MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 939 January, 1987

A Direct Method for Locating the Focus of Expansion

Shahriar Negahdaripour
Berthold K.P. Horn

Abstract: We address the problem of recovering the motion of a monocular observer relative to a rigid scene. We do not make any assumptions about the shapes of the surfaces in the scene, nor do we use estimates of the optical flow or point correspondences. Instead, we exploit the spatial gradient and the time rate of change of brightness over the whole image and explicitly impose the constraint that the surface of an object in the scene must be in front of the camera for it to be imaged.

Key Words: Focus of Expansion, Optical Flow, Motion Field, Passive Navigation, Structure from Motion, Correspondence Problem.

© Massachusetts Institute of Technology, 1987

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the National Science Foundation under grant No. DMC85-11966 and in part by Defense Advanced Research Projects Agency under Office of Naval Research contract No. N00014-75-C-0643.

1 Introduction

One of the primary tasks of a computer vision system is to reconstruct, from two-dimensional images, such three-dimensional properties of a scene as the shape, motion, and spatial arrangement of objects. In monocular vision, an important goal is to recover, from time-varying images, the relative motion between a viewer and the environment, as well as the so-called structure of the environment. The structure of the environment is usually taken to be collection of the relative distances of points on the surfaces in the scene from the viewer. In theory at least, absolute distances can be determined from the image data if the motion is known.

Three types of approaches, discrete, differential, and least-squares have been pursued in most of the earlier work in motion vision. Discrete methods establish correspondences between images of a point in the scene in a sequence of images in order to recover motion (see for example, Prazdny [1979], Roach & Aggarwal [1980], Longuet-Higgins [1981], Barnard & Thompson [1980], Mitiche [1984], Tsai & Huang [1984]). In the differential approach, the optical flow, an estimate of the velocity of the image of a point in the scene, as well as the first and second partial derivatives of the optical flow, are used to determine motion and the local structure of the surface of the scene (see Longuet-Higgins & Prazdny [1980], Waxman & Ullman [1983]). In the least-squares approach, motion parameters are found that are most consistent with the optical flow over the entire image (see Ballard and Kimball [1981], Bruss & Horn [1983], Adiv [1985]).

Amongst the shortcomings of the discrete methods are that they require the solution of point correspondence problems and that they are not very robust, since information from a small portion of the image is used. To overcome the first problem, methods have been suggested that only require line or contour correspondence (see for example, Tsai [1983], Yen & Huang [1983], and Aloimonos & Basu [1986]); however, the computation is still based on information in a relatively small portion of the image. Differential methods exploit only local information and, therefore, are sensitive to inherent ambiguities in the solution when data is noisy. In fact, since these methods essentially work with a vanishingly small field of view, they are unable to estimate all components of the motion (Horn & Weldon [1986]). Methods based on the least-squares approach are more robust, however, they make use of the unrealistic assumption that the computed optical flow is a good estimate of the true motion field. Also, the iterative algorithms for estimating an optical flow field are computationally expensive. This motivates investigation of methods that directly use brightness derivative information at every image point. Several special cases of the motion vision problem have already been addressed using this notion.

Negahdaripour [1986] investigates the problem of recovering motion directly from the time-varying image. He shows that the solution can be determined easily in certain special cases. For example, when the motion is purely rotational, one only has to solve three linear equations in three unknowns (Aloimonos & Brown [1984] apparently first reported a solution to this problem, followed by Horn & Weldon [1986], who also studied

its robustness). Another special case of interest is the one where the depth values of some points are known. The depth values at six image points are sufficient to recover the translational and rotational motion from six linear equations. In practice, to reduce the influence of measurement errors, the information from as many image points as possible should be used. If the variation in depth is negligible in comparison to the absolute distance of points on the surface, it can be assumed that the points are located at essentially at the same distance from the viewer, that is, the scene lies in a frontal plane. In this case, Negahdaripour [1986] shows that the six translational and rotational motion parameters can also be obtained from six linear equations.

When the scene is planar (but not necessarily a frontal plane) the results of the least-squares analysis of Negahdaripour & Horn [1987] can be applied. This approach leads to both iterative and closed-form solutions. Negahdaripour [1986] further presents iterative and closed-form solutions for quadratic surfaces. Through examples using synthetic data, he shows that the iterative method gives a better estimate than the analytical one in the case of quadratic surface, and that it is not as robust as the method that applies in the case of planar surfaces. He also addresses the lack of robustness of certain analytical methods published in the computer vision literature for recovering motion, and explains why the iterative method of Negahdaripour & Horn [1987] for planar surfaces happens to give the same estimate as the analytical method. Finally, Horn & Weldon [1986] give a treatment of several direct methods when the motion it is purely translational or purely rotational.

In this paper, we present a direct method for recovering the motion of a viewer without making any assumptions about the shapes of the surfaces in the scene. We only impose a simple physical constraint: Depth must be positive. That is, a point on a surface must be in front of the viewer in order for it to be imaged. Unfortunately, the problem is still rather difficult to solve when motion consists of both translation and rotation of the viewer. We therefore first address the problem of a translating observer and present two examples. We then explain how our method can be extended if the motion involves rotation as well as translation of the viewer. The general method requires considerably more computation than the special one, and the solution may not be unique given noisy data. This is because of the inherent difficulty in distinguishing between rotation about some axis parallel to the image plane and translation along an axis that is perpendicular to this rotational axis (Jerian & Jain [1983]). This problem is most apparent when the field of view is small (Horn & Weldon [1986]). We demonstrate some of these problems by means of an example.

