Vision Res. Vol. 32, No. 6, pp. 1177–1192, 1992
Printed in Great Britain. All rights reserved

0042-6989/92 $5.00 + 0.00

Recovering Heading for Visually-Guided Navigation

ELLEN C. HILDRETH*

Received 11 July 1991; in revised form 23 December 1991

We present a model for recovering the direction of heading of an observer who is moving relative to a scene that may contain self-moving objects. The model builds upon an algorithm proposed by Rieger and Lawton, based on earlier work by Longuet-Higgins and Prazdny. The algorithm uses velocity differences computed in regions of high depth variation to locate the *focus of expansion*, which indicates the observer's heading direction. We relate the behavior of the model to psychophysical observations regarding the ability of human observers to judge heading direction, and show how the model copes with self-moving objects in the environment.

Visual motion processing Egomotion Visually-guided navigation Motion perception Computational vision

INTRODUCTION

Relative movement in the changing visual image provides a primary cue to the three-dimensional (3-D) structure and motion of object surfaces, and the movement of the observer relative to the scene, allowing biological systems to navigate quickly and efficiently through the environment. This paper considers two aspects of the moving observer and environment that are critical to navigation: the recovery of the 3-D direction of heading of the observer relative to the scene and the segmentation of the scene into distinct objects on the basis of spatial discontinuities in motion. With regard to segmentation, we focus on the task of distinguishing between objects that are stationary with respect to the environment and those that undergo their own self-movement.

To motivate this work, consider an observer moving rapidly through a cluttered scene toward a moving or stationary target, while avoiding obstacles in his path. The observer must continually assess his 3-D direction of translation relative to the target, in order to make constant, correct adjustments of his heading direction to maintain a trajectory toward the target. In principle, either the absolute or relative directions of translation of the observer and target could be computed, but for the purpose of tracking, the observer must at least judge reliably whether he is heading to the left or right of the target. The observer must also monitor his heading relative to object surfaces in order to detect potential collisions with stationary or moving objects in the scene.

The judgment of relative 3-D heading alone is not sufficient to support navigation. It is also necessary to locate object boundaries from discontinuities in motion or other visual properties. Such boundaries are used in many ways. First, the rapid detection of motion discontinuities quickly draws the observer's attention to regions of the image containing objects that could collide with the observer and allows the segmentation of a target from a moving background. Second, the localization of object boundaries allows an assessment of the size and shape of relevant objects in the scene. If an object is moving directly toward the observer, this information is needed to determine an appropriate avoidance movement that steers the observer clear of the approaching object. If the object is a target being tracked, knowledge of its size and shape allows an assessment of its center of mass, which can serve as the focus of the observer's approach.

Finally, segmentation is essential for computing relative heading reliably and accurately, as it allows the observer to integrate only those motion measurements contained within single objects to compute their properties of motion. Without segmentation, the computation of 3-D motion parameters can be degraded by the inclusion of motion measurements from adjacent object surfaces undergoing different motions. Patterns of movement created by multiple objects undergoing self-motion can mimic velocity patterns that normally arise in critical situations such as a directly approaching object. For example, a set of objects positioned around a circle and moving away from the center of the circle mimic the pure expansion that is characteristic of an approaching object. The detection of object boundaries from motion discontinuities allows the distinction of these situations. For obstacle avoidance, it is further useful to distinguish whether an approaching surface is stationary relative to the background, or undergoing its own motion, because self-moving objects may undergo accelerative components of motion.

*Department of Computer Science, Wellesley College, Wellesley, MA 02181, U.S.A.

This paper focuses on the computation of the 3-D direction of translation of an observer relative to object surfaces. After presenting some theoretical preliminaries, we review existing perceptual literature regarding the ability of human observers to judge heading direction. We then consider existing algorithms for performing this computation in light of these perceptual observations. This analysis leads us to focus on a model proposed by Rieger and Lawton (1985) that exhibits some of the behavior observed in human judgments of heading. We present some modifications to Rieger and Lawton's model aimed at improving its performance in the presence of image noise and allowing it to cope with self-moving objects in the scene. The new model provides a partial segmentation of the image as a by-product of the 3-D heading computation. The results of computer simulations address the behavior of this model when applied to visual patterns similar to those used in perceptual studies and synthetic images of scenes containing self-moving objects. Finally, we list a number of questions that arise from this work that could form the basis for further perceptual experiments in this area. A more extended discussion of the relevance of this work to visually-guided navigation can be found in Hildreth (1990).

DERIVING 3-D DIRECTION OF TRANSLATION —THEORETICAL PRELIMINARIES

This section presents the equations relating image motion measurements to the parameters of translation and rotation of the observer relative to the scene. We assume that the observer is moving relative to a stationary scene, but the same geometric relationships hold locally for the case where an object is moving rigidly relative to the observer. We assume that a coordinate system is fixed with respect to the observer, with the Z-axis directed along the optical axis. The translation of the observer can be expressed in terms of translation along three orthogonal directions, which we denote by the vector $\mathbf{t} = (t_x, t_y, t_z)^T$, and the rotation of the observer can be expressed in terms of rotation around three orthogonal axes, which we denote by the vector $\mathbf{w} = (w_x, w_y, w_z)^T$. Let the position of a point P in space be given by the coordinate vector $\mathbf{r} = (X, Y, Z)^T$. Then the 3-D velocity of P in the observer's coordinate frame is given by:

$$\mathbf{V} = (\dot{X}, \dot{Y}, \dot{Z})^T = -\mathbf{t} - \mathbf{w} \times \mathbf{r}$$

where

$$\dot{X} = -t_x - w_y Z + w_z Y$$
$$\dot{Y} = -t_y - w_z X + w_x Z$$
$$\dot{Z} = -t_z - w_x Y + w_y X.$$

If we assume perspective projection of velocity $\mathbf{V}$ onto the image plane, with a focal length for the projection of 1, the projection of P onto the image (x,y) is given by:

$$x = \frac{X}{Z} \quad y = \frac{Y}{Z}.$$

The projected velocities in the image plane $(\dot{x}, \dot{y})$ are then given by:

$$\dot{x} = \frac{-t_x + x t_z}{Z} + w_x xy - w_y(x^2 + 1) + w_z y$$

$$\dot{y} = \frac{-t_y + y t_z}{Z} + w_x(y^2 + 1) - w_y xy - w_z x.$$

The first term represents the component of image velocity due to the translation of the observer and depends on the depth Z to each point in the scene. The remaining terms represent the component of velocity due to the observer's rotation and depend only on the rotation parameters and image location. The translational component alone yields a radial pattern of velocity, which in the case of forward translation, emanates from a single location in the image referred to as the *focus of expansion* (FOE), corresponding to the observer's direction of heading. This translational component depends on the ratios of the three translation parameters to depth Z, so it is not possible from motion information alone to recover the absolute translation and depth parameters.

THE PERCEPTION OF OBSERVER TRANSLATION

Although the image velocity field contains components of motion due to the observer's rotation and translation, psychophysical studies have concentrated on our ability to judge direction of translation. Navigation tasks impose severe demands on our ability to perform this computation. Cutting (1986) showed that under reasonable assumptions, we require an accuracy of about 1° of visual arc in our judgment of heading in order to avoid obstacles successfully while running and driving, as well as performing more challenging tasks such as downhill skiing and aircraft landing. This section reviews perceptual studies of the ability of human observers to judge their direction of translation, which suggest that the human system can achieve this degree of accuracy under the best conditions. We summarize some of these studies in detail, as they form the basis for computer simulations described later.

A series of experiments by Warren and his colleagues (Warren & Hannon, 1988, 1990; Warren, Morris & Kalish, 1988) measured the accuracy with which observers judge their heading direction in computer displays that simulate movement toward a planar surface or 3-D cloud of random dots. The first experiments simulated movement along a ground plane extending to a visible horizon. A target vertical line segment was located on the horizon and the subjects' task was to judge whether their direction of heading was to the left or right of the vertical target. In the initial experiments, the target was visible throughout the motion of the points, but in subsequent experiments, the target only appeared after the points stopped moving. Factors that were varied in these experiments include the orientation of the plane relative to the viewer, the observer's speed and direction of heading, dot density and the temporal extent of the motion. In later studies, movement was simulated relative to a 3-D volume of random dots.

The general conclusion of the studies by Warren and his colleagues is that human observers can judge their heading direction with an accuracy of 1–2° of visual angle, for a variety of surface types and under a range of experimental conditions. Performance is the same, regardless of whether the vertical target line is visible during the movement of the points. Observers perform better with higher speeds of translation, consistent with earlier observations by Johnston, White and Cumming (1973) and Carel (1961).

Warren and Hannon (1988, 1990) compared performance under three conditions: (1) the observer fixated a stationary marker and the displays only simulated pure translation of the observer; (2) the observer tracked a moving point, introducing a rotational component of motion; and (3) the display itself contained both translational and rotational components of motion and the observer maintained stationary fixation. For conditions (2) and (3), the same flow pattern appears on the surface of the eye, but in condition (2), rotational information could be derived from extraretinal eye movement signals, while in condition (3), such information must be derived from visual input alone. Subjectively, observers cannot distinguish between conditions (2) and (3), and in the latter case, there was a strong illusion of the eye actually moving. For the case of movement toward a ground plane or movement toward a cloud of random dots, there was essentially no difference in performance between these three conditions. When simulating translation perpendicular to a plane, however, performance still reached a high level of accuracy in the first two conditions, but was at chance for the third condition. Subjectively, observers perceived themselves as moving toward the point of fixation, which corresponds to a center of outflowing motion in this case. Similar observations regarding movement toward a frontoparallel plane were made in other studies (Llewellyn, 1971; Johnston *et al.*, 1973; Regan & Beverley, 1982; Rieger & Toet, 1985; Cutting, 1986). This observation suggests that extraretinal information regarding eye rotation is used in the analysis of heading direction, and that the passive decoupling of the rotational and translational components of motion from visual input alone requires differential motion produced by elements at different depths.

