PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Recovering Heading for Visually Guided Navigation in the Presence of Self-Moving Objects [and Discussion]

Ellen C. Hildreth, H. B. Barlow and H. C. Longuet-Higgins

Phil. Trans. R. Soc. Lond. B 1992 **337**, 305-313
doi: 10.1098/rstb.1992.0108

Recovering heading for visually guided navigation in the presence of self-moving objects

ELLEN C. HILDRETH

Department of Computer Science, Wellesley College, Wellesley, Massachusetts 02181, U.S.A. and MIT Department of Brian & Cognitive Sciences, Cambridge, Massachusetts 02139, U.S.A.

SUMMARY

I 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 the work of Rieger & Lawton (1985) and Longuet-Higgins & Prazdny (1981), whose approach uses velocity differences computed in regions of high depth variation to locate the focus of expansion that indicates the observer's heading direction. We present the results of computer simulations with natural and artificial images and relate the behaviour of the model to psychophysical observations regarding heading judgements.

1. INTRODUCTION

As an observer moves through the environment, the motion of features in the visual image is used to guide him successfully through a complex and constantly changing scene. Essential to the task of navigation is the recovery of the observer's direction of heading relative to the scene and segmentation of the visual image into regions corresponding to distinct objects that may undergo their own motion through space. Such objects may represent a target to be followed or obstacle to be avioded.

Many computational models have been proposed to solve the 'Observer motion problem', which refers to the recovery of the six parameters of the observer's instantaneous rotation and translation through space and the depth or orientation of the surface projected to each image location, from measurements of image motion or temporal change. Existing models for motion measurement often do not perform well in the vicinity of discontinuities in direction or speed of movement that can occur along object boundaries, and few models address the observer motion recovery in the presence of self-moving objects in the scene.

This paper extends a model for recovering heading direction from image motion developed by Longuet-Higging & Prazdny (1981) and Rieger & Lawton (1985). The basic approach considers the detection and analysis of object boundaries to be a central part of observer motion recovery. The next section reviews previous computational and perception studies of observer motion recovery. Section 3 summarizes the work of Longuet-Higgins & Prazdny (1981) and Rieger & Lawton (1985) and describes an extension that addresses the recovery of heading direction in the presence of self-moving objects. Computer simulations summarized in § 4 address the performance of the model on artificial and natural images and its relationship to the human recovery of heading direction. Section 5 considers further experimental questions that arise from this work.

2. BACKGROUND

When an observer moves relative to a stationary environment, he induces a pattern of image motion that depends on his translation and rotation, and the three-dimensional layout of the scene. This translation and rotation can each be expressed in terms of three parameters denoted by $\boldsymbol{T}=(T_x,T_y,T_z)^T$ and $\boldsymbol{R}=(R_x,R_y,R_z)^T$, respectively. Assuming perspective projection with a focal length of 1, the projected velocity $(\dot{x},\dot{y})$ of an image feature at location (x,y) is given by:

$$\dot{x}=\frac{-T_x+xT_z}{Z}+R_xxy-R_y(x^2+1)+R_zy$$

$$\dot{y}=\frac{-T_y+yT_z}{Z}+R_x(y^2+1)-R_yxy-R_zx$$

where $Z(x,y)$ denotes the depth of the surface point projecting to image location (x,y). The first term above represents the component of image velocity due to the observer's translation and depends on depth. The remaining terms represent the component due to the observer's rotation and depend only on $\boldsymbol{R}$ and image location. The translational component alone yields a radial pattern of image velocity that emanates from a single location called the focus of expansion (FOE), corresponding to the observer's direction of heading.

Recent perceptual studies have addressed the ability of the human system to recover heading direction from the changing image (Rieger & Toet

Phil. Trans. R. Soc. Lond. B (1992) **337**, 305–313
Printed in Great Britain

1985; Cutting 1986; Warren & Hannon 1988, 1990; Warren *et al.* 1988, 1991). In these studies, observers view displays of randomly positioned points or lines that undergo displacement over time that would normally be induced by a translation and rotation of the observer relative to a surface such as a ground plane, or one or more frontoparallel planes at different depths. From the observed judgements of heading direction, we can draw the following conclusions (see Hildreth (1991, 1992) for more discussion): human observers can achieve an accuracy of about 1–2° of visual angle at judging heading direction, with or without a visible target; performance improves with higher speeds of translation and a greater range of depth in the scene; extraretinal information regarding eye rotation is used in the recovery of heading direction; and heading can be judged reliably in the presence of significant noise in the image motions. For the case of pure translation, heading direction can be recovered accurately from only 2 or 3 frames, with accuracy increasing with time. Finally, heading direction can be recovered when the rotational and translational flows must be passively decoupled from visual input alone. This decomposition requires differential motion produced by elements at different depths, can be performed successfully with sparse, discontinuous flow fields, and requires a relatively small field of view if the rotational flow is small. These final observations are critical, as they consider how well observers recover heading direction in the presence of a significant rotational component of motion. Perceptual studies have not addressed the ability to recover heading direction in the presence of self-moving objects.