2 Brightness Change Constraint Equation

A viewer-centered coordinate system is chosen, the image is formed on a plane perpendicular to the viewing direction (which is along the z-axis), and the focal length is assumed to be unity, without loss of generality (Figure 1). Let $\mathbf{R} = (X, Y, Z)^T$ be a point in the

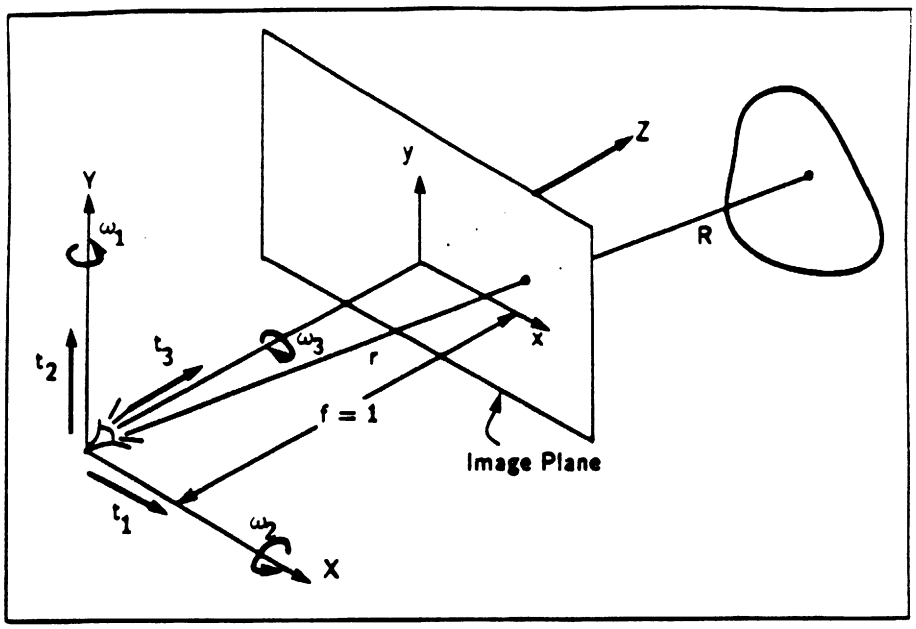

Figure 1. Viewer-centered coordinate system and perscpective projection.

scene that projects onto the point $\mathbf{r} = (x, y, 1)^T$ in the image. Assuming perspective projection, we have

$$\mathbf{r} = \frac{1}{\mathbf{R} \cdot \hat{\mathbf{z}}} \mathbf{R},$$

where $Z = \mathbf{R} \cdot \hat{\mathbf{z}}$ is the distance of the point $\mathbf{R}$ from the viewer, measured along the optical axis. This is referred to as the *depth* of the point.

Now, suppose the viewer moves with translational and rotational velocities $\mathbf{t}$ and $\boldsymbol{\omega}$ relative to a stationary scene. Then a points in the scene appears to move with respect to the viewer with velocity

$$\mathbf{R}_t = -\mathbf{R} \times \boldsymbol{\omega} - \mathbf{t}.$$

The corresponding point in the image moves with velocity (Negahdaripour & Horn [1987])

$$\mathbf{r}_t = -\Big(\hat{\mathbf{z}} \times \big(\mathbf{r} \times (\mathbf{r} \times \boldsymbol{\omega} - \frac{1}{\mathbf{R} \cdot \hat{\mathbf{z}}} \mathbf{t})\big)\Big).$$

The velocities of all image points, given by the above equation, taken collectively, define a two-dimensional vector field that we call the *image motion field.* This has also at times been referred to as the optical flow field (see Horn [1986] for a discussion of the distinction between optical flow and the motion field).

The brightness of the image of a patch on the surface of some object may change for a number of different reasons including changes in illumination or shading. Image brightness changes will, however, be dominated by the effects of the relative motion of the scene and the observer provided that the surfaces of the objects have sufficient texture

and the lighting conditions vary slowly enough both spatially and with time. In this case, brightness changes due to changing surface orientation and changing illumination can be neglected and we may assume that the brightness of a small patch on a surface in the scene remains essentially constant as it moves. Let $E(\mathbf{r}, t)$ denote the brightness of an image point $\mathbf{r}$ at time t. Then the constant brightness assumption allows us to write

$$\frac{d}{dt}E(\mathbf{r}, t) = E_{\mathbf{r}} \cdot \mathbf{r}_t + E_t = 0,$$

where E_t and $E_{\mathbf{r}} = (E_x, E_y, 0)^T$ denote the temporal and spatial derivatives of brightness respectively.