Warren and Hannon (1990) also examined the influence of dot density for simulated movement toward a 3-D cloud of dots. When the added rotational flow was generated by the subject tracking a dot on the display, there was no change in performance with dot density [confirming earlier observations by Warren *et al.* (1988)], but when rotational flow was added to the movements of the points, there was some degradation of performance with lower densities. Thus, observers could accurately judge heading direction when presented with a relatively sparse, discontinuous flow field.

The experiments by Warren and Hannon (1988, 1990) and Warren *et al.* (1988) used a total viewing time of about 3 sec, with image sequences of about 50 frames. It was later found that for pure translation of the observer, there is no deterioration in performance if the number of frames is reduced, until only 2–3 frames are presented (Warren, Blackwell, Kurtz, Hatsopoulos & Kalish, 1991). There is about 3° of accuracy for only two frames, with significant improvement when a third frame is added. When a rotational component is added to the motions of the points, more extended time may be needed to recover observer heading accurately (W. Warren, personal communication).

The visual system can also tolerate significant noise, with performance degrading smoothly with increased amounts of noise. Warren *et al.* (1991) found, for example, that in the case of pure translation of the observer, subjects could still judge heading direction with an average error of 2.6° when the directions of motion of individual points were randomly perturbed within an envelope of 90°. This result suggests that the heading computation may involve significant spatial pooling of image motion measurements.

Cutting (1986) examined observers' ability to determine their direction of translation toward a field of vertical lines placed on three frontoparallel planes whose separation in depth was varied. When the planes were at the same depth, subjects performed at chance, and heading accuracy improved with an increased separation of the planes in depth. The best accuracy achieved corresponds to a relative heading angle of about 1.25°.

Rieger and Toet (1985) measured subjects' ability to judge their heading direction relative to two frontoparallel planes of dense random dots placed at different depths. Translational and rotational components of motion were combined in the movements of the points on the display. The parameters that were varied in these experiments include the direction of translation, the separation in depth between the two planes, the magnitude of the rotational component of observer motion, and the size of the field of view. For the case of a single plane, performance degraded rapidly as the magnitude of simulated rotation was increased, similar to previous studies. When the points were placed at different depths, however, subjects could reliably judge heading direction over the range of angular rotations tested, with little degradation with the size of the field of view.

To summarize the perceptual experiments, we make the following observations regarding the human recovery of direction of translation:

- Human observers can achieve an accuracy of about 1–2° of visual angle at judging heading direction, with or without the presence of a target in the environment.
- Performance improves with higher speeds of translation.
- Performance improves when surfaces span a greater range of depth.
- Extraretinal information regarding eye rotation is used in the recovery of heading direction.

- Heading direction can be judged reliably in the presence of significant amounts of noise in the image motion measurements.
- For the case of pure translation, heading direction can be recovered accurately in a relatively short time of 2 or 3 frames, with accuracy increasing with time.
- Heading direction can also be recovered in a context where the rotational and translational flows must be passively decoupled from visual input alone. This decomposition
  (1) requires differential motion produced by elements at different depths,
  (2) can be performed successfully with sparse, discontinuous flow fields, and
  (3) requires only a relatively small field of view, at least as small as 10°.

The next section examines computational models for the recovery of observer motion in light of the above observations.

THE COMPUTATION OF DIRECTION OF TRANSLATION

Computational methods for recovering the direction of translation of an observer can be divided into two classes, depending on whether they use discrete or continuous image motion measurements. In the discrete approach, a set of isolated image features are tracked over time and their sequence of positions forms the input to a system of equations whose solution yields the parameters of 3-D structure and motion. The continuous approach uses an instantaneous 2-D velocity field at one or more instants of time, which together with spatial or temporal derivatives, are used to solve for 3-D structure and motion.

Many examples of the discrete approach present theoretical results regarding the minimal number of motion measurements required to compute 3-D structure and motion parameters uniquely (e.g. Ullman, 1979; Prazdny, 1980; Longuet-Higgins, 1981, 1984; Tsai & Huang, 1984a, b; Faugeras, Lustman & Toscani, 1987; Aloimonos & Brown, 1989; Weng, Huang & Ahuja, 1989). The direct application of the mathematical results suggests possible algorithms for recovering these parameters, but computer experiments indicate that they may be vulnerable to error in the image motion measurements. The ability of the human system to judge heading direction accurately for a few, sparse features in motion suggests that the underlying computation can derive movement parameters from discrete motion measurements, but unlike existing algorithms, the human system can tolerate large amounts of noise in these measurements. Algorithms that use discrete motion measurements over an extended time period exhibit better performance (Ullman, 1984; Broida & Challappa, 1986; Shariat, 1986; Faugeras *et al.*, 1987). Extended time appears to be necessary for the human system to decouple rotational and translational components of motion on the basis of visual input alone.

Approaches that use spatial derivatives of velocity require a locally continuous velocity field, or one that is sufficiently dense that interpolation can be used to approximate the continuous field (Longuet-Higgins & Prazdny, 1981; Koenderink & Van Doorn, 1976; Waxman & Ullman, 1985; Subbarao, 1988; Waxman & Wohn, 1988). Another approach based on the theory of planar dynamical systems uses the time evolution of the structure of the flow field in the vicinity of singularities (such as the FOE) to recover motion parameters (Verri, Girosi & Torre, 1989). This method may have difficulty with the sparse and discontinuous velocity fields used in perceptual studies. Some of these techniques also require accurate velocity measurements. Methods that rely directly on spatial and temporal derivatives of image intensity (Negahdaripour & Horn, 1987, 1989; Horn & Weldon, 1988; Heel, 1990a, b) may have difficulty coping with the impoverished displays of isolated dots used in perceptual studies.

Other velocity based approaches do not require a continuous velocity field (for example, Bruss & Horn, 1983; Ballard & Kimball, 1983; Jain, 1983; Lawton, 1983; Adiv, 1985; Burger & Bhanu, 1990; Heeger & Jepson, 1990). Some of these methods use an optimization approach, in which 3-D motion parameters are computed that yield a velocity field that best fits the observed image velocities in the least-squares sense, and integrate a large number of image motion measurements, yielding less sensitivity to error. The human system, however, does not require extensive spatial integration to compute heading direction accurately; in contrast, it copes with a small number of motion measurements and a relatively small field of view.

Finally, some methods make direct use of information about motion parallax, that is, the relative motion of features at different depths, to derive 3-D motion and structure (Longuet-Higgins & Prazdny, 1981; Rieger & Lawton, 1985; Cutting, 1986). The difference in velocity between two points that are nearby in the image, but separated in depth, depends largely on the translational parameters of observer motion and can be used directly to infer the direction of translation. The explicit reliance of these methods on depth variation in the scene makes them appealing from the perspective of the human system, which fails for the case of the perpendicular approach to a plane.

To summarize, it appears that most existing models do not exhibit the basic properties of the human recovery of direction of translation. None of these models have been shown to yield the accuracy of 1–2° of visual angle seen in human judgments of heading, over a range of viewing conditions. Some models could be modified to cope with some of the conditions considered in perceptual studies, but the need to cope with sparse, noisy and discontinuous motion fields, and the failure of the human system with the frontoparallel plane, seems to rule out many models on more fundamental grounds.

THE RIEGER AND LAWTON MODEL

This section describes the algorithm proposed by Rieger and Lawton (1985), which is based on earlier work by Longuet-Higgins and Prazdny (1981). This class of models begins with the observation that at the location of a discontinuity in depth, there will be a discontinuity in the translational component of the image velocity field because of the dependence of this component on depth, while the rotational component will be roughly constant across the boundary. Furthermore, if we construct a field of vectors that represent the differences in velocity across these boundaries, these vectors will be oriented approximately along the lines connecting their image location with the focus of expansion (the *translational field lines*), and therefore should all point to the FOE.

Longuet-Higgins and Prazdny suggested an algorithm based on the above observations that uses instantaneous spatial derivatives of velocity to recover the FOE. This original algorithm proved to be quite sensitive to error in the image velocity measurements. A robust algorithm that uses this observation to extract the FOE must take into account the fact that accurate velocity measurements may not be available immediately to either side of a depth discontinuity. Rieger and Lawton (1985) presented an algorithm that addresses this problem. The steps of the algorithm are as follows. First, the differences between each local image velocity and other velocities measured within a restricted neighborhood are computed. From the resulting distribution of velocity difference vectors, the dominant orientation of the vectors is computed and preserved only at locations where the distribution of velocity differences is strongly anisotropic. Such points typically arise where there is a strong depth variation in some direction. The result of this first stage is a set of directions at a number of points in the image that are all roughly aligned with the translational field lines. The FOE is then calculated as the best-fit intersection point for all the resulting vector directions. Once the FOE is determined, the direction of the translational component of motion is known at every location in the image, so that any motion in the original flow field that is perpendicular to this direction must be due to the rotation of the observer. From these perpendicular motions, the best rotational parameters are inferred (see also Burger & Bhanu, 1990). The full rotational flow field is then computed and subtracted from the original flow field to obtain the full translational component of the flow field. Finally, the relative depth at every point is computed from knowledge of the FOE and magnitude of the translational component of motion at each location.

