fixated point (vergence, or even possibly disparity signals) could also be used to inhibit or facilitate competing responses across the different maps. Finally, any vestibular information about instantaneous heading direction could also be used to inhibit or facilitate selectively different regions within each heading map and enhance performance. A number of physiological studies have shown oculomotor (Newsome, Wurtz & Komatsu, 1988), vestibular (Kawano *et al.*, 1984; Thiers & Erickson, 1992), and disparity (Roy & Wurtz, 1990; Roy, Kanatsu & Wurtz, 1992) signals in MST which could underlie this polysensory fusion of self-motion information.

*Egocentric reference frame*

A critical tacit assumption in our model is the premise that humans can effectively use information in an egocentric coordinate system (retinotopic coordinates) to navigate in an exocentric environment. Parietal neurons involved in saccadic programming appear to accomplish this transformation. They encode target location with respect to the head by having their retinotopic response weighted by an eye-position signal (Andersen, Essick & Siegel, 1985). We postulate that humans can convert heading from a retinotopic to a head-centered coordinate frame, possibly in a similar manner. In fact, there is recent preliminary evidence for eye-position weighting of MST responses (Bremmer & Hoffmann, 1993).

*Sampling heading, point distance, and fixation distance*

Because relevant psychophysical and physiological data were often unavailable, we were forced to make certain arbitrary choices. Heading sampling was based on a sensitivity criterion and resulted in denser sampling of directions close to the fovea. Unfortunately, we could not make the layout of the maps correspond to the topography of MST as such information is not available. In fact, if our model is correct, the topography of MST would not be a function of location in the visual field but of heading direction with all neurons responding to motion nearly anywhere in the visual field. Furthermore, we cannot set up our heading sampling according to psychophysics because, although the effect of eccentricity on heading estimation has been measured during pure translation (Crowell & Banks, 1993; Warren & Kurtz, 1992), no such information exists with regards to combined translation and rotation. The sampling of heading space could also have been based on consideration of what environments are commonly encountered. For example, we could have used visual angles that correspond to equal environmental units of distance measured along a ground plane (e.g. Johnston, 1986).

Point distances were sampled using a set of logarithmically spaced reference depth planes (equi-$z$). This was arbitrary. We could have just as easily sampled a set of logarithmically spaced concentric half-spheres (equi-$D$). Similarly, fixation distance was sampled arbitrarily along the iso-rotation-rate contours described in Fig. 4. We could have just as easily sampled a logarithmically spaced set of distances (much like the point-distance reference planes) and set up a different detector map for each fixation distance rather than for each eye-rotation rate. While these issues await future data for resolution, they do not detract from the overall

strengths of the general template framework proposed in this paper.

*Depth estimation*

The model is capable of simultaneously extracting relative depth as well as heading information. While the concept of extracting depth from the templates is straight-forward, the question of how such a process could occur physiologically is not. The problem is that the depth extraction process requires combining information from the sensor level (MT) with that from the detector maps (MST). The latter is needed to specify which of the detectors is the most active and consequently which of the sensors at the lower level contain the relevant depth information. The reciprocal connections between MT and MST (Maunsell & Van Essen, 1983a; Ungerleider & Desimone, 1986; Boussaoud, Ungerleider & Desimone, 1990) may provide the feedback necessary for solving this problem. The previously demonstrated inhibitory inputs to MT neurons from outside of their traditional receptive fields (Allman, Miezin & McGuiness, 1985) could be used to silence MT neurons that do not feed into the most active MST detector. However, at present, there is inadequate information about either the responses in MT during visually-simulated self-motion or the nature of the reciprocal connections to explore this issue further at this time.* Finally, disparity information available in both MT (Maunsell & Van Essen, 1983c) and MST (Roy & Wurtz, 1990; Roy *et al.*, 1992) may also contribute to depth extraction.

