SENSOR MOTION AND RELATIVE DEPTH FROM DIFFERENCE FIELDS OF OPTIC FLOWS

J. H. Rieger and D. T. Lawton

Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

Abstract

This paper develops a simple and robust procedure for recovering sensor motion parameters from image sequences induced by unconstrained sensor motion relative to a stationary environment. Difference vectors of optic flows approximate the orientations of the translational field lines in image areas in which there is depth variance between the corresponding environmental points and sufficient angular separation from the translational axis. This is developed into a procedure consisting of four steps: 1) locally computing difference vectors from an optic flow field; 2) thresholding the difference vectors; 3) minimizing the angles between the difference vector field and a set of radial field lines which correspond to a particular translational axis; and 4) extracting the translational and rotational component fields given the translational axis. This procedure does not require *a priori* knowledge about sensor motion or structure of the scene. It depends critically on sufficient variation in depth along some visual directions to endow the flow field with discontinuities. We present results of applying the procedure to sparse and low resolution displacement fields.

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 translation of the sensor induces a radial flow in the image with the intersection of the translational axis and image plane as its center. Sensor rotatation induces a rotational field in the image that is purely direction dependent (that is, a function of image position only).

The translational component (and its spatial and temporal derivative fields) contains, e.g., information about the shape of objects (Koenderink and van Doom 1977), about the relative depth properties of the environment (Lee 1980, Prazdny 1980), or about motion parameters for navigating along curved trajectories (Rieger 1983). Processing optic flows induced by observer/sensor motion can be done by decomposing a flow field into its rotational and translational components and then recovering the environmental information from the translational component. Techniques for doing this generally require high resolution image displacements as input and are sensitive to the noise and error that current techniques for determining image motions typically produce. They can also involve solving complex equations and require significant computation.

The recovery of sensor motion parameters can be simplified considerably by making use of the geometrical structure of optic flows in regions corresponding to environmental depth changes. In such regions the difference vectors that have been computed over some neighborhood are oriented approximately along translational field lines. This can be seen easily for the case of details that are located exactly in the same direction from an observer/sensor but are at different depths (such as points along occluding boundaries) as observed by Longuet-Higgins and Prazdny (1980), such points 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 and thus equal for flow vectors positioned at the same image point. The axis of sensor translation can then be obtained from the intersection of radial fieldlines which are determined by such difference vectors. Given the axis of translation, the rotational and translational component fields are strongly overdetermined.

There are significant difficulties in applying this observation to actual image sequences. 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. From images formed at discrete, successive instants we obtain image displacements and not instantaneous optic velocities. Thus the computation must be expressed in terms of discrete sensor motions. 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.

Difference Vectors from Spatially Separated Flow Vectors

Here we present results on the effects of using spatially separated image velocity vectors to determine difference vectors. A difference vector formed from spatially separated image velocity vectors can be decomposed into a signal component oriented along the correct translational field line and a noise component. We find that 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. The computation of difference vectors over the image does not require initially determining the location of occlusion boundaries or of image areas corrsponding to large visual slant.

Consider a sensor O moving relative to a static environment. As in figure 1 the point $\tilde{P}$ at the image position $\tilde{r} = (\tilde{x},\tilde{y}) = (x/z, y/z)$ corresponds to the environmental point P at the location $r = (x,y,z)$. We obtain the image velocity u at $\tilde{P}$ by differentiating wrt time

$$u = [(\dot{x} - \tilde{x}\dot{z})\, e_x + (\dot{y} - \tilde{y}\dot{z})\, e_y]\,/z\ .$$

Letting $v = (v_x, v_y, v_z)$ and $\omega = (\omega_x, \omega_y, \omega_z)$ denote the translational and rotational velocities of O the relative motion of P becomes

$$\dot{r} = -v - \omega \times r\ .$$

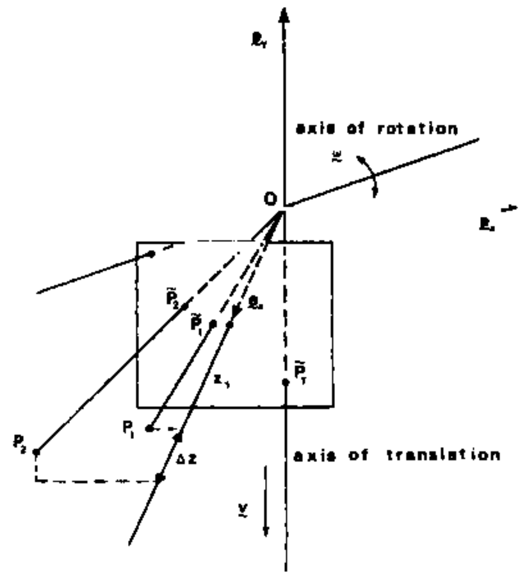

Figure 1

Eliminating $\dot{x}$, $\dot{y}$, and $\dot{z}$ between the above equations we can write the translational and rotational components of image velocity u separately

