Determining the Instantaneous Axis of Translation from Optic Flow Generated by Arbitrary Sensor Motion

J. H. Rieger and D. T. Lawton

COINS Technical Report 83-1

January 1983

Abstract

This paper develops a simple and robust procedure for determining the instantaneous axis of translation from image sequences induced by unconstrained sensor motion. The procedure is based upon the fact that difference vectors at discontinuities in optic flow fields generated by sensor motion relative to a stationary environment are oriented along translational field lines. This is developed into a procedure consisting of three steps: 1) locally computing difference vectors from an optic flow field; 2) thresholding the difference vectors; and 3) minimizing the angles between the difference vector field and a set of radial field lines which correspond to a particular translational axis. This method does not require *a priori* knowledge about sensor motion or distances in the environment. The necessary environmental constraints are rigidity and sufficient variation in depth along visual directions to endow the flow field with discontinuities. The method has been successfully applied to noisy, sparse, and low resolution flow fields generated from real world image sequences. Experiments are reviewed which indicate that the human visual system also utilizes discontinuities in optic flows in determining self-motion. In addition, due to the computational simplicity of the procedure, hardware realization for real-time implementation is possible.

1. Introduction

The motion of an observer/sensor is in general composed of a translation and a rotation. It generates an optic flow field in the image plane of the sensor due to changes of visual directions of details in the environment over time (Gibson et. al. 1955). The instantaneous translatory velocity of the sensor induces a radial image velocity field with the intersection of the translational axis and image plane as its center. Likewise the rotatational velocity field of the sensor induces a rotational velocity field in the image that is purely direction dependent (Longuet-Higgins and Prazdny 1980). In general an optical velocity field is the vector sum of translational and rotational fields (footnote 1).

The translational component (and its spatial and temporal derivative fields) contains, e.g., information about the shape of objects (Koenderink and van Doorn 1977), about the relative depth properties of the environment (Lee 1980, Prazdny 1980), or about motion parameters for navigating along curved trajectories (Rieger 1983). The rotational component, on the other hand, contains no environmental depth information and needs to be separated from the translational component. The main computational step toward extracting the translational field from optic flows containing translational and rotational components is

footnote: we refer to both the optical velocity field generated by continuous sensor motion and the optical displacement field generated by discrete sensor motion as optical flow fields. Which is being referred to should be obvious by context.

to determine the instantaneous axis of sensor translation.

There are several problems in using these and related formulations for determining camera motion parameters and environmental information from real world image sequences. The inference techniques generally require high resolution image displacements as input and are sensitive to the noise and errors that current techniques for determining image motions typically produce. They can also involve solving complex equations and require significant computation.

We show that the recovery of camera motion parameters can be simplified and performed robustly from noisy, low resolution, and sparse displacement fields by analyzing the image displacements at image positions where environmental depth changes occur. This procedure utilizes an observation made by Longuet-Higgins and Prazdny (1980) that details in the environment located in the same direction from an observer/sensor (i.e. along the same ray of projection), but are at different depths, will differ in their image velocity vectors by the difference of their translational components only. This is because the rotational components of optic flows are purely direction dependent. The axis of sensor translation is obtained from the intersection of radial fieldlines which are determined by such difference vectors. In cluttered environments we find details being located in the same visual direction but being separated in depth at occluding edges. Figure 1 shows an optical flow field induced by a sensor translating and rotating relative to two surfaces that are

separated in depth. We can see that the difference vectors of the vector pairs at the edge point toward the direction of translation indicated in the figure.

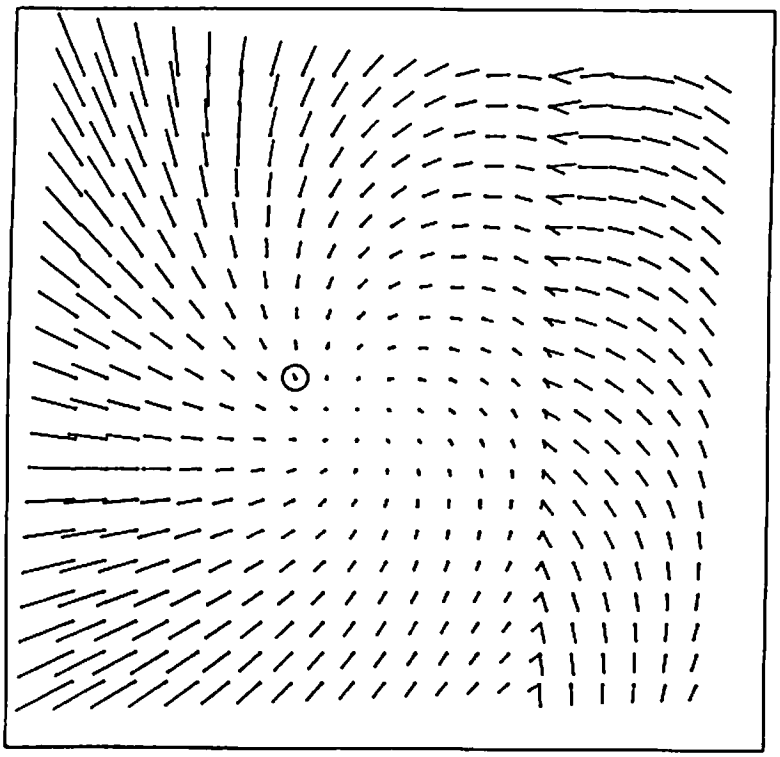

Fig. 1

There are significant difficulties in applying this observation to actual image sequences. We obtain image displacements and not instantaneous optic velocities from images formed at discrete, successive instants. Thus the computation must be expressed in terms of discrete sensor motions. Flow fields computed from actual image sequences are not arbitrarily dense and are in fact generally sparse so there will not be two distinct flow vectors positioned at the same image point. Thus it is necessary to perform the computation using difference vectors determined from image displacement vectors which are spatially separated. Also, real flow fields are noisy and errorful, especially near

occlusion boundaries because of the changes in image structure that occur there. Thus the procedure must be robust to such distortions in the determined difference vectors.