The algorithm proposed by Rieger and Lawton is appealing for a number of reasons. First, it provides an initial estimate of the direction of translation with minimal computation, independent of the rotation parameters and 3-D shape. Heading direction is a critical property of observer motion for navigation that must be computed with high accuracy and speed. It is also important to detect object boundaries from motion discontinuities as soon as possible, and these are precisely the locations that provide the best information for this algorithm. Another appealing aspect is its simplicity and reliance on primitive image motion information, such as velocity differences, that require little computation. The fact that it does not rely critically on the solution of optimization problems is also an advantage. Optimization can be used at each step of the algorithm, but the information being computed can be obtained to a close approximation with non-iterative techniques.

One question that arises regarding the Rieger and Lawton algorithm as it stands is whether it can achieve the degree of accuracy of human performance measured across the range of conditions used in perceptual studies. Simulations presented by Rieger and Lawton (1985) suggest that the resulting heading accuracy may be within a factor of 2 or 3 of the needed accuracy. An especially challenging aspect of human performance is its ability to cope with sparse displays. The average angular separation between points in Warren and Hannon's (1990) study is large compared to the neighborhood sizes used in Rieger and Lawton's simulations. Over larger distances, the assumptions of the model become less valid. Computer simulations presented later suggest that this algorithm can yield the desired accuracy for the particular conditions of the perceptual experiments, with reasonable assumptions about the available precision of image motion measurements.

BUILDING UPON THE RIEGER AND LAWTON MODEL

From a computational standpoint, the most severe limitation of Rieger and Lawton's model is that it does not cope with self-moving objects in the environment. The difference in velocity across the boundary between a self-moving object and stationary background, or between two self-moving objects, in general does not yield vectors that are oriented along the translational field lines that emanate from the true FOE. Combining these differences with those obtained along the boundaries between stationary surfaces can yield significant error in the computed FOE location, especially if self-moving objects cover a large part of the visual field. It is necessary to detect self-moving objects explicitly or to remove their influence on the FOE computation by some implicit means.

This section considers a different method for performing the FOE computation in Rieger and Lawton's model that allows self-moving objects to be present in the scene and helps to isolate the boundaries of such objects. We also discuss some additional modifications to other stages of the algorithm that improve its performance in the presence of error in the image motion measurements. The results of computer simulations with the algorithm described here are presented in the next section.

Previous methods for coping with self-moving objects

We first consider existing methods for detecting and coping with self-moving objects in the scene. One approach assumes that the camera is stationary, so that significant image motion indicates self-moving objects (Jain, Militzer & Nagel, 1977; Jain, Martin & Aggarwal, 1979; Anderson, Burt & van der Wal, 1985; Dinstein, 1988; Bouthemy & Lelande, 1990). A variation on this approach considered by Burt, Bergen, Hingorani, Kolczinski, Lee, Leung, Lubin and Schvaytser (1989) recovers global camera motion parameters by stabilizing regions of the image, analogous to eye tracking in the human system. Once the image motion due to the actual camera motion is largely removed, any significant motions that remain are likely to be due to self-moving objects. A second approach assumes that the camera undergoes pure translation, so that any self-moving objects violate the expected pure expansion of the image (e.g. Jain, 1984). If 3-D depth data is available, then inconsistency between image velocities, estimated observer motion and depth data can also signal self-moving objects (e.g. Thompson & Pong, 1990). Nelson (1990) shows that it is possible to detect such inconsistencies from partial information about image and observer motion. Nelson also notes that the motion of objects due to the observer's motion tends to change slowly over time, while self-moving objects can generate rapidly changing patterns of motion that can be used to detect their presence.

A more general strategy is to compute an initial set of observer motion parameters, either by combining all available data or by performing separate computations within limited image regions, and then to find areas of the scene that move relative to the observer in a way that is inconsistent with the global motion parameters (Heeger & Hager, 1988; Zhang, Faugeras & Ayache, 1988). If all motion information is used initially, the recovery of observer motion parameters can be degraded by the inconsistent motions of self-moving objects. On the other hand, the use of spatially local information can yield inaccuracy due to the limited field of view. Thompson, Lechleider and Stuck (1992) present a variation on this approach that uses a technique from robust statistics (Huber, 1981) to compute global motion parameters in the presence of "outliers", which are data that deviate significantly from consistency with the true parameters. Image motions resulting from self-moving objects are treated as outliers and the least median squares algorithm (Rousseeuw & Leroy, 1987; Meer, Mintz, Kim & Rosenfeld, 1991) is used to compute motion parameters in a way that detects potential outliers. Thompson *et al.* (1992) note that self-moving objects whose projected image motion is close to the motion that is expected from the observer's global translation and rotation are difficult to detect with this technique.

Modifying the Rieger and Lawton model to cope with self-moving objects

We present a strategy for detecting and coping with self-moving objects that builds on the Rieger and Lawton algorithm. We first summarize the strategy in general terms and then elaborate on the motivation and details. The scheme first computes local velocity differences and determines the dominant orientation of the distribution of velocity differences within a small neighborhood of each point, as in the Rieger and Lawton model. The orientations, θ_i, are preserved for the next stage of the computation only at points where the distribution of velocity differences is strongly anisotropic. Most of the θ_i measurements preserved at this stage arise from points on or near depth discontinuities, or along surfaces such as the ground plane, whose angle of slant relative to the image plane is large.

Some portion of the θ_i measurements will point roughly toward the true FOE, while θ_i measurements obtained in the vicinity of self-moving objects or those with high error will be oriented in arbitrary directions. Assuming that self-moving objects do not cover a large part of the visual field, we can obtain a good initial guess of the location of the FOE by looking for limited image regions for which a large percentage of the θ_i measurements point toward locations within the region. In particular, we consider how much evidence exists to support the FOE being located within each of a large set of image regions, choose the region (or regions) with maximum support and use the θ_i measurements that provide this maximum support to derive an FOE

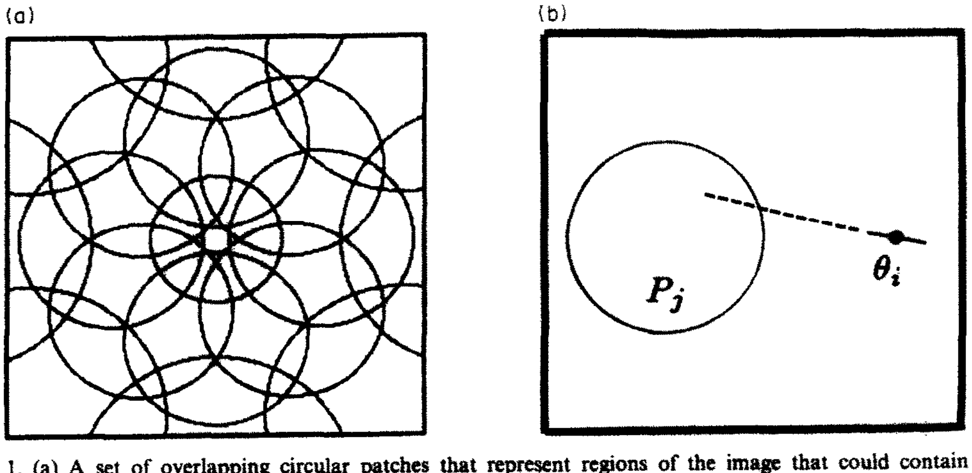

FIGURE 1. (a) A set of overlapping circular patches that represent regions of the image that could contain the FOE. (b) Positive evidence for the FOE being located within P_j is given by a measurement θ_i if a line from the point that contains the vector defined by θ_i intersects P_j.

estimate. [This strategy is based on the Hough transform used in computer vision (Ballard & Brown, 1982).]

In more detail, after the θ_i measurements are derived, the visual image is divided into a set of overlapping, circular patches that represent possible regions within which the FOE may be located, as shown in Fig. 1(a). For each patch P_j, all of the positive evidence for the FOE being located within P_j is collected. Positive evidence comes from points whose orientation θ_i lies along a line that intersects P_j, as shown in Fig. 1(b). If the true FOE is located within the patch P_j, then velocity differences computed within or along the boundaries of stationary objects should yield positive evidence. Points at which the orientation θ_i does not yield positive evidence either lie within or near the boundaries of self-moving objects, or they yield large error in the computation of θ_i. If the true FOE is not located within P_j, there may be points that yield an orientation θ_i that incorrectly provides positive evidence for an FOE in P_j, but the percentage of points yielding such false positive evidence should be substantially reduced. For each patch P_j, if a large percentage of the available θ_i yield positive evidence for the FOE being located within P_j, then the set of θ_i estimates yielding this positive evidence is used to generate a hypothesized FOE location. If multiple FOE hypotheses remain after this stage, they are reconciled to obtain a single FOE location by considering the extent of the positive evidence in their support, their goodness-of-fit to the computed θ_i, and the proximity of the multiple hypotheses.

