International Journal of Computer Vision, 2, 51–76 (1988)

Direct Methods for Recovering Motion*

BERTHOLD K.P. HORN AND E.J. WELDON JR.
Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, Hawaii 96822‡

Abstract

We have developed direct methods for recovering the motion of an observer in a static environment in the case of pure rotation, pure translation, and arbitrary motion when the rotation is known. Some of these methods are based on the minimization of the difference between the observed time derivative of brightness and that predicted from the spatial brightness gradient, given the estimated motion. We minimize the square of the integral of this difference taken over the image region of interest. Other methods presented here exploit the fact that surfaces have to be in front of the observer in order to be seen.

We do not establish point correspondences, nor do we estimate the optical flow. We use only first-order derivatives of the image brightness, and we do not assume an analytic form for the surface. We show that the field of view should be large to accurately recover the components of motion in the direction toward the image region. We also demonstrate the importance of points where the time derivative of brightness is small and discuss difficulties resulting from very large depth ranges. We emphasize the need for adequate filtering of the image data before sampling to avoid aliasing, in both the spatial and temporal dimensions.

1 Introduction

In this paper we consider the problem of determining the motion of a monocular observer moving with respect to a rigid, unknown world. We use a least-squares, as opposed to a discrete, method of solving for the motion parameters; our method uses all of the points in a two-image sequence and does not attempt to establish correspondence between the images. Hence the method is relatively robust to quantization error, noise, illumination gradients, and other effects.

So far, we can determine the observer motion in two special cases:

- when the motion is pure rotation,
- when the motion is pure translation or when the rotational component of the motion is known.

At this writing we have not developed a direct method that is applicable to arbitrary motion.

1.1 Earlier Work

In the continuous or least-squares approach to motion vision, motion parameters are found that are consistent with the observed motion of the entire image. Bruss and Horn [1] use this approach to calculate motion parameters assuming that the optical flow is known at each point. Adiv [2] uses the approach of Bruss and Horn to segment the scene into independently moving planar objects; he shows that given the optical flow, segmentation can be performed and the motion calculated. Nagahdaripour and Horn [3] eschew the use of optical flow and calculate the observer's motion

*This research was supported by the National Science Foundation under Grant No. DMC85-11966. Additional support was provided by NASA (Grant No. GSFC 5-1162) and by the Veteran's Administration.

‡BKPH on leave from the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

directly from the spatial and temporal derivatives of the image brightness, assuming a planar world. The advantage of this direct approach, which we also use here, is that certain computational difficulties inherent in the calculation of optical flow are avoided. In particular, it is not necessary to make the usual assumption that the optical flow field is smooth; an assumption that is violated near object boundaries, necessitating flow segmentation.

Waxman and Ullman [4] and Waxman and Wohn [5] also avoid the discrete approach to motion vision; their techniques make use of first and second derivatives of the optical flow to compute both the motion parameters and the structure of the imaged world. In the interests of developing methods that can be implemented, the techniques presented in this paper avoid the use of second- and higher-order derivatives.

1.2 Summary of the Paper

One of our approaches to the motion vision problem can be summarized as follows: Given the observer motion and the spatial brightness function of the image one can predict the time derivative of brightness at each point in the image. We find the motion that minimizes the integral of the square of the difference between this predicted value and the observed time derivative. The integral is taken over the image region of interest, which, in the discussion here, is usually taken to be the whole image.

We use auxiliary vectors derived from the derivatives of brightness and the image position that occur in the basic brightness change constraint equation. Study of the distribution of the directions of these vectors on the unit sphere suggests specific algorithms and also helps uncover relationships between accuracy and parameters of the imaging situation.

We have developed a simple robust algorithm for recovering the angular velocity vector in the case of pure rotation. This algorithm involves solving three linear equations in the three unknown components of the rotation vector. The coefficients of the equations are moments of components of one of the axuiliary vectors over the given image region. We show that the accuracy of the recovered component of rotation about the direction toward the image region is poor relative to the other components, unless the image region subtends a substantial solid angle.

We have developed several algorithms for recovering the translational velocity in the case of pure translation. These algorithms exploit the constraint that objects have to be in front of the camera in order to be imaged. This constraint leads to a nonlinear constrained optimization problem. The performance of these algorithms depends on a number of factors including:

- the angle subtended by the image, i.e., the field of view,
- the direction of motion relative to the optical axis,
- the depth range,
- the distribution of brightness gradients,
- the noise in the estimated time derivative of brightness,
- the noise in the estimated spatial gradient of brightness, and
- the number of picture cells considered.

We have not yet been able to select a "best" algorithm from the set developed, since one may be more accurate under one set of circumstances while another is better in a different situation. Also, the better algorithms tend to require more computation, and some do not lend themselves to parallel implementation. Further study using real image data will be needed to determine the range of applicability of each algorithm.

We found a strong dependence of the accuracy of recovery of certain components of the motion on the size of the field of view. This is in concert with other reports describing difficulties with small fields of view, such as references [2] and [5].

1.3 Comments on Sampling, Filtering, and Aliasing

Work with real image data has demonstrated the need to take care in filtering and sampling. The estimates of spatial gradient and time derivatives are sensitive to aliasing effects resulting from inadequate low-pass filtering before sampling. This

is easily overlooked, particularly in the time direction. It is usually a mistake, for example, to simply pick every nth frame out of an image sequence. At the very least, n consecutive frames should be averaged before sampling in order to reduce the high-frequency components. One may object to the "smearing" introduced by this technique, but a series of widely separated snapshots typically do not obey the conditions of the sampling theorem, and as a result the estimates of the derivatives may contain large errors.

This, of course, is nothing new, since the same considerations apply when one tries to estimate the optical flow using first derivatives of image brightness (Horn and Schunck [6]). It is important to remember that the filtering must be applied before sampling—once the data has been sampled, the damage has been done.

2 The Brightness-Change Constraint Equation

Following Longuet-Higgins and Prazdny [7] and Bruss and Horn [1] we use a viewer-based coordinate system. Figure 1 depicts the system under consideration. A world point

$$\mathbf{R} = (X,Y,Z)^T \tag{1}$$

is imaged at

$$\mathbf{r} = (x,y,1)^T \tag{2}$$

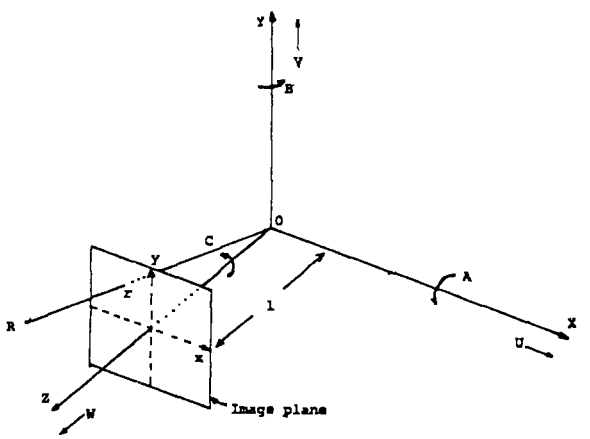


Fig. 1. The viewer-centered coordinate system. The translational velocity of the camera is $\mathbf{t} = (U,V,W)^T$, while the rotational component is $\omega = (A,B,C)^T$.

That is, the image plane has equation $Z = 1$. The origin is at the projection center and the Z-axis runs along the optical axis. The X- and Y-axes are parllel to the x- and y-axes of the image plane. Image coordinates are measured relative to the principal point, the point $(0,0,1)^T$ where the optical axis pierces the image plane. The points $\mathbf{r}$ and $\mathbf{R}$ are related by the perspective projection equation

$$\mathbf{r} = (x,y,1)^T = \left(\frac{X}{Z}, \frac{Y}{Z}, \frac{Z}{Z}\right)^T = \frac{\mathbf{R}}{\mathbf{R}\cdot\hat{\mathbf{z}}} \tag{3}$$

with

$$Z = \mathbf{R}\cdot\hat{\mathbf{z}} \tag{4}$$

and where $\hat{\mathbf{z}}$ denotes the unit vector in the Z direction.

Suppose the observer moves with instantaneous translational velocity $\mathbf{t} = (U,V,W)^T$ and instantaneous rotational velocity $\omega = (A,B,C)^T$ relative to a fixed environment, then the time derivative of the vector $\mathbf{R}$ can be written as

$$\mathbf{R}_t = -\mathbf{t} - \omega \times \mathbf{R} \tag{5}$$

The motion of the world point $\mathbf{R}$ results in motion of the corresponding image point; the value of this *motion field* is given by

$$\mathbf{r}_t = \frac{d\mathbf{r}}{dt} = \frac{d}{dt}\left(\frac{\mathbf{R}}{\mathbf{R}\cdot\hat{\mathbf{z}}}\right) = \frac{\mathbf{R}_t(\mathbf{R}\cdot\hat{\mathbf{z}}) - (\mathbf{R}_t\cdot\hat{\mathbf{z}})\mathbf{R}}{(\mathbf{R}\cdot\hat{\mathbf{z}})^2} \tag{6}$$

This can also be expressed as

$$\mathbf{r}_t = \frac{\hat{\mathbf{z}}\times(\mathbf{R}_t\times\mathbf{r})}{\mathbf{R}\cdot\hat{\mathbf{z}}} \tag{7}$$

since $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{c}\cdot\mathbf{a})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}$. Substituting equation (5) into othis result gives (see Negahdaripour and Horn [3]):

$$\mathbf{r}_t = -\hat{\mathbf{z}}\times\left[\mathbf{r}\times\left(\mathbf{r}\times\omega - \frac{\mathbf{t}}{\mathbf{R}\cdot\hat{\mathbf{z}}}\right)\right] \tag{8}$$

In component form this can be expressed as

$$\mathbf{r}_t = \begin{pmatrix} x_t \\ y_t \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{-U + xW}{Z} + Axy - B(x^2 + 1) + Cy \\ \dfrac{-V + yW}{Z} - Bxy + A(y^2 + 1) - Cx \\ 0 \end{pmatrix} \tag{9}$$

a result first obtained by Longuet-Higgins and Prazdny [7].

This shows how, given the world motion, the motion field can be calculated for every image point. If we assume that the brightness of a small surface patch is not changed by motion, then expansion of the total derivative of brightness E leads to

$$\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \tag{10}$$

(The applicability of the constant brightness assumption is discussed in Appendix A.) Denoting the vector $(\partial E/\partial x, \partial E/\partial y, 0)^T$ by $E_{\mathbf{r}}$ and $\partial E/\partial t$ by E_t, permits us to express this result more compactly in the form

$$E_{\mathbf{r}} \cdot \mathbf{r}_t + E_t = 0 \tag{11}$$

Substituting equation (8) into this result and rearranging gives

$$E_t - \{[(E_{\mathbf{r}} \times \hat{\mathbf{z}}) \times \mathbf{r}] \times \mathbf{r}\} \cdot \boldsymbol{\omega} + \frac{[(E_{\mathbf{r}} \times \hat{\mathbf{z}}) \times \mathbf{r}] \cdot \mathbf{t}}{\mathbf{R} \cdot \hat{\mathbf{z}}} = 0 \tag{12}$$

To simplify this expression we let

$$\mathbf{s} = (E_{\mathbf{r}} \times \hat{\mathbf{z}}) \times \mathbf{r} \tag{13}$$

and

$$\mathbf{v} = -\mathbf{s} \times \mathbf{r} \tag{14}$$

so equation (12) reduces to the *brightness change constraint equation* of Negahdaripour and Horn [3], namely

$$\mathbf{v} \cdot \boldsymbol{\omega} + \frac{\mathbf{s} \cdot \mathbf{t}}{\mathbf{R} \cdot \hat{\mathbf{z}}} = -E_t \tag{15}$$

The vectors $\mathbf{s}$ and $\mathbf{v}$ can be expressed in component form as

$$\mathbf{s} = \begin{pmatrix} -E_x \\ -E_y \\ xE_x + yE_y \end{pmatrix} \quad \text{and}$$

$$\mathbf{v} = \begin{pmatrix} +E_y + y(xE_x + yE_y) \\ -E_x - x(xE_x + yE_y) \\ yE_x - xE_y \end{pmatrix} \tag{16}$$

Note that $\mathbf{s} \cdot \mathbf{r} = 0$, $\mathbf{v} \cdot \mathbf{r} = 0$ and $\mathbf{s} \cdot \mathbf{v} = 0$. These three vectors thus form an orthogonal triad. The vectors $\mathbf{s}$ and $\mathbf{v}$ are inherent properties of the image. Note that the projection of $\mathbf{s}$ into the image plane is just the (negative) gradient of the image. Also, the quantity $\mathbf{s}$ indicates the directions in which translation of a given magnitude will contribute maximally to the temporal brightness change of a given picture cell. The quantity $\mathbf{v}$ plays a similar role for rotation.

