Handbook of Brain Theory and Neural Networks : Table of Contents : Motion Perception: Navigation
«« Previous Next »»

Motion Perception: Navigation

Constance S. Royden and Ellen C. Hildreth

Introduction

When an observer moves through the world, the resulting image motion on the retina, known as *optical flow*, can inform the observer about his own motion through space and about the three-dimensional (3D) structure and motion of objects in the scene. This information is essential for tasks such as the visual guidance of locomotion through the environment and the manipulation and recognition of objects.

This article focuses on the recovery of observer motion from optical flow. We include strategies for detecting moving objects and avoiding collisions, discuss how this information may be used to control actions, and describe the neural mechanisms underlying heading perception.

The Image Flow Field

This section describes the relationship between two-dimensional (2D) image motion and the 3D translation and rotation of the observer relative to the scene. The mechanisms for deriving the 2D image velocities that result when the observer or objects move are described elsewhere (Hildreth and Koch, 1987; Mitiche and Bouthemy, 1996; see also MOTION PERCEPTION, ELEMENTARY MECHANISMS).

Consider an observer moving relative to a stationary scene, with a coordinate system fixed to the observer and the *z*-axis directed along the optical axis. The instantaneous translation of the observer can be expressed in terms of translation along three orthogonal directions, given by the vector **T** = $(T_x, T_y, T_z)^T$. Observer rotation can be expressed in terms of rotation around each of these axes, given by **R** = $(R_x, R_y, R_z)^T$. Let **P** = $(X, Y, Z)^T$ be the position of a point in space, as shown in Figure 1. The 3D velocity of **P** in the observer's coordinate frame is given by

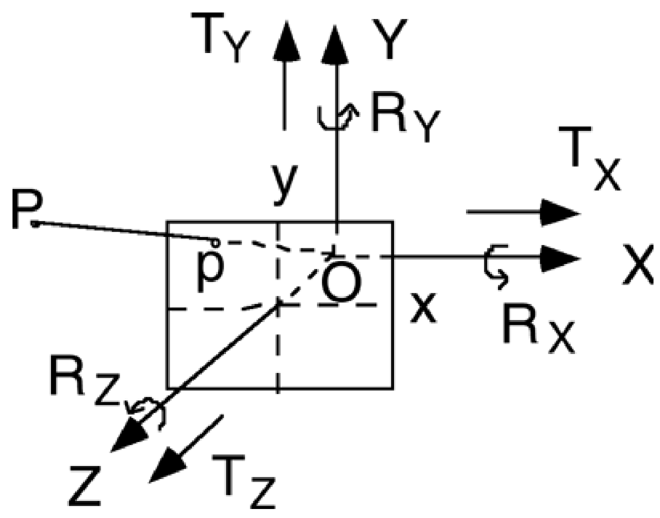

Figure 1. Coordinate system for a moving observer who is located at the origin.

$$\mathbf{V} = (\dot{X}, \dot{Y}, \dot{Z})^T = -\mathbf{T} - \mathbf{R} \times \mathbf{P}$$

where

$$\begin{aligned} \dot{X} &= -T_x - R_y Z + R_z Y \\ \dot{Y} &= -T_y - R_z X + R_x Z \\ \dot{Z} &= -T_z - R_x Y + R_y X \end{aligned}$$

If we assume perspective projection onto the image plane, using a focal length of 1, the projection of **P** onto the image (*x*, *y*) is given by

$$x = X/Z, \quad y = Y/Z$$

The projected velocities in the image plane (v_x, v_y) are therefore

$$v_x = (-T_x + xT_z)/Z + R_x xy - R_y(x^2 + 1) + R_z y$$
$$v_y = (-T_y + yT_z)/Z + R_x(y^2 + 1) - R_y xy - R_z x$$

The first term represents the component of image velocity due to observer translation and depends on the depth Z of each point in the scene. The remaining terms represent the component of image velocity due to the observer's rotation and do not depend on depth. The translational component yields a radial pattern of velocity (Figure 2*A*) that emanates from a single location in the image, called the *focus of expansion* (FOE). The FOE corresponds to the observer's heading and occurs at the location (T_X/T_Z, T_Y/T_Z) in the image. In contrast, the image flow field that results from a pure rotation of the observer is nearly constant over this region of the image. The image flow field for combined translation and rotation of the observer (Figure 2*B*) is the vector sum of the two flow fields from translation and rotation.