If we substitute the formula for the motion field into this equation we arrive at the *brightness change constraint equation* for the case of rigid body motion (Negahdaripour & Horn [1987]),

$$E_t + \mathbf{v} \cdot \boldsymbol{\omega} + \frac{1}{\mathbf{R} \cdot \hat{\mathbf{z}}} \mathbf{s} \cdot \mathbf{t} = 0,$$

where, for conciseness, we have defined

$$\mathbf{s} = (E_{\mathbf{r}} \times \hat{\mathbf{z}}) \times \mathbf{r} \quad \text{and} \quad \mathbf{v} = \mathbf{r} \times \mathbf{s}.$$

In component form, $\mathbf{s}$ and $\mathbf{v}$ are given by

$$\mathbf{s} = \begin{pmatrix} -E_x \\ -E_y \\ xE_x + yE_y \end{pmatrix} \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} xyE_x + (y^2+1)E_y \\ -(x^2+1)E_x - xyE_y \\ yE_x - xE_y \end{pmatrix}.$$

A useful immediately consequence of the way the vectors $\mathbf{r}$, $\mathbf{s}$, and $\mathbf{v}$ are defined is that they form an orthogonal triad, that is

$$\mathbf{r} \cdot \mathbf{s} = 0, \quad \mathbf{r} \cdot \mathbf{v} = 0, \quad \text{and} \quad \mathbf{s} \cdot \mathbf{v} = 0.$$

Note that the brightness change constraint equation is not altered if we scale both $Z = \mathbf{R} \cdot \hat{\mathbf{z}}$ and $\mathbf{t}$ by the same factor, k say. We conclude that we can determine only the direction of translation and the relative depth of points in the scene; this well-known ambiguity is here referred to as the *scale-factor ambiguity* of motion vision.

The brightness change constraint equation shows how the motion of the observer, $\{\boldsymbol{\omega}, \mathbf{t}\}$, and the depth of a point in the scene, Z, impose a constraint on the spatial and temporal derivatives of the image brightness corresponding to a point in the scene. Unfortunately, we cannot recover both depth and motion using this constraint equation alone. To show this, we solve the constraint equation for Z, in terms of the true motion parameters $\{\boldsymbol{\omega}, \mathbf{t}\}$, to obtain

$$Z = -\frac{\mathbf{s} \cdot \mathbf{t}}{c + \mathbf{v} \cdot \boldsymbol{\omega}}.$$

Now, for an arbitrary motion $\{\omega', \mathbf{t}'\}$, depth values that satisfy the brightness change constraint equation can be determined using

$$Z' = -\frac{\mathbf{s} \cdot \mathbf{t}'}{c + \mathbf{v} \cdot \omega'},$$

(provided that the denominator is not zero). This may suggest that, for any choice of the pair $\{\omega', \mathbf{t}'\}$, we can determine depth values such that the brightness change equation is satisfied at every image point. Clearly an infinitely number of solutions is possible since the motion parameters can be chosen arbitrarily.

3 Positiveness of Depth

The depth values of points on the visible portions of a surface in the scene are constrained to be positive; that is, only points in front of the viewer are imaged. In theory, any motion pair $\{\omega', \mathbf{t}'\}$ that gives rise to negative depth values cannot be the correct one. Thus, the problem is to determine the pair $\{\omega, \mathbf{t}\}$ that gives rise to positive depth values ($Z > 0$) over the whole image. One may well ask whether there is a unique solution; that is, given that the brightness change equation is satisfied for the motion $\{\omega, \mathbf{t}\}$ and the surface $Z > 0$, is there another motion $\{\omega', \mathbf{t}'\}$ and another surface $Z' > 0$ that satisfies the brightness change equation at every point in the image? In general, this is possible since, for example, an image of uniform brightness could correspond to an arbitrary uniform surface moving in an arbitrary way. Hence, the brightness gradients (or lack of brightness gradients) can conspire to make the problem highly ambiguous. In practice, given a sufficiently textured scene, it is more likely that we have the opposite problem: There is no solution because of noise in the images and the error in estimating brightness derivatives; that is, every possible set of motion parameters, including the correct ones, lead to some negative depth values. So we have to invent a method for selecting a solution that comes closest to being consistent with the image data.

The problem is rather difficult when both rotation and translation are unknown. Therefore, we first restrict attention to the special case when either rotation is zero or is at least known. We then show how the procedure may be extended to deal with the general case.

4 Pure Translation or Known Rotation

Suppose the rotational component of motion is known. Then we can write the brightness change equation in the form

$$\tilde{c} + \frac{1}{Z}(\mathbf{s} \cdot \mathbf{t}) = 0,$$

where $\tilde{c} = c + \mathbf{v} \cdot \omega$. For simplicity, we will from now on write c where $\tilde{c}$ should appear. The problem is still under-constrained if we restrict ourselves to the brightness change constraint equation alone. At each point, we have one constraint equation. Given n

image points we have therefore n constraint equations, but $n + 2$ unknowns (n depth values and two independent parameters required to specify the direction of translation). Most of these "solutions," however, are inconsistent with the physical constraint that $Z > 0$ for every point on the visible parts of the surfaces imaged. If we impose this additional constraint we may have many, only one, or no solution depending on the variety of brightness gradient directions in the image and the amount of noise in the data, as mentioned earlier. Note that we need to use constraint from a whole image region since the problem remains underconstrained if we restrict ourselves to information from a small number of points or a line.

Before we discuss the general method, we show how a simplified constraint can be used to recover motion provided that so-called stationary points can be identified. We then present a more general procedure for locating the *focus of expansion* (FOE) and consequently the direction of motion.