3 Solving the Brightness Change Constraint Equation

Equation (15) relates observer motion $(\mathbf{t},\boldsymbol{\omega})$, the depth of the world $\mathbf{R} \cdot \hat{\mathbf{z}} = Z(x,y)$ and certain measurable quantities of the image $(\mathbf{s},\mathbf{v})$. In general, it is not possible to solve for the first two of these given the last. Some interesting special cases are addressed in this paper and in Negahdaripour and Horn [3]; these are:[1]

i. Known depth: In section 3.1 we show that given Z, $\mathbf{s}$, and $\mathbf{v}$, the quantities, $\mathbf{t}$ and $\boldsymbol{\omega}$ can be calculated in closed form using a least-squares method.
ii. Pure rotation ($\|\mathbf{t}\| = 0$): In section 3.2 we show that given $\mathbf{v}$, the rotation vector $\boldsymbol{\omega}$ can be calculated in closed form.
iii. Pure translation or known rotation: In section 3.3 we present a least-squares method for determining $\mathbf{t}$. Once $\mathbf{t}$ is known, the brightness change constraint equation can be used to

[1] We do not discuss here related methods using optical flow, such as those of Bruss and Horn [1].

find the depth at each picture cell:

$$Z = \mathbf{R} \cdot \hat{\mathbf{z}} = -\frac{\mathbf{s} \cdot \mathbf{t}}{E_t + \mathbf{v} \cdot \omega} \tag{17}$$

iv. Planar world: Negahdaripour and Horn [3] present a closed-form solution for $\mathbf{t}$, ω, and the normal $\mathbf{n}$ of the world plane.

v. Quadratic patches: Negahdaripour [8] gives a closed-form solution in the case that a portion of the world can be represented as a quadratic form.

In this paper we consider various integrals over an image region thought to correspond to a single rigid object in motion relative to the viewer. In the simplest case, the observer is moving relative to a static environment and the whole image can be used. The size of the field of view has a strong effect on the accuracy of the determination of the components of motion along the optical axis. When we need to estimate this accuracy, we will, for convenience, assume a circular image of radius r_v. This corresponds to a conical field of view with half angle θ_v, where $r_v = \tan\theta_v$, since we have assumed that the focal length equals one. (We assume that $0 < \theta_v < \pi/2$).

We will show that the field of view should be large. Although orthographic projection usually simplifies machine vision problems, this is one case in which we welcome the effects of perspective "distortion"!

3.1 Depth Known

When depth is known, it is straightforward to recover the motion. (Depth may have been obtained using a binocular stereo system or some kind of range finder.) We cannot, in general, find a motion to satisfy the brightness change constraint equation at every picture cell, because of noise in the measurements. Instead we minimize

$$\iint [E_t + \mathbf{v} \cdot \omega + (1/Z)\mathbf{s} \cdot \mathbf{t}]^2 \, dx\, dy \tag{18}$$

Differentiating with respect to ω and $\mathbf{t}$ and setting the results equal to zero leads to the pair of vector equations:

$$\begin{aligned}
&\left[\iint (1/Z)^2 \mathbf{s}\mathbf{s}^T \, dx\, dy\right]\mathbf{t} \\
&\quad + \left[\iint (1/Z)\mathbf{s}\mathbf{v}^T \, dx\, dy\right]\omega \\
&\quad = -\iint E_t (1/Z)\mathbf{s} \, dx\, dy \\
&\left[\iint (1/Z)\mathbf{v}\mathbf{s}^T \, dx\, dy\right]\mathbf{t} \\
&\quad + \left[\iint \mathbf{v}\mathbf{v}^T \, dx\, dy\right]\omega \\
&\quad = -\iint E_t \mathbf{v} \, dx\, dy
\end{aligned} \tag{19}$$

This is a set of six linear equations in six unknowns with a symmetric coefficient matrix. (The equations can be solved by partitioning in order to reduce the computational effort.) The coefficients are all integrals of products of components of $(1/Z)\mathbf{s}$ and $\mathbf{v}$. It may be useful to note that

$$\text{trace}(\mathbf{s}\mathbf{v}^T) = \text{trace}(\mathbf{v}\mathbf{s}^T) = \mathbf{s} \cdot \mathbf{v} = 0 \tag{20}$$

We could have obtained slightly different equations for ω and $\mathbf{t}$ if we had chosen to weight the integrand in equation (18) differently. We study the special case in which $\|\mathbf{t}\| = 0$ and the special case in which $\|\omega\| = 0$ later.

One application of the above result is to "dynamic stereo." A binocular stereo system can provide disparity estimates from which $1/Z$ can be calculated. The above equations can then be used to solve for the motion, provided estimates of the derivatives of image brightness are also supplied. The correspondence problem of binocular stereo has, unfortunately, been found to be a difficult one. It would represents the major computational burden in a dynamic stereo system. We hope that motion vision research will eventually lead to simpler methods for recovering depth than those used for binocular stereo—although they are likely to be relatively inaccurate when based only on instantaneous translational and rotational velocity estimates.

Fig. 2. Shown here are the (a) 10th, (b) 20th, (c) 30th, and (d) 40th image out of a 40-image sequence obtained when a CCD camera mounted on a tripod was (manually) rotated about its vertical axis. After the initial acceleration, the image motion in the center is between 7 and 8 picture cells between successive frames. Image motion between frames is slightly larger in the corners of the image.

3.2 Pure Rotation

When $\|\mathbf{t}\| = 0$, the brightness change constraint equation reduces to

$$\mathbf{E}_t + \mathbf{v} \cdot \boldsymbol{\omega} = 0 \tag{21}$$

We wish to find the value of $\boldsymbol{\omega}$ that minimizes the sum of the squares of the errors in the time derivative of brightness, that is, we want to minimize

$$\iint [E_t + \mathbf{v} \cdot \boldsymbol{\omega}]^2 \, dx\, dy \tag{22}$$

Differentiating with respect to $\boldsymbol{\omega}$ and setting the result equal to zero gives us

$$2 \iint [E_t + \mathbf{v} \cdot \boldsymbol{\omega}] \mathbf{v} \, dx\, dy = 0 \tag{23}$$

Since $(\mathbf{v} \cdot \boldsymbol{\omega})\mathbf{v} = \mathbf{v}(\mathbf{v} \cdot \boldsymbol{\omega}) = (\mathbf{v}\mathbf{v}^T)\boldsymbol{\omega}$, we can write this in the form

$$\left[\iint \mathbf{v}\mathbf{v}^T \, dx\, dy \right] \boldsymbol{\omega} = - \iint E_t \mathbf{v} \, dx\, dy \tag{24}$$

which is just a special case of equation (19). This is a set of three linear equations in the three un-

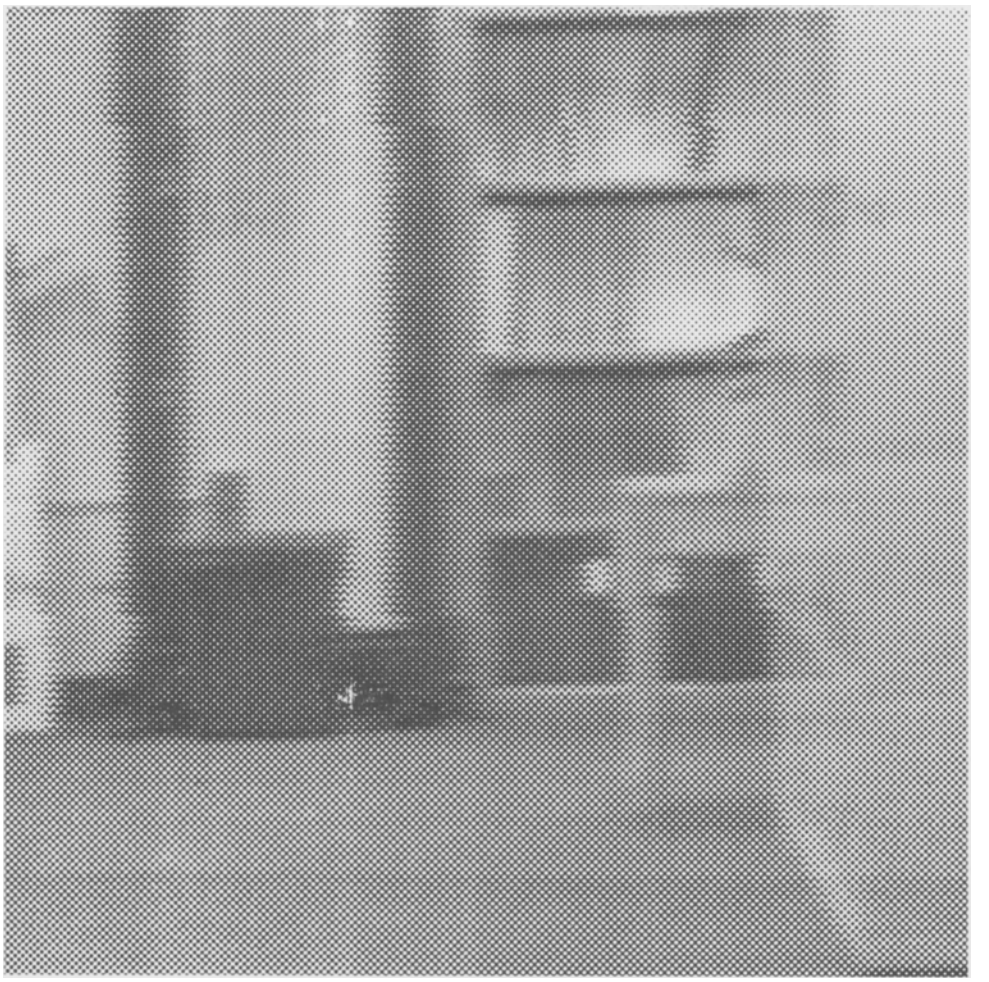
known components of ω, namely, A, B, and C. The coefficient matrix is symmetric and only the right-hand side depends on the time derivative of brightness.

One can, by the way, tell whether one is dealing with a case of pure rotation or not. In the presence of a translational component, equation (21) will not be a good approximation and so the integral in formula (22) will not be small. Experiments show that this simple method of determining rotation is robust and easy to implement. Slight variations are possible by weighting the intergrand differently. This method is reminiscent of the optical flow based method of Bruss and Horn [1] and very similar to a method developed by Aloimonos and Brown [9], to which our attention was drawn after we wrote this paper.

Shown in figure 2 is every tenth frame out of a 40-frame sequence taken with a tripod-mounted CCD camera rotated manually about its vertical axis. The vertical component of the computed rotational velocity is shown in figure 3 as a function of the frame number. The units along the vertical axis are picture cells per time step in the center of the image (rather than say radians/second). After the initial acceleration, image components near the center of the image move by about 7 to 8 picture cells between successive frames. Three

curves are given for varying amounts of image low-pass filtering and subsampling. The lowest curve (A) corresponds to the raw image data, and shows that for this particular scene at least, a motion of 7 to 8 picture cells is too much for accurate recovery of the angular velocity. The computed velocity appears to "saturate" at around 4 picture cells per time step. The next higher curve (B) corresponds to image compression by low-pass filtering and subsampling by a factor of two in each direction. In the compressed image sequence, the motion is in effect only about 3 to 4 picture cells per time step. The top curve (C) was obtained using images that were low-pass filtered and subsampled a second time to reduce them by a total of a factor of four in each direction. In this doubly compressed sequence, motion in the center of the image amounts to only about 1.5 to 2 picture cells per time step, and the angular velocity is accurately recovered. Further filtering and subsampling leads to velocity estimates that are essentially the same as the ones obtained with this sequence.

3.2.1 Distribution of the Directions of $\mathbf{v}$. To understand the properties of the algorithm for recovering the instantaneous rotational velocity, one needs to study the matrix obtained by integrating $\mathbf{v}\mathbf{v}^T$. We can think of the direction of $\mathbf{v}$ as identifying a point on the unit sphere and of $\|\mathbf{v}\|$ as the mass of a particle placed there. The collection of vectors corresponding to an image region then can be thought of as a set of particles on the unit sphere. The integral of $\mathbf{v}\mathbf{v}^T$ is the symmetric 3×3 matrix whose elements are integrals of the nine



pair-wise products of components of $\mathbf{v}$. This matrix is related to the inertia matrix of this set of particles. If the particles were spread uniformly over the surface of the sphere, this matrix would be the total mass times the identity matrix. As we show next, the particles are confined to a band, so this matrix, while diagonal on average, is not a multiple of the identity matrix.