These problems are addressed in this paper. In section 2 we analyze the error introduced by computing difference vectors from spatially separated pairs of vectors. This will lead to a simple method of determining sensor translation from sparse image displacement fields in section 3. In section 4 the results of applying the procedure to simulated data will be compared with its predicted behavior. Results obtained using displacement fields generated from actual image sequences will be described in section 5. Finally, we will discuss in section 6 the relation between the method and psychophysical experiments.

2.1. Components of difference vectors between spatially separated optic flow vectors

In this section we decompose a difference vector formed from spatially separated image velocity vectors into a signal component oriented along the correct translational field line and a noise component. The signal component increases for difference vectors formed at image locations where large depth changes occur in the corresponding environmental positions. It also increases with increasing distance between the difference vector and the intersection of the translational axis with the image plane. To the extent that these conditions are satisfied for an optic flow field, its difference vector field will approach the corresponding set of correct translational field lines. Note

that this does not require knowledge about the location of occlusion boundaries or of image areas corrsponding to large visual slant [del*(1/distance)].

Let us consider a difference vector $\overrightarrow{\Delta F}$ at a point $\tilde{P}_1$ in the image, which has been computed by subtracting the flow vector at a point $\tilde{P}_2$ that is separated from $\tilde{P}_1$ by some distance (d_x, d_y) in the image plane from the flow vector at $\tilde{P}_1$ (see figure 2).

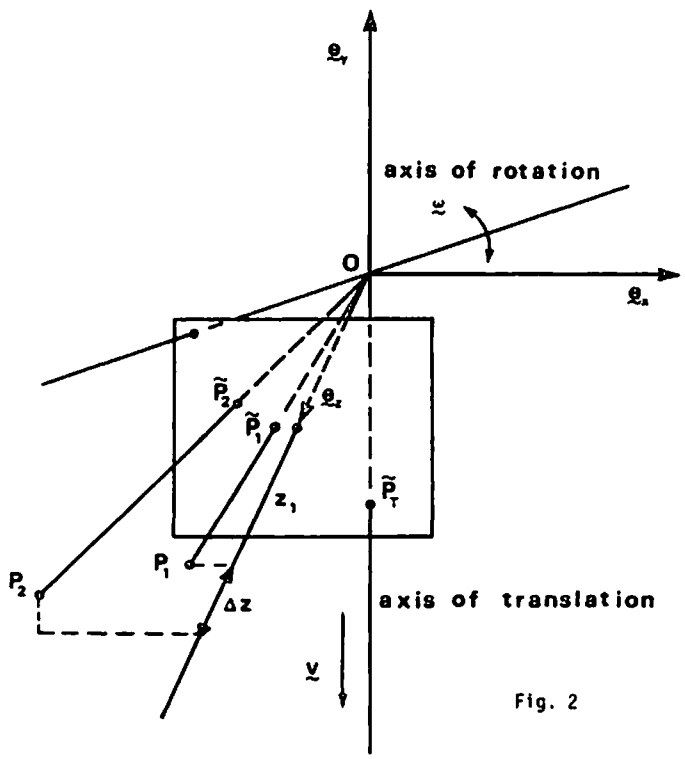

Fig. 2

We obtain a component $\overrightarrow{\Delta F_R}$ that is due to changes in the rotational component field [the field induced by the rotational velocity $(\omega_x, \omega_y, \omega_z)$ of the sensor]

$$\overrightarrow{\Delta F_R} = (d_y\omega_z - d_x^2\omega_y + d_xd_y\omega_x)\underset{\sim}{e}_x + (-d_x\omega_z - d_xd_y\omega_y + d_y^2\omega_x)\underset{\sim}{e}_y .$$

If we let $\tilde{P}_T$ denote the intersection of the translational axis with the image plane, V_Z the translational velocity of the sensor along the z-axis, Z_1 the depth of P_1, and ΔZ the depth difference between P_1 and P_2, the translational component of the difference vector reads

$$\overrightarrow{\Delta F_T} = \frac{V_Z}{Z_1 + \Delta Z} \left\{ \overrightarrow{\tilde{P}_1\tilde{P}_2} - \frac{\Delta Z}{Z_1} \overrightarrow{\tilde{P}_T\tilde{P}_1} \right\} .$$

We can rewrite $\overrightarrow{\Delta F}$ as consisting of a component along a translational fieldline and a noise component

$$\overrightarrow{\Delta F} = \left[\frac{-V_Z \Delta Z}{Z_1 Z_2} \overrightarrow{\tilde{P}_T\tilde{P}_1} \right]_{\text{Signal}} + \left[\frac{V_Z}{Z_2} \overrightarrow{\tilde{P}_1\tilde{P}_2} + \overrightarrow{\Delta F_R} \right]_{\text{Noise}} .$$

For difference vectors with sufficient angular separation from the translatory axis and separation in depth $\overrightarrow{\Delta F}_{\text{Signal}} >> \overrightarrow{\Delta F}_{\text{Noise}}$.