Computational methods for recovering observer motion can be divided into two classes, depending on whether they rely on discrete or continuous image motion measurements. In the discrete approach, isolated image features are tracked over time and their sequence of positions is used to solve for the parameters of three-dimensional structure and motion (Ullman 1979; Prazdny 1980; Longuet-Higgins 1981; Tsai & Huang 1984; Aloimonos & Brown 1989; Weng *et al.* 1989). In the continuous approach, a two-dimensional velocity yield, sometimes together with spatial or temporal derivatives, is used to solve for three-dimensional structure and motion (Koenderink & Van Doorn 1976; Longuet-Higgins & Prazdny 1981; Bruss & Horn 1983; Lawton 1983; Adiv 1985; Subbarao 1988; Waxman & Wohn 1988; Verri *et al.* 1989; Heeger & Jepson 1990). Perrone (1992) proposed a model based on measurements from spatio-temporal filters that reflect the early motion measurement mechanisms in biological systems.

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 most existing algorithms, the human system can tolerate significant noise in these sparse measurements. Recent discrete algorithms that use motion measurements over an extended time exhibit better performance (Ullman 1984; Broida & Challappa 1986; Shariat 1986; Faugeras *et al.* 1987). Many continuous approaches also require accurate velocity measurements, which can make them sensitive to noise. Methods that rely directly on spatial and temporal derivatives of image intensity (Negahdaripour & Horn 1989; Horn & Weldon 1988; Heel 1990) may have difficulty coping with the impoverished displays used in perceptual studies.

Some methods for observer motion analysis can cope with self-moving objects. One approach assumes a stationary camera, with significant image motion indicating self-moving objects (Jain *et al.* 1979; Anderson *et al.* 1985; Bouthemy & Lelande 1990). A variation on this approach uses image stabilization (Burt *et al.* 1989). A second approach assumes that the camera undergoes pure translation, so that self-moving objects violate the expected pure expansion of the image (Jain 1984). If three-dimensional depth data are available, then inconsistency between image velocities, estimated observer motion and depth data can signal self-moving objects (Thompson & Pong 1990, Nelson 1990).

A more general strategy computes initial observer motion parameters and then finds areas of the scene that move in a way that is inconsistent with these parameters (Heeger & Hager 1988; Zhang *et al.* 1988, Thompson *et al.* 1992). If all motion information is used initially, the recovery of motion parameters can be degraded by the inconsistent motions of self-moving objects, especially if they cover a significant portion of the visual field. On the other hand, using spatially local information can yield inaccuracy due to the limited field of view.

3. BUILDING UPON THE RIEGER & LAWTON MODEL

The model of Rieger & Lawton (1985), based on the work of Longuet-Higgins & Prazdny (1981), begins with the observation that along a depth discontinuity, there will be a discontinuity in the translational component of velocity because of its dependence on depth, while the rotational component will be roughly constant across the boundary. Furthermore, the field of vectors representing the differences in velocity across these boundaries will be oriented approximately along the lines connecting their location with the focus of expansion (the translational field lines), and therefore should point to the FOE.

Longuet-Higgins & Prazdny suggested an algorithm that uses spatial derivatives of velocity to recover the FOE, which is very sensitive to error. Rieger & Lawton (1985) developed a more robust algorithm that consists of the following steps. First, the differences between each image velocity and other velocities within a neighborhood are computed. From the resulting distribution of velocity difference vectors, the dominant orientation of the vectors is computed and preserved at locations where this distribution is strongly anisotropic. Such points typically arise where there is strong depth variation. The result of this stage is a set of directions at different image locations that are roughly aligned with the translational field lines.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The FOE is then computed as the 'best-fit' intersection of these vector difference directions. Once the FOE is determined, the direction of the translational component of motion is known at every image location, so that motion perpendicular to this direction must be due to the observer's rotation. From these perpendicular motions, the best rotational parameters are inferred. The full rotational flow field is computed and subtracted from the original flow field to obtain the full translational component, and relative depth at each point is computed from the FOE and magnitude of the translational component of motion.

Rieger & Lawton's algorithm is appealing for several reasons. First, it relies critically on the presence of depth differences that are essential for the human recovery of heading. Second, it provides a rough initial estimate of heading direction, independent of the rotation parameters and three-dimensional structure. Third, it is 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. Finally, it is simple and relies on primitive motion information, such as velocity differences. A question that arises is whether this algorithm can achieve the accuracy of human judgements across the range of conditions considered in perceptual studies. A challenging aspect of human performance is the ability to cope with sparse displays. Over larger distances, the observations that the model relies upon are less valid. This question is addressed in the next section.

The most severe limitation of Rieger & Lawton's model is that it does not cope with self-moving objects. 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 oriented along the translational field lines. Combining these differences with those obtained in regions of stationary surfaces yields error in the computed FOE, especially if self-moving objects cover a large part of the visual field. It is necessary to detect self-moving objects or to remove their influence on the FOE computation in some way.