The reasoning behind this strategy is that by combining only those θ_i measurements that yield positive evidence for an FOE being located within restricted patches, we reduce the degradation in the FOE computation that results from the presence of self-moving objects and large errors in the θ_i estimates. When patches that contain the true FOE are considered, self-moving objects and large errors in the θ_i are likely to result in θ_i estimates that do not yield positive evidence and hence do not enter into the FOE computation. Patches that do not contain the true FOE are likely to yield significantly less positive evidence and therefore do not lead to an FOE hypothesis. The remainder of this section elaborates on details of the individual steps of the algorithm.

As shown in Fig. 1a, the circular patches may increase in size with distance from the center of the image. This serves to minimize the total number of patches needed to cover the image and to allow the FOE to be computed more accurately when it is located toward the center of the image. Reducing the total number of patches reduces the amount of computation required to test the set of patches for possible FOE locations. The desire to compute the FOE more accurately toward the center of the image arises in part from properties of human vision. Human observers judge heading direction most accurately when their eyes are pointed in the direction of heading, and the spatial resolution of processing increases toward the center of the eye. Thus heading direction is derived most accurately when the FOE lies near the center of the visual image.

The determination of whether a particular measurement θ_i is consistent with the FOE being located within a patch P_j requires a simple computation. One can either determine whether the orientation θ_i falls within a cone of directions defined by the two lines running through the underlying point and tangent to the circular boundary of P_j, or whether the perpendicular distance from the center of P_j to the line containing the vector in the direction θ_i is less than the radius of P_j. The measurements of θ_i obtained from points within P_j are not included in the positive evidence for P_j, because the size of the translational component of velocity is small in the vicinity of the FOE, yielding unreliable velocity differences. We also limit the overall extent of the region from which θ_i measurements are considered for P_j, because the range of consistent orientations θ_i becomes too small for points distant from P_j, requiring too much accuracy in their estimate.

After the set of θ_i that yield positive evidence for a given patch are computed, we determine whether there is sufficient evidence to combine these θ_i measurements to derive an FOE hypothesis. The percentage of all θ_i measurements that yield positive evidence is compared to a threshold. This threshold must be large enough to minimize the number of false hypotheses generated from patches that do not contain the true FOE, while allowing a significant portion of the visual field to contain self-moving objects. The choice of threshold here is governed in part by what percentage of points yielding θ_i measurements are expected to be within or near the boundaries of self-moving objects, and in part by what percentage of points from stationary regions of the scene are expected to yield false positive evidence for inappropriate FOE locations. With regard to the first factor, we note that if too much of the visual field contains self-moving surfaces, human observers do not judge their heading correctly.

Figure 2 addresses the second factor mentioned above. We consider a patch P at the center of the image, as shown in Fig. 2(a), and determine the positive evidence that could be obtained for FOE locations within P, for different true locations of the FOE. Evidence is considered from all points lying within a circular region surrounding P, and we assume that every point in the image yields a measurement of θ_i that is directed along translational field lines emanating from the true FOE. The graph in Fig. 2(b) shows the percentage of θ_i measurements that represent positive evidence for FOE locations within P for the true FOE locations indicated with solid circles in Fig. 2(a). When the true FOE is located within P, 100% of all θ_i measurements yield positive evidence, but as the true FOE moves outside P, the percentage of points that could yield positive evidence for FOE locations inside P drops rapidly. (For the simulations presented later, we required that 40–50% of the θ_i measurements yield positive evidence for a patch P_j, in order to generate an FOE hypothesis from P_j.) Figure 2(c) shows a map of the points that could yield positive evidence for the FOE being located within P when the true FOE is located outside P, as described in the figure legend. If, in a particular scene, all of the available θ_i measurements fall in the regions shown in black in Fig. 2(c), it would appear that there is significant positive evidence for an FOE within P and the set of θ_i measurements would be used to generate an FOE hypothesis. If the true FOE is located outside P, the estimate obtained may not have as good a fit to the θ_i measurements as the FOE hypothesis generated from a correct patch. In general, however, a skewed spatial distribution of the available θ_i measurements can yield an inappropriate FOE estimate.

Self-moving objects can also yield false positive evidence for an FOE being located within a given patch P_j, especially if an object undergoes a significant translation toward or away from P_j. If the true FOE is not located within P_j, then the added θ_i measurements from self-moving objects are likely to yield an FOE hypothesis that does not yield a good fit to the θ_i measurements. Even for the patch that contains the true FOE, self-moving objects with significant translation near but not along the true translational field lines can distort the computation of the FOE location. We assume that this situation is rare, and note that when it does occur, it is unlikely to persist for an extended period of time, or over an extended region of the image.

Due in part to the overlap of adjacent patches [see Fig. 1(a)], valid FOE hypotheses may emerge from multiple patches. If there is a single FOE location that both accounts for a significantly larger percentage of the θ_i measurements and yields a significantly better goodness-of-fit to these measurements, then this FOE location is considered to be the best current guess (the simulations presented later required a 20% difference in these two properties). Multiple FOE locations that are close to one another can be averaged together to yield a current estimate. If there are multiple FOE hypotheses with strong support that are distant from one another, it may be possible to resolve the global FOE through an analysis of self-moving objects in the scene, which we consider next.

If there is significant positive evidence for the FOE being located within a patch P_j, then points that do not yield positive evidence can be used to detect self-moving objects. In particular, extended, connected groups of such points can signal a self-moving object. Isolated points or small groups of points yielding negative evidence are

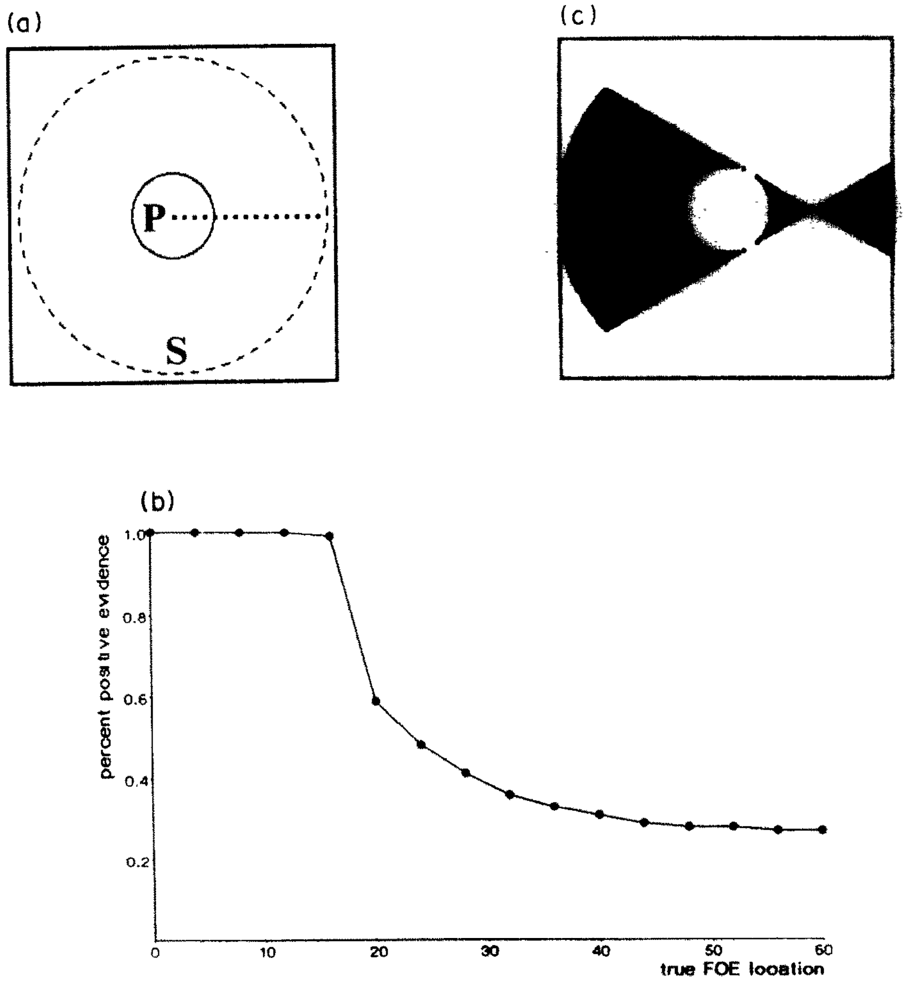

FIGURE 2. (a) We consider the positive evidence for the FOE being located within the central patch P from θ_i measurements that could be obtained within the larger annular region S, for the set of true FOE locations indicated by the solid dots. The radius of P is 16 pixels, and radius of S is 64 pixels. (b) Graph of the percentage of θ_i measurements that would provide positive evidence for the FOE being located within P as a function of the true location of the FOE. (c) Given the patch P with radius 16, and a true FOE located 32 pixels to the right of the center of P, the points that could yield positive evidence for the FOE being located within P are shown in black.

likely to be the consequence of error in the θ_i computation. Some points within or near the boundaries of self-moving objects will yield false positive evidence for an FOE within P_j, but if these points are connected to an extended region of points yielding negative evidence, we assume that they represent a continuation of a self-moving object and generate a new FOE hypothesis with these points removed.

Finally, we note that a coarse-to-fine strategy can be used, in which larger patch sizes are used first to obtain a rough estimate of the region (or regions) likely to contain the global FOE, and the size of the patches is successively reduced to refine the estimated FOE location. At each scale, a current estimate could be obtained and smaller patches could then be centered on the current estimate. Such a coarse-to-fine strategy provides a rapid assessment of the rough FOE location and reduces the total amount of computation required to obtain a more precise estimate.