We know that $\mathbf{v} \cdot \mathbf{r} = 0$ and that the possible directions of $\mathbf{r}$ lie within a cone defined by the field of view. For a particular value of $\mathbf{r}$, the equation $\mathbf{v} \cdot \mathbf{r} = 0$ defines a plane that cuts the unit sphere in a great circle (see figure 4). The vector $\mathbf{v}$ must point in a direction corresponding to a point on this great circle. Since $\mathbf{r}$ lies inside a cone of directions with half-angle θ_v, these great circles have axes that lie in this cone also. The collection of great circles lies in a band around the unit sphere of width equal to the total width of the visual field.

We can obtain the same result algebraically as follows: Let $\hat{\mathbf{r}}$, $\hat{\mathbf{v}}$, and $\hat{\mathbf{s}}$ be unit vectors in the directions $\mathbf{r}$, $\mathbf{v}$, and $\mathbf{s}$. Then, since $\mathbf{r}$, $\mathbf{s}$, and $\mathbf{v}$ are mutually orthogonal,

$$(\hat{\mathbf{z}} \cdot \hat{\mathbf{r}})^2 + (\hat{\mathbf{z}} \cdot \hat{\mathbf{v}})^2 + (\hat{\mathbf{z}} \cdot \hat{\mathbf{s}})^2 = 1 \tag{25}$$

while

$$(\hat{\mathbf{x}} \cdot \hat{\mathbf{v}})^2 + (\hat{\mathbf{y}} \cdot \hat{\mathbf{v}})^2 + (\hat{\mathbf{z}} \cdot \hat{\mathbf{v}})^2 = 1 \tag{26}$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are unit vectors in the X, Y, and Z directions. Subtracting the two equalities we obtain

$$(\hat{\mathbf{x}} \cdot \hat{\mathbf{v}})^2 + (\hat{\mathbf{y}} \cdot \hat{\mathbf{v}})^2 = (\hat{\mathbf{z}} \cdot \hat{\mathbf{r}})^2 + (\hat{\mathbf{z}} \cdot \hat{\mathbf{s}})^2 \tag{27}$$

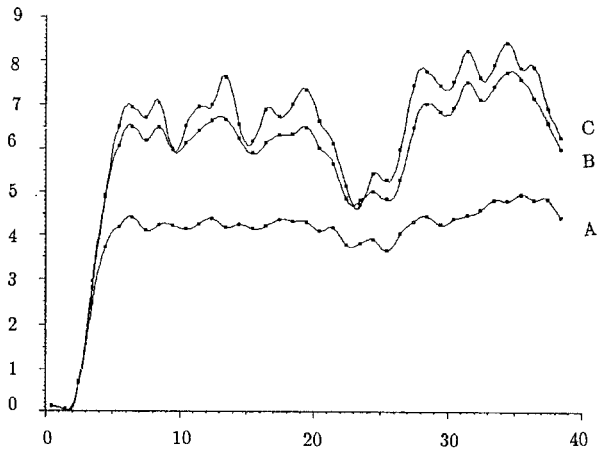


Fig. 3. Recovered vertical component of the angular velocity vector as a function of the frame number. The angular velocity is given in picture cells of image displacement at the center of the image per time step. Curve A was obtained using the raw image data, curve B from a low-pass filtered and subsampled image sequence, and curve C from an image sequence that was low-pass filtered and subsampled twice. Further low-pass filtering and subsampling produces essentially the same curve.

which, since

$$(\hat{\mathbf{z}} \cdot \hat{\mathbf{r}})^2 \geqslant \cos^2 \theta_v \quad \text{and} \quad (\hat{\mathbf{z}} \cdot \hat{\mathbf{s}})^2 \geqslant 0 \qquad (28)$$

tells us that

$$(\hat{\mathbf{x}} \cdot \hat{\mathbf{v}})^2 + (\hat{\mathbf{y}} \cdot \hat{\mathbf{v}})^2 \geqslant \cos^2 \theta_v \qquad (29)$$

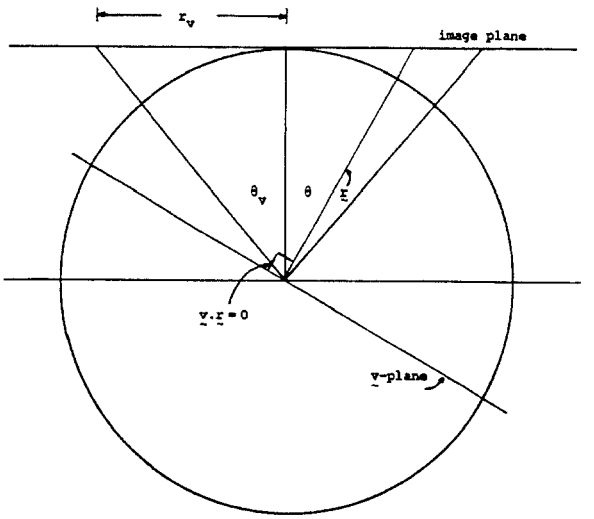


Fig. 4. A cross-section through the v-sphere defined by the image point r.

Thus the directions of v lie within an angle θ_v of the "equator" of the unit sphere. We call this band (shown in figure 5) the *permissible band.*

3.2.2 Estimate of the Condition Number. It is important to determine under what circumstances the recovery of the rotational velocity is ill-conditioned, and whether the different com-

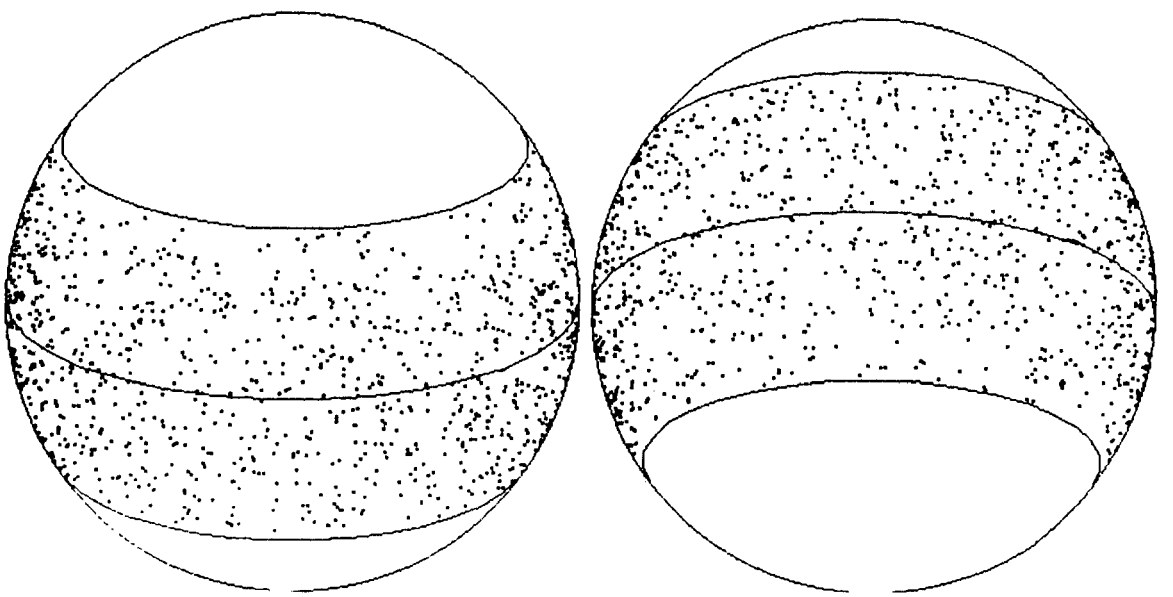


Fig. 5. The permissible band on the v-sphere (front and rear views).

ponents of the rotation vector are equally affected by noise in the data. To study these issues, one needs to estimate some of the properties of the coefficient matrix of the equations. We can get a rough idea of what the integral of $\mathbf{v}\mathbf{v}^T$ looks like by assuming that the particles corresponding to $\mathbf{v}$ are spread *uniformly* within the permissible band. In Appendix B we show that the area of this band is

$$K = 4\pi \sin\theta_v \tag{30}$$

The mixed inertia terms (such as the integral of the product of x and y) are all zero because of the symmetry of the band. The resulting 3×3 matrix is thus diagonal. Furthermore, if we let I_{xx}, I_{yy}, and I_{zz} be the diagonal terms then it can be shown (in Appendix B) that

$$I_{xx} = I_{yy} = (K/3)[1 + (1/2)\cos^2\theta_v] \tag{31}$$

while

$$I_{zz} = (K/3)\sin^2\theta_v \tag{32}$$

Thus I_{xx}/K and I_{yy}/K vary little with θ_v, while I_{zz}/K changes dramatically. The (L^2) condition number of this matrix—the ratio of the largest to the smallest eigenvalue—is just

$$\frac{I_{xx}}{I_{zz}} = \frac{1 + (1/2)\cos^2\theta_v}{\sin^2\theta_v} = \frac{3/2}{\sin^2\theta_v} - \frac{1}{2} \tag{33}$$

This is very large when the field of view is small. When the condition number is very large, small errors in the right-hand sides of the equations, or in the coefficients of the matrix itself, can lead to large errors in the solution.

In fact, the particles are *not* spread uniformly within the permissible band and do not have mass independent of position, so the above is only an estimate. We obtain the exact result in section 3.2.4.

3.2.3 Stability of the Solution Method. The numerical stability of the solution for ω is reduced when the condition number is large. In practice, the elements of the matrix of equation (24) will be corrupted by noise in the measurements, as will the right-hand side vector in this equation. The estimate of the third component of ω will be affected more by these perturbations than the other two. Experiments confirm that the component of rotation about the optical axis is distributed more by noise than the others. The ratio of the errors grows roughly as the inverse of the size of the field of view. This is not a peculiarity of our method, but applies in general.

It is intuitively obvious why this should be. Rotations about the x- and y-axes produce motion fields that vary but little over the image. A small field of view can be used to estimate these components with almost the same accuracy as can a large field of view (provided the same number of picture cells are used.) Rotation about the z-axis, on the other hand, produces a motion field that varies directly with distance from the principal point. Thus the maximum velocity depends on the size of the field of view. With a small field of view, the maximum velocity in the image will be small and relative errors in measurements correspondingly larger.

If an image region is used that is smaller than the whole field of view and perhaps off center, the analysis becomes more complex. In this case, the component of rotation about the direction toward the center of the region is less accurately known; the accuracy again decreasing with the size of the image region. This shows the futility of approaches based on data from small image patches, or higher derivatives of brightness at one point in the image. When working with very small image regions, the best one can do is to estimate the optical flow—there is no point in trying to recover the "rotation" about the center of the region.

3.2.4 Ensemble Average of the Integral of $\mathbf{v}\mathbf{v}^T$. The integral of $\mathbf{v}\mathbf{v}^T$ varies from image to image. It has already been suggested, however, that it will be approximately diagonal. We can obtain a more precise answer by averaging over an ensemble of images with all possible directions for the brightness gradient at each image point. We assume that different directions for the brightness gradient are equally likely. The result so obtained can be viewed in another way: it is the integral obtained in the limit from a textured image as the scale of the texture is made smaller and smaller. In this case we can arrange for every direction of the brightness gradient to be found in any small

patch of the image. By suitable choice of the texture we can arrange that no direction of the brightness gradient is favoured—all directions occur with equal frequency.

If we take into account the distribution of directions of $\mathbf{v}$ and the weights $\|\mathbf{v}\|$, we find (in Appendix B) that

$$\overline{\iint \mathbf{v}\mathbf{v}^T \, dx\, dy} = k_v \begin{pmatrix} 1 + r_v^2/2 + r_v^4/6 & 0 & 0 \\ 0 & 1 + r_v^2/2 + r_v^4/6 & 0 \\ 0 & 0 & r_v^2/2 \end{pmatrix} \tag{34}$$

where r_v is the diameter of the image and the constant k_v depends on the size of the field of view as well as the distribution of magnitudes of the brightness gradient. In practice we can easily find k_v since

$$\text{trace}\left(\iint \mathbf{v}\mathbf{v}^T \, dx\, dy\right) = \iint \text{trace}(\mathbf{v}\mathbf{v}^T) \, dx\, dy = \iint \mathbf{v} \cdot \mathbf{v} \, dx\, dy \tag{35}$$

so

$$2k_v(1 + 3r_v^2/4 + r_v^4/6) = \overline{\iint \mathbf{v} \cdot \mathbf{v} \, dx\, dy} \tag{36}$$

Note that the condition number is

$$\frac{1 + r_v^2/2 + r_v^4/6}{r_v^2/2} = \frac{2}{r_v^2} + 1 + \frac{r_v^2}{3} \tag{37}$$

It attains a minimum of $1 + 2\sqrt{2/3} = 2.633\ldots$ when $r_v = \sqrt{\sqrt{6}} = 1.565\ldots$. Thus the component of rotation about the optical axis is not recovered as accurately as the other two components, no matter how large the field of view. Also, as far as rotation is concerned, there is little advantage to making the field of view wider than a half-angle $\theta_v = \tan^{-1} \sqrt{\sqrt{6}} = 57.42\ldots$ degrees, since the condition number reaches its minimum there.