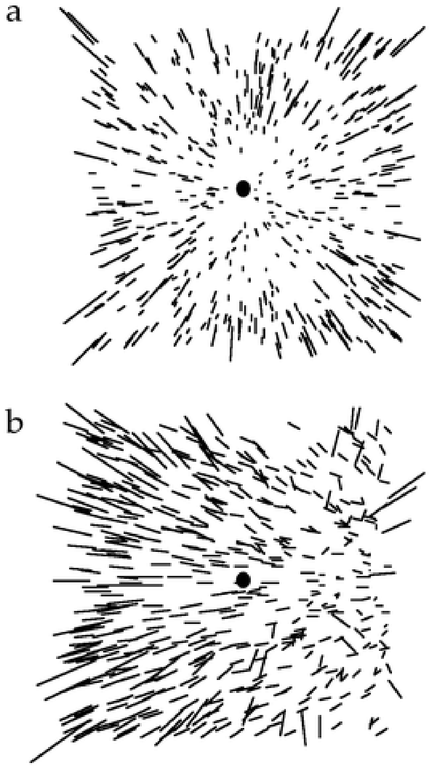

Figure 2. *A*, Radial flow field resulting from an observer moving in a straight line through a 3D cloud of dots. *B*, Flow field resulting from an observer translating and rotating (about a vertical axis) through a 3D cloud of dots. The solid circles indicate the direction of observer translation.

For observer translation, one can locate the FOE by finding the point of intersection of lines through the velocity vectors in the image. This simple strategy fails for combined translation and rotation, which occurs when the observer moves along a curved path or rotates his eyes or head, because the additional rotation components of velocity eliminate the FOE. This strategy also fails for nonrigid scenes that contain moving objects whose paths of motion deviate from the radial translational flow lines.

The Perception of Heading

Perceptual studies show that people judge 3D motion accurately under many conditions (Hildreth and Royden, 1998; Warren, 1998a; van den Berg, in Lappe, 2000). When translating toward a stationary scene, people exhibit discrimination thresholds as low as 0.2° when the heading is near the line of sight. Thresholds rise with more peripheral headings. Heading judgments are performed successfully with sparse, discontinuous flow fields, and require a relatively small field of view if the rotational flow is small. For pure translation, people recover heading with moderate accuracy from only a 90 ms presentation, but improve up to about 300 ms of viewing time. Heading judgments remain accurate in the presence of moderate amounts of noise in the image flow field. The addition of other static and stereoscopic depth cues can enhance the accuracy of heading judgments in the presence of added noise or observer rotations.

People judge their translational heading accurately in the presence of small rotational rates generated by slow eye movements. Faster eye movements require extraretinal information about the speed of eye movement for accurate heading recovery (Hildreth and Royden, 1998; Warren, 1998a). In contrast, people accurately judge their motion along curved paths at low and high rotation rates (Warren, 1998a; van den Berg, in Lappe, 2000).

Under many conditions, a moving object has no effect on heading judgments. However, when the object crosses the observer's path, small biases in observer heading judgments result (Royden and Hildreth, 1996; Warren, 1998a). The ability to judge heading does not deteriorate when observers attend a second, object-related task (Hildreth and Royden, 1998).

These observations suggest that the human mechanism for judging heading from visual stimuli is remarkably robust and performs well under a variety of nonoptimal conditions.

Models of Observer Motion Recovery

Computational approaches for recovering heading can be divided into several categories, as described in this section. Many models fit more than one category; we present individual models within a category that best represents the particular approach.

Discrete Models

In discrete models, image features are tracked over time. Their sequence of positions forms the input to a system of equations whose solution yields the parameters of 3D structure and motion, assuming that the features move in a rigid configuration relative to the observer. Computer experiments indicate that such algorithms are vulnerable to error in the image motion measurements, although the use of motion measurements over an extended time can yield better performance (Martin and Aggarwal, 1988).