Work in the area of robust statistics provides techniques for deriving global parameters in the presence of significant outliers in the data (Rousseeuw & Leroy, 1987; Meer *et al.*, 1991). Similar to the scheme proposed by Thompson *et al.* (1992), the θ_i measurements derived from self-moving objects could be considered outliers and techniques such as least median squares could be applied to the full set of θ_i measurements to compute an FOE estimate and detect the "outlying" self-moving objects. The approach presented here takes better advantage of the geometrical relationship between θ_i measurements obtained from stationary and self-moving objects and requires much less computation.

Other modifications of the Rieger and Lawton model

This section considers additional modifications aimed at improving the performance of the Rieger and Lawton algorithm in the presence of error in the image motion measurements. These modifications include temporal smoothing of the image velocities, a different strategy for computing the dominant orientations, θ_i, that filters the local distributions of velocity differences, and a method for refining the θ_i measurements at a later stage.

If the errors in the 2-D velocities of moving features are uncorrelated from one moment to the next, then smoothing or averaging of the velocity measurements over time can improve their quality. This temporal smoothing should be limited in time, as the observer's heading can change over a long time interval. In the simulations presented later, velocity measurements with added noise that were obtained at two different times were averaged together. This smoothing took place prior to the computation of velocity differences and significantly improved the quality of these difference estimates.

The local distribution of velocity differences can be computed in one of two ways. First, the difference between the velocity of a point p_i and that of each neighboring point p_j within some distance of p_i can be computed to obtain a set of velocity differences associated with p_i. If p_i has n neighbors, then the distribution will contain at most n differences. A second option is to consider fixed neighborhoods distributed over the image and to compute the difference in velocity between every pair of points that falls within each neighborhood. In this case, if there are n points within a given neighborhood, then there will be at most $n(n-1)/2$ velocity differences computed. Both strategies were used in the simulations described later. For the simulations with sparse dot patterns, all pairs of points within fixed neighborhoods were used to obtain the local distributions of velocity differences, while the simulations with images on dense grids used only the differences between single locations and their neighbors.

To obtain estimates of the dominant orientations, θ_i, note that the distribution of velocity differences computed at a point or within a neighborhood that lies in the vicinity of a depth discontinuity or on a surface with a substantial slant in depth will typically cover a range of directions, as shown in Fig. 3(a). Differences between the velocity of two points that lie at significantly

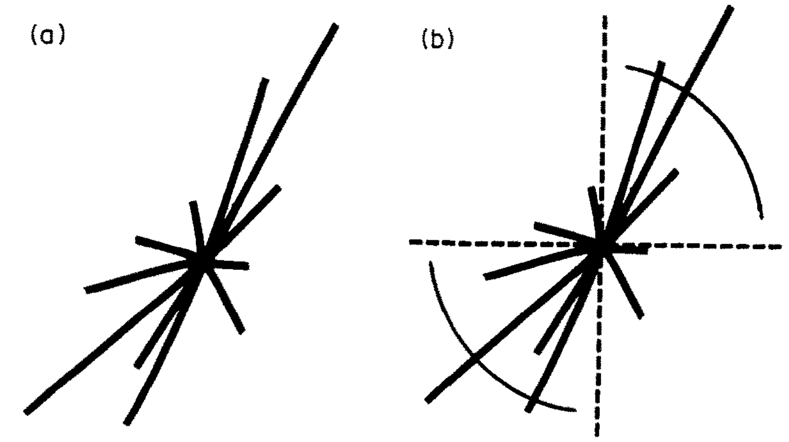

FIGURE 3. (a) A typical distribution of velocity differences obtained at a point that is near a depth discontinuity or located on a highly slanted surface. The larger vectors represent the difference in velocity between this point and other points lying at significantly different depths, and are directed roughly along the translational field line. Other vectors represent the difference between the velocity at this point and that of other points located at similar depths. The aim is to compute the dominant direction of these differences. (b) We find two opposite 90° ranges of orientations that separate the differences in a way that maximizes the ratio between the sum of the lengths of the velocity differences lying within and outside of these ranges.

different depths will be larger and oriented roughly along the translational field line directed toward the FOE. There will be some deviation from the true translational field line, due to error in the velocity measurements or to the spatial separation between the two points, yielding added differences in velocity due to the rotation of the observer. Differences obtained from pairs of points at a similar depth will be smaller and have directions that are randomly distributed over a 360° range. These latter difference measurements can degrade the computation of the dominant orientation if all of the difference measurements are considered together. To reduce this degradation, we only combine velocity differences within two opposite ranges of 90°, as shown in Fig. 3(b), and choose the particular ranges that yield the largest ratio between the overall weight of the differences obtained within and outside of these ranges. Estimates of θ_i are preserved only at locations at which this ratio is above a threshold, indicated a strong anisotropy in the directions of the velocity differences. The θ_i themselves are computed by finding a line that represents a best least-squares fit to the set of difference vectors.

Finally, the θ_i estimates can be improved after an initial FOE estimate is obtained. An initial FOE estimate yields a set of predicted translational field lines, along which local velocity differences should lie. The local distributions of velocity differences can then be filtered to emphasize differences whose direction is closer to the orientation of the translational field lines. A new FOE location can be computed based on the computation of new dominant orientations of the filtered local velocity differences. In principle, the same strategy can be applied over time. The location of the FOE can change over time, so it is necessary to estimate the rotational component of motion as well, in order to predict the displacement of the FOE in the image due to the observer's rotation. This can be done, for example, in the way that Rieger and Lawton (1985) propose. At each new moment in time, the current estimate of the location of the FOE can be used to weigh local velocity differences in the computation of a new FOE. A better estimate of the FOE should then result in a better estimate of the rotational component of motion, yielding progressive improvement over an extended sequence of images.

COMPUTER SIMULATIONS

This section presents the results of computer simulations that consider aspects of the human recovery of heading direction and the use of the algorithm for computer vision systems.

Simulations with the model applied to perceptual displays

This section summarizes the results of simulations with our extension of the Rieger and Lawton (1985) model, applied to visual patterns similar to those used in the perceptual studies described earlier. We used synthetic image data corresponding to displays of discrete points whose image motion is determined by the translation and rotation of an observer relative to a random-dot surface in space. The motions of the dots on the image plane were computed analytically and these movements, with or without added noise, formed the input to the model for heading recovery.

The following conditions of the perceptual experiments by Warren and his colleagues were approximately simulated here:

- *Observer's translation*: the observer translates in the horizontal plane, with a heading direction spanning a range within 6° to the left and right of straight ahead. For most experiments, translational speed was 1.9 m/sec.
- *Observer's rotation*: the typical range of simulated angular velocity of the eye was 0.3–0.7°/sec, covering the full range of 2-D directions.
- *Field of view*: 40° horizontal × 32° vertical.
- *Temporal extent*: most experiments used a total viewing time of about 3 sec, with a frame rate of 15 frames/sec. The simulations presented here, however, used average displacements computed from only the first three image frames.
- *Ground plane*: the observer's simulated eye height was 1.6 m and points covered a plane extending 37.3 m in front of the observer. The spatial distribution of the points was uniform on the plane, creating a non-uniform distribution in the image, due to perspective projection.
- *3-D cloud*: points were placed randomly within a depth range of 6.9–37.3 m.
- *Frontoparallel plane*: a plane was placed at a distance of 9.3 m in front of the observer.
- *Number of dots*: in most experiments, there was an average of 63 dots at the beginning of the movement.

In these experiments, observers were asked to judge only the horizontal component of motion. Additional error in the perception of the vertical component of heading would indicate a larger overall heading error. The accuracy of 1–2° measured in perceptual experiments refers to the horizontal component alone.

The simulations also considered the following conditions: (1) points placed on two frontoparallel planes, whose absolute and relative depths were varied; (2) variation in the absolute and relative range of depth for the 3-D cloud; (3) wider heading angles ranging up to 30° to the left and right of straight ahead; (4) larger rotational components, corresponding to an angular velocity of the eye up to 10°/sec; and (5) a smaller field of view of 20°. Some of these issues were motivated by the studies of Rieger and Toet (1985) and Cutting (1986). Note that with a very large rotational component, the relative difference between the velocities at nearby locations due to the translational component becomes small, reducing the signal available for recovering the direction of heading.

Thresholds were imposed on the absolute image velocity and on the velocity differences that were considered detectable. The threshold used for absolute velocity was 1°/sec and the threshold on velocity differences was 10% (Nakayama, 1985). Values falling below these thresholds did not enter into the computation of heading direction. There will be noise in the velocity estimates, but it is not clear what is a reasonable level of noise to expect for the visual system. In the simulations, we first determined the level of noise in the velocity measurements that yields a heading accuracy of about 2–3°, for the case of translation relative to the ground plane and the overall conditions of the perceptual experiments summarized above. (It is expected that the greater heading accuracy of 1–2° measured for the human system could be obtained by extending the heading computation further in time.) We found that this accuracy could be achieved with an average error in speed of about 25% and average error in the direction of velocity of about 25°. Error was introduced as Gaussian distributed perturbations of the direction and speed of velocity. An average error in speed of 25% and in velocity direction of 25° was then used throughout the remaining simulations. Limited temporal smoothing was performed to reduce the overall sensitivity of the algorithm to error in the initial velocities.