Another critical issue in depth extraction is resolution. If we only use the peak activity in the set of five sensors, the resolution would be limited to the five reference planes but a simple interpolation scheme could be used to obtain greater resolution (Perrone & Stone, 1992a). However, it is not clear at this point just how sensitive the model needs to be. There is little psychophysical data pertaining to how well humans can extract depth information from motion alone while translating and tracking. The special case of pure translation when the heading direction is 90 deg to the line of sight has received much attention (e.g. Braunstein & Tittle, 1988; Ono, Rivest & Ono, 1986; Rogers & Graham, 1979) but other situations have largely been ignored.

*MST neurons as heading detectors*

We have shown that the detectors show responses to the standard repertoire of flow-field stimuli that resemble those of MST neurons. There are detectors that respond to nearly pure planar and radial stimuli or to combinations of planar and radial and even roll stimuli. For simplicity, we have however chosen explicitly to neglect roll and therefore the model does not include sensors that respond specifically to pure roll although many such roll detectors can be found in MST (e.g. Duffy & Wurtz, 1991a). Extending the model to deal with a limited set of roll situations could be easily accomplished and the original general model (Perrone, 1992) showed how such detectors could be made and used. Psychophysical studies must be performed to determine the extent to which humans can estimate heading in the presence of roll. Furthermore, the proposed role of MST in self-motion perception does not preclude its role in other related functions (e.g. postural or oculomotor control) in which roll detectors may play an important part. In particular, roll detectors would be necessary for any visually-driven torsional stabilization of the eye as standard foveal gaze-control mechanisms would be ineffective.

Recent evidence seems to support our original proposal that MST analyzes the global pattern of image motion using neurons acting as templates looking for specific instances of combined translation and rotation (Perrone, 1987, 1992; Stone & Perrone, 1991) rather than acting as processors decomposing the flow field into its translational and rotational components (e.g. Rieger & Lawton, 1985). Orban *et al.* (1992) showed that MST neurons do not appear to respond to a particular component of flow independently of the presence of other flow components. They presented data that suggest neurons do not "decompose" the flow field but rather are indeed tuned to specific combinations of flow components, i.e. they act as templates. As suggested by Saito, Tanaka and colleagues (Saito *et al.*, 1986; Tanaka *et al.*, 1986), Verri *et al.* (1992) were indeed able to model MST neuronal responses using simple linear integration from an appropriately organized set of MT input units.

A number of investigators have reported that many MST receptive fields show position or translational invariance (Tanaka *et al.*, 1989; Duffy & Wurtz, 1991b; Andersen *et al.*, 1991; Orban *et al.*, 1992), meaning that their responses appear largely invariant to simple shifts in the position where the stimulus is presented within the receptive field. At first glance, this property appears antithetical to MST neurons playing a role in heading estimation. The argument is that, in the case of expansion, the shifts represent changes in heading. Therefore, if the MST responses code for heading, they should be sensitive to stimulus position shifts. First of all, the data show broad tuning to changes in position, not true invariance (see e.g. radial responses in Figs 7 and 8 of Duffy & Wurtz, 1991b). However broad tuning for heading by individual MST neurons is not incompatible with the precise coding of heading. For example, orientation tuning of primate V1 neurons is quite broad [$\sim$40 deg (Schiller, Finlay & Volman, 1976; De Valois, Yund & Hepler, 1982)] yet human orientation discrimination can be quite precise [$\sim$1 deg (Blakemore & Nachmias, 1971; Thomas & Gille, 1979)]. Second, the stimuli used to test for position invariance were likely suboptimal: relatively small patches of either 2-D flow components (e.g. pure radial, planar, or roll stimuli) or combinations thereof rather than true flow fields generated by simulating 3-D motion. The position tuning might be much narrower for the optimal stimulus. In the case of our detectors, they are clearly sensitive to shifts

*A crucial question is whether this inhibitory surround is caused by feedforward, local, or feedback cortical pathways.

in the position of their preferred stimuli yet small patches of pure expansion or roll can generate similar responses throughout their receptive field. Third, in a recent preliminary study of the pure translation case, a systematic exploration of the effect of changing the FOE of a purely radial field, while keeping the overall stimulus size large and constant, did indeed find that some MST neurons appear tuned to a specific heading or FOE position (Duffy & Wurtz, 1993). Furthermore, neurons representing the full range of heading directions were found. Although this result needs to be extended to combined translational and rotational flow fields, it does however dispel the suggestion that MST neurons cannot code for heading.