We consider a strategy for detecting and coping with self-moving objects that builds upon the Rieger & Lawton algorithm. The scheme first computes local velocity differences and determines the dominant orientation of the distribution of velocity differences within a neighborhood of each point, as in the Rieger & Lawton model. The orientations, θ_i, are preserved at points where the distribution of velocity differences is strongly anisotropic. Most of the θ_i measurements preserved are derived 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 location, whereas those 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 FOE location 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 a large set of possible image regions, then choose the region (or regions) with maximum support and use the θ_i measurements that provide this support to derive an FOE estimate. This strategy is similar to the Hough transform (Ballard & Brown 1982).

In more detail, the visual image is divided into overlapping, circular patches that represent possible regions within which the FOE may be located. For each patch P_j, we collect positive evidence for the FOE being located within P_j, which comes from points whose orientation θ_i lies along a line that intersects P_j. If the true FOE is located within P_j, then velocity differences derived from regions of stationary surfaces 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 are the result of error in the θ_i computation. If the true FOE is not located within P_j , there will be some points that yield an orientation θ_i that incorrectly provides positive evidence for an FOE in P_j, but the percentage of points should be substantially reduced. If a sufficiently 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 by considering the extent of the positive evidence in their support and the proximity of the multiple hypotheses (Hildreth 1991, 1992).

The reasoning behind this strategy is that by combining only these θ_i measurements that yield positive evidence for an FOE being located within restricted patches, we reduce the degradation that results from the presence of self-moving objects and from large errors in the θ_i estimates. When patches that contain the true FOE are considered, self-moving objects and large errors 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 do not lead to an FOE hypothesis. Self-moving objects with significant translation near but not along the true translational field lines can distort the computation of the FOE location, but this situation is unlikely to persist for an extended time period, or over an extended image region.

Valid FOE hypotheses may emerge from multiple patches. If there is one FOE hypothesis that accounts for a significantly larger percentage of the θ_i measurements, then this FOE is considered to be the best guess. Multiple FOE locations that are close to one another can be averaged together. 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 particular, extended, connected groups of points that yield negative evidence for the FOE being located