Differential Models

These models recover 3D motion and structure parameters from first and second spatial derivatives of the image flow field. One approach uses the *differential invariants* of the flow field, divergence (expansion/contraction), curl (rotation), and two components of deformation, dilation and shear (for references see Hildreth and Royden, 1998). Divergence and deformation depend only on the observer’s translation and surface slant, and are invariant under observer rotations. In principle, these measures can be used to recover the observer’s translation and the 3D shape of object surfaces. Most models that use differential invariants require a continuous, smooth optical flow field. In contrast, the human system can recover heading reliably from a few, sparse features that are sampled from a discontinuous flow field.

Motion Parallax Models

Motion parallax models use the fact that the translational components of the image velocities depend on the depth of the points in the scene, while the rotational components are independent of depth (Longuet-Higgins and Prazdny, 1980). Consequently, subtracting the image velocities from two points located at a depth discontinuity eliminates the rotational components. One can locate the translational heading using the resulting “difference vectors” by calculating the best point of intersection of lines through these vectors.

These models provide a method for quickly assessing heading, independent of the recovery of observer rotation and 3D scene structure. Because they combine information from multiple velocity vectors, they work fairly well in the presence of noisy velocity inputs. Simulations with a motion parallax model developed by Hildreth (1992) show behavior consistent with that observed in earlier perceptual studies. Motion differences computed by neurons in the middle temporal (MT) area of the primate visual system may be used to compute observer translation in the presence of rotations (Royden, 1997).

Error Minimization Models

Error minimization models compute observer motion and 3D structure parameters that yield a flow field that best fits the measured optical flow. For example, Bruss and Horn use this approach to derive observer motion parameters and surface structure that best account for the measured flow field in a least-squares sense (see Hildreth and Royden, 1998). The error minimization strategy, together with spatial pooling of motion measurements over an extended image region, allows the algorithm to tolerate substantial error in the individual image motion measurements. Many models proposed for recovery of observer motion incorporate some form of error minimization. Notably, Heeger and Jepson (1992) presented an error minimization model that has been implemented in a neural network form by Lappe and Rauschecker (in Lappe, 2000).

Template Models

Template models use special-purpose computational mechanisms, such as a family of templates, tailored to detect patterns of optical flow corresponding to specific observer motion parameters. For example, a template for detecting forward motion along the line of sight would respond optimally to a radially expanding pattern of image velocities whose FOE is located at the center of the visual field. Template models deal effectively with noise in the input velocity measurements by integrating over a large area. Perrone and Stone (1994) proposed a template model that computes heading using components that respond to motion similarly to neurons in the primate visual area MT.

Eye Movement Models

In addition to retinal information, the human visual system can use the oculomotor signal to obtain information about eye rotation, either from an efference copy of the signal or from proprioceptive feedback from the extraocular muscles. Royden, Crowell, and Banks suggest that this information is essential for recovering heading accurately in the presence of fast eye rotations. The flow field corresponding to a known eye rotation could be subtracted from the overall flow field before the observer’s heading is calculated (Hildreth and Royden, 1998). Lappe (in Lappe, 2000) and van den Berg and Beintema (see Lappe, in Lappe, 2000) have developed neural models that explicitly incorporate eye movement signals.

Cutting (1986) noted that when an observer fixates a point in space, the most rapidly moving objects in the vicinity of the fixation point can be used to judge heading relative to the fixation direction. One can locate the heading with successive fixations on objects in the scene. This model requires little computation from the flow field itself; however, it fails for certain configurations of scenes and eye fixations. It also requires multiple saccades to locate heading, something that is not essential for human heading judgments.

Neural Network Models

Several neural network models use training algorithms to learn to compute heading from optic flow input (Lappe, in Lappe, 2000). Hatsopoulos and Warren created a two-layer neural network that is trained using the Widrow-Hoff learning rule to recognize the correct translational heading for an observer moving along a straight line. The input layer consists of units tuned to direction and speed of motion. After training, the weights connecting the input and output layers adapt so that the output neurons detect radial patterns of motion corresponding to particular headings. The network only interprets flows derived from pure observer translation. Zemel and Sejnowski developed a learning network that segments the scene according to the motion of objects relative to the observer. Heading can be estimated from the resulting encoding.