Unfortunately there is inadequate information about MST neurons to determine whether our model is truly a good explanation of their response properties and function. First, as yet, no causal link has been established between self-motion estimation and MST. Lesion, correlation, and stimulation studies along the lines of those done by Newsome *et al.* for MT (Newsome, Wurtz, Dursteler & Mikami, 1985; Newsome, Britten & Movshon, 1989; Salzman, Britten & Newsome, 1990) will be necessary in order to establish such a link although clinical observations support the contention that parieto-temporal cortex is involved in the processing of visual orientation within the environment (Holmes, 1918). Furthermore, a role for MST in self-motion perception should not be construed to deny a role in postural control or oculomotor control or object-motion perception (e.g. Tanaka, Sugita, Moriya & Saito, 1993). In particular, MST has been clearly shown to contribute to smooth pursuit (Dursteler, Wurtz & Newsome, 1987; Newsome *et al.*, 1988; Dursteler & Wurtz, 1988; Komatsu & Wurtz, 1989). Given that our heading detectors also report the eye velocity, they could use this information to contribute to the generation of smooth eye movements. In addition, MST appears to have at least two functionally different subregions: one lateroventral and one dorsal (Komatsu & Wurtz, 1988). Our detectors are meant to model the responses of those neurons in dorsal MST or MSTd (Duffy & Wurtz, 1991a). Finally, other cortical areas downstream from MST, such as the fundus of the superior temporal area (FST), the superior temporal polysensory area (STP), or area 7a (Boussaoud *et al.*, 1990), also have neurons whose receptive field properties appear aptly suited to be involved in self-motion estimation (Bruce, Desimone & Gross, 1986; Steinmetz, Motter, Duffy & Mountcastle, 1987; Hikosaka, Iwai, Saito & Tanaka, 1988; Erickson & Dow, 1989). A more thorough exploration of their properties will be required before any model can be fully evaluated as a descriptor of extrastriate cortical processing.

Our primary contribution is that we have provided a theoretical framework in which specific physiological and psychophysical experiments can be designed. More specifically, we have developed a set of useful stimuli (the template patterns) that could be used to test our hypotheses. For example, if MST neurons indeed act like our detectors, then they should respond nearly identically to random-dot stimuli generated by motion towards different layouts for the same rotation/translation combination. In the past, many studies of MST have used flow-fields that are essentially 2-D stimuli (that do not correspond to a realistic 3-D situation) in an effort to categorize mathematically MST responses to idealized flow components. Our template model provides a clear set of 2-D flow fields that are essentially 3-D stimuli (projected onto an image plane) that we believe are better suited to explore the role of MST or any other cortical area in self-motion perception.

## CONCLUSIONS

We have presented a model of putative self-motion processing in primate extrastriate cortex that can account for many of the physiological responses of MST neurons and for many aspects of human heading perception. We do not presume to have determined how or even if MST plays a role in primate visual self-motion perception nor do we claim to provide the first or most complete solution to the self-motion estimation problem. We merely conclude that our template approach is a viable framework for explaining the processing of self-motion information within primate extrastriate cortex and is consistent with much of what is known to date. The specific details as to how such templates are constructed and how the different maps are arranged are not critical to this conclusion and, if a template algorithm is indeed being implemented in MST, the exact map arrangement will likely turn out to be quite different from those specified here. However, as the response properties of higher order neurons in the visual cortex become increasingly complex, an organized collaborative effort between modelers, psychophysicists, and physiologists will be necessary to sort things out. The significance of our model is that it provides a robust, quantitative, and most of all testable, hypothesis which links previous psychophysical and physiological results and will hopefully help guide future psychophysical and physiological studies.

## REFERENCES

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

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

Albright, T. D., Desimone, R. & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology, 51*, 16–31.

Allman, J., Miezin, F. & McGuiness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review of Neuroscience, 8*, 407–430.

Andersen, R. A., Essick, G. K. & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230*, 456–458.