4.1 Stationary Points

An image point where $c = 0$ will be referred to as a *stationary point* (Horn & Weldon [1986]). In the case of pure translation, ($\boldsymbol{\omega} = \mathbf{0}$), a stationary point is one where the time derivative of brightness, E_t, is zero. In order to exclude regions of uniform brightness from consideration, we restrict attention to points with non-zero brightness gradient ($E_\mathbf{r} \neq \mathbf{0}$). When $c = 0$, the brightness change equation reduces to

$$\frac{1}{Z}(\mathbf{s} \cdot \mathbf{t}) = 0,$$

and, if the depth is finite, this immediately implies that

$$(\mathbf{s} \cdot \mathbf{t}) = 0.$$

(We assume a finite depth range here—background regions at essentially infinite depth have to be detected and removed—see Horn & Weldon [1986].) Since Z drops out of the equation, we conclude that the depth value cannot be computed at a stationary point. These points, however, do provide strong constraints on the location of the FOE.

In fact, with perfect data, just two non-parallel vectors $\mathbf{s}_1$ and $\mathbf{s}_2$, at two stationary points, provide enough information to recover the translational vector $\mathbf{t}$. We note that $\mathbf{t}$ is perpendicular to both $\mathbf{s}_1$ and $\mathbf{s}_2$ and so must be parallel to the cross-product of these two vectors. That is,

$$\mathbf{t} = k\,(\mathbf{s}_1 \times \mathbf{s}_2),$$

where k is some constant that cannot be determined from the image brightness gradients alone because of the scale-factor ambiguity.

This approach can be interpreted directly in terms of quantities in the image plane: The brightness gradient at a stationary point is orthogonal to the direction to the FOE,

or, equivalently, the tangent of the iso-brightness contour at a stationary point passes through the FOE. Intersecting the tangents of the iso-brightness contours at two different stationary points allows us to determine the FOE (see the appendix for more details).

In practice it will be better to apply least-squares techniques to information from many stationary points. Because of noise in the images, as well as quantization error, the constraint equation ($\mathbf{s} \cdot \mathbf{t} = 0$) will not be satisfied exactly. This suggests minimizing the sum of the squares of the errors at every stationary point; that is, we minimize

$$\sum_{i=1}^{n} (\mathbf{s}_i \cdot \mathbf{t})^2 = \mathbf{t}^T \Big(\sum_{i=1}^{n} \mathbf{s}_i \mathbf{s}_i^T \Big) \mathbf{t}.$$

(In the above we have used the identity $\mathbf{s} \cdot \mathbf{t} = \mathbf{s}^T \mathbf{t}$.) Note that the resulting quadratic form can not be negative.

Because of the scale-factor ambiguity we can only determine the direction of $\mathbf{t}$, not its magnitude, so we have to impose the constraint $\|\mathbf{t}\|^2 = 1$ (otherwise we immediately get the trivial solution $\mathbf{t} = \mathbf{0}$). This leads to a constrained optimization problem. We can create an equivalent unconstrained optimization problem, with a closed-form solution, by introducing a Lagrange multiplier. We find that we now have to minimize

$$J = \mathbf{t}^T \Big(\sum_{i=1}^{n} \mathbf{s}_i \mathbf{s}_i^T \Big) \mathbf{t} + \lambda (1 - \mathbf{t}^T \mathbf{t}).$$

The necessary conditions for stationary values of J are

$$\frac{\partial J}{\partial \mathbf{t}} = \mathbf{0} \quad \text{and} \quad \frac{\partial J}{\partial \lambda} = 0.$$

Executing the indicated differentiations we arrive at

$$\Big(\sum_{i=1}^{n} \mathbf{s}_i \mathbf{s}_i^T \Big) \mathbf{t} = \lambda \mathbf{t} \quad \text{and} \quad \mathbf{t}^T \mathbf{t} = 1.$$

This is an eigenvalue-eigenvector problem; that is, $\{\mathbf{t}, \lambda\}$ is an eigenvector-eigenvalue pair of the 3×3 matrix

$$\sum_{i=1}^{n} \mathbf{s}_i \mathbf{s}_i^T.$$

This real symmetric matrix generally will have three eigenvalues and these eigenvalues will be non-negative since the quadratic form we started off with was non-negative definite. It is easy to see that J is minimized by the eigenvector associated with the smallest eigenvalue, since substitution of the solution yields

$$J = \mathbf{t}^T (\lambda \mathbf{t}) + \lambda (1 - \mathbf{t}^T \mathbf{t}) = \lambda \mathbf{t}^T \mathbf{t} + \lambda - \lambda \mathbf{t}^T \mathbf{t} = \lambda.$$

It should be noted that with just two stationary points, the 3×3 matrix has rank two since it is the sum of two dyadic products. The solution then is the eigenvector corresponding to the zero eigenvalue. Geometrically, this is the vector normal to the plane formed by $\mathbf{s}_1$ and $\mathbf{s}_2$, as discussed earlier. By the way, if $\mathbf{t}$ is an eigenvector, so is $-\mathbf{t}$. While these two possibilities correspond to the same FOE, it may be desirable to distinguish between them. This can be done by choosing the one that makes most depth values positive rather than negative (see Horn & Weldon [1986]).

The least-squares method just described can be interpreted in terms of quantities in the image plane also. At each stationary point, the tangent to the iso-brightness contour provides us with a line on which the FOE would lie if there was no measurement error. In practice these lines will not intersect in a common point due to noise. The position of the FOE may then be estimated by finding the point with the minimum (weighted) sum of squares of distances from the lines (see the appendix for more details).