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

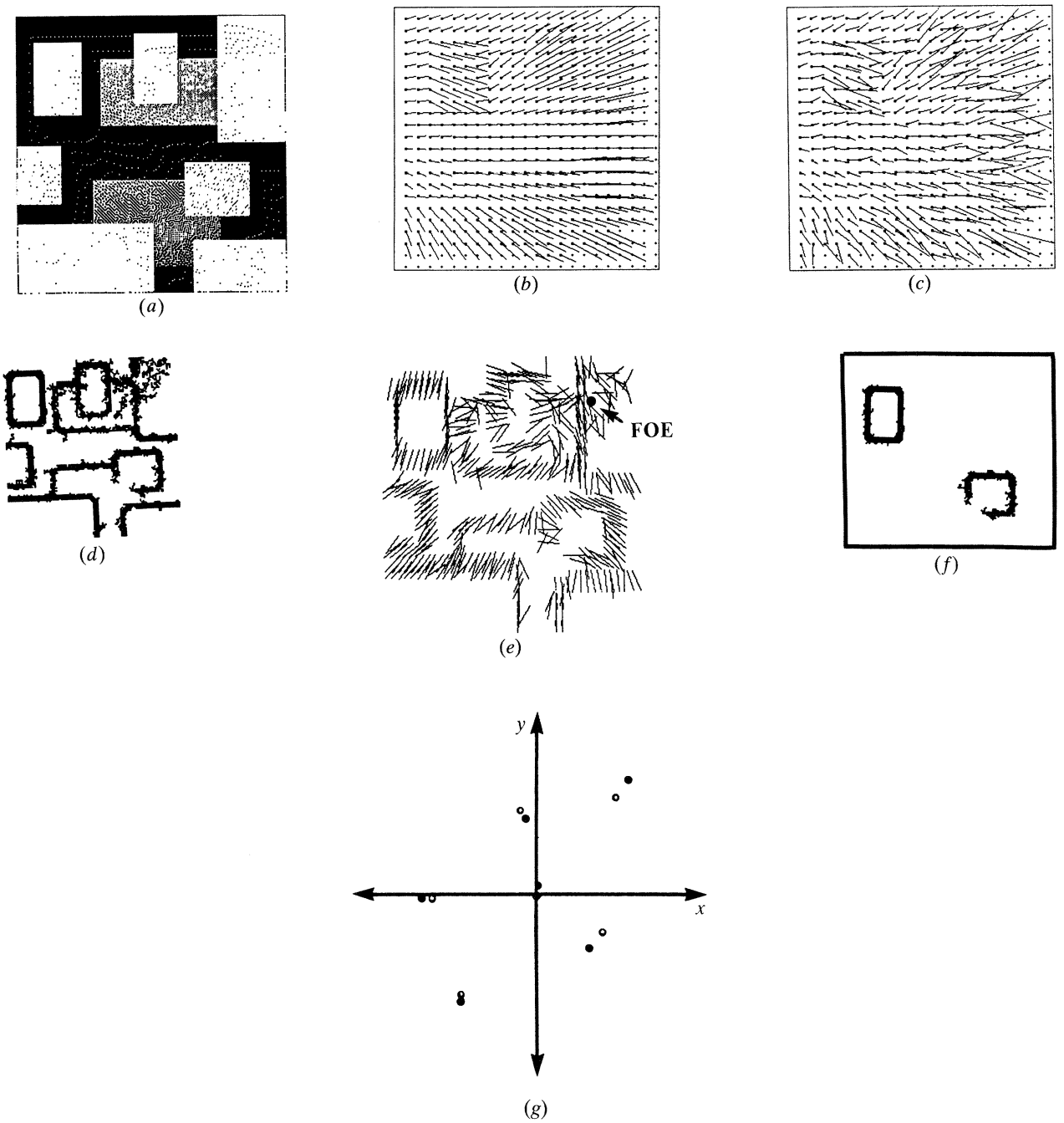


Figure 1. (*a*) A synthetically generated depth map, with brightness encoding depth. (*b*) An ideal velocity field obtained from the known depth map and known observer and object motion parameters. (*c*) The velocity field with added noise. (*d*) A map of extended groups of locations where θ_i were derived from velocity difference distributions with strong anisotropy. (*e*) A sampling of the dominant orientations, θ_i. The true FOE is located in the upper right corner. (*f*) Locations where θ_i measurements were obtained that indicate self-moving objects. (*g*) 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. (From Hildreth 1992.)

within the patch P_j can signal a self-moving object. Isolated points or small groups of points yielded negative evidence are 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 such points are connected to an extended region of points yielding negative evidence, they may represent a continuation of a self-moving object and a new FOE hypothesis can be derived with these points removed.

Some additional modifications were made to the Rieger & Lawton algorithm aimed primarily at improving its performance 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 orien-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

tations, θ_i, that filters the local distributions of velocity differences, and a method for refining the θ_i measurements at a later stage (Hildreth 1991).

4. COMPUTER SIMULATIONS

This section presents the results of computer simulations that consider the performance of the algorithm on artificial and natural images. In the first example, a velocity field was first generated analytically from a known depth map and movement parameters for the observer and two self-moving objects. Noise was added to the image velocities, in the form of Gaussian distributed perturbations of their speed and direction. The algorithm was applied to the noisy velocity field to recover the FOE and to detect self-moving objects.

The initial depth map is shown in figure 1*a*. Brightness encodes depth, with darker objects located further from the observer. (A dithered image is shown, with the density of black and white dots conveying different brightness levels.) The scene consists of planar surface patches of different three-dimensional orientations positioned over a distance of 75–250 units from the observer. From this known depth map and known observer motion parameters, an image velocity field was computed. An example is shown in figure 1*b*. Two of the objects underwent their own motion and the true FOE is located near the upper right corner of the image. Noise was added to yield velocity fields such as that shown in figure 1*c*.

The distribution of velocity differences was computed for each image location, consisting of the differences in velocity between this location and other locations within a neighbourhood of radius 4 pixels. The dominant orientation, θ_i, of this distribution was computed and preserved at locations where the distribution of velocity differences was strongly anisotropic. Isolated θ_i measurements were removed, assuming that the most appropriate θ_i estimates to use for the FOE computation should occur in the vicinity of extended boundaries. The locations of the θ_i that remain after this step are shown in figure 1*d*. 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 1*e* shows the dominant orientations that were computed at a sample of the image locations. There is significant error in the θ_i measurements, as those vectors not located near the self-moving objects should point toward the FOE.

To compute the FOE location, the image was divided into overlying circular patches with a radius of 24 pixels centered every 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 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 and the proximity of the multiple hypotheses. Figure 1*g* shows the true (solid circles) and computed (open circles) FOE locations for six different choices of the observer translation parameters, and for rotation parameters, $\boldsymbol{R} = (0.0, 2.0, 0.0)$ (these parameters were used to generate the velocity field shown in figure 1*b*). 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 possibly indicating self-moving objects. Figure 1*f* shows the final locations hypothesized to arise from self-moving objects, which correspond correctly to the two self-moving objects in the scene.

Figure 2*a* shows a natural image from a sequence taken from a camera that was translated toward the centre of the image while rotating slightly around the vertical axis. Three objects in the scene, indicated by the arrows, underwent their own horizontal motion. An image velocity field was derived from two images using an algorithm that integrates initial motion measurements obtained at the locations of the zero-crossings of the images filtered with a Laplacian-of-Gaussian function, which assumes that the motion of small image patches can be approximated by pure translation (Hildreth 1984). A sampling of the resulting velocity field is shown in figure 2*b*. These image velocities formed the input to the model for heading recovery, which derived the FOE location highlighted with a large dot in figure 2*b*.

The algorithm was also applied to patterns similar to those used in the perceptual studies cited earlier. These simulations used synthetic images of discrete points whose image motion is determined by the motion of an observer relative to a random-dot surface in space. The image motions were computed analytically, and with added noise, formed the input to the model for heading recovery. The perceptual studies used extended image sequences, but the simulations used velocity information obtained from only three frames. A threshold of $1^\circ\,s^{-1}$ was imposed on the absolute image velocity and a threshold of 10% was imposed on the velocity differences. Noise in the form of Gaussian distributed perturbations of the direction and speed of motion was added to the velocities. The simulations examined what level of noise yields results within the range of human performance.