Andersen, R. A., Graziano, M. & Snowden, R. (1990). Translational invariance and attentional modulation of MST cells. *Society for Neuroscience Abstracts, 16*, 7.

Baloh, R. W., Beykirch, Honrubia, V. & Yee, E. D. (1988). Eye movements induced by linear acceleration on a parallel swing. *Journal of Neurophysiology, 60*, 2000 2013.

Bandopadhay, A., Chandra, B. & Ballard, D. H. (1986). Active navigation: Tracking an environmental point considered beneficial. *Proceedings of the Workshop on Motion: Representation and Analysis* (pp. 23 29). Kiawah Island.

Barnes, G. R. & Smith, R. (1981). The effect on visual discrimination of image movement across the stationary retina. *Aviation, Space, and Environmental Medicine, 52*, 466 476.

Blakemore, C. B. & Nachmias, J. (1971) The orientation specificity of two-visual after-effects. *Journal of Physiology, London, 213*, 157 174.

Boussaoud, D., Ungerleider, L. G. & Desimone, R. (1990). Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *Journal of Comparative Neurology, 296*, 462 495.

Braunstein, M. L. & Tittle, J. S. (1988). The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 582 590.

Bremmer, F. & Hoffmann, K. P. (1993). Pursuit related activity in macaque visual cortical areas MST and LIP is modulated by eye-position. *Society for Neuroscience Abstracts, 19*, 1283.

Bruce, C. J., Desimone, R. & Gross, C. G. (1986). Both striate cortex and superior colliculus contribute to visual properties of neurons in superior temporal polysensory area of macaque monkey. *Journal of Neurophysiology, 55*, 1057 1075.

Buizza, A., Leger, A., Droulez, J., Berthoz, A. & Schmid, R. (1980). Influence of otolithic stimulation by horizontal linear acceleration on optokinetic nystagmus and visual motion perception. *Experimental Brain Research, 39*, 165 176.

Busettini, C., Miles, F. A. & Schwarz, Y. (1991). Ocular responses to translation and their dependence on viewing distance. II. Motion of the scene. *Journal of Neurophysiology, 66*, 865–878.

Calvert, E. S. (1954). Visual judgements in motion. *Journal of the Institute of Navigation, London, 7*, 233 251.

Cappozzo, A. (1981). Analysis of the linear displacement of the head and trunk during walking at different speeds. *Journal of Biomechanics, 14*, 411 425.

Collewijn, H. (1977). Gaze in freely moving subjects. In Baker, R. & Berthoz, A. (Eds), *Control of gaze by brain stem neurons* (pp. 13 22). Amsterdam: Elsevier/North-Holland.

Crowell, J. A. & Banks, M. S. (1993). Perceiving heading with different retinal regions and types of optic flow. *Perception & Psychophysics, 53*, 325 337.

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

Cutting, J. E., Springer, K., Braren, P. A. & Johnson, S. H. (1992). Wayfinding on foot from information in retinal, not optical, flow. *Journal of Experimental Psychology: General, 121*, 41 72.

De Bruyn, B. & Orban, G. A. (1990). The role of direction information in the perception of geometric optic flow components. *Perception & Psychophysics, 47*, 433 438.

De Valois, R. L., Yund, E. W. & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22*, 531 544.

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

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

Duffy, C. J. & Wurtz, R. H. (1993). MSTd neuronal responses to the center-of-motion in optic flow fields. *Society for Neuroscience Abstracts, 19*, 1283.

Dursteler, M. R. & Wurtz, R. H. (1988). Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST. *Journal of Neurophysiology, 60*, 940 965.

Dursteler, M. R., Wurtz, R. H. & Newsome, W. T. (1987). Directional pursuit deficits following lesions of the foveal representation within the superior temporal sulcus of the macaque monkey. *Journal of Neurophysiology, 57*, 1263 1287.

Erickson, R. G. & Dow, B. M. (1989). Foveal tracking cells in the superior temporal sulcus of the macaque monkey. *Experimental Brain Research, 78*, 113–131.

Ferman, L., Collewijn, H., Jansen, T. C. & Van den Berg, A. V. (1987). Human gaze stability in the horizontal, vertical and torsional direction during voluntary head movements, evaluated with a three-dimensional scleral induction coil technique. *Vision Research, 27*, 811–828.