Although the scene consisted of a single rigid surface, we used the strategy described in the previous section for computing the FOE location in the presence of self-moving objects, to reduce the sensitivity of the FOE computation to error in the θ_i estimates. Three circular and overlapping patches representing possible locations of the FOE were centered on heading directions located at 6, 0 and $-6°$ from straight ahead, and each covered an area of radius 6°. Thus heading angles computed by the algorithm could cover a range from -12 to 12° in the horizontal direction. Preliminary simulations suggested that image patches outside of the regions covered by these three patches yield significantly less positive evidence, and therefore need not be included in this analysis. If more than one patch yielded a predicted FOE location, we first checked whether one estimate was significantly better than the others, in that it had significantly more positive evidence and better fit to the θ_i measurements (about 20% difference in both cases). If this was not the case, then the multiple predictions were averaged together to yield a final estimate.

TABLE 1. The results of simulations with the Rieger and Lawton model, applied to images generated by an observer moving along a ground plane. Average errors, in deg, are given for the horizontal component of heading. The top entry gives results for the following parameters: observer speed of 1.9 m/sec; 40° field of view; 60 points; 6° heading range; 0.3–0.7°/sec rotation range; 25% average error in image speed; and 25° average error in the direction of image velocity

Parameters	Horizontal
Initial parameters	2.5
7.6 m/sec	2.2
20° field of view, 60 points	2.6
40° field of view, 30 points	4.0
20° field of view, 30 points	2.7
40% average speed error, 40° average direction error	3.9
5–10°/sec rotation range	4.4

The results of simulations with the ground plane are summarized in Table 1. Each data point represents an average of the results from 100 different random configurations of points. The full set of parameters used for the first example (top entry in Table 1) is given in the legend; other entries indicate only the value of the parameter that differed from the first example. A ground speed for the observer of 1.9 m/sec and presentation rate of 15 frames/sec corresponds to 0.127 m/frame of observer translation. Similarly, an angular velocity range of 0.3–0.7°/sec for the simulated eye rotation corresponds to a range of 0.02–0.05° per frame. This range of angular velocities used in perceptual studies is small. We also conducted simulations with rotations in the range from 5–10°/sec. The field of view is defined as the total width of the field in the horizontal direction. For each configuration of points, a simulated heading direction was chosen randomly from the range of 6° to the left and right of straight ahead. Velocity differences were computed for any pair of velocity measurements falling within a neighborhood of 6° of one another.

From this initial set of simulations, it can be seen that direction judgments improve with higher speed of observer translation and higher density of points, and degrade with higher error in the velocity differences and a higher angular velocity of the eye. If the density of points is kept relatively constant, the field of view has little effect on heading accuracy. These factors interact with one another. For example, with the limited field of view, higher angular rotations yield significant degradation in the direction computation, but if the field of view and number of points is increased, a more accurate heading direction can be obtained for higher rotation speeds. Most simulation results reported in the literature use fairly large rotational components, which often yields significant error; such rotations may also yield larger error in human judgments of heading. Overall, the heading accuracy remains high for the range of conditions explored here.

In general, as the velocity difference errors increase, there can be substantial error in the local computations of the dominant orientation of the distribution of velocity differences within image neighborhoods. If these measurements are distributed over a large field, however, the overall computation of the FOE can still be accurate. There is a characteristic asymmetry in the pattern of errors obtained over the visual field. The directions of the dominant orientation of local velocity differences usually point to the right of the FOE in the right half of the visual field and to the left of the FOE in the left half of the visual field. With a roughly uniform distribution of points in the horizontal direction, these errors effectively cancel one another out in the overall computation of the FOE. The same observation holds true in the vertical direction. An implication of this observation is that if the distribution of θ_i measurements is strongly skewed within the visual field, a characteristic error in the heading computation can result.

TABLE 2. The results of simulations with the Rieger and Lawton model, applied to images generated by an observer moving toward a 3-D cloud of points or two frontoparallel planes separated in depth. Unless specified above, parameters were as follows: observer speed 1.9 m/sec; 40° field of view; 80 points; 6° heading range; 0.3–0.7°/sec rotation range; 25% average error in image speed; and 25° average error in the direction of image velocity

Parameters	Horizontal
3-D cloud, depth range 7–40 m	2.3
3-D cloud, depth range 15–32 m	4.0
3-D cloud, depth range 7–40 m, 10 points	5.0
Two planes, 5 and 25 m	1.5
Two planes, 10 and 20 m	2.6
Two planes, 20 and 40 m	3.7
Two planes, 5 and 25 m, 6–12° heading range	1.8

The results of some additional simulations with the 3-D cloud and two planes of dots are shown in Table 2. For all of these simulations, the field of view was a square of size 40°, which is somewhat larger than the 40 x 32° field of view used in the perceptual experiments. The results of simulations with the ground plane suggest that the density of points is a critical factor in determining the accuracy of recovered heading. Because of the somewhat larger field of view used in the simulations here, we used displays of 80 points, rather than 60, in order to keep the density of points similar to that used in the perceptual experiments. Other parameters used in these simulations are listed in the legend for Table 2. Overall, similar heading accuracy can be obtained for the 3-D cloud and two planes. Accuracy degrades as absolute depth is increased, but improves as the overall range of depth is increased. Errors increase slightly for more oblique heading directions. In general, heading direction is underestimated, in that it is closer to straight ahead relative to the true direction of heading. An increased field of view can reduce the errors for more oblique headings. Errors increase significantly for sparse patterns containing only 10 points, largely because the image neighborhoods over which the velocity differences are computed contain very few pairs of points from which to compute the θ_i measurements. For the case of the frontoparallel plane, the errors were very large. For headings chosen within a 6° range of directions around straight ahead, the average heading error in this case was 5.0° in the horizontal direction.

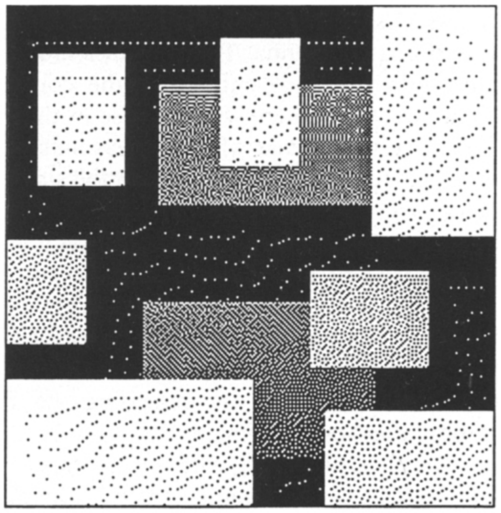

FIGURE 4. A synthetically generated depth map, with brightness encoding depth (a dithered image is shown, so that the density of black and white dots conveys different brightness levels). Depths range from 75 to 250 units.

Simulations with self-moving objects

This section presents the results of simulations with the algorithm applied to synthetic image sequences containing multiple objects, some of which undergo their own self-motion. For each example, a known velocity field was first generated from a known depth map and movement parameters for the observer and objects. Noise was added to the image velocities, in the form of Gaussian distributed perturbations of their speed and direction. The algorithm was then applied to the noisy velocity field to recover the location of the FOE and to detect self-moving objects.

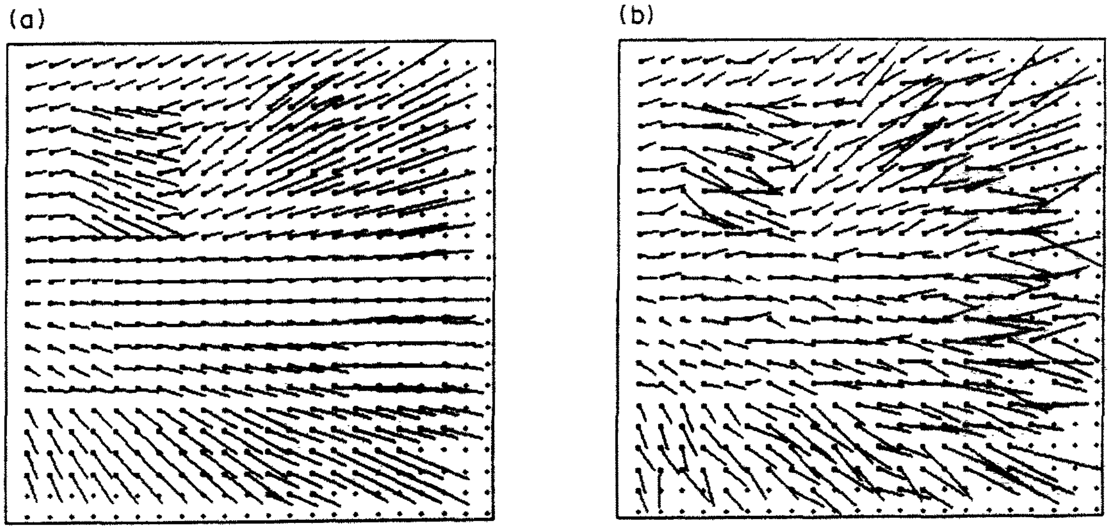

FIGURE 5. (a) An ideal velocity field obtained from the known depth map shown in Fig. 4 and known observer motion parameters. (b) The velocity field with added noise.