The simulation results are summarized in table 1. Each data point represents an average heading error obtained from 100 different random configurations of points. The parameters used are given in the legend. From the results, it can be seen that direction judgements 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 a small field of view, higher angular rotations yield significant degradation in the direction computation, but if the field of view and number of points are increased, a more accurate heading direction can be obtained for higher rotation speeds.

Simulation results with the three-dimensional cloud and two planes of dots indicate that accuracy

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

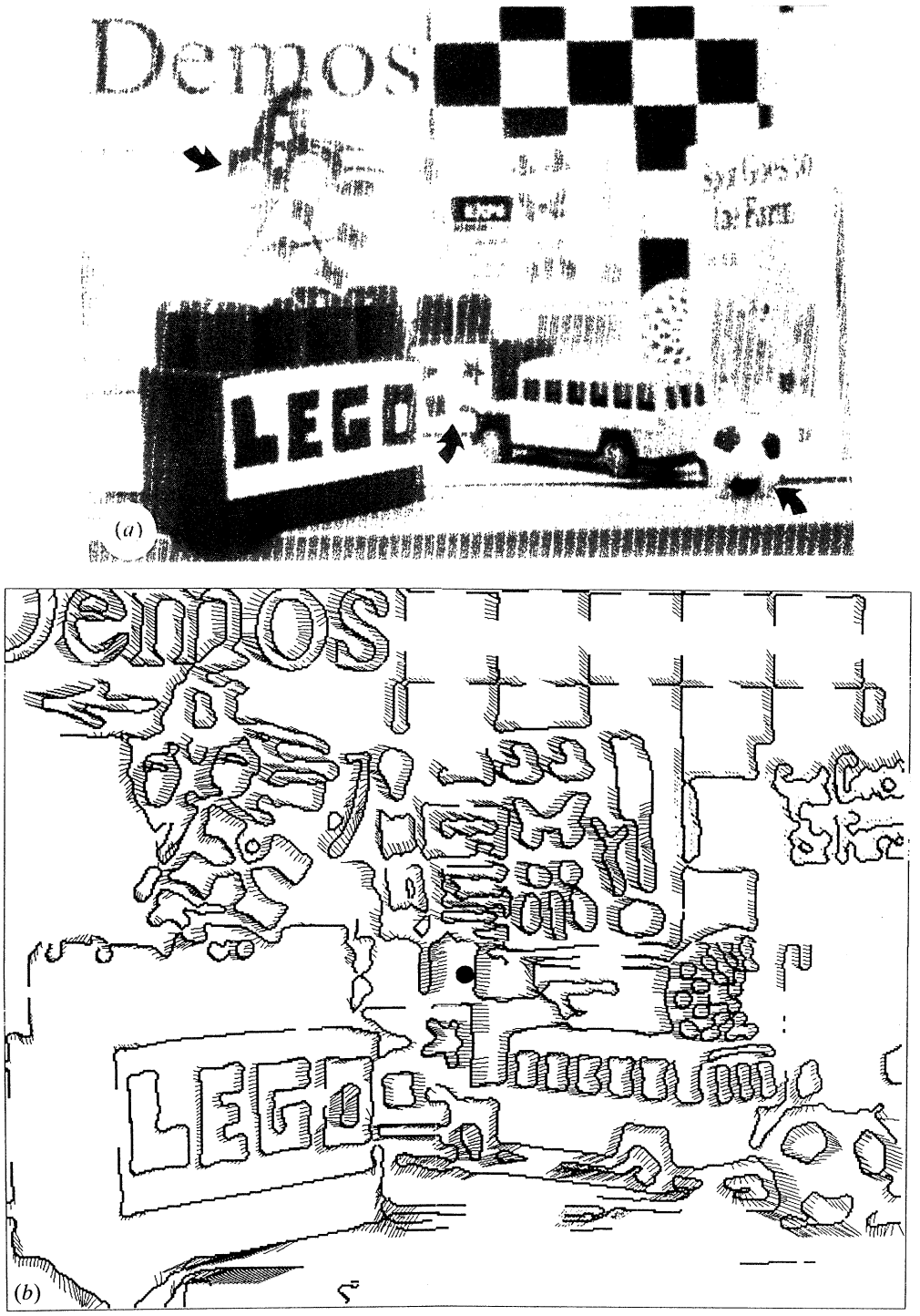


Figure 2. (*a*) A natural image from a sequence taken from a camera that was translated toward the center of the image while rotating slightly around the vertical axis. The three objects indicated by the arrows underwent their own horizontal motion. (*b*) A sampling of the computed image velocity field. The final computed FOE is highlighted with a large dot.

degrades as absolute depth is increased, but improves as the range of depth is increased. A walking speed of 1.9 m s^{-1} was used, so that the time to impact the closer surfaces used in these simulations is 3–4 s. Errors increase slightly with more oblique heading directions. In general, heading direction is underestimated, in that it is closer to straight ahead relative to the true heading. An increased field of view reduces the errors for more oblique headings. Errors increase for sparse patterns, largely because the image neighbourhoods over which the velocity differences are computed contain few pairs of points from which to compute the θ_i measurements. For the case of the frontoparallel plane, the average heading error was 5.0°.