Fuchs, A. F., Kaneko, C. R. S. & Scudder, C. A. (1985). Brainstem control of saccadic eye movements. *Annual Review of Neuroscience, 8*, 307 337.

Gellman, R. S., Carl, J. R. & Miles, F. A. (1990). Short-latency ocular following responses in man. *Visual Neuroscience, 5*, 107 122.

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

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, Mass.: Houghton Mifflin.

Gibson, J. J., Olum, P. & Rosenblatt, F. (1955). Parallax and perspective during aircraft landings. *American Journal of Psychology, 68*, 372 385.

Glünder, H. (1990). Correlative velocity estimation: Visual motion analysis, independent of object form, in arrays of velocity-tuned bilocal detectors. *Journal of the Optical Society of America A, 7*, 255 263.

Gresty, M. A. & Bronstein, A. M. (1992). Visually controlled spatial stabilisation of the human head: Compensation for the eye's limited ability to roll. *Neuroscience Letters, 140*, 63 66.

Grossman, G. E. & Leigh, R. J. (1990). Instability of gaze during locomotion in patients with deficient vestibular function. *Annals of Neurology, 27*, 528–532.

Grossman, G. E., Leigh, R. J., Abel, L. A., Lanska, D. J. & Thurston, S. E. (1988). Frequency and velocity of rotational head perturbations during locomotion. *Experimental Brain Research, 70*, 470 476.

Grossman, G. E., Leigh, R. J., Bruce, E. N., Huebner, W. P. & Lanska, D. J. (1989). Performance of the human vestibuloocular reflex during locomotion. *Journal of Neurophysiology, 62*, 264–272.

Hatsopoulos, N. & Warren, W. H. (1991). Visual navigation with a neural network. *Neural Networks, 4*, 303–317.

Heeger, D. J. & Jepson, A. D. (1990). Visual perception of three-dimensional motion. *Neural Computation, 2*, 129 137.

Heeger, D. J. & Jepson, A. D. (1992). Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer, Vision, 7*, 95–177.

Henn, V., Cohen, B. & Young, L. R. (1980). Visual-vestibular interaction in motion perception and the generation of nystagmus. *Neurosciences Research Program Bulletin, 18*, 556 567.

Hikosaka, K., Iwai, E., Saito, H. & Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *Journal of Neurophysiology, 60*, 1615–1637.

Hildreth, E. C. (1992). Recovering heading, for visually-guided navigation. *Vision Research, 32*, 1177 1192.

Holmes, G. (1918). Disturbances of visual orientation. *British Journal of Ophthalmology, 2*, 449–468 and 506–516.

Hubel, D. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.

Hubel, D. H. & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology, 158*, 267–294.

Hubel, D. H., Wiesel, T. N. & Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *Journal of Comparative Neurology, 177*, 361–380.

Israël, I. & Berthoz, A. (1989). Contribution of the otoliths to the calculation of linear displacement. *Journal of Neurophysiology, 62*, 247–263.

Johnston, A. (1986). A spatial property of the retino-cortical mapping. *Spatial Vision, 1*, 319–331.

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

Kawano, K. & Miles, F. A. (1986). Short-latency ocular following responses in monkey. II. Dependence on a prior saccadic eye movement. *Journal of Neurophysiology, 56*, 1355–1380.

Kawano, K. & Sasaki, M. (1984). Response properties of neurons in posterior parietal cortex of monkey during visual-vestibular stimulation. II. Optokinetic neurons. *Journal of Neurophysiology, 52*, 352–360.

Kawano, K., Sasaki, M. & Yamashita, M. (1984). Response properties of neurons in posterior parietal cortex of monkey during visual-vestibular stimulation. I. Visual tracking neurons. *Journal of Neurophysiology, 51*, 340–351.

Keller, E. L. & Heinen, S. J. (1991). Generation of smooth-pursuit eye movements: Neuronal mechanisms and pathways. *Neuroscience Research, 11*, 79–107.

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.

Komatsu, H. & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements I. Location and visual properties of neurons. *Journal of Neurophysiology, 60*, 580–603.