Some simplifications of the method for recovering the rotational velocity based on the above analysis are discussed by us in reference [10].

3.2.6 The v-Bar Projection. We know that the directions of the vectors $\mathbf{v}$ lie in the permissible band. But what about the vectors

$$\bar{\mathbf{v}} = -E_t \mathbf{v} \tag{38}$$

occurring in the integral on the right-hand side of Eq. (24)? We know that in the case of pure rotation $E_t = -\mathbf{v} \cdot \boldsymbol{\omega}$, so $\bar{\mathbf{v}} = (\mathbf{v} \cdot \boldsymbol{\omega})\mathbf{v}$. We conclude that

$$\bar{\mathbf{v}} \cdot \boldsymbol{\omega} = (\mathbf{v} \cdot \boldsymbol{\omega})^2 \geqslant 0 \tag{39}$$

Thus the directions of the vectors $\bar{\mathbf{v}}$ are confined to a hemisphere with $\boldsymbol{\omega}$ at its pole or "navel." We call this the *compatible hemisphere* for the case of pure rotation.

If the vectors $\bar{\mathbf{v}}$ covered this hemisphere uniformly, we could easily estimate $\boldsymbol{\omega}$ by finding the center of mass of the particles on the unit sphere corresponding to the values of $\bar{\mathbf{v}}$. The center of mass of a hemisphere of uniform density is at a point midway between the center of the sphere and the navel of the hemisphere, so we could use the formula

$$\boldsymbol{\omega} \approx \frac{2 \iint \bar{\mathbf{v}} \, dx\, dy}{\iint \|\bar{\mathbf{v}}\| \, dx\, dy} \tag{40}$$

Unfortunately, the vectors $\bar{\mathbf{v}}$ do not cover the whole compatible hemisphere, since they are confined to the permissible band, just as are the vectors $\mathbf{v}$. In fact, the vectors $\bar{\mathbf{v}}$ lie in the intersection of the permissible band and the compatible hemisphere, as shown in figure 6.

We can now see in another way why a small field of view reduces the accuracy with which we can estimate the component of rotation about the optical axis. If the field of view is small, the permissible band will be narrow, a mere ring. Our task is to guess which hemisphere cut the ring in half. This is easy when we are dealing with a band that covers almost the whole sphere—when it is very narrow, however, there is some uncertainty. A hemisphere claimed to provide a solution can easily be rotated about the line connecting the ends of the cut ring without significantly changing the intersection of the hemisphere and the ring as illustrated in figure 7. Thus the z-

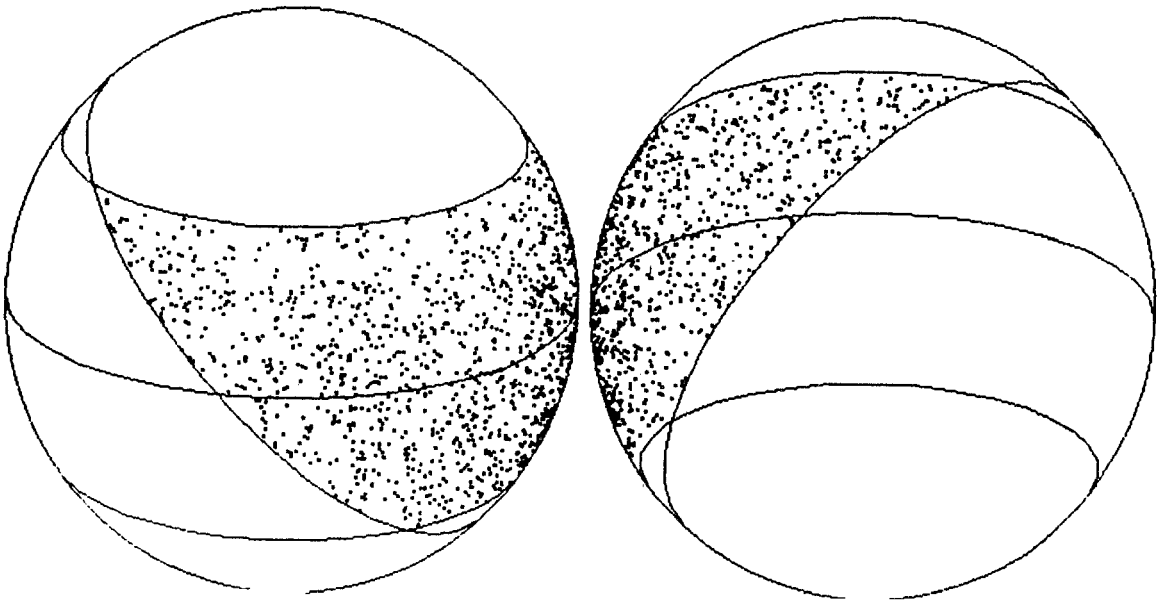


Fig. 6. The v-sphere showing the intersection of the permissible band and the compatible hemisphere, in the case when the field of view is wide (front and rear views).

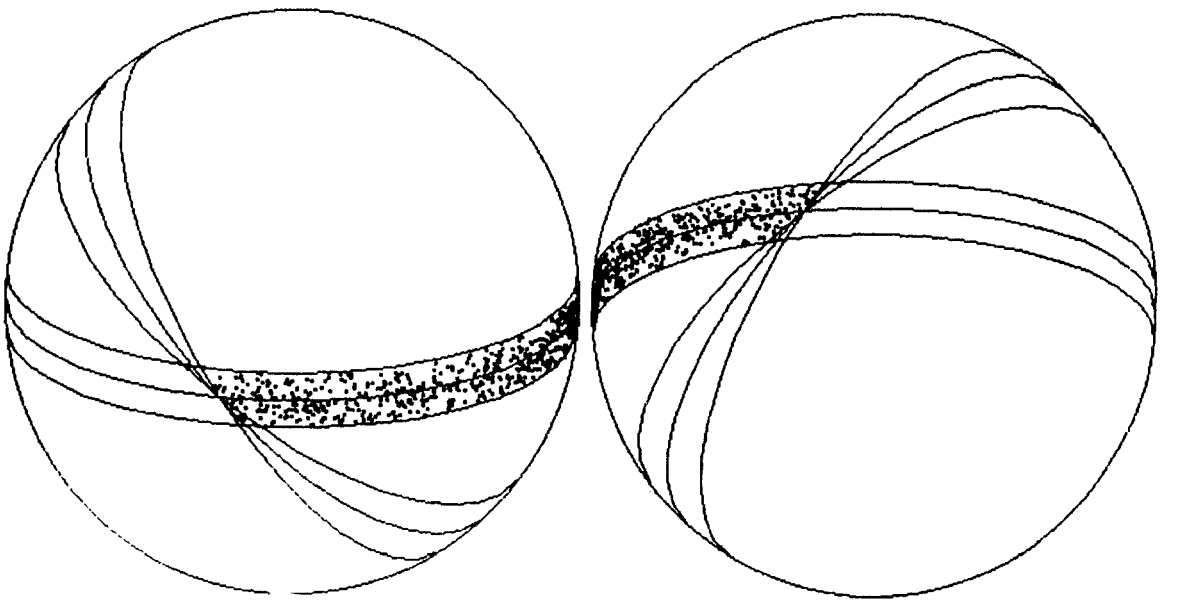


Fig. 7. The intersection of the permissible band with three different compatible hemispheres when the field of view is narrow.

component of the direction to the navel of the hemisphere is uncertain in the presence of noise, or, when measurements from only a small number of picture cells are available.

We could use the geometric insight that the vectors $\mathbf{v}$ all lie in the intersection of the permissible band and the compatible hemisphere to construct algorithms for recovering the direction of the vector ω. Some other method would then have to be found to estimate the magnitude of ω. However, we do not need to approach the problem in this way, in light of the least-squares solution presented above. This geometric approach, however, *will* be fruitful in the case of pure translation where we find a similar geometric constraint and have no need to find the magnitude of the motion vector.

If there is a translational component to the motion, by the way, the points $\bar{\mathbf{v}}$ will not be confined to a hemisphere. This provides a convenient test to see whether the method presented above can be applied or not.

3.3 Rotation Known or Pure Translation

If the rotation is known, perhaps measured by some other instrument, but depth is not, then the general problem reduces to the problem of pure translation. We can write

$$E_t' + (1/Z)\mathbf{s} \cdot \mathbf{t} = 0 \tag{41}$$

where

$$E_t' = E_t + \mathbf{v} \cdot \omega \tag{42}$$

In the remainder of this section we do not distinguish between E_t and E_t'.

Note that equation (41) is not altered if we replace Z by kZ and $\mathbf{t}$ by $k\mathbf{t}$. Thus we can recover motion and depth only up to a scale factor. In the sequel we will set $\|\mathbf{t}\| = 1$ when convenient.

First of all, note that, unlike the case in which depth is known (section 3.1), we cannot obtain a useful result by simply minimizing

$$\iint [E_t + (1/Z)\mathbf{s} \cdot \mathbf{t}]^2 \, dx \, dy \tag{43}$$

since the integrand can be trivially made equal to zero at each point by the choice

$$Z = -\frac{\mathbf{s} \cdot \mathbf{t}}{E_t} \tag{44}$$

(This may, however, produce negative values for Z, a fact that we exploit later.) Given the correct value of $\mathbf{t}$, the above equation provides a means for recovering depth, as already mentioned.

Equation (44) and the fact that depth must be positive, by the way, lead to a simple upper bound on the depth at a particular point even when the direction of translational velocity is not known. Since $Z > 0$, we can write

$$Z = |Z| = \frac{|\mathbf{s} \cdot \mathbf{t}|}{|E_t|} \tag{45}$$

and so

$$Z \leqslant \frac{\|\mathbf{s}\| \|\mathbf{t}\|}{|E_t|} \tag{46}$$

The right-hand side here is the depth computed on the assumption that $\mathbf{s}$ is parallel to $\mathbf{t}$. Of course this is only an upper bound, since Z will be much smaller if $\mathbf{s}$ happens to be nearly orthogonal to $\mathbf{t}$. The bound is particularly poor, as a result, where $\mathbf{r}$ is nearly parallel to $\mathbf{t}$, that is, near the focus of expansion (or the focus of compression) in the image.

3.3.1 Depth Known—The Case of Pure Translation. If we know the depth, as earlier, we can minimize the total error in the time derivative of brightness:

$$\iint [E_t + (1/Z)\mathbf{s} \cdot \mathbf{t}]^2 \, dx \, dy \tag{47}$$

by differentiating with respect to $\mathbf{t}$. Setting the result equal to zero gives

$$\left[\iint (1/Z)^2 \mathbf{s}\mathbf{s}^T \right] \mathbf{t} \, dx \, dy = -\iint (1/Z) E_t \mathbf{s} \, dx \, dy \tag{48}$$

which is just equation (19) with $\|\omega\| = 0$. This is a set of three linear equations in the three components of $\mathbf{t}$ (U, V, and W). The coefficient matrix is symmetric and only the right-hand side depends on the time derivative of brightness. Note that in equation (48) we attach less weight to information from points where Z is large.

The method is accurate if the correct values of depth are given. If estimates are used, the quality

of the result will depend on the quality of the estimates. The accuracy of the result also depends on the size of the field of view, as we show later.

We get slightly different results if we weight the integrand in equation (47) differently. Multiplying by Z, for example, gives $[ZE_t + \mathbf{s} \cdot \mathbf{t}]$ for the integrand and

$$\left[\iint \mathbf{s}\mathbf{s}^T \, dx\, dy\right]\mathbf{t} = -\iint ZE_t \mathbf{s} \, dx\, dy \tag{49}$$

for the solution. Alternatively multiplying by Z/E_t gives $[Z + (1/E_t)\mathbf{s} \cdot \mathbf{t}]$ for the integrand and

$$\left[\iint (1/E_t)^2 \mathbf{s}\mathbf{s}^T \, dx\, dy\right]\mathbf{t} = -\iint (Z/E_t)\mathbf{s} \, dx\, dy \tag{50}$$

for the solution. In this case we are minimizing the error in depth, rather than the error in the time derivative of brightness, as in equation (48).