$$u_T = [(\tilde{x}v_z - v_x)\, e_x + (\tilde{y}v_z - v_y)\, e_y]\,/z\ ,$$

$$u_R = (-\omega_y + \tilde{y}\omega_z - \tilde{x}^2\omega_y + \tilde{x}\tilde{y}\omega_x)\, e_x + (-\tilde{x}\omega_z + \omega_x - \tilde{x}\tilde{y}\omega_y + \tilde{y}^2\omega_x)\, e_y\ .$$

Two image points $\tilde{P}_1$ and $\tilde{P}_2$ that are separated by $\tilde{r}_2 - \tilde{r}_1 = (d_x, d_y)$ differ in their rotational flow vectors by

$$\begin{aligned}\Delta u_R &= u_{R2} - u_{R1} \\ &= [d_y\omega_z - d_x(2\tilde{x}_1 + d_x)\,\omega_y + (\tilde{y}_1 d_x + \tilde{x}_1 d_y + d_x d_y)\,\omega_x]\, e_x \\ &\quad + [-d_x\omega_z + d_y(2\tilde{y}_1 + d_y)\,\omega_x - (\tilde{y}_1 d_x + \tilde{x}_1 d_y + d_x d_y)\,\omega_y]\, e_y\ .\end{aligned}$$

If $\tilde{r}_T = (v_x/v_z, v_y/v_z)$ denotes the intersection of the translational axis with the image plane we can rewrite the translational flow vector as $u_T = v_z(\tilde{r} - \tilde{r}_T)/z$. Then the difference vector of two translational flow vectors at separated image positions $\tilde{P}_1$ and $\tilde{P}_2$ becomes

$$\begin{aligned}\Delta u_T &= u_{T2} - u_{T1} \\ &= \frac{v_z}{z_1 + \Delta z}\left[\tilde{r}_2 - \tilde{r}_1 + \frac{\Delta z}{z_1}(\tilde{r}_1 - \tilde{r}_T)\right],\end{aligned}$$

where $\Delta z = z_2 - z_1$ is the depth separation of the environmental details P_1 and P_2 that correspond to $\tilde{P}_1$ and $\tilde{P}_2$ in the image.

Now we can rewrite Δu as consisting of a component along a translational fieldline and a noise component

$$\Delta u = \Delta u_T + \Delta u_R =$$

$$\left[\frac{v_z \Delta z}{z_1 z_2}(\bar{r}_1 - \bar{r}_T) \right]_{Signal} + \left[\frac{v_z}{z_2}(\bar{r}_2 - \bar{r}_1) + \Delta u_R \right]_{Noise}.$$

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

Recovery of Motion Parameters and Depth

In order to compute difference vectors from image displacement fields formed over discrete time intervals (as opposed to continuous intantaneous image velocity fields), 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. Thus, given image displacements DI and D2 at positions PI and P2, the difference vector between points 1 and 2 is obtained by subtracting D2 from DI and positioning the resulting vector at PI + DI.

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. For a given difference vector and a set of radial field lines, the *error angle* is the angle between the difference vector and the fieldline at that position.

We have found that reducing the number of difference vectors by increasing the length threshold and decreasing the separation threshold improves the fit of the difference vector field to the set of correct field lines up to a certain degree. This is because short difference vectors (compared to the local average magnitude) are more likely to deviate from the correct field lines and computing difference vectors over larger neighborhoods increases the noise components. If, however, thresholding eliminates too many difference vectors the fit detonates, since the signal of the difference vector field becomes less distinguished for a decreasing number of vectors.

For each image displacement vector a set of difference vectors of sufficient length is determined over its neighborhood. For the resulting field of difference vectors, processing involves finding a translational axis and the corresponding set of radial field lines which minimizes the sum of the magnitude of the error angles. The procedure used is basically that used in Lawton (1982, 1983) to determine the translational axis from noisy displacement fields induced by rectilinear sensor motion. The error measure is defined on a half sphere, where points on the half sphere are possible candidates for the translational axis. The advantage of using a sphere as a domain is that it allows for a uniform, global sampling of the error function. The search process itself consists of a global sampling of the error measure to determine its rough shape using a generalized Hough transform (Ballard 1980, O' Rourke 1981) followed by a local search to find a minimum.