Motion on a Curvilinear Path

When an observer moves along a curvilinear path, his instantaneous translation and rotation are the same as those for an observer pursuing straight-line motion with eye movement, resulting in an ambiguity. Distinguishing these situations requires an analysis of the flow field over an extended time. Alternatively, eye movement information may be used to disambiguate these conditions. Human observers distinguish curved from straight paths with high accuracy and judge the path curvature well. This finding suggests that the visual system computes both translation and rotation components of observer motion (Warren, 1998a; van den Berg, in Lappe, 2000).

Coping with Moving Objects

For most models, the presence of moving objects in the scene can adversely affect the derivation of observer motion. Image points associated with moving objects may move in a direction inconsistent with the observer’s motion, causing errors in the heading estimate. Some models first detect moving objects and then compute heading from the remaining stationary components of the scene. Another approach computes an initial estimate of observer motion by combining all available data or by performing separate computations within limited image regions. One can then identify moving objects by finding areas of the scene for which the image motion differs significantly from that expected from these initial motion parameters (e.g., Hildreth, 1992). See Hildreth and Royden (1998) for a review of models of moving object detection.

Visuomotor Transformations for Navigation

Successful navigation requires that visual information be used to control motor actions to move through the world. This requires a transformation between the retinocentric heading coordinates computed by the models described in the previous section to a body-centered coordinate system, taking into account eye and head movements. It seems likely that the visual system uses extraretinal information, such as eye movement and vestibular signals, to account for rotations of the head and eyes. Neurons in the medial superior temporal (MST) area of visual cortex may combine these extraretinal signals with visual information to compute the body-centric heading (see Andersen et al., in Lappe, 2000).

The mechanisms for transformation of the visual information into motor control commands are not yet understood, but several approaches have been described. In one approach, motor planning takes place based on the computed heading of the observer. For example, to reach a desired goal, the motor system could initiate turning commands that minimize the error between the computed heading and the direction to the goal. Visual feedback allows constant refinement of the motor strategy to keep errors in heading from accumulating (Warren, 1998b).

Another approach is based on specific tasks the observer must perform to navigate through the environment. Such tasks include steering toward a goal, pursuing prey, braking, avoiding obstacles, or computing time to contact with an approaching surface. It has been suggested that each of these tasks may be accomplished through a task-specific subsystem that uses only the information in the flow field necessary to complete the task (Aloimonos, 1997; Warren, 1998b). For example, time to contact can be computed from the ratio, τ, given as the ratio of size/(rate of size change). The coupling between the task-specific information and the resulting action can be modeled as a nonlinear dynamical system. For example, when steering toward a goal, visual information provides the angle, β, between the FOE and the direction of the goal. This angle can be used to control the observer’s rate of turning. The result is a system with a stable fixed point at $\beta = 0$, corresponding to the observer heading toward the goal. Complex motor behavior may emerge through interactions between loosely coupled subsystems underlying different tasks (Warren, 1998b).

Neural Mechanisms of Heading Computation

In primates, visual area MST is probably involved in computing heading (Duffy, in Lappe, 2000). Neurons in MST respond well to large motion patterns and receive direct input from cells in area MT, which is known to process motion (see also Motion Perception, Elementary Mechanisms). Some MST neurons prefer expanding or contracting radial patterns of motion, as would be generated by an observer moving in a straight line forward or backward. These cells have different preferred centers of expansion, so they could be involved in finding the FOE in an optical flow field. Other cells respond well to uniform motion in a single direction, and yet others respond to rotating patterns of motion. Many cells respond to some combination of these.

It is unclear how these cell responses contribute to the computation of heading in the presence of rotations; however, several models have been developed that could explain this. Hatsopoulos and Warren (Warren, 1998a) and Perrone and Stone (1994) proposed template models that use components that behave similarly to neurons in area MT in their response to motion. In both models, these components connect to another layer of cells with properties similar to the cells in MST. The connection patterns are such that the cells in the second layer respond to spatial patterns that mimic the flow fields that result from particular observer motion parameters. The Hatsopoulos and Warren model deals only with pure observer translation. Perrone and Stone used templates that deal with combinations of translations and the rotations that result when the observer makes eye movements to track an object in the scene. This model also recovers the relative depths of surfaces in the scene.