5. SUMMARY AND CONCLUSIONS

This paper considered the computation of the three-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *The results of simulations with the Rieger & Lawton model, applied to images generated by an observer moving along a ground plane or toward a pair of frontoparallal planes*

(Average errors, in degrees, are given for the horizontal component of heading. Unless specified, the simulations used the following parameters: observer speed of 1.9 m s^{-1}, 40° field of view, 6° heading range, 0.3–0.7° s^{-1} rotation range, 25% average error in image speed, and 25° average error in the direction of image velocity. Simulations with the ground plane (top six entries) used 60 points and the simulations with the frontoparallel planes used 80 points. (Adapted from Hildreth 1992.))

experimental conditions	heading error
initial parameters, ground plane	2.5
7.6 m s^{-1}	2.2
20° field of view, 60 points	2.6
20° field of view, 30 points	2.7
40% average speed error	3.9
40° average direction error	
5°–10° s^{-1} rotation range	4.4
three dimensional cloud, depth range 7–40 m	2.3
three dimensional cloud, depth range 15–32 m	4.0
three dimensional cloud, depth range 7–40 m	5.0
10 points	
two planes, 5 m and 25 m	1.5
two planes, 10 m and 20 m	2.6
two planes, 20 m and 40 m	3.7
two planes, 5 m and 25 m	1.8
6°–12° heading range	

dimensional direction of translation of an observer relative to the scene, in the presence of self-moving objects. Driven in part by perceptual observations, we pursued the model of Rieger & Lawton (1985) and proposed extensions 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. Simulation results with visual patterns similar to those used in perceptual studies suggest that it exhibits behavior similar to that of the human system. This work strengthens the idea that local velocity differences can be computed and integrated spatially to recover the observer's heading direction. The particular methods used here to compute the two-dimensional velocities and the dominant direction of their differences are not intended to be biologically plausible. However, velocity sensitive mechanisms in primate area MT have receptive fields that are organized in a centre-surround fashion that yields a sensitivity to relative motion (see, for example, Allman *et al.* (1985)). Furthermore, cells sensitive to global expansions have been found in area MST (see, for example, Tanaka & Saito (1989); Tanaka *et al.* (1989)), and the responses of these cells could, in principle, be derived from velocities, velocity differences, or some other measure of temporal change.

A number of additional questions regarding the human perception of heading direction arise from this analysis. Does heading accuracy decrease with more oblique headings, and is there a tendency to underestimate oblique headings? Is the size of the field of view more critical for the judgement of oblique headings? Is there degradation of heading judgements when larger angular rotations are simulated, and is the size of the field of view critical in this case? Is there a systematic degradation in heading accuracy with a smaller depth range and larger absolute depth? How much of the image must contain significant depth variation? 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 judgements? Is there any difference in performance, depending on whether the boundaries of a self-moving object yield immediately perceptable motion discontinuities? 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? The answers to these questions can provide further cues to the underlying mechanisms for heading recovery.

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.

REFERENCES

Adiv, G. 1985 Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. Patt. Anal. Machine Intell.* **PAMI-7**, 384–401.

Allman, J., Miezin, F. & McGuiness, E. 1985 Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *A. Rev. Neurosci.* **8**, 407–430.

Aloimonos, J. & Brown, C.M. 1989 On the kinetic depth effect. *Biol. Cybern.* **60**, 445–455.

Anderson, C.H., Burt, P.J. & van der Wal, G.S. 1985 Change detection and tracking using pyramid transform techniques. In *Proc. SPIE Conf. on Intelligent Robots and Computer Vision*, pp. 300–305, Boston: SPIE.

Ballard, D.H. & Brown, C.M. 1982 *Computer vision*. Englewood Cliffs, New Jersey: Prentice-Hall.

Bouthemy, P. & Lalande, P. 1990 Detection and tracking of moving objects based on a statistical regularization method in space and time. In *Proc. First European Conf. Comp. Vision* (ed. O. Faugeras), pp. 307–311. Berlin: Springer-Verlag.

Broida, T.J. & Challappa, R. 1986 Estimation of object motion parameters from noisy images. *IEEE Trans. Patt. Anal. Machine Intell.* **PAMI-8**, 90–99.

Bruss, A.R. & Horn, B.K.P. 1983 Passive navigation. *Comp. Vis. Graph. Image Proc.* **21**, 3–20.

Burt, P.J., Bergen, J.R., Hingorani, R., Kolczinski, R., Lee,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

W.A., Leung, A., Lubin, J. & Shvaytser, H. 1989 Object tracking with a moving camera, an application of dynamic motion analysis. In *Proc. IEEE Workshop on Visual Motion*, pp. 2–12, IEEE Computer Society Press.

Cutting, J.E. 1986 *Perception with an eye towards motion*. Cambridge, Massachusetts: The MIT Press.

Faugeras, O.D., Lustman, F. & Toscani, G. 1987 Motion and structure from motion from point and line matches. In *Proc. First Int. Conf. on Comp. Vision*, pp. 25–34. IEEE Computer Society Press.