Komatsu, H. & Wurtz, R. H. (1989). Modulation of pursuit eye movements by stimulation of cortical areas MT and MST. *Journal of Neurophysiology, 62*, 31–47.

Lappe, M. & Rauschecker, J. P. (1993). A neural network for the processing of optic flow from ego-motion in man and higher mammals. *Neural Computation, 5*, 374–391.

Lee, D. N. (1974). Visual information during locomotion. In Macleod, R. B. & Pick, H. (Eds), *Perception: Essays in honor of J. J. Gibson* (pp. 250–267). Ithaca, N.Y.: Cornell University Press.

Leigh, R. J. & Brandt, T. (1993). A reevaluation of the vestibulo-ocular reflex: New ideas for its purpose, properties, neural substrate and disorders. *Neurology, 47*, 1288–1295.

Lisberger, S. G., Morris, E. J. & Tyschen, L. (1987). Visual motion processing and sensory motor integration for smooth pursuit eye movements. *Annual Review of Neuroscience, 10*, 97–129.

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

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature, 293*, 133–135.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London B, 208*, 385–387.

Maunsell, J. H. R. & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience, 10*, 363–401.

Maunsell, J. H. R. & Van Essen, D. C. (1983a). The connections of the middle temporal area (MT) and their relationship to cortical hierarchy in the macaque monkey. *Journal of Neuroscience, 3*, 2563–2586.

Maunsell, J. H. R. & Van Essen, D. C. (1983b). Functional properties of neurons in the middle temporal visual area of the Macaque monkey. I. Selectivity for stimulus direction, speed, orientation. *Journal of Neurophysiology, 49*, 1127–1147.

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

Miles, F. A. & Lisberger, S. G. (1981). Plasticity in the vestibulo-ocular reflex: A new hypothesis. *Annual Review of Neuroscience, 4*, 273–299.

Miles, F. A., Kawano, K. & Optican, L. M. (1986). Short-latency ocular following responses of monkey. I. Dependence on temporo-spatial properties of visual input. *Journal of Neurophysiology, 56*, 1321–1354.

Movshon, J. A., Newsome, W. T., Gizzi, M. S. & Levitt, J. B. (1988). Spatio-temporal tuning and speed sensitivity in macaque visual cortical neurons. *Investigative Ophthalmology and Visual Science, 29*, 327.

Murphy, B. J. (1978). Pattern thresholds for moving and stationary gratings during smooth eye movement. *Vision Research, 18*, 521–530.

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

Newsome, W. T., Britten, K. H. & Movshon, J. A. (1989). Neuronal correlates for a perceptual decision. *Nature, 341*, 52–54.

Newsome, W. T., Wurtz, R. H. & Komatsu, H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. II Differentiation of retinal from extraretinal inputs. *Journal of Neurophysiology, 60*, 604–620.

Newsome, W. T., Wurtz, D. H., Dursteler, M. R. & Mikami, A. (1985). Deficits in visual motion processing following ibotenic acid lesions of the middle: Temporal visual area of the macaque monkey. *Journal of Neuroscience, 5*, 825–840.

Ono, M. E., Rivest, J. & Ono, H. (1986). Depth perception as a function of motion parallax and absolute-distance information. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 331–337.

Orban, G. A., Laae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H. & Torre, V. (1992). First-order analysis of optical flow in monkey brain. *Proceedings of the National Academy of Science U.S.A., 89*, 2595–2599.

Paige, G. D. (1989). The influence of target distance on eye movement responses during vertical linear motion. *Experimental Brain Research, 77*, 585–593.

Paige, G. D. & Tomko, D. L. (1991a). Eye movement responses to linear head motion in the squirrel monkey: I. Basic characteristics. *Journal of Neurophysiology, 65*, 1170–1182.

Paige, G. D. & Tomko, D. L. (1991b). Eye movement responses to linear head motion in the squirrel monkey: II. Visual-vestibular interactions and kinematic considerations. *Journal of Neurophysiology, 65*, 1183–1196.