2.2. Global behavior of difference vector field, possible filter operations

From the previous discussion it is now possible to make some predictions about deviations between the orientations of the difference vector field and the corresponding translational field. The orientation of the noise component of each difference vector is affected by the relative positions of its orginal vector pair. If we assume a random distribution for the positions of image velocity vectors, the difference vector field fits the correct set of translational field lines better as the number of difference vectors is increased since $\overrightarrow{\Delta F}_{\text{Signal}}$ and $\overrightarrow{\Delta F}_{\text{Noise}}$ are additive. An additional factor affecting the fit of the difference vector field to the correct translational field

lines is the length of the difference vectors. If a difference vector is small compared to the local average magnitude of the difference vector field, it is more probable that its orientation will be different from the correct translational fieldline. We therefore filtered the difference vector fields in the experiments described in the next sections by thresholding on difference vector length. As was expected the fit of the difference vectors improved up to a certain magnitude of the length threshold and then detoriated again due the smaller number of difference vectors. So far the value for the threshold was found by trial and error, for future experiments it would useful to adjust the threshold automatically to an optimal value given information about the difference vector field like density, average magitude and dispersion of vectors.

3. Implementation

Before describing the optimization procedure, it is necessary to develop the calculation in terms of discrete sensor motions and describe how image difference vectors are determined.

If we are dealing with displacements of details in images formed over discrete time intervals instead of an instantaneous optic velocity field, we have to be careful to describe all quantities with respect to the same reference system. Suppose two environmental points lie along the same ray of projection in an image at time t. Translating and rotating the sensor will displace the projections of these points to new positions in the

image at time t+1. In the image at time t+1, the image points will be separated due to the translational component of the sensor motion (unless they are located on the translational axis). The separated image points and the intersection of the translational axis with the image plane will be collinear at time t+1. This is the discrete analog of the fact that difference vectors at discontinuities of an instantaneous optic velocity field are oriented along translational field lines.

3.1. Determining Difference Vectors from Sparse Flow Fields

Given image displacements D1 and D2 at positions P1 and P2, the difference vector between points 1 and 2 is obtained by subtracting D2 from D1 and positioning the resulting vector at P1 + D1. Two thresholds are used in evaluating difference vectors. The separation threshold determines the maximal allowable distance between displacement vectors in determining difference vectors. The neighborhood of a given displacement vector contains all other displacement vectors which lie within a distance determined by the separation threshold. The length threshold determines the minimal allowable length for a difference vector.

Since it is not necessary to determine occlusion boundaries before forming the difference vectors, the procedure is applied homogenously with respect to a displacement field. Each vector determines a set of difference vectors in its neighborhood which are of sufficient length. The resulting set of difference

vectors determined for all the displacement vectors are then used by the optimization procedure.

3.2. Optimization Procedure

The procedure used to determine a translational axis from a set of difference vectors is similar to that used in (Lawton 1982) to determined a translational axis from a noisy displacement field generated by translational motion. It involves finding a translational axis and the corresponding set of radial field lines which minimizes the measure

$$\sum_{i=1}^{n} (1.0 - \text{abs}(\cos\theta_i))$$

where θ_i is the angle between the ith difference vector and the radial field line at that position in the image.

The error measure is defined on a unit sphere with each point corresponding to a translational axis. The sphere has significant advantages over the image plane as the domain since it allows for a uniform, global sampling of the error function. Using a spherical coordinate system (r, θ, ϕ) where

$$x = r \sin\phi \sin\theta$$

$$y = r \cos\phi$$

$$z = r \sin\phi \cos\theta$$

each axis of translation is defined by some point (θ, ϕ) or $(\theta+\pi, \phi+\pi)$ on the unit sphere. Due to this redundancy, we restrict the error function to the hemisphere determined by the bounds $-\pi/2 \leq \theta \leq \pi/2$ and $0 \leq \phi \leq \pi$.

The search process consists of a global sampling of the error measure to determine its rough shape followed by a local search to find a minimum. The global search is an instance of a generalized Hough Transform (Ballard 1980, O' Rourke 1981) inwhich each difference vector votes against a particular translational axis by the term in the error measure above. The results of the global search are stored in a global error histogram which is indexed by positions seperated by regular intervals on the unit sphere. In the experiments below the global sampling was performed densely with respect to the unit hemisphere to display the error function. The local search utilizes steepest descent with diminishing step-size and is intitialized where the minimum value is determined by the global sampling. Below, the local search was defined with respect to pixel coordinates in the image plane with the finest resolution set to one pixel. Subpixel interpolation was not used but could be without difficulty.

4. Experiments with Simulated Data

Several experiments have been performed with simulated displacement fields to understand the effects of such factors as resolution, neighborhood size, noise, and environmental depth variance. Two are presented here. The first shows the effects of using low resolution displacement fields. The second shows the behaviour of the procedure as environmental depth variance is increased.

4.1. Experiment 1

The flow field in figure 3a shows image displacements positioned at pixel positions having coordinates which are multiples of 8 from a 128x128 pixel field. The components of the displacement vectors were stored as 8 bit integers. Fewer bits were actually required since the maximal displacment was less than 16 pixels in length. The environment consisted of a surface at depth of 10 units along the z axis and a background surface at a depth of 30 units along the z axis. The obvious discontinuities in the flow field in figure 3a indicate the boundary of the nearer surface. The sensor motion consisted of an intial rotation of 0.1 radians about the (1,1,1) axis followed by a translation of 2 units along (0,0,1). The separation threshold was set to 1 pixel since there was a displacement vector at each pixel location. The length threshold was set to 3 pixels. The resulting error function is shown in figure 3b (Darker in the figure corresponds to less error; also recall that this is a plot of a hemisphere in theta,