Royden (1997) developed a model that makes use of the motion-opponent properties of MT neurons to deal with observer rotations. Many neurons in MT have both excitatory and inhibitory regions within their receptive fields. The Royden model uses operators with this receptive-field layout to eliminate the observer rotation at the initial processing stage. These cells project to a second layer of cells, similar to those in MST, that are tuned to radial patterns of input. As with the motion parallax models described earlier, the centers of these radial patterns correspond to observer headings.

Finally, Lappe (in Lappe, 2000) and van den Berg and Beintema (cited by Lappe, in Lappe, 2000) developed neural models that explicitly incorporate eye movement signals to deal with rotations generated by eye movements. In Lappe's model, extraretinal input compensates for the image motion induced by eye movements. In van den Berg and Beintema's model, the responses of template cells tuned to retinal flow are multiplied by a "rate-coded" measure of eye velocity, producing a layer of cells that have a preferred flow field that changes dynamically to compensate for eye movements.

Currently, there is insufficient physiological or psychophysical data to distinguish among these models of neural computation of heading. It seems likely that some compensation for eye movements occurs in area MST (see Andersen et al., in Lappe, 2000); however, this compensation could be incorporated into the models that do not currently use it. The models are all reasonably consistent with the known behavior of MT and MST cells. Determination of which, if any, most accurately describes the neural computation awaits further experimentation.

Discussion

People judge heading well under many conditions; however, it is still uncertain how the visual system accomplishes this task. Superficially, most of the models cited here exhibit general biological plausibility, in that they can be implemented by a network of simple, local processing mechanisms operating in parallel. Physiological observations reveal the general properties of the representation of optic flow information and provide some indication that heading computations take place in areas MT and MST of the primate visual system. It remains a challenge to incorporate all of the important aspects of recovery of 3D observer motion into a neuronal model that exhibits a broad range of human behavior and incorporates the details of physiological observations.

Road Map: Vision

Related Reading: Motion Perception, Elementary Mechanisms ◊ Robot Navigation

References

Aloimonos, Y., Ed., 1997, *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, Mahwah, NJ: Erlbaum.

Cutting, J. E., 1986, *Perception with an Eye for Motion*, Cambridge, MA: Bradford/MIT Press.

Heeger, D. J., and Jepson, A. D., 1992, Subspace methods for recovering rigid motion: I. Algorithm and implementation, *Int. J. Comput. Vision*, 7:95–117.

Hildreth, E. C., 1992, Recovering heading for visually-guided navigation, *Vision Res.*, 32:1177–1192.

Hildreth, E. C., and Koch, C., 1987, The analysis of visual motion: From computational theory to neuronal mechanisms, *Annu. Rev. Neurosci.*, 10:477–533.

Hildreth, E. C., and Royden, C. S., 1998, Computing observer motion from optic flow, in *High-Level Motion Processing: Computational, Neurobiological and Psychophysical Perspectives* (T. Watanabe, Ed.), Cambridge, MA: MIT Press, pp. 269–293. ◆

Lappe, M., Ed. 2000, *Neuronal Processing of Optic Flow, Int. Rev. Neurobiol*. vol. 44. (special issue),

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

Martin, W. N., and Aggarwal, J. K., Eds., 1988, *Motion Understanding: Robot and Human Vision*, Boston: Kluwer.

Mitiche, A., and Bouthemy, P., 1996, Computation and analysis of image motion: A synopsis of current problems and methods, *Int. J. Comput. Vision*, 19:29–55. ◆

Perrone, J. A., and Stone, L. S., 1994, A model of self-motion estimation within primate extrastriate visual cortex, *Vision Res.*, 34:2917–2938.

Royden, C. S., 1997, Mathematical analysis of motion-opponent mechanisms used in the determination of heading and depth, *J. Opt. Soc. Am. A*, 14:2128–2143.

Royden, C. S., and Hildreth, E. C., 1996, Human heading judgments in the presence of moving objects, *Percept. Psychophy.*, 58:836–856.

Warren, W. H., 1998a, The state of flow, in *High-Level Motion Processing: Computational, Neurobiological and Psychophysical Perspectives* (T. Watanabe, Ed.), Cambridge, MA: MIT Press, pp. 315–358. ◆

Warren, W. H., 1998b, Visually controlled locomotion: 40 years later, *Ecol. Psychol.*, 10:177–219.

«« Previous   Next »»