Perrone, J. A. (1986). Anisotropic responses to motion toward and away from the eye. *Perception & Psychophysics, 39*, 1–8.

Perrone, J. A. (1987). Extracting 3-D egomotion information from a 2-D flow field: A biological solution? *Optical Society of America Technical Digest Series, 22*, 47.

Perrone, J. A. (1990). Simple technique for optical flow estimation. *Journal of the Optical Society of America A, 7*, 264–278.

Perrone, J. A. (1992). Model for the computation of self-motion in biological systems. *Journal of the Optical Society of America, A, 9*, 177–194.

Perrone, J. A. & Stone, L. S. (1991). The perception of egomotion: Global versus local mechanisms. *Investigative Ophthalmology and Visual Science (Suppl.), 32*, 957.

Perrone, J. A. & Stone, L. S. (1992a). A possible role for the speed tuning properties of MT cells in the recovery of depth from motion. *Investigative Ophthalmology and Visual Science (Suppl.), 33*, 1141.

Perrone, J. A. & Stone, L. S. (1992b). Using the properties of MT neurons for self-motion estimation in the presence of eye-movements. *Perception (Suppl. 2), 21*, 64.

Peterka, R. J. (1992). Response characteristics of the human torsional oculomotor reflex. *Annals of the New York Academy of Sciences, 656*, 877–879.

Pozzo, T., Berthoz, A. & Lefort, L. (1990). Head stabilization during various oculomotor tasks in humans. *Experimental Brain Research, 82*, 97–106.

Regan, D. & Beverley, K. I. (1978). Looming detectors in the human visual pathway. *Vision Research, 18*, 415–421.

Regan, D. & Beverly, K. I. (1979). Visually guided locomotion: Psychophysical evidence for a neural mechanism sensitive to flow patterns. *Science, 205*, 311–313.

Rieger, J. H. & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A, 2*, 354–360.

Rieger, J. H. & Toet, L. (1985). Human visual navigation in the presence of 3D rotations. *Biological Cybernetics, 52*, 377–381.

Robinson, D. A. (1981). Control of eye movements. In Brooks (Ed.), *Handbook of physiology, Section 1* (Chap. 28, pp. 1275–1320). Bethesda, Md: Wilkins & Wilkins.

Rogers, B. & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*, 125–134.

Roy, J. P. & Wurtz, R. H. (1990). The role of disparity-sensitive cortical neurons in signalling the direction of self-motion. *Nature (London), 348*, 160–162.

Roy, J. P., Kamatsu, H. & Wurtz, R. H. (1992). Disparity sensitivity of neurons in monkey extrastriate area MST. *Journal of Neurocience, 12*, 2478–2492.

Royden, C. S., Banks, M. S. & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature, 360*, 583–585.

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*, 145–157.

Salzman, C. D., Britten, K. H. & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature, 346*, 174–177.

Sandini, G. & Tistarelli, M. (1990). Active tracking strategy for monocular depth inference over multiple frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 13–27.

Sandini, G., Tagliasco, V. & Tistarelli, M. (1986). Analysis of object motion and camera motion in real scenes. *Proceedings of the IEEE Conference on Robotics and Automation* (pp. 627–633). San Francisco, Calif.

Schiller, P. H., Finlay, B. L. & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. *Journal of Neurophysiology, 39*, 1320–1333.

Schwarz, U. & Miles, F. A. (1991). Ocular responses to translation and their dependence on viewing distance. I. Motion of the observer. *Journal of Neurophysiology, 66*, 851–864.

Schwarz, U., Busettini, C. & Miles, F. A. (1989). Ocular responses to linear motion are inversely proportional to viewing distance. *Science, 245*, 1394–1396.

Smith, R. (1985). Vergence eye-movement responses to whole-body linear acceleration stimuli in man. *Ophthalmology and Physiological Optics, 5*, 303–311.

Solomon, D. & Cohen, B. (1992). Stabilization of gaze during circular locomotion in light I. Compensatory head and eye nystagmus in the running monkey. *Journal of Neurophysiology, 67*, 1146–1157.

Sparks, D. L. & Mays, L. E. (1990). Signal transformations required for the generation of saccadic eye movements. *Annual Review of Neuroscience, 13*, 309–336.