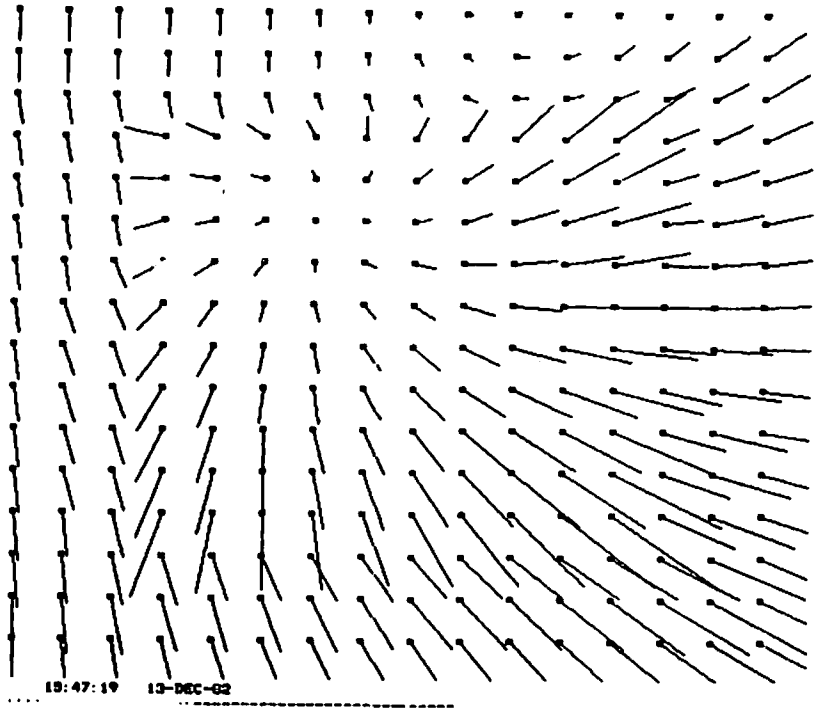

Figure 3a

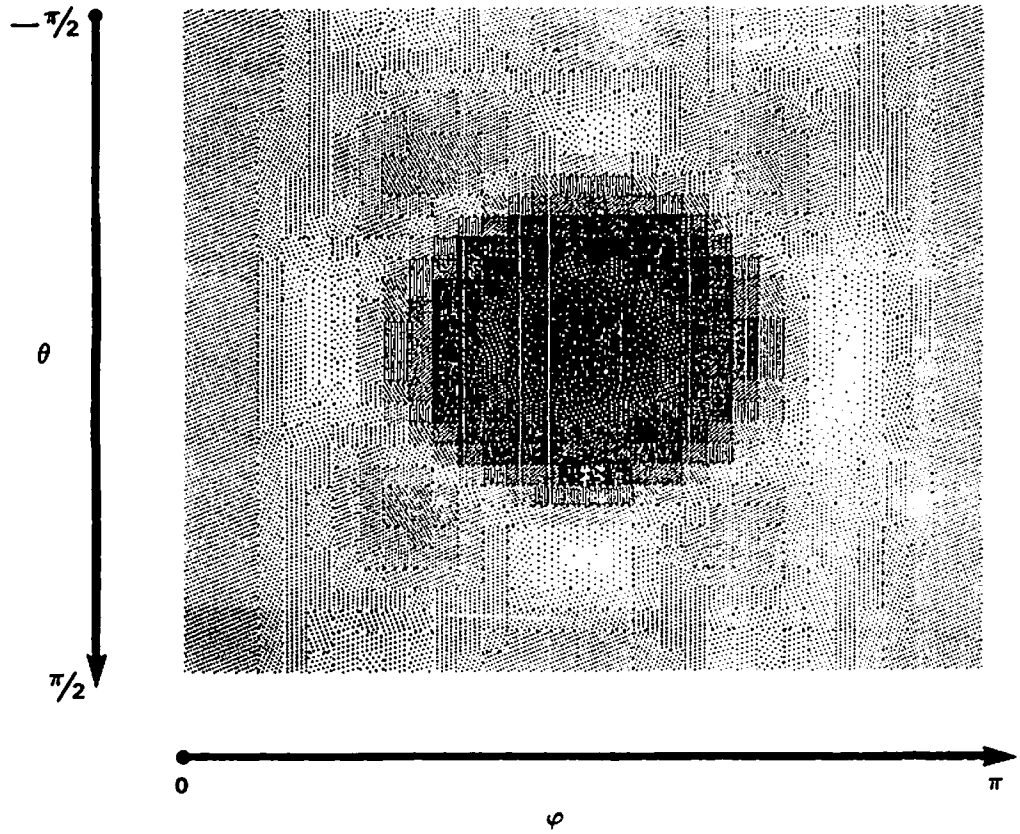

Figure 3b

phi coordinates and not the image plane). As can be seen, it is strongly unimodal. The minimum in the global histogram corresponded to the image position (60.28, 60.28). The local search determined the minimum to be at (63, 63). The correct, subpixel, position was (63.5, 63.5).

4.2. Experiment 2

The flow fields in figures 4a and 5a are 32x32 pixel fields using 32 bit real numbers for their components. They were produced using the same motion as in experiment 1 except the translational displacement along the z-axis was only 1 unit. To see the effects of depth variance, the field in figure 4a was produced by moving relative to a plane 20 units away perpendicular to the z-axis (no depth variance) while the field in figure 5a was produced by moving relative to environmental points with random depths between 20 and 120 units. The separation threshold was set to one pixel because of the density of the field and the length threshold was set to 0.0. The associated error functions are shown in figures 4b and 5b (the error function plots were normalized independently; the error values in figure 5b are smaller than those in figure 4b). For the case of no depth variance (figures 4a and 4b), the minimum in the global histogram was determined to correspond to the image position (19.97, 7.897). The local search determined the minima to be at (20, 9). The correct position was (15.5, 15.5). As is expected from the discussion in section 2, the minimun is incorrect and the error function unsharp. For the case of increased depth variance