The computation of the sensor rotation (scaled by focal length) from the original flow field and the radial (translational) fieldlines is straightforward. Note that the components of the flow perpendicular to the radial fieldlines are induced by sensor rotation. Introducing, for convenience, a polar coordinate system (r,8) in the image plane centered at PT we have a system of overconstrained linear equations of the type $u_R \cdot e_\theta = u \cdot e_\theta$ in the three unknowns ω_x, ω_y, and ω_z.

Knowing the rotational parameters yields the translational and rotational component fields of the orginal flow field. The translational component is directly related to the relative depth of a scene (i.e. the depth scaled by the sensor displacement in depth 8z) by the relation $z/\delta z = | \bar{r} - \bar{r}_T | / | u_T |$, where uT is a translational flow vector in the image. If the frame rate is known the relative depth of an environmental point corresponds to its temporal separation from the sensor (under constant approach velocity). Biological systems seem to exploit this optical relation for a variety of navigational tasks (Lee 1980, Wagner 1982).

Experiments

The flow field in figure 2a shows image displacements at pixel positions having coordinates which are multiples of 8 from a 128 x 128 pixel field. The components of the displacement vectors were stored as 8 bit integers. The environment consisted of a spherical surface patch at depth of 10 units along the z axis and a background spherical surface patch at a depth of 30 units along the z axis. The obvious discontinuities in the flow field in figure 2a 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 and the length threshold was set to 3 pixels. Figure 2b shows the average difference vectors which exceeded the length threshold. Note their occurrance along the occlusion boundary and their strong radial character. The resulting error function is shown in figure 2c (Darker in the figure corresponds to less error; also recall that this is a plot of a hemisphere in polar 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, 60). The local search determined the minimum to be at (63, 63). The correct, subpixel, position was (63.5, 635). The determined rotational and translational components are shown in figures 2d and 2e respectively. The relative depth map determined from the translational component field is shown in figure 2f encoded by intensity (darker means closer to the observer).