4.2 Constraints Imposed by Brightness Gradient Vectors

We first assume that two translational motions and two surfaces satisfy the brightness change equation; that is, we have

$$c + \frac{1}{Z}(\mathbf{s} \cdot \mathbf{t}) = 0 \quad \text{and} \quad c + \frac{1}{Z'}(\mathbf{s} \cdot \mathbf{t}') = 0.$$

Here, $\{Z > 0, \mathbf{t}\}$ denotes the true solution and $\{Z' > 0, \mathbf{t}'\}$ denotes a spurious (or assumed) solution. We will show that we must have $Z = kZ'$ and $\mathbf{t} = k\,\mathbf{t}'$, for some non-zero constant k, provided that there is sufficient texture and that we consider a large enough region of the image. This means that the solution is unique up to the scale-factor ambiguity.

Solving for Z and Z' we obtain

$$Z = -\frac{1}{c}(\mathbf{s} \cdot \mathbf{t}) \quad \text{and} \quad Z' = -\frac{1}{c}(\mathbf{s} \cdot \mathbf{t}').$$

The depth value cannot be computed at a point where $c = 0$; that is, at a stationary point. We already know how to exploit the information at these points and so exclude them from further consideration, that is, we assume from now on that $c \neq 0$.

Since Z is the true solution, we are guaranteed that $Z > 0$. If $\{Z', \mathbf{t}'\}$ is to be an acceptable solution, we must also have $Z' > 0$ and so

$$ZZ' = \frac{1}{c^2}(\mathbf{s} \cdot \mathbf{t})(\mathbf{s} \cdot \mathbf{t}') > 0.$$

Now the focus of expansion (FOE) is the intersection of the translational velocity vector $\mathbf{t}$ and the image plane $z = 1$. It lies at

$$\tilde{\mathbf{t}} = \frac{\mathbf{t}}{\mathbf{t} \cdot \hat{\mathbf{z}}},$$

provided that $\mathbf{t} \cdot \hat{\mathbf{z}} \neq 0$ (otherwise, it is at infinity in the direction given by the vector $\mathbf{t}$). We can similarly write

$$\tilde{\mathbf{t}}' = \frac{\mathbf{t}'}{\mathbf{t}' \cdot \hat{\mathbf{z}}},$$

for the focus of expansion corresponding to the assumed translational velocity $\mathbf{t}'$ (provided again that $\mathbf{t}' \cdot \hat{\mathbf{z}} \neq 0$). We can write $\mathbf{s} = (E_\mathbf{r} \times \hat{\mathbf{z}}) \times \mathbf{r}$ in the form

$$\mathbf{s} = (\mathbf{r} \cdot E_\mathbf{r})\hat{\mathbf{z}} - (\mathbf{r} \cdot \hat{\mathbf{z}})E_\mathbf{r},$$

and, noting that $\mathbf{r} \cdot \hat{\mathbf{z}} = 1$, we obtain

$$\mathbf{s} = (\mathbf{r} \cdot E_\mathbf{r})\hat{\mathbf{z}} - E_\mathbf{r}.$$

Therefore, we have

$$\mathbf{s} \cdot \mathbf{t} = (\mathbf{r} \cdot E_\mathbf{r})(\mathbf{t} \cdot \hat{\mathbf{z}}) - E_\mathbf{r} \cdot \mathbf{t},$$

that is,

$$\mathbf{s} \cdot \mathbf{t} = (\mathbf{t} \cdot \hat{\mathbf{z}})\left((\mathbf{r} - \tilde{\mathbf{t}}) \cdot E_\mathbf{r}\right).$$

Similarly, we obtain

$$\mathbf{s} \cdot \mathbf{t}' = (\mathbf{t}' \cdot \hat{\mathbf{z}})\left((\mathbf{r} - \tilde{\mathbf{t}}') \cdot E_\mathbf{r}\right).$$

Substituting these into the inequality $ZZ' > 0$ we arrive at

$$(\mathbf{t} \cdot \hat{\mathbf{z}})(\mathbf{t}' \cdot \hat{\mathbf{z}})\left((\mathbf{r} - \tilde{\mathbf{t}}) \cdot E_\mathbf{r}\right)\left((\mathbf{r} - \tilde{\mathbf{t}}') \cdot E_\mathbf{r}\right) > 0.$$

If $(\mathbf{t} \cdot \hat{\mathbf{z}})$ and $(\mathbf{t}' \cdot \hat{\mathbf{z}})$ have the same sign, we must have

$$\left((\mathbf{r} - \tilde{\mathbf{t}}) \cdot E_\mathbf{r}\right)\left((\mathbf{r} - \tilde{\mathbf{t}}') \cdot E_\mathbf{r}\right) > 0.$$

For convenience, we denote the term on the left-hand side of the inequality p from here on. So for $ZZ' > 0$ we must have $p > 0$. (Note that if $(\mathbf{t} \cdot \hat{\mathbf{z}})$ and $(\mathbf{t}' \cdot \hat{\mathbf{z}})$ have opposite signs, the inequality is reversed.) Without loss of generality, we assume from now on that the above constraint holds—the proof is similar in the opposite case, as we will indicate.