The two alternate solutions given in equations (49) and (50) have the advantage that the depth Z does not appear in the integrals on the left-hand side. This means that they are particularly well suited for iterative schemes where Z is reestimated on each cycle. The solution of equation (49) has the further advantage that neither Z nor E_t appear on the left-hand side. This makes it easy to compute an ensemble average for this integral.

3.3.2 Distribution of the Directions of $\mathbf{s}$. To understand the properties of the above algorithms for recovering $\mathbf{t}$, we must examine the matrix obtained by integrating multiples of $\mathbf{s}\mathbf{s}^T$. Once again, we can think of the direction of $\mathbf{s}$ as identifying a point on the unit sphere and of a multiple of $\|\mathbf{s}\|$ as the mass of a particle placed there. The integral considered is related to the inertia matrix of the set of particles on the unit sphere.

Now just as the directions of $\mathbf{v}$ lay in a band of width equal to the width of the field of view, because $\mathbf{v} \cdot \mathbf{r} = 0$, so do the directions of $\mathbf{s}$, since $\mathbf{s} \cdot \mathbf{r} = 0$. The distribution of points within the band is not quite the same, but we will ignore such details for now. First of all, assuming again a uniform distribution within the permissible band, we get the same estimate of the condition number as in section 3.2.2, namely

$$\frac{1 + (1/2)\cos^2 \theta_v}{\sin^2 \theta_v} = \frac{3/2}{\sin^2 \theta_v} - \frac{1}{2}$$

Accuracy in the determination of W, the Z component of $\mathbf{t}$, will be reduced relative to that of the other two components when the field of view is small. Experiments confirm that for small fields of view, the estimate of the component of translation along the optical axis is disturbed more by noise than the other two. Hence a wide field of view is called for.

The integral of $\mathbf{s}\mathbf{s}^T$ varies from image to image. In order to better understand the matrix defined by $\mathbf{s}\mathbf{s}^T$, we would like to examine a typical image. Since it is difficult to define such an image, instead, as in section 3.2.4, we can take an average over an ensemble of images containing all possible directions for the brightness gradient at each image point. If we take into account the distribution of directions of $\mathbf{s}$ and the weights $\|\mathbf{s}\|$, we find (in Appendix B) that

$$\overline{\iint \mathbf{s}\mathbf{s}^T \, dx\, dy} = k_s \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & r_v^2/2 \end{pmatrix} \tag{51}$$

where r_v is the radius of the image and the constant k_s depends on the size of the field of view as well as the distribution of magnitudes of the brightness gradient. In practice we can find k_s by noting that

$$\begin{aligned} &\text{trace}\left(\iint \mathbf{s}\mathbf{s}^T \, dx\, dy\right) \\ &\quad = \iint \text{trace}\,(\mathbf{s}\mathbf{s}^T)\, dx\, dy \\ &\quad = \iint \mathbf{s} \cdot \mathbf{s} \, dx\, dy \end{aligned} \tag{52}$$

so

$$2k_s(1 + r_v^2/4) = \overline{\iint \mathbf{s} \cdot \mathbf{s} \, dx\, dy} \tag{53}$$

Note that the condition number is just $\min(r_v^2/2, 2/r_v^2)$ which reaches a minimum of 1 when $r_v = \sqrt{2}$. In the case of pure translation, the component of translation along the optical axis is

found with more accuracy than the other two when the field of view has a half-angle wider than $\theta_v = \tan^{-1}\sqrt{2} = 54.74\ldots$ degrees, since $r_v^2/2$ then is larger than one.

Some simplifications of the method for recovering the translational velocity based on the above analysis are discussed by us in reference [10].

3.4 Translation with Rotation Known

In this section we deal with the problem of determining the direction of translation and depth $Z(x,y)$ given the rotation vector ω.

3.4.1 The Importance of a Wide Field of View. In the general case, the need for a wide field of view

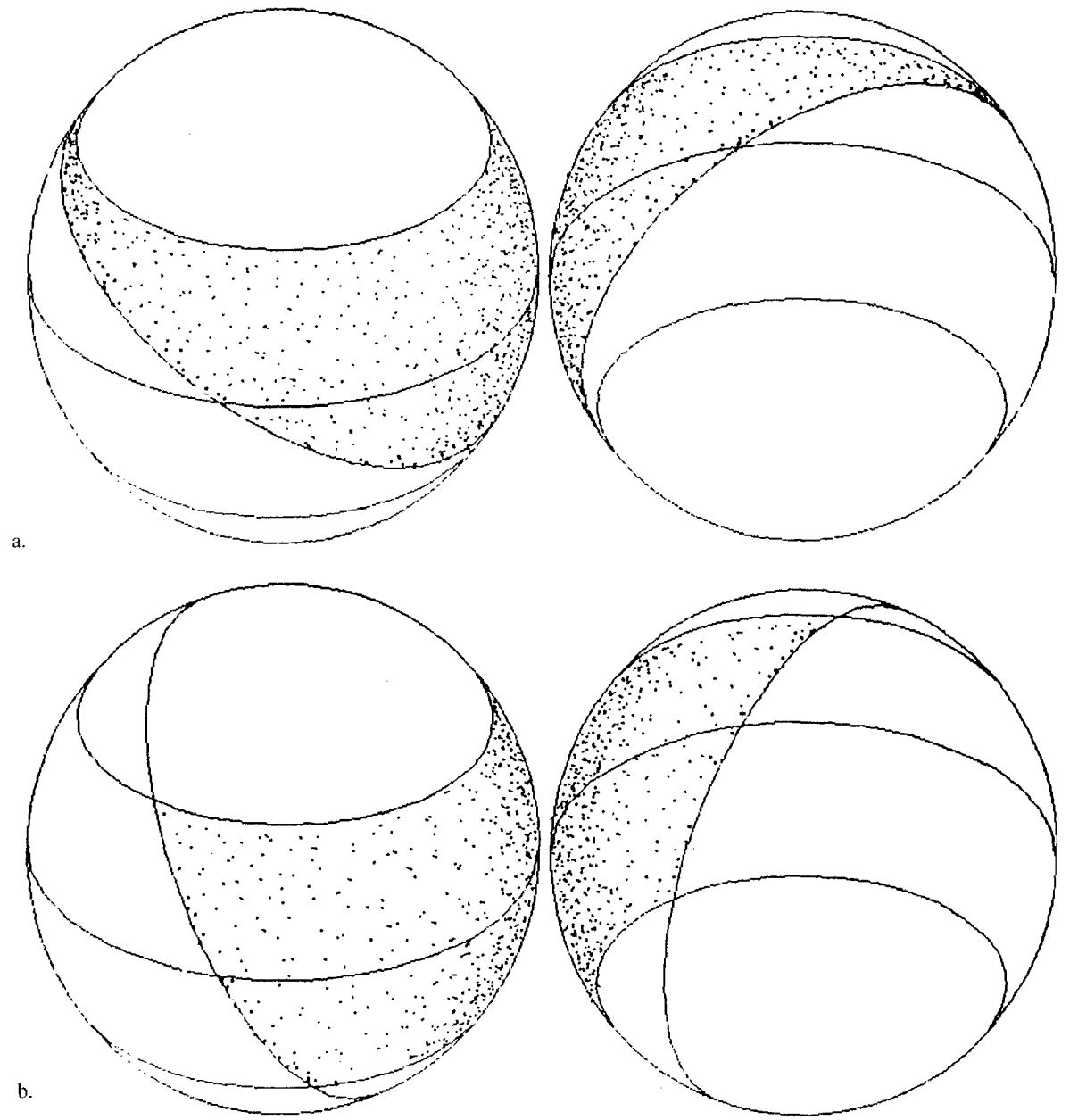


Fig. 8. The intersection of the permissible band on the s-shape and the compatible hemisphere for two cases. (a) Focus of expansion within field of view. (b) Focus of expansion outside field of view.

is very clear. In a small image region near the center of the image, for example, rotation about the y-axis looks the same as translation along the x-axis, while rotation about the x-axis looks the same as translation along the (negative) y-axis. As is well known in stereo-photogrammetry, a large field of view is needed to separate these components of the transformation between two camera positions [11, 12].

If we take note of this ambiguity, and the uncertainty with which the components of rotation and translation along the optical axis can be determined, we see that locally, out of six parameters, only two combinations can be estimated. These two quantities are just the components of the motion field. The same argument can be made for points at some distance from the principal point of the image.

There is a difference between the case when the motion is predominantly along the optical axis and the case where it is predominantly parallel to the image plane. The transition between the two situations occurs when the direction of the vector $\mathbf{t}$ moves outside the cone of directions of the field of view, that is, when the focus of expansion (or compression) moves outside the image. When the focus of expansion is inside the image, then the great circle defined by $\mathbf{s} \cdot \mathbf{t} = 0$ lies entirely inside the permissible band on the unit sphere (figure 8a). The measured values of $\mathbf{s}$ then provide constraint all the way around the great circle. Conversely, when the focus of expansion lies outside the image, the great circle cuts the permissible band (figure 8b). In this case the known values of $\mathbf{s}$ provide constraint only along two segments of the great circle. These segments get shorter and shorter as the vector $\mathbf{t}$ becomes more and more parallel to the image plane. It should be clear that in this case the direction of the vector $\mathbf{t}$ can be determined with somewhat lower accuracy than when the focus of expansion is near the principal point.

3.4.2 The $\bar{\mathbf{s}}$-Bar Projection. The integrals on the right-hand side of the equations for $\mathbf{t}$ developed in section 3.3.1 contain positive multiples of the vector

$$\bar{\mathbf{s}} = -\text{sign}\,(E_t)\mathbf{s} \tag{54}$$

(Here we only care about the directions of the vectors, so we ignore scale factors.) Now in the case of translation with known rotation, we have (from equation (41))

$$E_t = -(1/Z)\mathbf{s} \cdot \mathbf{t}$$

and

$$\begin{aligned}\bar{\mathbf{s}} \cdot \mathbf{t} &= (1/Z)\ \text{sign}\,(\mathbf{s} \cdot \mathbf{t})\mathbf{s} \cdot \mathbf{t} \\ &= (1/Z)|\mathbf{s} \cdot \mathbf{t}| \geqslant 0\end{aligned} \tag{55}$$

since $Z > 0$. We are only interested at this point in the sign of $\bar{\mathbf{s}} \cdot \mathbf{t}$, so we can use any convenient positive multiple of $\bar{\mathbf{s}}$ such as

$$-(1/E_t)\mathbf{s}, \quad -\text{sign}\,(E_t)\mathbf{s}, \quad \text{or} \quad -E_t\mathbf{s}$$

in the discussion that follows.

Equation (55) states that $\bar{\mathbf{s}}$ can only lie in the hemisphere that has $\mathbf{t}$ as its navel. We call this the *compatible hemisphere* in the case of translation with known rotation. Since $\bar{\mathbf{s}}$ is a multiple of $\mathbf{s}$, it must also lie in the permissible band. Thus $\bar{\mathbf{s}}$ can only lie in the intersection of the permissible band and the compatible hemisphere. We will exploit this geometric insight shortly.

Our task can be viewed as that of finding the hemisphere that contains all of the directions specified by the vectors $\bar{\mathbf{s}}$ derived from the image. Note that the solution may not be unique and that there may not be any solution. Later we will modify the problem definition somewhat to deal with these possibilities.

If there is a rotational component to the motion, by the way, the points $\bar{\mathbf{s}}$ will not be confined to a hemisphere. This provides a convenient test to see whether the methods presented here can be applied or not.

3.4.3 Motion Determination as a Linear Programming Problem. We wish to find a vector $\mathbf{t}$ that makes $\bar{\mathbf{s}} \cdot \mathbf{t} \geqslant 0$ at all image points. We can think of this as a gigantic linear programming problem.[2] There are three unknowns and one inequality for every picture cell. (Actually, since we do not care about the magnitude of $\mathbf{t}$, there are only two degrees of freedom.)

[2]We do not wish to suggest, by the way, that linear programming algorithms could be fruitfully applied to this problem.