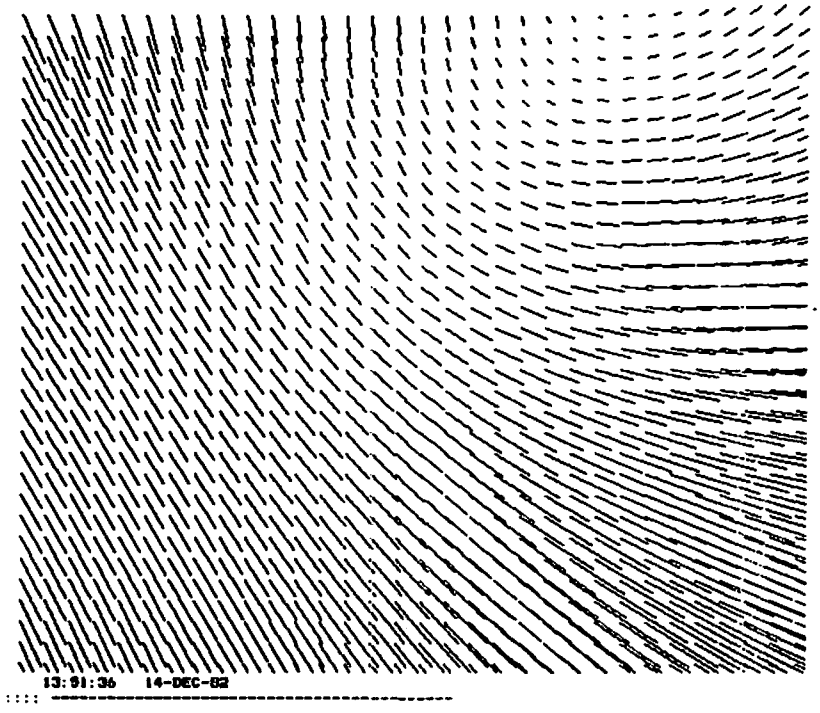

Figure 4a

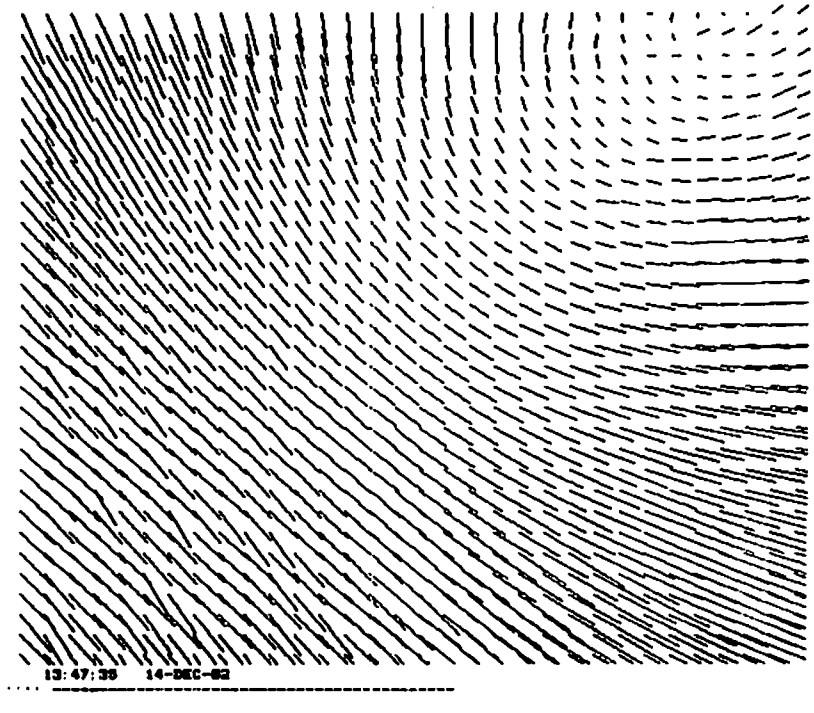

Figure 5a

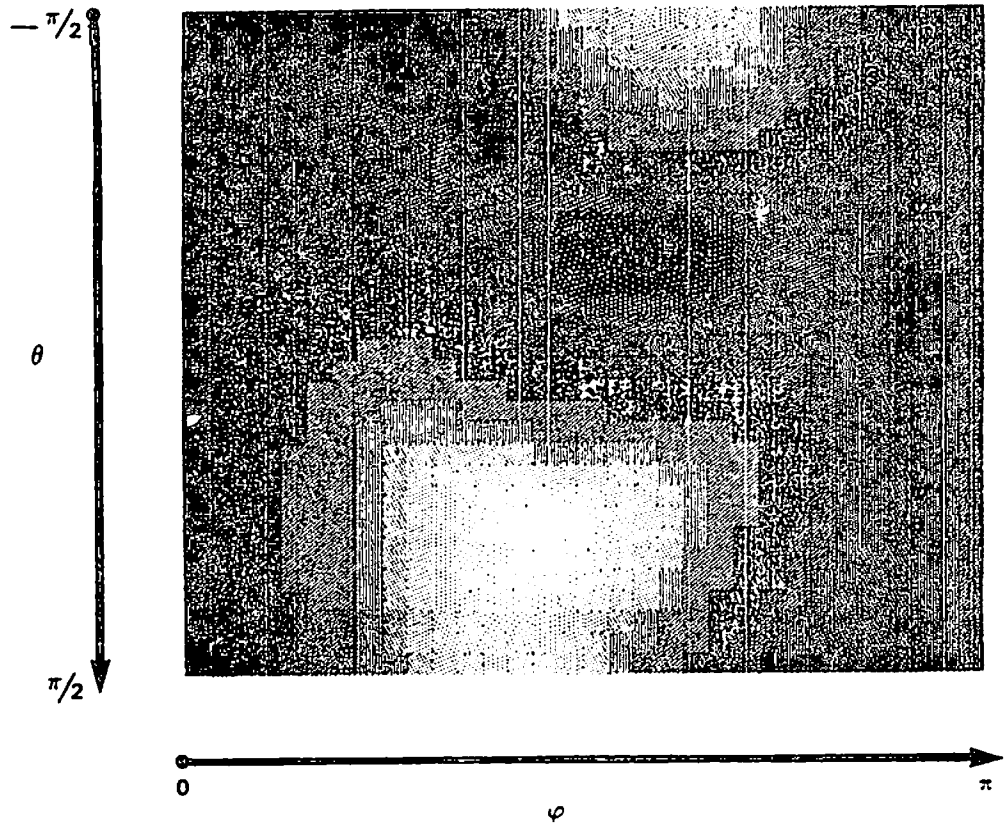

Figure 4b

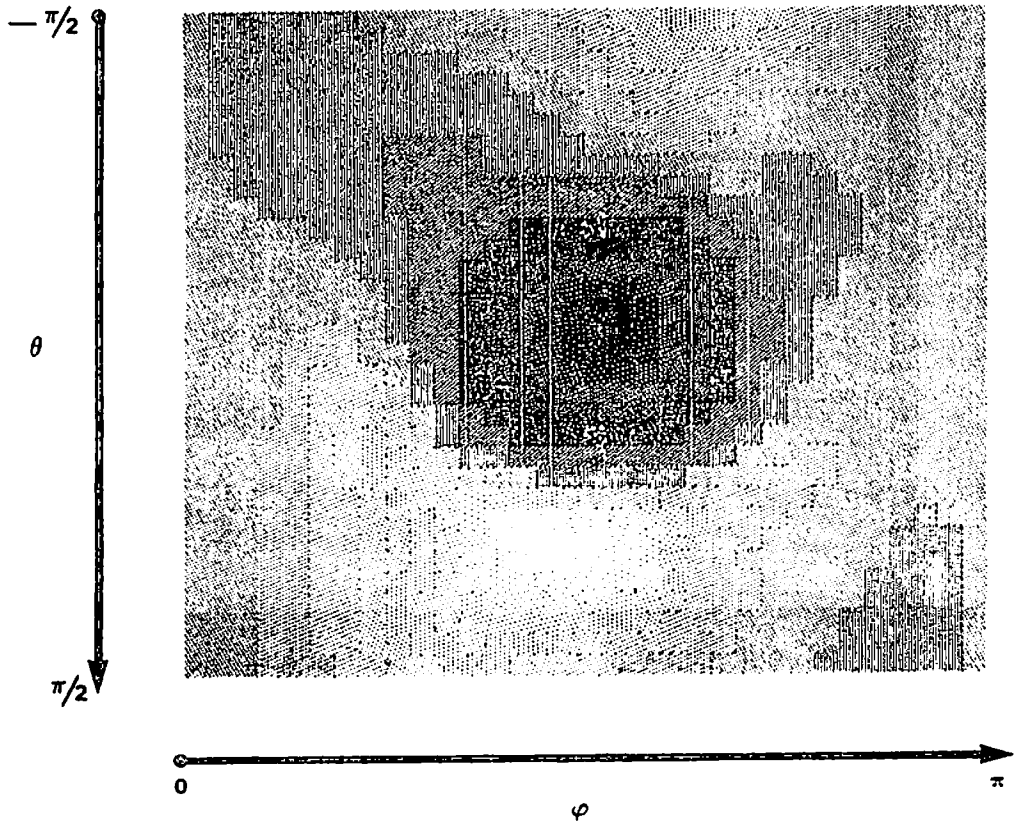

Figure 5b

(figures 5a and 5b), the minimum in the global histogram was determined to be at (16.29, 14.71). The local search determined the minimum to be at (17,15), showing the expected improvement with increased variance in environmental depth.

5. Real Data

Figures 6a and 6b are 128x128 pixel images with 256 intensity levels taken from a GE TN2200 solid state camera. The camera was displaced roughly in the general direction of its z-axis between two textured objects towards a textured background and then rotated about its y axis a few degrees. Figure 6c shows the displacements determined for a set of interesting points extracted from the image in figure 6a using the interest operator described in (Lawton 1983). The displacements were found by correlating 5x5 pixel windows centered at these positions in the first image with 5x5 pixel windows positioned at locations within +/- 15 pixels in the x and y directions in the succeeding image. Displacements for points within 10 pixels of the image boundary were ignored.

The separation threshold was set to 10 pixels and the length threshold was set to 3 pixels. A plot of the error function produced using these threshold values in shown in figure 6d. The local search found a minimum at (52, 75). The correct position of the intersection of the translational axis and the image plane for the second image was determined to be at (57.97, 74.58). Since the focal length was rather long, the determined