Steinmetz, M. A., Motter, B. C., Duffy, C. J. & Mountcastle, V. B. (1987). Functional properties of parietal visual neurons: Radial organization of directionalities within the visual field. *Journal of Neuroscience, 7*, 177–191.

Stone, L. S. & Lisberger, S. G. (1989). Synergistic action of complex and simple spikes in the monkey flocculus in the control of smooth pursuit eye movement. *Experimental Brain Research (Suppl.), 17*, 299–312.

Stone, L. S. & Perrone, J. A. (1991). Human heading perception during combined translational and rotational self-motion. *Society for Neuroscience Abstracts, 17*, 847.

Stone, L. S. & Perrone, J. A. (1993). Human heading perception cannot be explained using a local differential motion algorithm. *Investigative Ophthalmology and Visual Science, 34*, 1229.

Takahashi, M., Hoshikawa, H., Tsujita, N. & Akiyama, I. (1988). Effect of labarinthine dysfunction upon head oscillation and gaze during stepping and running. *Acta Otolaryngolica, 106*, 348–353.

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

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.

Tanaka, K., Sugita, Y., Moriya, M. & Saito, H. (1993). Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *Journal of Neurophysiology, 69*, 128–142.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y. & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience, 6*, 134–144.

Thiers, P. & Erickson, R. G. (1992). Vestibular input to visual-tracking neurons in area MST of awake rhesus monkeys. *Annals of the New York Academy of Science, 656*, 960–963.

Thomas, J. P. & Gille, J. (1979). Bandwidths of orientation channels in human vision. *Journal of the Optical Society of America, 69*, 652–660.

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S. & Imbert, M. (1992). Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science, 257*, 1279–1281.

Ungerleider, L. G. & Desimone, R. (1986). Cortical connections of area MT in the macaque. *Journal of Comparative Neurology, 248*, 190–222.

Van Essen, D. C., Munsell, J. H. R. & Bixby, J. L. (1981). The middle temporal visual area in the macaque: Myeloarchitecture, connections, functional properties and topographic representation. *Journal of Comparative Neurology, 199*, 293–326.

Verri, A., Straforini, M. & Torre, V. (1992). A model of spatial organisation of the receptive fields of neurons in the middle superior temporal area. *Perception (Suppl. 2), 21*, 64.

Vieville, T. & Masse, D. (1987). Ocular counter-rolling during active head tilting in humans. *Acta Oto-Laryngologica, 103*, 280–290.

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

Warren, W. H. & 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 A, 7*, 160–168.

Warren, W. H. & Kurtz, K. J. (1992). The role of central and peripheral vision in perceiving the direction of self-motion. *Perception & Psychophysics, 51*, 443–454.

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

Waters, R. L., Morris, J. & Perry, J. (1973). Translational motion of the head and trunk during normal walking. *Journal of Biomechanics, 6*, 167–172.

Weisel, T. N., Hubel, D. H. & Lam, D. M. K. (1974). Autoradiographic demonstration of ocular-dominance columns in the monkey striate cortex by means of transneuronal transport. *Brain Research, 79*, 273–279.

Westheimer, G. & McKee, S. P. (1975). Visual acuity in the presence of retinal-image motion. *Journal of the Optical Society of America, 65*, 847–851.

Wilson, V. J. & Mellville-Jones, G. (1979). *Mammalian vestibular physiology*. New York: Plenum Press.

Zacharias, G. L., Caglayan, A. K. & Sinacori, J. B. (1985). A visual cueing model for terrain-following applications. *Journal of Guidance, 8*, 201–207.

Zeki, S. M. (1980). The response properties of cells in the middle temporal area (area MT) of owl monkey visual cortex. *Proceedings of the Royal Society of London B, 207*, 239–248.

*Acknowledgements*—The authors thank Drs Preeti Verghese, Brent Beutter and Irv Statler for their critical reading of an earlier draft and helpful comments. This work was supported by NASA RTOPs Nos 199-12-06-12-24, 199-16-12-37, 505-64-53 and NASA NCC 2-307.