Heeger, D.J. & Hager, G. 1988 Egomotion and the stabilized world. In *Proc. 2nd Int. Conf. Comp. Vision*, pp. 435–440. IEEE Computer Society Press.

Heeger, D.J. & Jepson, A. 1990 Visual perception of three-dimensional motion. *MIT Media Lab. Tech. Rep.* 124.

Heel, J. 1990 Direct Dynamic motion vision. In *Proc. IEEE Conf. Robotics and Automation*, pp. 1142–1147. IEEE Computer Society Press.

Hildreth, E.C. 1984 *The measurement of visual motion*. Cambridge, Massachusetts: The MIT Press.

Hildreth, E.C. 1991 The recovery of heading for visually-guided navigation. *MIT Artif. Intell. Lab. Memo 1297.*

Hildreth, E.C. 1992 The recovery of heading for visually-guided navigation. *Visual Res.* (In the press.)

Horn, B.K.P. & Weldon, E.J. 1988 Direct methods for recovering motion. *Int. J. comp. Vis.* **2**, 51–76.

Jain, R.C. 1983 Direct computation of the focus of expansion. *IEEE Trans. Patt. Anal. Machine Intell.* **PAMI-5**, 58–63.

Jain, R.C. 1984 Segmentation of frame sequences obtained by a moving observer. *IEEE Trans. Patt. Anal. Machine Intell.* **PAMI-6**, 624–629.

Jain, R., Martin, W.N. & Aggarwal, J.K. 1979 Extraction of moving object images through change detection. In *Proc. Sixth Int. Joint Conf. Artif. Intell.*, pp. 425–428. IJCAI.

Koenderink, J.J. & Van Doorn, A.J. 1976 Local structure of movement parallax of the plane. *J. opt. Soc. Am.* **66**, 717–723.

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

Longuet-Higgins, H.C. & Prazdny, K. 1981 The interpretation of moving retinal images. *Proc. R. Soc. Lond.* B **208**, 385–397.

Negahdaripour, S. & Horn, B.K.P. 1989 A direct method for locating the focus of expansion. *Comp. Vis. Graph. Image Proc.* **46**, 303–326.

Nelson, R.C. 1990 Qualitative detection of motion by a moving observer. *Univ. Rochester Comp. Science Tech. Rep. 341*, April.

Perrone, J.A. 1992 Model for the computation of self-motion in biological systems. *J. opt. Soc. Am.* A **9**, 177–194.

Prazdny, K. 1980 Egomotion and relative depth map from optical flow. *Biol. Cyber.* **36**, 87–102.

Rieger, J.H. & Lawton, D.T. 1985 Processing differential image motion. *J. opt. Soc. Am.* A **2**, 354–360.

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

Shariat, H. 1986 The motion problem: How to use more than two frames. Ph.D. thesis, University of Southern Calif.

Subbarao, M. 1988 *Interpretation of visual motion: a computational study. Research notes in artificial intelligence*. San Mateo: Morgan Kaufmann.

Tanaka, K. & Saito, H. 1989 Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the Medial Superior Temporal area of the Macaque monkey. *J. Neurophysiol.* **62**, 626–641.

Tanaka, K., Fukada, Y. & Saito, H. 1989 Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the Medial Superior Temporal area of the Macaque monkey. *J. Neurophysiol.* **62**, 642–656.

Thompson, W.B., Lechleider, P. & Stuck, E.R. 1992 Detecting moving objects using the rigidity constraint. *IEEE Trans. Patt. Anal. Machine Intell.* (In the press.)

Thompson, W.B. & Pong, T.C. 1990 Detecting moving objects. *Int. J. comp. Vis.* **4**, 39–57.

Tsai, R.Y. & Huang, T.S. 1984 Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. Patt. Anal. Machine Intell.* **PAMI-6**, 13–27.

Ullman, S. 1979 *The interpretation of visual motion*. Cambridge, Massachusetts: The 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. *J. opt. Soc. Am.* A **6**, 69–712.

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. *Biol. Cybern.* **65**, 311–320.

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

Warren, W.H. & Hannon, D.J. 1990 Eye movements and optical flow. *J. opt. Soc. Am.* A **7**, 160–169.

Warren, W.H., Morris, M.W. & Kalish, M. 1988 Perception of translational heading from optical flow. *J. exp. Psychol.: Human Percept. Perf.* **14**, 646–660.

Waxman, A.M. & Wohn, K. 1988 Image flow theory: A framework for 3-D inference from time-varying imagery. In *Advances in computer vision* (ed. C. Brown), pp. 165–224. 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 Trans. Patt. Anal. Machine Intell.* **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. In *Proc. 2nd Int. Conf. Comp. Vision*, pp. 177–186. IEEE Computer Society Press.

Discussion

H. B. Barlow (*Physiological Laboratory, University of Cambridge, U.K.*). In Professor Hildreth's diagrams the arrows on the straight edge segments indicate the true direction of motion. How was this computed, and would the use of the orthogonal components alone make the detection of the FOE more difficult or impossible?