For $\mathbf{t}'$ to be a possible translational motion, the inequality developed above must hold for every point $\mathbf{r}$ in the image region under consideration, that is, $p > 0$. At each point, $E_\mathbf{r}$ is constrained to lie in a direction that guarantees that $\left((\mathbf{r} - \tilde{\mathbf{t}}) \cdot E_\mathbf{r}\right)$ and $\left((\mathbf{r} - \tilde{\mathbf{t}}) \cdot E_\mathbf{r}\right)$ have the same sign. In practice, a sufficiently large image region will contain some image brightness gradients that violate this constraint unless $\tilde{\mathbf{t}} = \tilde{\mathbf{t}}'$. We will estimate the probability that an arbitrarily-chosen brightness gradient will violate this constraint. This probability varies spatially and we show that there is a line segment in the image along which the probability of violating the constraint becomes one. Furthermore we exploit the distribution in the image of places where $Z' < 0$ to obtain an estimate of the true FOE.

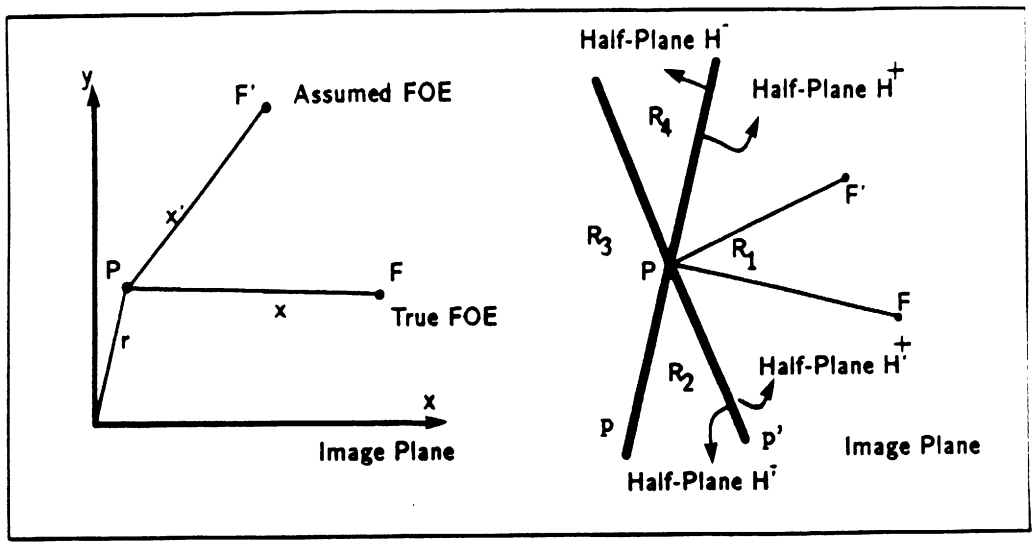

Figure 2. True and spurious (assumed) FOEs define the positive and negative half-planes.

4.2.1 Permissible and Forbidden Ranges

Define

$$\mathbf{x} = (\tilde{\mathbf{t}} - \mathbf{r}) \quad \text{and} \quad \mathbf{x}' = (\tilde{\mathbf{t}}' - \mathbf{r}).$$

The vectors $\mathbf{x}$ and $\mathbf{x}'$ represent the line segments from a point P in the image, with coordinates $\mathbf{r} = (x, y, 1)^T$, to the true and spurious FOEs, respectively. These are the line segments PF and PF' in Figure 2a. Note that the scalar product $(\mathbf{x} \cdot E_\mathbf{r})$ is positive if the angle between $\mathbf{x}$ and the brightness gradient vector at point P is less than $\pi/2$, and it is negative when the angle is greater than $\pi/2$. It is zero when $\mathbf{x}$ is orthogonal to the gradient vector. Similarly, the dot product $(\mathbf{x}' \cdot E_\mathbf{r})$ is positive, negative, or zero when the angle between $\mathbf{x}'$ and the gradient vector at point P is less than, greater than or equal to $\pi/2$.

We have, from the discussion in the previous section, the constraint $p > 0$ or,

$$(\mathbf{x} \cdot E_\mathbf{r})(\mathbf{x}' \cdot E_\mathbf{r}) > 0$$

(provided that, as assumed, $(\mathbf{t} \cdot \hat{\mathbf{z}})$ and $(\mathbf{t}' \cdot \hat{\mathbf{z}})$ have the same sign). Now suppose that we define two directions in the image plane orthogonal to the vectors $\mathbf{x}$ and $\mathbf{x}'$ as follows (see Figure 2b):

$$\mathbf{p} = \mathbf{x} \times \hat{\mathbf{z}} \quad \text{and} \quad \mathbf{p}' = \mathbf{x}' \times \hat{\mathbf{z}}.$$

The vector $\mathbf{p}$ gives the direction of a line that divides the possible directions of $E_\mathbf{r}$ into two ranges with differing signs for $(\mathbf{x} \cdot E_\mathbf{r})$. Similarly, the vector $\mathbf{p}'$ gives the direction of a line that divides the possible directions of $E_\mathbf{r}$ into two ranges with differing signs for $(\mathbf{x}' \cdot E_\mathbf{r})$