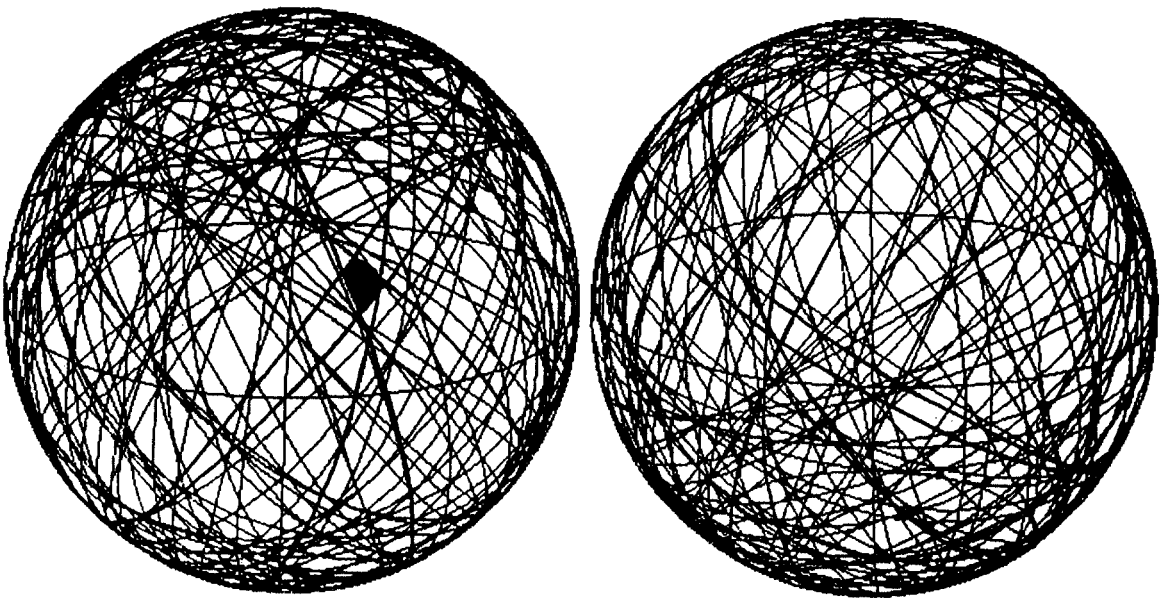


Fig. 9. The s-sphere showing great circles for many image points. Circles for critical points (emphasized) constrain location of **t**.

Since we do not have a criterion function to be extremized, we will have an infinite number of solutions—if there are any solutions at all. All of these solutions will lie in a convex polygon on the unit sphere. The sides of this polygon are portions of great circles corresponding to constraints which we will call *critical constraints* (see figure 9). With data from a large number of cells we expect this solution polygon to be small. We may choose its center as the "best" solution.

Typically, the solution polygon will have relatively few sides. Thus data from a small number of *critical picture cells* fully constrain the solution. First of all, note that each side of this polygon corresponds to an equality of the form $\bar{\mathbf{s}} \cdot \mathbf{t} = 0$ for some picture cell. From the brightness change constraint equation we know that $E_t = 0$ when $\bar{\mathbf{s}} \cdot \mathbf{t} = 0$. Thus the critical constraints are provided by picture cells where E_t is small (and $\bar{\mathbf{s}}$ is not). This is an important observation, which can be used to reduce the size of the linear programming problem; we simply disregard the inequalities arising from picture cells where E_t is large.

(There is a class of points for which $\bar{\mathbf{s}} \cdot \mathbf{t}$ is arbitrary, even though $\mathbf{E}_t$ is small and $\bar{\mathbf{s}}$ is not; these are image points for which Z is large. Such points provide false constraints on **t**. For a practical system, some means must be found for identifying these points. One way of doing this, for images with large depth range, is based on the following observation. In a real image, regions for which Z is large, that is, the background, tend to encompass a significant area with *all* points in the area having $E_t \approx 0$. On the other hand, points with $\mathbf{s} \cdot \mathbf{t} = 0$ and $\bar{\mathbf{s}}$ large are usually isolated and surrounded by regions for which $E_t \neq 0$. The above difficulty appears in all of the methods of determining motion; it is harder to determine **t** when the depth range is large.)

We observe in passing that the points most useful in constraining the translational motion vector are the very same points where it is difficult to calculate depth accurately! One may also make the observation that a method for segmenting the scene into foreground and background regions would be very useful in the case of general motion, since the background regions can then be used to recover the rotational component. The known rotational velocity can then, in turn, be used to recover the translational component from the foreground regions.

The linear programming method of determining $\mathbf{t}$ discussed above uses relatively little of the image data. In fact, only points at the edge of the compatible hemisphere influence the solution at all. While this is a sensible procedure if the data is trustworthy, it will be quite sensitive to noise. For noisy—that is, real images, it may be worthwhile to consider other points, such as those in a band for which E_t is less than small cutoff value.

3.4.4 The Perceptron "Learning" Algorithm. One way of finding the solution of a large number of homogeneous inequalities is by means of the iterative perceptron "learning" algorithm (Minsky and Papert [13]; Duda and Hart [14]). Given a set of vectors $\{\bar{\mathbf{s}}_i\}$, this procedure is guaranteed to find a vector $\mathbf{t}$ that satisfies $\bar{\mathbf{s}}_i \cdot \mathbf{t} \geqslant 0$, *if* such a vector exists. It even does this in a finite number of steps, provided there exists some ε such that $\bar{\mathbf{s}}_i \cdot \mathbf{t} > \varepsilon$ for all $\bar{\mathbf{s}}_i$ in the given set (which almost always happens when the set is finite).

The idea is to start with some nonzero vector $\mathbf{t}^0$ and to test whether the inequalities are satisfied. (A reasonable choice for $\mathbf{t}^0$ is one of the vectors $\bar{\mathbf{s}}_i$.) If the inequality is not satisfied for a particular vector in the set, then the smallest adjustment is made to make the dot-product zero. (Note that this may disturb inequalities that have passed the test already.) Suppose that the present estimate for the direction of the translation vector is the direction of $\mathbf{t}^n$. We now test the dot-product $\bar{\mathbf{s}}_i \cdot \mathbf{t}^n$. If it is negative, we adjust our estimate of the vector $\mathbf{t}$ according to the rule

$$\mathbf{t}^{n+1} = \mathbf{t}^n + \delta\mathbf{t}^n \tag{56}$$

where

$$\delta\mathbf{t}^n = -\frac{\mathbf{s}_i \cdot \mathbf{t}^n}{\mathbf{s}_i \cdot \mathbf{s}_i}\mathbf{s}_i \tag{57}$$

Note that $\bar{\mathbf{s}}_i \cdot \mathbf{t}^{n+1} = 0$ and that the magnitude of $\mathbf{s}_i$ does not matter. (Also, the test above can be replaced with a test that checks whether $-\mathbf{s}_i \cdot \mathbf{t}^n$ has the same sign as E_t.)

If the inequalities are inconsistent, that is, if the $\bar{\mathbf{s}}_i$ are not confined to a hemisphere (or nearly so), as will happen in practice due to noise, the algorithm will not converge. Furthermore, there is no guarantee that the guess at any stage is particularly good. We discuss several simple refinements that can help in this case in reference [10].

The vector $\mathbf{t}^n$ in the perceptron "learning" algorithm is obviously a linear combination of vectors drawn from the set $\{\bar{\mathbf{s}}_i\}$. Vectors in this set have directions that correspond to points in the permissible band. Now suppose that this band is very narrow. Then, to build a vector with a significant z-component one has to add many of these vectors. In order to keep the x- and y-components small, these vectors must almost come in pairs from opposite ends of the narrow band. Not surprisingly, the algorithm performs rather poorly in this situation; it is much happier with vectors sprinkled uniformly in direction over a full hemisphere.

It should also be noted that in a real-time application, we do not expect the velocity estimates to change rapidly. Thus the previous value of the velocity is likely to be an excellent first estimate for the current value of $\mathbf{t}$. This means that very few iterations will be needed to get an acceptable new value. A considerable amount of computation can be saved this way, just as it can in the computation of the optical flow (Horn and Schunck [6]). We discuss a parallel perceptron algorithm in reference [10].

3.4.5 Minimizing the Integral of Z^2. In this section we assume that the depth range Z_{max}/Z_{min} is finite. This will generally be the case in robotic applications. The method discussed in this section can also be applied to images in which the background has very large Z, if, as discussed in section 3.4.3, these regions are excised from the image before the motion vector is calculated.

We have seen that we can compute depth when the motion $\mathbf{t}$ is known using equation (44)

$$Z = -(1/E_t)\mathbf{s} \cdot \mathbf{t}$$

Now if we use the *wrong* value $\mathbf{t}'$ in this formula, we get the wrong depth value:

$$Z' = -(1/E_t)\mathbf{s} \cdot \mathbf{t}' = Z(\mathbf{s} \cdot \mathbf{t}')/(\mathbf{s} \cdot \mathbf{t}) \tag{58}$$

We expect only positive values for Z, but this formula may give us negative values, since $(\mathbf{s} \cdot \mathbf{t}')$ may be negative where $(\mathbf{s} \cdot \mathbf{t})$ is positive and vice versa. More interestingly, we may obtain very large values for Z (both positive and negative).

since $(\mathbf{s} \cdot \mathbf{t})$ may be almost zero while $(\mathbf{s} \cdot \mathbf{t}')$ is not. That is, the magnitude of Z will often be very large near points where $E_t \approx 0$. We may conclude that we could determine the correct value for $\mathbf{t}$ by minimizing the integral of Z^2 over the image, that is by minimizing the quadratic form

$$\iint (1/E_t)^2(\mathbf{s} \cdot \mathbf{t})^2 \, dx\, dy$$

$$= \mathbf{t}^T \left[\iint (1/E_t^2)\mathbf{s}\mathbf{s}^T \, dx\, dy \right] \mathbf{t} \tag{59}$$

subject to the constraint $\|\mathbf{t}\| = 1$. The solution is the eigenvector of the real symmetric 3×3 matrix

$$M = \iint (1/E_t^2)\mathbf{s}\mathbf{s}^T \, dx\, dy \tag{60}$$

associated with the smallest eigenvalue. We can prove this by minimizing the sum

$$S = \mathbf{t}^T M \mathbf{t} + \lambda(1 - \mathbf{t}^T\mathbf{t}) \tag{61}$$

where λ is a Lagrangian multiplier. Then

$$\frac{\partial S}{\partial \mathbf{t}} = 2M\mathbf{t} - 2\lambda\mathbf{t} = 0 \tag{62}$$

which yields

$$M\mathbf{t} = \lambda\mathbf{t} \tag{63}$$

Thus λ is an eigenvalue of M, and $\mathbf{t}$ is the corresponding eigenvector. Substituting equation (63) into equation (61) gives the result $S = \lambda$. Thus $\mathbf{t}^T M \mathbf{t}$ is minimized by taking the smallest of the three eigenvalues of M for λ.

To minimize problems due to noise, we can add a small positive constant to E_t^2 commensurate with the expected noise in E_t^2. That is, we take as our solution the eigenvector of

$$M' = \iint \frac{1}{E_t^2 + n^2} \mathbf{s}\mathbf{s}^T \, dx\, dy \tag{64}$$

associated with the smallest eigenvalue.

If $\mathbf{e}$ is an eigenvector, so is $-\mathbf{e}$. But we want Z to be positive. Rather than test this condition at a single point, we compute an average like

$$\bar{\mathbf{s}}_0 = -\iint (1/E_t)\mathbf{s} \, dx\, dy \qquad \text{or}$$

$$\bar{\mathbf{s}}_0 = -\iint \frac{E_t}{E_t^2 + n^2} \mathbf{s} \, dx\, dy \tag{65}$$

and check whether

$$\bar{\mathbf{s}}_0 \cdot \mathbf{t} > 0 \tag{66}$$

If it is not, we simply change the sign of the solution $\mathbf{t}$.

As before, we may choose to weight the integral of equation (59) according to some measure of how trustworthy are the data from each picture cell.

The method presented in this section produces an estimate of the translation vector $\mathbf{t}$ in closed form and with high accuracy. Of course, a cubic must be solved to obtain the eigenvalues—but there is an analytic method for doing that. The corresponding eigenvectors can then be found by taking cross products of two rows of a 3×3 matrix.