A depth map for the scene that formed the basis of these experiments is shown in Fig. 4. Brightness encodes depth, with darker objects located further from the observer. (A dithered image is shown, so that the density of black and white dots conveys different brightness levels.) The scene consists of planar surface patches of different 3-D orientations positioned over a distance of 75–250 units from the observer. From this known depth map and a set of known parameters for the observer's rotation and translation, an image velocity field was computed. An original velocity field is shown in Fig. 5(a). The velocities are sampled from an array of size 128 × 128. Noise was then added to yield velocity fields such as that shown in Fig. 5(b). Before computing the velocity differences, the velocities were averaged spatially over a neighborhood of size 3 × 3 pixels, to reduce the sensitivity to noise of the subsequent velocity differences.

The distribution of velocity differences was then computed for each image location. The distribution at a given location consisted of the differences in velocity between this location and every other location within a

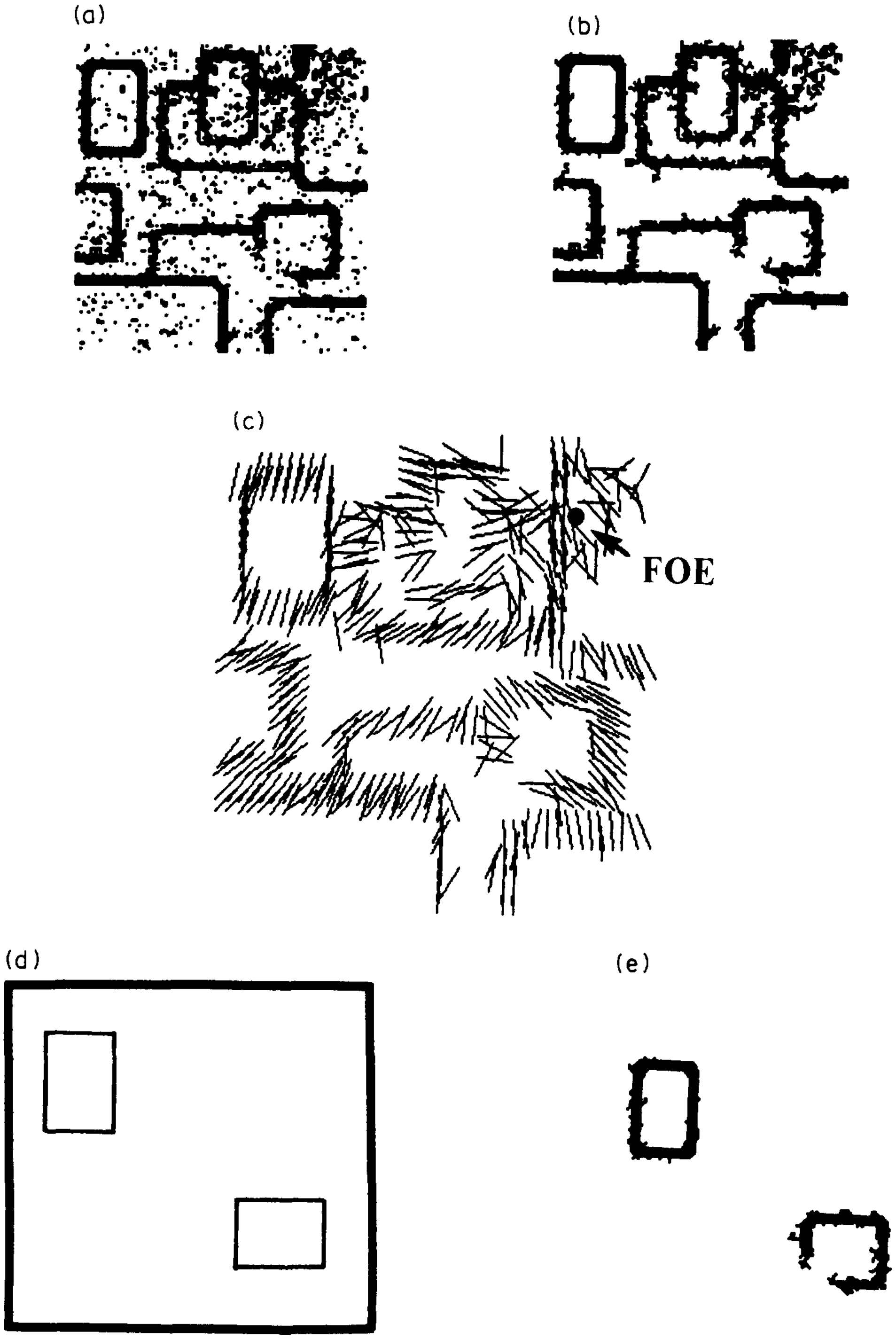

FIGURE 6. (a) A map of all the locations where θ_t were derived from local velocity difference distributions with strong anisotropy. (b) Isolated θ_t measurements are removed. (c) A sampling of the dominant orientations, θ_t. The true FOE is located in the upper right corner. (d) The locations of two objects in the scene that are self-moving. (e) Locations where θ_t measurements were obtained that indicate self-moving objects.

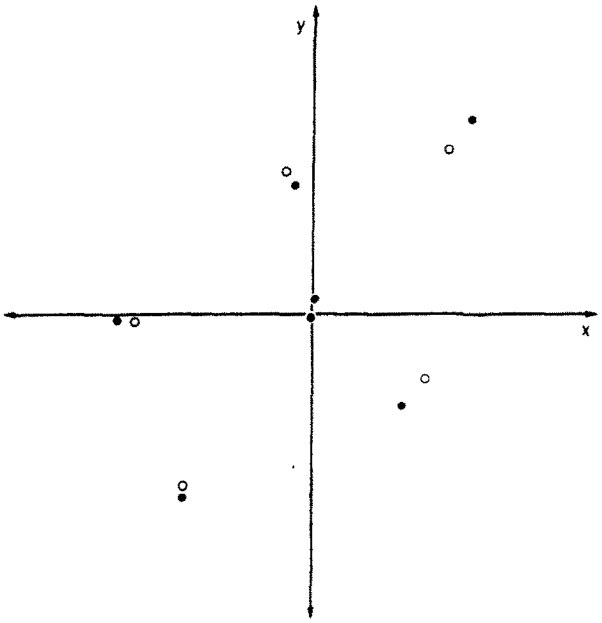

FIGURE 7. True FOE locations (solid circles) are compared to the FOE locations derived from the algorithm (open circles) for six choices of the observer translation parameters. The full extent of the horizontal and vertical axes corresponds to an image distance of 128 pixels.

neighborhood of radius 4 pixels. The dominant orientation, θ_i, of this distribution was computed using the scheme described in the previous section, and these θ_i measurements were preserved at locations where the distribution of local velocity differences was strongly anisotropic. For one set of observer and object motion parameters, a map of the locations at which the θ_i were initially preserved is shown in Fig. 6(a). Isolated θ_i measurements that do not belong to a connected patch of at least 10 pixels were then removed, assuming that the most appropriate θ_i estimates to use for the FOE computation would occur along extended boundaries. The locations of the θ_i that remain after this filtering step are shown in Fig. 6(b). These measurements are concentrated around the locations of boundaries and over the surface of the object in the upper right corner of the image, which has a large slant. Figure 6(c) shows the dominant orientations computed at a sample of the image locations. The true FOE is located near the upper right corner of the image, and the two objects highlighted in Fig. 6(d) are self-moving. There is significant error in the θ_i measurements, as those vectors in Fig. 6(c) that are not located in the vicinity of the two self-moving objects should all point toward the FOE.

To compute the location of the FOE, the image was carved up into overlapping circular patches with a radius of 24 pixels, centered at locations spaced by 24 pixels. For each patch P_j, the set of θ_i measurements yielding positive evidence for the FOE being located within P_j was then determined. If at least 50% of the θ_i measurements yielded positive evidence, a hypothesized FOE was computed. If multiple FOE hypotheses emerged, they were reconciled to obtain a single FOE location by considering the extent of the positive evidence in their support, their goodness-of-fit to the computed θ_i, and the proximity of the multiple hypotheses. Figure 7 shows the true (solid circles) and computed (open circles) FOE locations for 6 different choices of the observer translation parameters, and for rotation parameters, $(w_x, w_y, w_z) = (0.0, 2.0, 0.0)$ (these rotation parameters were used to generate the velocity fields shown in Fig. 5). The error in the final FOE estimates is small, given the large error in the input velocity fields and the θ_i estimates.

Once an initial estimate for the FOE location was obtained, extended regions yielding negative evidence were isolated as indicating self-moving objects. For the example shown in Fig. 6, the patch that yielded the most positive evidence is located in the upper right corner of the image. The θ_i measurements that were not directed toward this patch were isolated, and extended, connected groups of such measurements were hypothesized to correspond to self-moving objects. Figure 6(e) shows the final self-moving objects detected, which correspond correctly to the two self-moving objects in the scene.

SUMMARY AND CONCLUSIONS

This paper addressed the computation of the 3-D direction of translation of an observer relative to object surfaces. Consideration of perceptual observations regarding the human recovery of heading direction and existing computational models led us to examine the model proposed by Rieger and Lawton (1985) in more detail. We explored some extensions to the Rieger and Lawton model that yield improvement of its performance in the presence of error in the image motion measurements and allow it to cope with scenes containing multiple moving surfaces. The results of computer simulations with this modified model applied to visual patterns similar to those used in perceptual studies suggest that it exhibits much of the basic behavior of the human system.