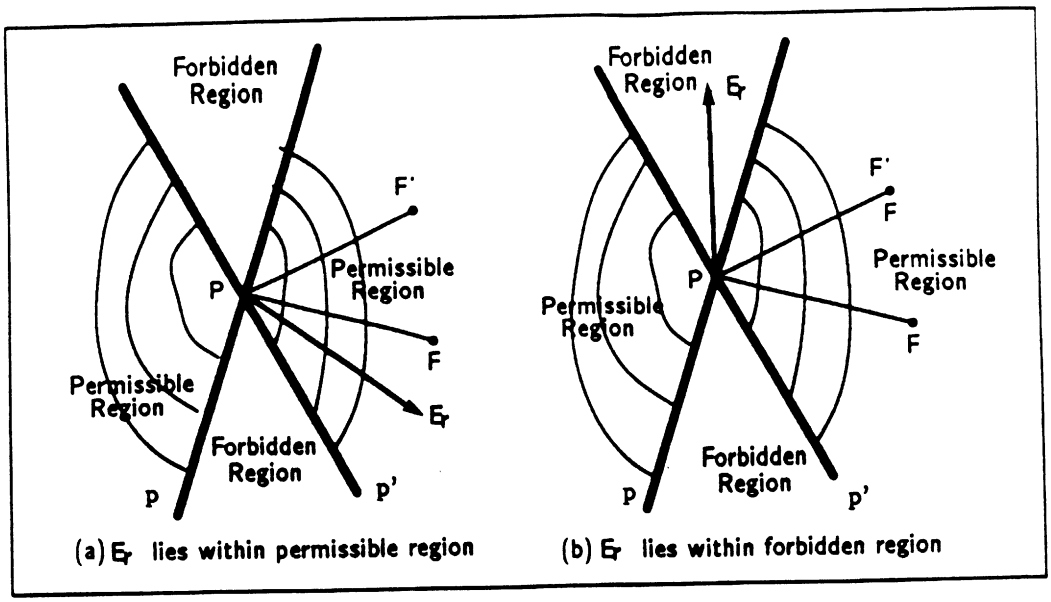

Figure 3. Permissible and forbidden ranges for the brightness gradient direction.

Unless $\mathbf{p}$ happens to be parallel to $\mathbf{p}'$, we can express an arbitrary gradient vector $E_\mathbf{r}$ in the form

$$E_\mathbf{r} = \alpha\mathbf{p} + \beta\mathbf{p}',$$

for some constants α and β. Then

$$(\mathbf{x} \cdot E_\mathbf{r})(\mathbf{x}' \cdot E_\mathbf{r}) = -\alpha\beta \left\|\mathbf{x} \times \mathbf{x}'\right\|^2.$$

We see that the product denoted p is positive when $E_\mathbf{r}$ lies between $\mathbf{p}$ and $-\mathbf{p}'$ ($\alpha > 0$ and $\beta < 0$) and when $E_\mathbf{r}$ lies between $-\mathbf{p}$ and $\mathbf{p}'$ ($\alpha < 0$ and $\beta > 0$). The union of these two ranges is called the *permissible range* for $E_\mathbf{r}$ since it leads to positive depth values. Conversely, the product will be negative when $E_\mathbf{r}$ lies between $\mathbf{p}$ and $\mathbf{p}'$ ($\alpha > 0$ and $\beta > 0$) and when $E_\mathbf{r}$ lines between $-\mathbf{p}$ and $-\mathbf{p}'$ ($\alpha < 0$ and $\beta < 0$). The union of these two ranges is called the *forbidden range* for $E_\mathbf{r}$ since it leads to negative depth values.

Denoting the half-planes separated by the line parallel to $\mathbf{p}$ by H^+ and H^- and those separated by the line parallel to $\mathbf{p}'$ by H'^+ and H'^-, we define regions $R_1, \ldots, R_4$ as follows:

$$R_1 = H^+ \cap H'^+, \quad R_3 = H^- \cap H'^-,$$

and

$$R_2 = H^+ \cap H'^-, \quad R_4 = H^- \cap H'^+.$$

We see that $R_1 \cup R_3$ is the permissible range for $E_\mathbf{r}$, because $E_\mathbf{r}$ has to lie in this region in order to satisfy the constraint $Z' > 0$. Conversely, the region $R_2 \cup R_4$ is the forbidden range for $E_\mathbf{r}$ since $Z' < 0$ when $E_\mathbf{r}$ lies in this region (see Figure 3). (Note

that the permissible range will be the region consisting of R_2 and R_4, and the forbidden range will consist of R_1 and R_3, when $(\mathbf{t} \cdot \hat{\mathbf{z}})(\mathbf{t}' \cdot \hat{\mathbf{z}}) < 0$.)

We now show that if $\widetilde{\mathbf{t}} \neq \widetilde{\mathbf{t}}'$, then the vector $E_{\mathbf{r}}$ has to lie in the forbidden region (and, therefore, $Z' < 0$) for some image points. Therefore, we must have $\widetilde{\mathbf{t}} = \widetilde{\mathbf{t}}'$ to guarantee that $Z' > 0$ for every image point. In this case, $Z = kZ'$ for some non-zero constant k. This implies that

$$\frac{1}{\mathbf{t} \cdot \hat{\mathbf{z}}}\mathbf{t} = \frac{1}{\mathbf{t}' \cdot \hat{\mathbf{z}}}\mathbf{t}'$$

or $\mathbf{t} = k\,\mathbf{t}'$. Since this means that we can recover the translational motion up to a scale factor, we conclude that the solution is unique up to the scale-factor ambiguity.