The preceding method of calculating $\mathbf{t}$ has another justification that some readers might find more persuasive. From equation (41) we know that $\mathbf{s} \cdot \mathbf{t} \approx 0$ for points with $E_t \approx 0$ (again ignoring background points). Thus we are basically looking for a vector $\mathbf{t}$ that makes $\mathbf{s} \cdot \mathbf{t} \approx 0$ whenever $E_t \approx 0$. The points where the time derivatives are small provide most constraint, as already discussed. We could try to minimize something like

$$\iint_C (\mathbf{s} \cdot \mathbf{t})^2 \tag{67}$$

where C is the set of image points where $E_t \approx 0$. Rather than use a strict cutoff on E_t, we may consider a weighting scheme in an integral like

$$\iint w(\mathbf{s} \cdot \mathbf{t})^2 \, dx\, dy \tag{68}$$

over the whole image where the weighting function w is chosen to emphasize points where $E_t \approx 0$. A reasonable choice, $w = 1/(E_t^2 + n^2)$, leads to integral given in equation (64). The eigenvector corresponding to the smallest eigenvalue is a normal of the plane that best fits the weighted set of points (see figure 10).

If there is a rotational component of the motion, by the way, the vectors $\bar{\mathbf{s}}$ where E_t is small will not lie near a great circle. In this case the smallest eigenvalue will not be small. This provides a convenient test. We discuss a method that avoids the need to find eigenvalues and eigenvectors in reference [10]. Related methods for finding the

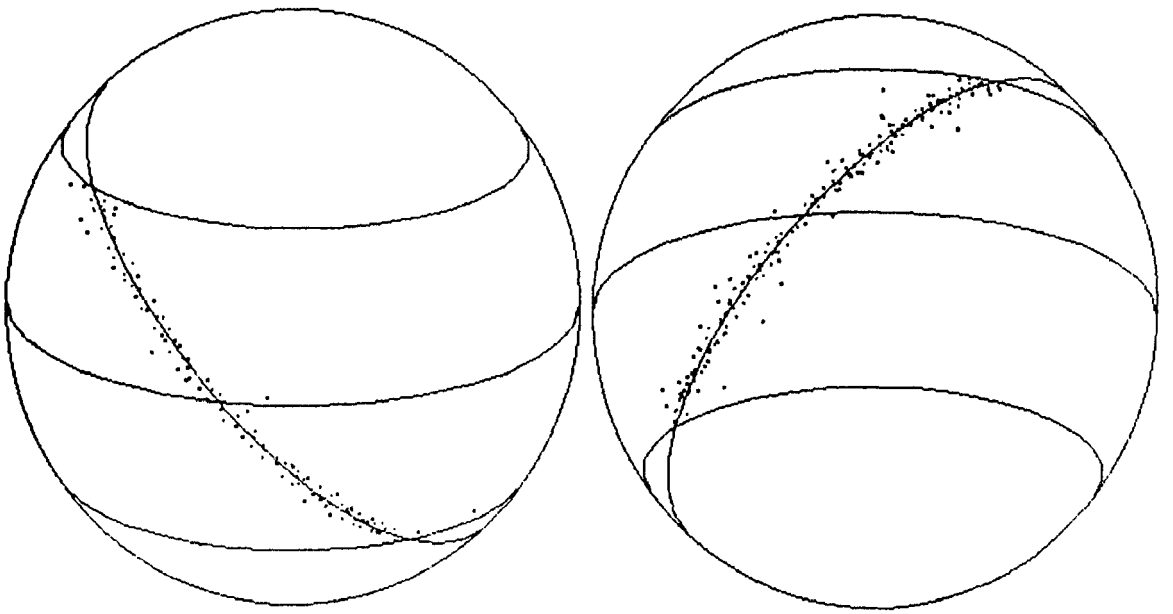


Fig. 10. The great circle corresponding to the motion **t** which best fits the points on the **s**-sphere for which $E_t \approx 0$.

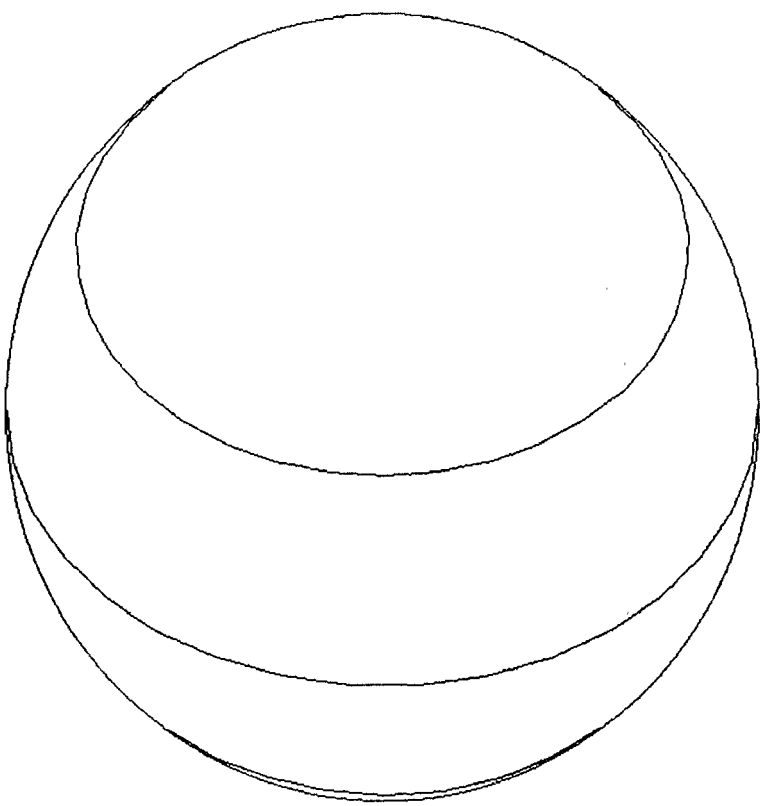


Fig. 11. Plot of several noisy estimates of the translation vector **t** on the **s**-sphere (200 pixels/estimate, 1% noise in brightness measurements).

focus of expansion are discussed in reference [15].

Figure 11 shows a scatter plot of positions on the unit sphere for **t** recovered from noisy synthetic data. Each estimate is based on brightness gradient at 200 picture cells with 1% noise in the derivatives of brightness. Note the elongation of the cluster of points in a direction parallel to the optical axis. When tens of thousands of picture cells are used, instead of hundreds, the algorithm can tolerate considerably more noise. While further experimentation is called for, we found that this algorithm behaves at least as well, if not better, than the others we have investigated.

4 Conclusions

We have developed methods for recovering motion directly from the first derivatives of brightness in an image region in the cases of pure rotation and pure translation (and general motion when the rotational component is known). We have tested these methods on synthetic image data and, to a limited extent, on some kinds of real-image sequences. In the case of pure rotation we give an exact simple solution to the obvious least-squares problem. In the case of pure translation we give several methods with different tradeoffs between accuracy, noise-sensitivity and computational expense. While we have preliminary ideas about the relative merits of these methods, detailed conclusions will have to await further careful experimentation with real images. Some further results on both synthetic and real-image data are reported in references [10] and [15].

We show that it is trivial to recover depth when the motion is known and that it is trivial to recover the motion when depth is known. We emphasize the importance of a large field of view and point out difficulties arising in the pure translation case when there is a very large depth range. We also note that image points where the brightness derivative is small provide most constraint on the translation vector while the depth at these points is hard to recover. We show that it is difficult to recover the translational motion toward, and rotational motion about, the line connecting the projection center to the image region of interest, when that region is small. We emphasize the need for adequate low-pass filtering in both spatial and time dimension before sampling in order to ensure that estimates of derivatives are accurate.

The discussion is facilitated by introduction of the auxiliary vectors **s** and **v**. The directions of these vectors have been shown to be constrained to lie in the intersection of a *permissible band* and a *compatible hemisphere* on the unit sphere. These geometric concepts help lend intuitive support to the algebraic results.

5 Acknowledgement

The authors would like to thank the other members of the Waikiki Surfing and Computer Vision Society, and especially Shahriar Negahdaripour, for their contributions to this paper.

References

1. A.R. Bruss and B.K.P. Horn, "Pasive navigation," *Computer Vision, Graphics, and Image Processing,* 21, pp. 3–20, 1983.
2. G. Adiv, "Determining 3-D motion and structure from optical flow generated by several moving objects," COINS TR 84-07, Computer and Information Science, University of Massachusetts, Amherst, MA, April 1984.
3. S. Negahdaripour and B.K.P. Horn, "Direct passive navigation," *IEEE Trans. PAMI*-9(1), pp. 168–176, January 1987.
4. A.M. Waxman and S. Ullman, "Surface structure and 3-D motion from image flow: A kinematic analysis," *Intern. J. Robotics Res.,* 4(3), pp. 72–94, Fall 1985.
5. A.M. Waxman and K. Wohn, "Contour evolution, neighborhood deformation, and global image flow: Planar surfaces in motion," *Intern. J. Robotics Res.* 4(3), pp. 95–108, Fall 1985.
6. B.K.P. Horn and B.G. Schunck, "Determining optical flow," *Artificial Intelligence,* 17, pp. 185–203, 1981.
7. H.C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proc. Roy. Soc. London* B 208, pp. 385–397, 1980.
8. S. Negahdaripour, "Direct methods for structure from motion," Ph.D. Thesis, Mechanical Engineering, MIT, September 1986.
9. Y. Aloimonos and C. Brown, "Direct processing of curvilinear motion from a sequence of perspective images," *Proc. Workshop on Computer Vision, Representation and Control,* Annapolis, Maryland, 1984.
10. B.K.P. Horn and E.J. Weldon Jr., "Computationally effi-

cient methods of recovering translational motion," *Intern. Conf. Computer Vision,* London, England, June 8-11, 1987.
11. P.R. Wolf, *Elements of Photogrammetry,* McGraw-Hill: New York, 1983.
12. S.K. Gosh, *Theory of Stereophotogrammetry.* Ohio State University Bookstores, 1972.
13. M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry.* MIT Press: Cambridge, MA, 1969.
14. R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis.* Wiley: New York, 1973.
15. S. Negahdaripour and B.K.P. Horn, "Using depth-is-positive constraint to recover translational motion," *IEEE Workshop on Computer Vision,* Miami Beach, Florida, 1987.

Appendix A: The Brightness Change Constraint Equation

The brightness change constraint equation is based on the assumption that the brightness of the image of a patch on the surface does not change as the observer moves relative to the surface. Expansion of the total derivative in the equation $dE/dt = 0$ by means of the chain-rule leads to the constraint equation

$$\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \qquad \text{(A1)}$$

or

$$uE_x + vE_y + E_t = 0 \qquad \text{(A2)}$$

where E_x, E_y, and E_t are the partial derivatives of brightness with respect to x, y, and t, while u and v are the time derivatives of x and y.

In practice, the brightness of a patch rarely remains exactly the same. The brightness change constraint equation is nevertheless a useful approximation, as long as the change in brightness at an image point due to the motion is much larger than the change in brightness due to other effects, such as change in viewing direction or illumination. This will be the case as long as there is good contrast at high spatial frequencies, as will be shown next.

Suppose that the brightness of a patch does in fact change due to changes in viewing direction or changes in illumination. In most cases the rate of change of brightness will be relatively small. Let us say that $dE/dt = \varepsilon$, and so

$$E_t = -(uE_x + vE_y) + \varepsilon \qquad \text{(A3)}$$

Consider now a simple grating pattern in the image that, at a particular time, is described by the equation

$$E = E_0[1 + \sin(ax + by)] \qquad \text{(A4)}$$

Then the components of the brightness gradients are

$$E_x = aE_0 \cos(ax + by)$$
$$E_y = bE_0 \cos(ax + by) \qquad \text{(A5)}$$

Consequently

$$uE_x + vE_y = (au + bv)E_0 \cos(ax + by) \qquad \text{(A6)}$$

It is clear that the error in E_t, the rate of change of brightness at a point in the image, resulting from changes in the brightness of the surface, is relatively small, as long as $(au + bv)E_0$ is large compared to ε. (This term, $(au + bv)E_0$, will be zero when the image motion happens to be parallel to the ridges of the grating. In practice, however, surface markings will contain many spatial frequency components and most of these will not be aligned in this special way.) We conclude that the relative error in the rate of change of brightness with time is small, as long as there is significant contrast at the higher spatial frequencies.

The approximation breaks down when the surface markings are weak and the changes of brightness due to changes in viewing direction or illumination are rapid. This happens, for example, in the rare situation where a specular surface momentarily lines up exactly to reflect the light from a point source toward the viewer. It also happens when an object moving relative to a point source enters a cast shadow.