Some navigational tasks require rapid sensing and response by the moving observer. The demands of such tasks may compel the human visual system to use specialized routines that use only partial or qualitative information regarding motion in the image or in the scene that can be computed reliably with minimal computation, and which is critical to performing a specific task. In the model presented here, simple measurements of velocity differences within local image neighborhoods are used to compute only the direction of observer heading, independent of the observer's rotation or scene layout. Velocity differences in regions of significant depth variation provide a direct cue to the observer's heading that can be exploited with relatively little computation. This partial information about heading direction can then be used directly by routines that detect potential collisions or track objects in the scene. Furthermore, because velocity differences will be significant along discontinuities in depth that occur along the boundaries of stationary and self-moving objects, they can also be used to detect these boundaries. We have shown that the heading computation itself can embody a strategy for detecting the boundaries of self-moving objects. This boundary information can also be used by routines that detect potential collisions, to determine the overall size and shape of relevant objects.

A number of additional questions regarding the human perception of heading direction arise from the analysis of the model presented here, which can be explored through further perceptual experiments. Among these are the following:

- Does accuracy in judging heading direction decrease with more oblique headings, and is there a general tendency to underestimate oblique headings? Is the size of the field of view more critical for the accurate judgment of oblique headings?
- Is there degradation of heading judgments when larger angular rotations are simulated, and is the size of the field of view critical in this case?
- Does an asymmetric spatial distribution of points yield characteristic errors in heading judgments, as suggested by the simulations?
- Is there a systematic degradation in heading accuracy with a smaller depth range and larger absolute depth?

It would be useful to examine the accuracy of our judgment of the vertical component of heading direction, to assess our overall precision at performing the heading computation. Other experimental questions arise regarding the recovery of observer heading in the presence of motion discontinuities and self-moving objects:

- What is the effect of self-moving objects in the field of view on the accuracy of heading judgments?
- Is there any difference in performance, depending on whether the boundaries of a self-moving object yield immediately perceptable motion discontinuities?
- How much of the image must contain significant depth variation? Suppose, for example, that the image contains a single object (a small frontoparallel plane) in front of a larger frontoparallel plane in the background. How large must the closer object be, and how much does it need to be separated in depth from its background, in order to yield accurate heading judgments?
- How much deviation in direction of image motion must a self-moving object undergo, relative to the motion direction expected from the observer's motion alone, in order to detect its presence?

Further experimental work that addresses these questions is critical to assessing the appropriateness of a model of the type explored here as a description of the recovery of heading direction by the human system.

REFERENCES

Adiv, G. (1985). Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-7,* 384–401.

Aloimonos, J. & Brown, C. M. (1989). On the kinetic depth effect. *Biological Cybernetics, 60,* 445–455.

Anderson, C. H., Burt, P. J. & van der Wal, G. S. (1985). Change detection and tracking using pyramid transform techniques. *Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision,* Boston, Mass., pp. 300–305.

Ballard, D. H. & Brown, C. M. (1982). *Computer vision.* Englewood Cliffs, N.J.: Prentice-Hall.

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

Bouthemy, P. & Lalande, P. (1990). Detection and tracking of moving objects based on a statistical regularization method in space and time. In Faugeras, O. (Ed.), *Proceedings of the First European Conference on Computer Vision* (pp. 307–311.) Berlin: Springer.

Broida, T. J. & Challappa, R. (1986). Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8,* 90–99.

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

Burger, W. & Bhanu, B. (1990). Estimating 3-D egomotion from perspective image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12,* 1040–1058.

Burt, P. J., Bergen, J. R., Hingorani, R., Kolczinski, R., Lee, W. A., Leung, A., Lubin, J. & Shvaytser, H. (1989). Object tracking with a moving camera, an application of dynamic motion analysis. *Proceedings of the IEEE Workshop on Visual Motion,* Irvine, Calif., pp. 2–12.

Carel, W. L. (1961). *Visual factors in the contact analog* (pp. 1–65). Ithaca, N.Y.: General Electric Advanced Electronics Center Publishers R61 ELC60.

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

Dinstein, I. (1988). A new technique for visual motion alarm. *Pattern Recognition Letters, 8,* 347.

Faugeras, O. D., Lustman, F. & Toscani, G. (1987). Motion and structure from motion from point and line matches. *Proceedings of the International Conference on Computer Vision,* London, June 1987, pp. 25–34.

Heeger, D. J. & Hager, G. (1988). Egomotion and the stabilized world. *Proceedings of the 2nd International Conference on Computer Vision,* Tampa, Fla., pp. 435–440.

Heeger, D. J. & Jepson, A. (1990). Visual perception of three-dimensional motion. *MIT Media Laboratory Technical Report* No. 124.

Heel, J. (1990a). Direct dynamic motion vision. In *Proceedings of the IEEE Conference on Robotics and Automation* (pp. 1142–1147). IEEE Computer Society Press.

Heel, J. (1990b). Direct estimation of structure and motion from multiple frames. *MIT Artificial Intelligence Laboratory Memo* No. 1190.

Hildreth, E. C. (1990). Recovering heading for visually-guided navigation. *MIT Artificial Intelligence Laboratory Memo* No. 1297.

Horn, B. K. P. & Weldon, E. J. (1988). Direct methods for recovering motion. *International Journal of Computer Vision, 2,* 51–76.

Huber, P. J. (1981). *Robust statistics.* New York: Wiley.

Jain, R. C. (1983). Direct computation of the focus of expansion. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5,* 58–63.

Jain, R. C. (1984). Segmentation of frame sequences obtained by a moving observer. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6,* 624–629.

Jain, R., Martin, W. N. & Aggarwal, J. K. (1979). Extraction of moving object images through change detection. *Proceedings of the 6th International Joint Conference on Artificial Intelligence,* Tokyo, pp. 425–428.

Jain, R., Militzer, D. & Nagel, H. H. (1977). Separating non-stationary from stationary scene components in a sequence of real world TV images. *Proceedings of the 5th International Joint Conference on Artificial Intelligence,* Cambridge, Mass., pp. 425–428.

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

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

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

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. (1984). Visual ambiguity of a moving plane. *Journal of the Optical Society of America A, 1,* 1215.

Longuet-Higgins, H. C. & Prazdny, K. (1981). The interpretation of moving retinal images. *Proceedings of the Royal Society of London Series B, 208,* 385–397.

Meer, P., Mintz, D., Kim, D. Y. & Rosenfeld, A. (1991). Robust regression methods for computer vision: A review. *International Journal of Computer Vision, 6,* 59–70.

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

Negahdaripour, S. & Horn, B. K. P. (1987). Direct passive navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9,* 168–176.

Negahdaripour, S. & Horn, B. K. P. (1989). A direct method for locating the focus of expansion. *Computer Vision, Graphics and Image Processing, 46,* 303–326.

Nelson, R. C. (1990). Qualitative detection of motion by a moving observer. *University Rochester Computer Science Technical Report* No. 341.

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

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

Rieger, J. H. & 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.

Rousseeuw, P. & Leroy, A. (1987). *Robust regression and outlier detection.* New York: Wiley.

Shariat, H. (1986). The motion problem: How to use more than two frames. PhD. thesis, Department of Electrical Engineering, University Southern California.

Subbarrao, M. (1988). Interpretation of visual motion: A computational study. *Research notes in artificial intelligence.* San Mateo, Calif.: Morgan Kaufmann.

Thompson, W. B. & Pong, T. C. (1990). Detecting moving objects. *International Journal of Computer Vision, 4,* 39–57.

Thompson, W. B., Lechleider, P. & Stuck, E. R. (1992). Detecting moving objects using the rigidity constraint. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* In press.

Tsai, R. Y. & Huang, T. S. (1984a). Estimating three-dimensional motion parameters of a rigid planar patch: III. Finite point correspondences and the three-view problem. *IEEE Transactions on Acoustic Speech Signal Processing, ASSP-32,* 213–220.

Tsai, R. Y. & Huang, T. S. (1984b). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6,* 13–27.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, Mass.: MIT Press.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and rubbery motion. *Perception, 13,* 255–274.

Verri, A., Girosi, F. & Torre, V. (1989). Mathematical properties of the two-dimensional motion field: From singular points to motion parameters. *Journal of the Optical Society of America A, 6,* 698–712.

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,* 160–169.

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.

Warren, W. H., Blackwell, A. W., Kurtz, K. J., Hatsopoulos, N. G. & Kalish, M. L. (1991). On the sufficiency of the velocity field for perception of heading. *Biological Cybernetics, 65,* 311–320.

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

Waxman, A. M. & Wohn, K. (1988). Image flow theory: A framework for 3-D inference from time—varying imagery. In Brown, C. (Ed.), *Advances in computer vision.* New Jersey: Erlbaum.

Weng, J., Huang, T. S. & Ahuja, N. (1989). Motion and structure from two perspective views: Algorithms, error analysis and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-11,* 451–476.

Zhang, Z., Faugeras, O. D. & Ayache, N. (1988). Analysis of a sequence of stereo scenes containing multiple moving objects using rigidity constraints. *Proceedings of the 2nd International Conference on Computer Vision,* Tampa, Fla., pp. 177–186.

Acknowledgements—I thank Shimon Ullman and Eric Grimson for valuable comments on a draft of this paper. This paper describes research done at the Artificial Intelligence Laboratory and the Center for Biological Information Processing at the Massachusetts Institute of Technology. Support for the A. I. Laboratory's research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124. The Center's support is provided in part by the Office of Naval Research, Cognitive and Neural Sciences Division, the National Science Foundation (IRI-8719394 and IRI-8657824) and the McDonnell Foundation.