Figure 6a

Figure 6b

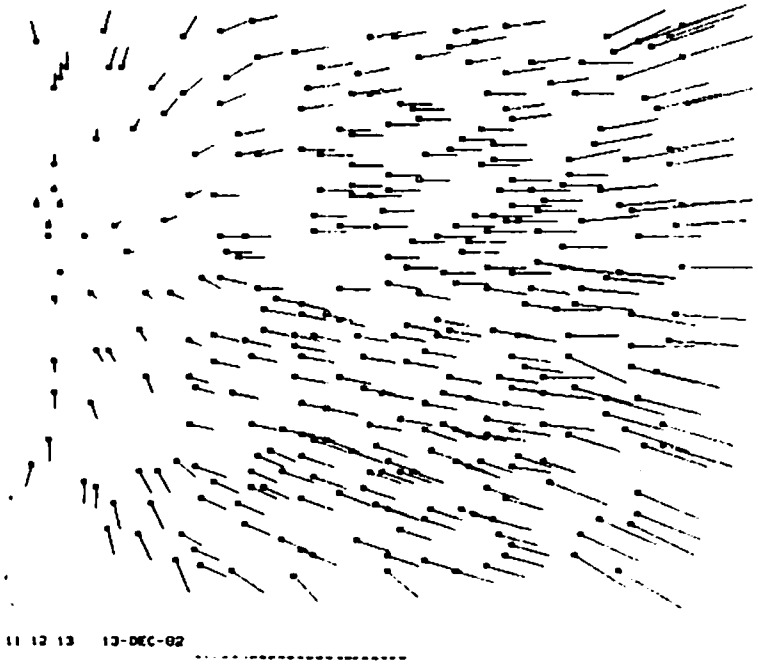

Figure 6c

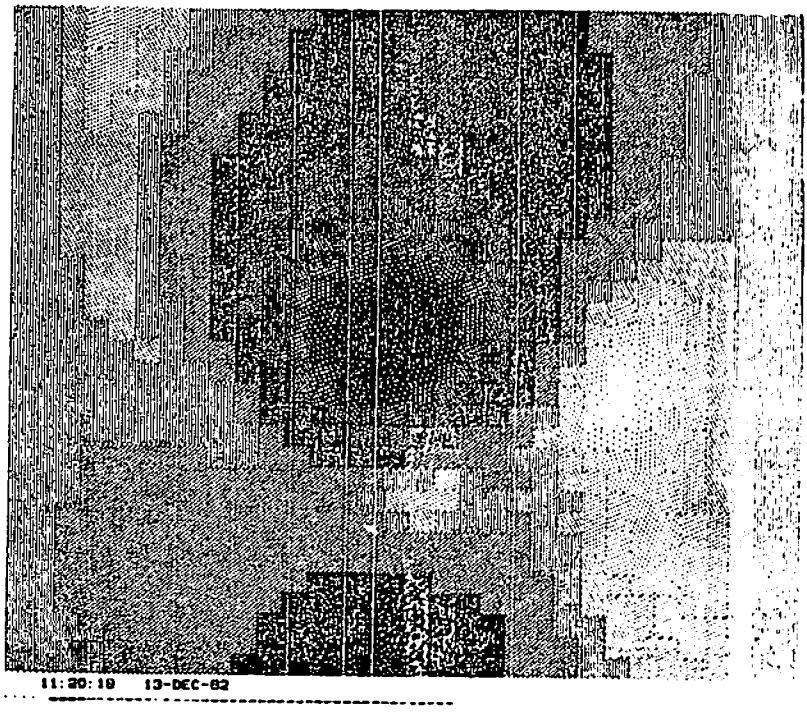

Figure 6d

translational axis was well within 5 degrees of the actual one.

Increasing the separation threshold smoothed the error function and made it more stongly unimodal. It also moved the determined intersection of translational axis and the image plane away from the correct position towards the center of the image. Increasing the length threshold significantly decreased the number of difference vectors.

6. Discussion

We have developed a procedure that determines the translational axis of sensor motion from sparse optic displacement fields given some variation in depth of the surroundings. The simplicity of the procedure makes it attractive for utilizing special hardware architectures for real-time implementation. The inference of camera motion parameters is considerably simplified once the translational axis has been determined. In particular, given a sequence of images we can apply the procedure to the displacement fields obtained from successive pairs of images. If the translational axis has been computed for each successive displacement of the sensor, the sensor rotations between the frames can be obtained using at least three displacement vectors in each displacement field that are located on two perpendicular lines intersecting at the current translational axis. Given the sensor rotation between each pair of successive images, the computation of the translational and rotational component fields is straightforward.

6.1. Determining direction of motion from optic flow: results from human vision

Studies of the accuracy with which humans can determine their direction of translation from optic flow were initiated by Gibson, who was the first to point out the possible importance of the 'focus of expansion' (see footnote) for navigation (Gibson et. al. 1955, Gibson 1958). In these and subsequent studies errors up to 10 degrees of visual angle were found (Llewellyn 1971, Johnston et. al. 1973, Regan and Beverley 1982). This led many researchers to conclude that optic flows do not, at least in humans, play an important role in navigation. On closer inspection all three experiments have in common the following: they all simulated an approach toward a plane perpendicular to the axis of translation, and the performance of the observers detoriated as the separation of the directions of gaze and translation increased (i.e. the rotational component of the optic flow increased -- in the first two experiments this was not an independent variable but a side effect of increasing deviation of translatory axis from the screen center). These errors in judging the direction of translation are as one would expect from the procedure described in this paper. Additional evidence for the importance of depth seperation comes from Warren (1976) who simulated the case of an observer moving along an axis of translation parallel to the ground, i.e. $\Delta Z/Z_1$ .ne.0, and found

footnote: The 'focus of expansion' is the center of the radial fieldlines of the translational component field of optic flows. However, it does in general not coincide with a focus of expansion of a flow field in the mathematical sense, namely a maximum of the divergence of the field (Koenderink and van Doorn 1981).

somewhat smaller errors (in the average 5 degrees of visual angle). Only recently Cutting did an experiment with systematic variation of the separation of details in depth along visual directions and of the angle between directions of gaze and translation (Cutting submitted). Indeed, the performance of human observers changed dramatically as $\Delta Z/Z_1$ changed: for $\Delta Z/Z_1 = 0$ angles between axes of translation and gaze of nearly 20 degrees could not be distinguished -- in contrast: for $\Delta Z/Z_1 \approx 1.5$ the differentiable visual angle was 37.5 minutes of arc! However, in this experiment the observers had to fixate some detail located on the second of three transparent planes. Thus deviations of the axes of gaze and translation are also indicated by a sign reversal of image velocities between details in front and behind the fixated one. It is therefore not possible to tell if the observers relied in their judgements on local discontinuities or sign reversals of the optical flow. An experiment that isolates the two kinds of information could be designed easily. If human observers had to rely on the information given by sign reversals of image velocities, they could not determine the translational axis immediately --- instead they had to change their direction of gaze until it coincided with the translational axis, in which case the sign reversal vanished.