A number of additional factors help keep the relative error in E_t small. First of all, some surfaces have the property that they appear equally bright from all viewing directions. A Lambertian

surface is a very special case of this, where the brightness varies as the cosine of the incident angle. The image brightness is not affected at all by the motion of the observer when a surface of this type is fixed relative to the light source. While most real surfaces do not appear equally bright from all viewing directions, brightness typically varies slowly with changes in observer position (slowly enough that we are usually not aware of any such changes).

Brightness variations resulting from changes in surface orientation are most severe when there is a single point source. These variations are reduced when there are multiple sources or an extended source. In the extreme case of a scene illuminated from all sides, for example, image brightness does not depend on surface orientation at all, even if the surface is specular!

Similarly, a lens occupying a large solid angle, as seen from the object, will smooth out changes in brightness resulting from changes in viewer position. One can see this easily in the extreme case of a glossy reflection, which will be seen only over a small range of positions if a small lens of pin hole is used. A large entrance aperture on the other hand will smear out the highlight effect over a larger range of viewing positions. This is not a big help in many imaging situations, however, since objects are far from the sensor relative to the size of the sensor, except in the case of microscopy.

To summarize: The brightness of the image of a patch may change somewhat as the observer moves relative to the surface. The brightness change constraint equation nevertheless provides a good way of estimating the rate of change of brightness with respect to time at a point in the image. The relative error in this estimate will be small when there is significant contrast in the surface markings at higher spatial frequencies. (There will be no error at all when the surface appears equally bright from all viewing directions and the object does not move relative to the light source.)

Appendix B: Ensemble Averages

Some of the integrals that appear in this paper, while functions of the scene content, tend to lie close to average values when evaluated over sufficiently large textured regions. These average values can be useful in two different enterprises:

- analyzing the relative stability of the components of the solution, and
- developing simplified methods for recovering the solution (as shown in reference [10]).

In order to compute these averages we have to make some assumptions about the probability distribution of brightness gradients. We assume here that this distribution is rotationally symmetric and independent of image position. That is, on average we see the same brightness gradients at every image point, and all directions of the brightness gradient are equally likely. The distribution of magnitude of this gradient is left arbitrary, however, since it does not directly affect the main results.

B.1 Moment Integrals for the Uniform Band

Before we start, let us quickly obtain the equivalent results under the assumption that data points ($\mathbf{s}$ or $\mathbf{v}$) are *uniformly* distributed over the permissible band. Let η denote the latitude and ξ the longitude on the unit sphere. We see that the area of the band is just

$$K = \int_{-\pi}^{\pi} \int_{-\theta_v}^{\theta_v} \cos\eta \, d\eta \, d\xi = 4\pi \sin\theta_v \qquad \text{(B1)}$$

The Cartesian coordinates of a point on the unit sphere are given by $\mathbf{p} = (x,y,z)^T$ where

$$\begin{aligned} x &= \cos\eta \cos\xi \\ y &= \cos\eta \sin\xi \\ z &= \sin\eta \end{aligned} \qquad \text{(B2)}$$

Let the integral of $\mathbf{p}\mathbf{p}^T$ be

$$\int_{-\pi}^{\pi} \int_{-\theta_v}^{\theta_v} \mathbf{p}\mathbf{p}^T \, d\eta \, d\xi = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} \qquad \text{(B3)}$$

Then

$$\begin{aligned} I_{xx} &= \int_{-\pi}^{\pi} \int_{-\theta_v}^{\theta_v} x^2 \cos\eta \, d\eta \, d\xi \\ &= \int_{-\pi}^{\pi} \cos^2\xi \, d\xi \int_{-\theta_v}^{\theta_v} \cos^3\eta \, d\eta \end{aligned} \qquad \text{(B4)}$$

that is,

$$I_{xx} = \frac{\pi}{3}[3 \sin\theta_v + (1/3)\sin 3\theta_v]$$

$$= \frac{4\pi}{3}\sin\theta_v[1 + (1/2)\cos^2\theta_v] \quad \text{(B5)}$$

while

$$I_{zz} = \int_{-\pi}^{\pi}\int_{-\theta_v}^{\theta_v} z^2 \cos\eta \, d\eta \, d\xi$$

$$= \int_{-\pi}^{\pi} d\xi \int_{-\theta_v}^{\theta_v} \sin^2\eta \cos\eta \, d\eta \quad \text{(B6)}$$

that is,

$$I_{zz} = \frac{4\pi}{3}\sin^3\theta_v \quad \text{(B6)}$$

Also, by symmetry, $I_{yy} = I_{xx}$ and the off-diagonal terms, I_{xy}, I_{yz}, and I_{zx}, are all zero. We have

$$I_{xx} = (K/3)[1 + (1/2)\cos^2\theta_v] = I_{yy} \quad \text{(B7)}$$

and

$$I_{zz} = (K/3)\sin^2\theta_v \quad \text{(B8)}$$

so

$$I_{xx} + I_{yy} + I_{zz} = K \quad \text{(B9)}$$

The moment matrix is diagonal and so I_{xx}, I_{yy}, and I_{zz} are the three eigenvalues. The condition number is the ratio of the largest to the smallest or

$$\frac{1 + (1/2)\cos^2\theta_v}{\sin^2\theta_v} \quad \text{(B10)}$$

These results give us a quick estimate of the ensemble averages of the intergrals is $\mathbf{s}\mathbf{s}^T$ and $\mathbf{v}\mathbf{v}^T$. To do better, we have to take into account the actual distribution of $\mathbf{s}$ and $\mathbf{v}$ in the permissible band.

B.2 Ensemble Average of the Integral of ss^T

It is convenient to use polar coordinates in the case of a circular image. We have

$$\begin{aligned} x &= r\cos\theta \qquad -\pi \leqslant \theta \leqslant +\pi \\ y &= r\sin\theta \qquad 0 \leqslant r \leqslant r_v \end{aligned} \quad \text{(B11)}$$

Similarly, we may use polar coordinates for the brightness gradient

$$\begin{aligned} E_x &= \rho\cos\phi \qquad -\pi \leqslant \phi \leqslant +\pi \\ E_y &= \rho\sin\phi \qquad 0 \leqslant \rho \end{aligned} \quad \text{(B12)}$$

Let the probability of seeing a brightness gradient with magnitude lying between ρ and $\rho + \delta\rho$ be $2\pi\rho P(\rho)\delta\rho$. Now

$$\mathbf{s} = \begin{pmatrix} -E_x \\ -E_y \\ xE_x + yE_y \end{pmatrix}$$

$$= \rho \begin{pmatrix} -\cos\phi \\ -\sin\phi \\ r\cos(\theta - \phi) \end{pmatrix} \quad \text{(B13)}$$

Consequently

$$\mathbf{s}\cdot\mathbf{s} = \rho^2[1 + r^2\cos^2(\theta - \phi)] \quad \text{(B14)}$$

Consider first the integral of $\mathbf{s}\cdot\mathbf{s}$:

$$\int_{-\pi}^{\pi}\int_0^{r_v} (\mathbf{s}\cdot\mathbf{s}) r \, dr \, d\theta \quad \text{(B15)}$$

To obtain the desired ensemble average we integrate over ρ and ϕ as follows:

$$\int_{-\pi}^{\pi}\int_0^{\infty} P(\rho)\left[\int_{-\pi}^{\pi}\int_0^{r_v} (\mathbf{s}\cdot\mathbf{s}) r \, dr \, d\theta\right] \rho \, d\rho \, d\phi \quad \text{(B16)}$$

This integral can be split into two parts:

$$\int_0^{\infty} \rho^3 P(\rho) \, d\rho \int_{-\pi}^{\pi} d\phi \int_0^{r_v} r \, dr \int_{-\pi}^{\pi} d\theta = 2\pi^2 P_2 r_v^2 \quad \text{(B17)}$$

and

$$\int_0^{\infty} \rho^3 P(\rho) \, d\rho \int_{-\pi}^{\pi} d\phi \int_0^{r_v} r^3 \, dr \int_{-\pi}^{\pi} \cos^2\theta' \, d\theta' = \frac{\pi^2}{2} P_2 r_v^4 \quad \text{(B18)}$$

where $\theta' = (\theta - \phi)$ and

$$P_2 = \int_0^{\infty} \rho^3 P(\rho) \, d\rho \quad \text{(B19)}$$

We note in passing that P_2 is a measure of the average squared magnitude of the brightness gradient. Combining the two parts above we find that

$$\iint \mathbf{s} \cdot \mathbf{s}\, dx\, dy = 2\pi^2 P_2 r_v^2 (1 + r_v^2/4) \qquad \text{(B20)}$$

Similarly we obtain

$$\begin{aligned} I_{xx} &= \int_{-\pi}^{\pi}\int_{0}^{\infty} P(\rho) \\ &\quad \left[\int_{-\pi}^{\pi}\int_{0}^{r_v} \rho^2 \cos^2\phi\, r\, dr\, d\theta\right] \rho\, d\rho\, d\phi \\ &= \int_{0}^{\infty} \rho^3 P(\rho)\, d\rho \\ &\qquad \int_{-\pi}^{\pi} \cos^2\phi\, d\phi \int_{0}^{r_v} r\, dr \int_{-\pi}^{\pi} d\theta \\ &= \pi^2 P_2 r_v^2 \end{aligned} \qquad \text{(B21)}$$

and

$$\begin{aligned} I_{zz} &= \int_{-\pi}^{\pi}\int_{0}^{\infty} P(\rho)\left[\int_{-\pi}^{\pi}\int_{0}^{r_v} \rho^2 r^2 \right. \\ &\qquad \left. \cos^2(\theta - \phi) r\, dr\, d\theta\right] \rho\, d\rho\, d\phi \\ &= \int_{0}^{\infty} \rho^3 P(\rho)\, d\rho \int_{-\pi}^{\pi} d\phi \\ &\qquad \int_{0}^{r_v} r^3\, dr \int_{-\pi}^{\pi} \cos^2\theta'\, d\theta' \\ &= \frac{\pi^2}{4} P_2 r_v^4 \end{aligned} \qquad \text{(B22)}$$

while $I_{yy} = I_{xx}$.

The moment matrix is diagonal, so I_{xx}, I_{yy}, and I_{zz} are the eigenvalues. The condition number is the ratio of the largest to the smallest or

$$\begin{aligned} &\frac{2}{r_v^2} \quad \text{for} \quad r_v^2 \leqslant 2 \quad \text{and} \\ &\frac{r_v^2}{2} \quad \text{for} \quad r_v^2 \geqslant 2 \end{aligned} \qquad \text{(B23)}$$

This result does not depend on P_2, as stated earlier.

B.3 Ensemble Average of the Integral of $\mathbf{v}\mathbf{v}^T$

Here we proceed much as in the previous section with

$$\begin{aligned} \mathbf{v} &= \begin{bmatrix} +E_v + y(xE_x + yE_y) \\ -E_x - x(xE_x + yE_y) \\ yE_x - xE_y \end{bmatrix} \\ &= \rho \begin{bmatrix} +\sin\phi + r^2 \sin\theta \cos(\theta - \phi) \\ -\cos\phi - r^2 \cos\theta \cos(\theta - \phi) \\ r \sin(\theta - \phi) \end{bmatrix} \end{aligned} \qquad \text{(B24)}$$

and

$$\mathbf{v} \cdot \mathbf{v} = \rho^2(1 + r^2)[1 + r^2 \cos^2(\theta - \phi)] \qquad \text{(B25)}$$

which follows from $\mathbf{v} = \mathbf{s} \times \mathbf{r}$, $\mathbf{s} \cdot \mathbf{r} = 0$, $\mathbf{r} \cdot \mathbf{r} = 1 + r^2$ and

$$\mathbf{s} \cdot \mathbf{s} = \rho^2[1 + r^2 \cos^2(\theta - \phi)] \qquad \text{(B26)}$$

After some tedious manipulations, similar to those in the last section, we find

$$\iint \mathbf{v} \cdot \mathbf{v}\, dx\, dy = \frac{\pi^2}{2} P_2 r_v^2 \left(1 + \frac{3r_v^2}{4} + \frac{r_v^4}{6}\right) \qquad \text{(B27)}$$

and

$$I_{xx} = \pi^2 P_2 r_v^2 \left(1 + \frac{r_v^2}{2} + \frac{r_v^4}{6}\right) = I_{yy} \qquad \text{(B28)}$$

and

$$I_{zz} = \pi^2 P_2 \left(\frac{r_v^4}{2}\right) \qquad \text{(B29)}$$

Again, the matrix is diagonal and so I_{xx}, I_{yy}, and I_{zz} are the eigenvalues. The condition number is just

$$\frac{2}{r_v^2} + 1 + \frac{r_v^2}{3} \qquad \text{(B30)}$$

which is independent of P_2 once more.