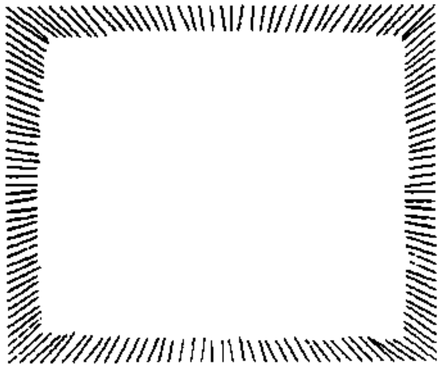

Figure 2b

Figure 2c

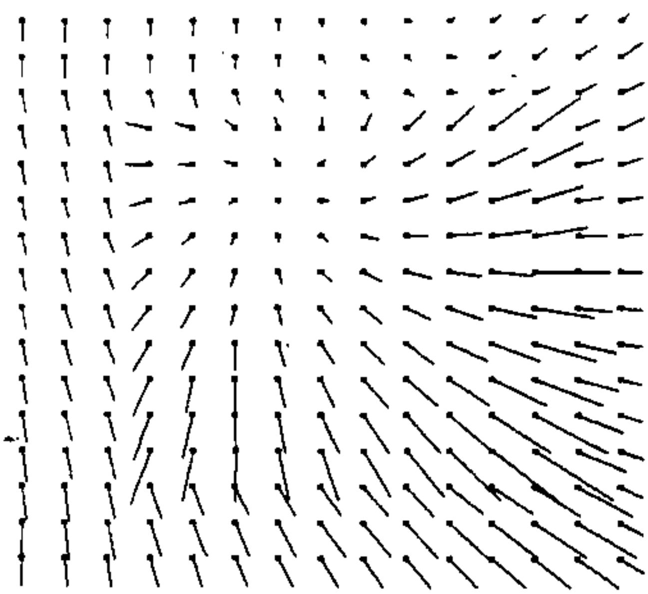

Figure 2a

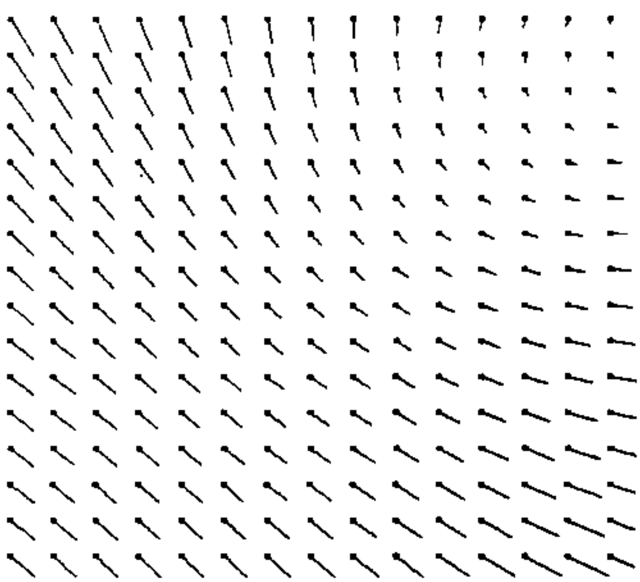

Figure 2d

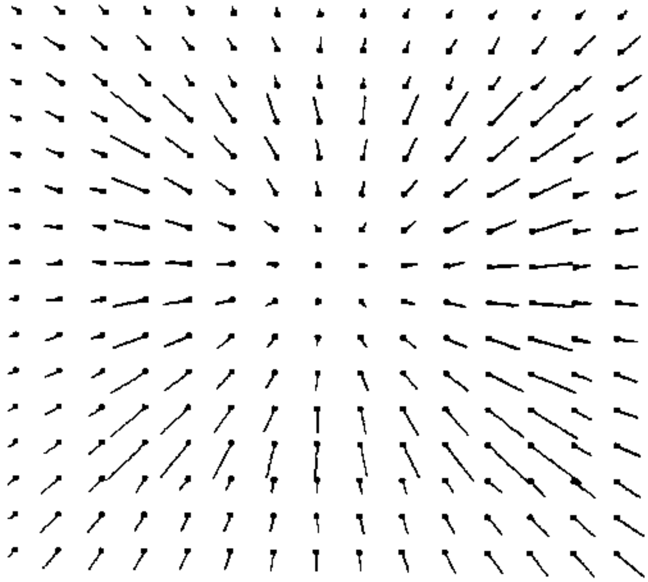

Figure 2e

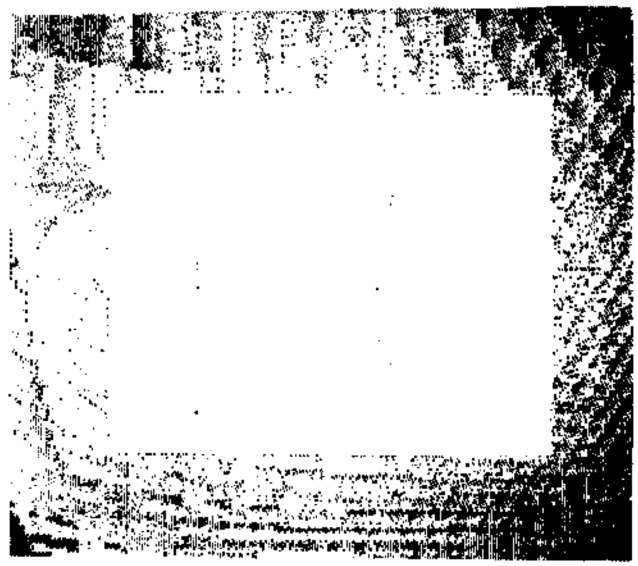

Figure 2f

Several experiments with simulated displacement fields have confirmed the expected effects of environmental depth variance and displacement vector density and demonstrated robust performance with respect to various kinds of noise. The procedure has also successfully determined the translational axis from displacement fields obtained from low resolution image sequences from a solid state camera. For such image sequences with large environmental depth variances the translational axis has been determined within a few degrees of visual angle (Rieger and Lawton 1983).

We thank Frank Glazer for pointing out a bug in this paper. This research was supported by DARPA grant N0000014-82-K-04064 to the MOTION group at UMASS.

References

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

Gibson, J. J., Olum, P., and Rosenblatt, F., "Parallax and Perspective During Aircraft Landings", *Am. Journ. Psychol., vol. 68,* pp. 372-385, 1955.

Koenderink, J. J., and van Doom, 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.

Lawton, D. T., "Motion Analysis via Local Transnational 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. Land. B., vol 290,* pp. 169-179, 1980.

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.

Rieger, J. H., Information in Optical Flows Induced by Curved Paths of Observation", *J. Opt. Soc. Am., vol. 73,* pp. 339-344, 1983.

Rieger, J. H., and Lawton, D. T., "Determining the Axis of Translation from Optic Flow Generated by Arbitrary Sensor Motion", *COINS Technical Report,* 83-01,1983.

Wagner, H., "Flow-field Variables Trigger Landing in Flies", *Nature, vol. 297, pp.* 147-148, 1982.