E. C. Hildreth. The orthogonal component of velocity was computed first, by comparing the spatial and temporal gradients of the filtered image at the locations of the zero-crossing contours. These orthogonal components were then combined over limited segments of the zero-crossing contours to compute a two-dimensional velocity vector, assuming that these segments underwent a pure translation across the image. For the purpose of this simulation, velocity

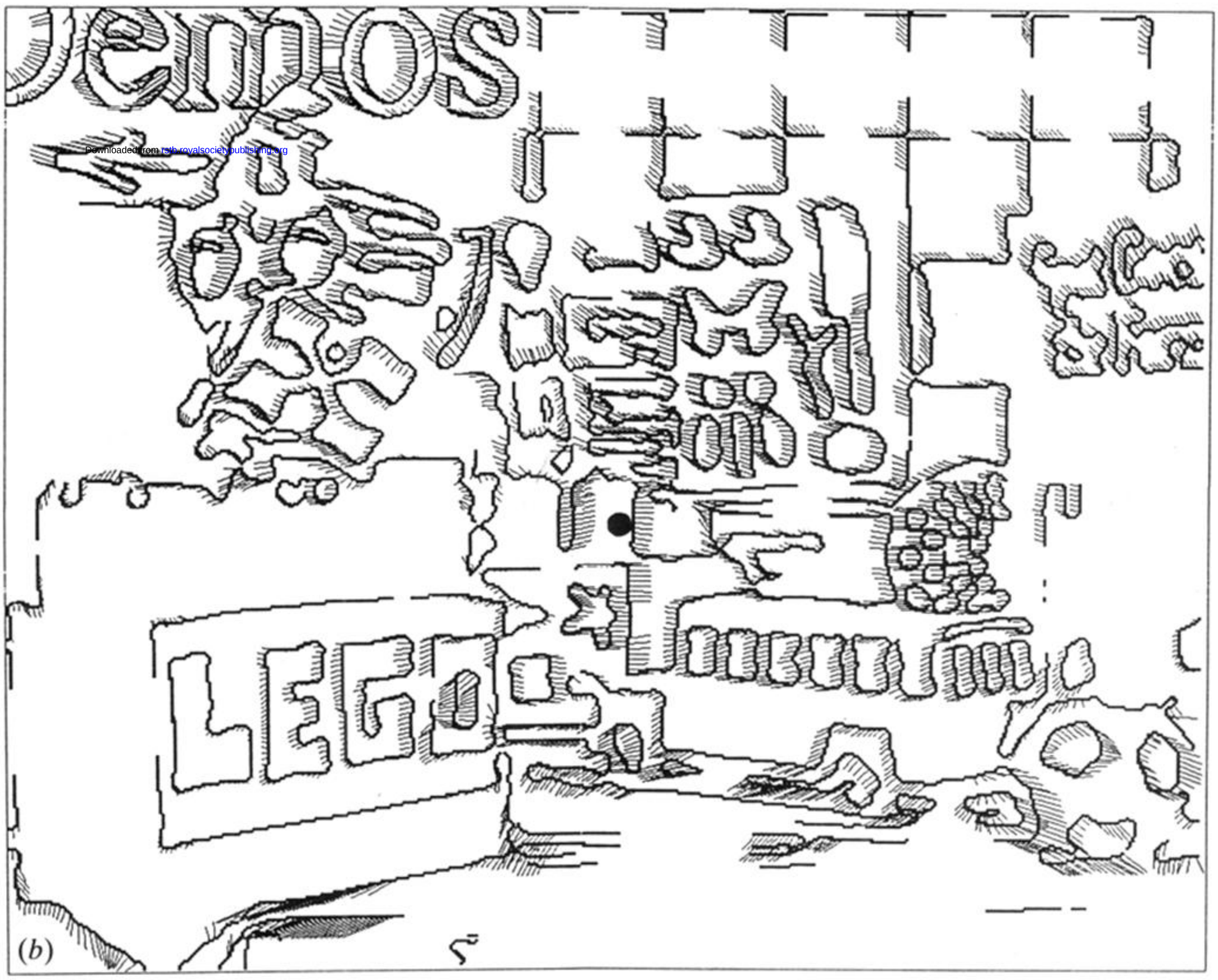
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

vectors were only derived in regions where the zero-crossing contours assumed a range of two-dimensional orientations. The idea of using the orthogonal components directly for the FOE computation has not been considered in this model, but it has been addressed in recent work by Perrone (1992).

H. C. LONGUET-HIGGINS (*Laboratory of Experimental Psychology, University of Sussex, U.K.*). When a scene comprises a number of objects in relative motion, can these methods be used for determining the relative positions of more than one object at a time?

E. C. HILDRETH. Yes, one can think of multiple moving objects as providing evidence for multiple FOE locations that each reflect the relative direction of translation between the observer and an object in the scene. The FOE location with strongest support may correspond to the direction of translation of the observer relative to stationary components of the scene, but other FOE hypotheses that have strong support from a coherent region of the image can be used both to isolate self-moving objects and to determine their direction of translation relative to the observer.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

igure 2. (*a*) A natural image from a sequence taken from a camera that was translated toward the center of the nage while rotating slightly around the vertical axis. The three objects indicated by the arrows underwent their wn horizontal motion. (*b*) A sampling of the computed image velocity field. The final computed FOE is highlighted ith a large dot.