ACKNOWLEDGEMENTS

We would like to thank Steve Epstein for introducing us and Kate Greenspan for feeding us. The image sequences we have been working with could not have been obtained without the UMASS Robotics group, Bill Nugent, and especially Gerry Pocock. Jeff Walker rigged up an interesting camera mount. This research was supported by DARPA grant N00014-82-K-0464 and NIH grant 5 R01 NS14971-04.

BIBLIOGRAPHY

Ballard, D. H., "Parameter Networks: Towards a Theory of Low-Level Vision", Proc. of 7th IJCAI, Vancouver, British Columbia, pp. 1068-1078, 1981.

Cutting, J. E., "Motion Parallax and Visual Flow: How to Determine Direction of Locomotion", submitted, 1982.

Gibson, J. J., Olum, P., and Rosenblatt, F., "Parallax and Perspective During Aircraft Landings", Am. Journ. Psychol., vol. 68, pp. 372-375, 1955.

Gibson, J. J., "Visually Controlled Locomotion and Visual Orientation in Animals", Br. Journ. of Psychol., vol. 49, pp. 182-194, 1958.

Johnston, I. R., White, G. R., Cumming, R. W., "The Role of Optical Expansion Patterns in Locomotor Control", Am. Journ. Psychol., vol. 86, no. 2, pp. 311-324, 1973.

Koenderink, J. J., and van Doorn, A. J., "How an Ambulant Observer can Construct a Model of the Environment from the Geometrical Structure of the Visual Inflow", Kybernetik 1977, G. Hauske and E. Butenandt, editors, R. Oldenbourg Verlag, Munich, 1978.

Koenderink, J. J., and van Doorn, A. J., "Exterospecific Component of the Motion Parallax Field", J. Opt. Soc. Am., vol. 71, pp. 953-957, 1981.

Lawton, D. T., "Motion Analysis via Local Translational Processing", IEEE Workshop on Computer Vision: Representation and Control, pp. 59-72, 1982.

Lawton, D. T., "Processing Translational Motion Sequences", Computer Graphics and Image Processing, in press, 1983.

Lee, D. N., "The Optic Flow Field: the Foundation of Vision", Phil. Trans. R. Soc. Lond. B., vol 290, pp. 169-179, 1980.

Llewellyn, K. R., "Visual Guidance of Locomotion", Journ. Exper. Psychol., vol. 91, no. 2, pp. 245-261, 1971.

Longuet-Higgins, H. C. and Prazdny, K., "The Interpretation of a Moving Image", Proc. R. Soc. Lond. B., vol 208, pp. 385-397, 1980.

O'Rourke, J., "Motion Detection Using Hough Techniques", Proceedings of PRIP. pp. 82-87, 1981.

Prazdny, K., "Egomotion and Relative Depth Map from Optical Flows", Biol. Cybernet., vol. 36, pp. 87-102, 1980.

Regan, D. and Beverly, K. I., "How do We Avoid Confounding the

Direction We are Looking and the Direction We are Moving?", Science, Vol. 215, pp. 194-196, 1982.

Rieger, J. H., "Information in Optical Flows Induced by Curved Paths of Observation, J. Opt. Soc. Am., vol. 73, in press, 1983.

Warren, R., "The Perception of Egomotion", Journ. Exper. Psychol: Human Percep. and Perform., vol. 2, no. 3, pp. 448-456, 1976.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: COINS TR 83-1
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): DETERMINING THE INSTANTANEOUS AXIS OF TRANSLATION FROM OPTIC FLOW GENERATED BY ARBITRARY SENSOR MOTION

5. TYPE OF REPORT & PERIOD COVERED: INTERIM

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): J.H. Rieger and D.T. Lawton

8. CONTRACT OR GRANT NUMBER(s): DARPA N00014-82-K-0464

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Office of Naval Research
Arlington, Virginia 22217

12. REPORT DATE: 1/83

13. NUMBER OF PAGES: 25

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

Processing unrestricted camera motion
Psychophysical implications

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

This paper develops a simple and robust procedure for determining the instantaneous axis of translation from image sequences induced by unconstrained sensor motion. The procedure is based upon the fact that difference vectors at discontinuities in optic flow fields generated by sensor motion relative to a stationary environment are oriented along translational field lines. This is developed into a procedure consisting of

three steps: 1) locally computing difference vectors from an optic flow field; 2) thresholding the difference vectors; and 3) minimizing the angles between the difference vector field and a set of radial field lines which correspond to a particular translational axis. This method does not require a priori knowledge about sensor motion or distances in the environment. The necessary environmental constraints are rigidity and sufficient variation in depth along visual directions to endow the flow field with discontinuities. The method has been successfully applied to noisy, sparse, and low resolution flow fields generated from real world image sequences. Experiments are reviewed which indicate that the human visual system also utilizes discontinuities in optic flows in determining self-motion. In addition, due to the computational simplicity of the procedure, hardware realization for real-time implementation is possible.