4.2.2 Distribution of Points Violating the Inequality Constraint

Suppose now that the point P lies along the line passing through F and F', which we refer to as a *FOE constraint line*. Then we have

$$\mathbf{r} = (1 - \gamma)\widetilde{\mathbf{t}} + \gamma\widetilde{\mathbf{t}}',$$

for some γ. We see that $0 < \gamma < 1$ when the point P lies on the segment between the points F and F'. Also $\gamma < 0$, if P lies on the ray emanating from F (segment FX) and $\gamma > 1$, if P it lies on the ray emanating from F' (segment $F'X'$). For points on the FOE constraint line, we have

$$\mathbf{x} = \widetilde{\mathbf{t}} - \mathbf{r} = \gamma(\widetilde{\mathbf{t}} - \widetilde{\mathbf{t}}') \quad \text{and} \quad \mathbf{x}' = \widetilde{\mathbf{t}}' - \mathbf{r} = (\gamma - 1)(\widetilde{\mathbf{t}} - \widetilde{\mathbf{t}}').$$

The product of interest to us here, p, is then given by

$$(\mathbf{x} \cdot E_{\mathbf{r}})(\mathbf{x}' \cdot E_{\mathbf{r}}) = \gamma(\gamma - 1)\big((\widetilde{\mathbf{t}} - \widetilde{\mathbf{t}}') \cdot E_{\mathbf{r}}\big)^2.$$

It is clear that p will be negative when $0 < \gamma < 1$, unless the gradient vector is orthogonal to FF' (note that FF' is the vector $(\widetilde{\mathbf{t}} - \widetilde{\mathbf{t}}')$). The point P is a stationary point if the gradient vector is orthogonal to the line FF', and we have excluded such points from consideration. This implies that, for points on the line segment FF', the depth values Z' are guaranteed to be negative (unless the point happens to be a stationary point). The product p is positive when $\gamma < 0$ or $\gamma > 1$. So in this case the depth values are guaranteed to be positive for points along the rays FX and $F'X'$, unless the point is a stationary point. (The situation is reversed when $(\mathbf{t} \cdot \hat{\mathbf{z}})(\mathbf{t}' \cdot \hat{\mathbf{z}}) < 0$, with positive depth values along FF' and negative ones along the rays FX and $F'X'$.)

A probability value can be assigned to each image point as a measure of the likelihood that $Z' < 0$ at that image point. Since $Z' < 0$ if the gradient vector lies outside the permissible range, we can conclude that the probability distribution function depends on θ, the angle between the vectors $\mathbf{x}$ and $\mathbf{x}'$, as well as on the distribution of the brightness gradient vectors.

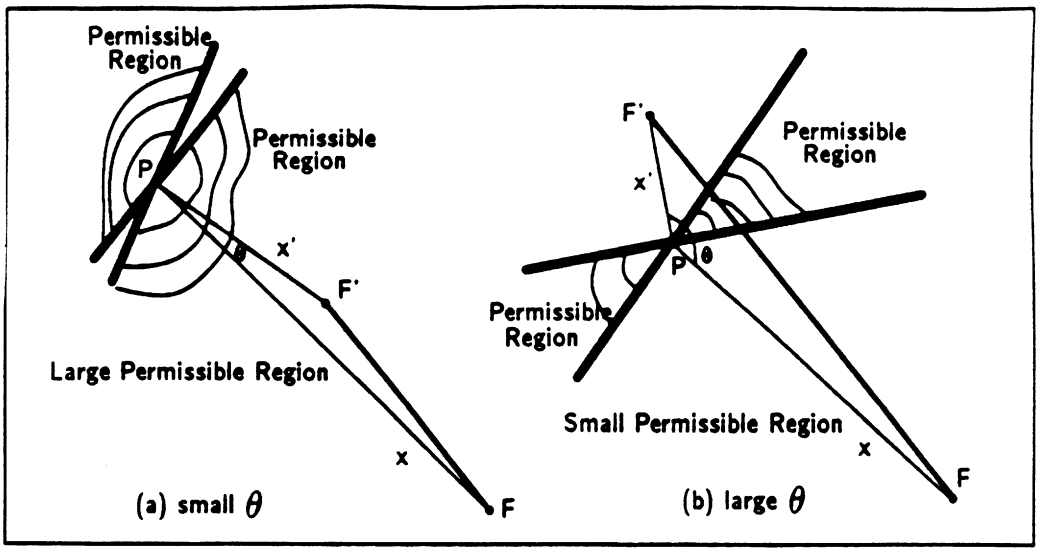

Figure 4. Relationship between the size of the permissible range and the relative position of an image point with respect to the FOE constraint line.

When θ is small, the permissible range for E_r consists of a large set of allowed directions (see Figure 4a). Therefore, the points where θ is small are likely to have positive depth values even for an incorrect translational vector $\mathbf{t}'$. These are points that are either at some distance laterally from the FOE constraint line or are in the vicinity of the two rays FX and $F'X'$.

Conversely, when θ is large, the permissible range for E_r comprises a small set of directions (see Figure 4b). Therefore, it is more likely that the brightness gradient lies outside this range, giving rise to negative depth values. In the extreme case when $\theta = \pi$ (that is, the point lies along FF') the depth values are guaranteed to be negative (unless the point is a stationary point). The forbidden range for a point on FF' contains all possible directions for E_r excluding only the line orthogonal to FF'.

Suppose that the probability distribution of the gradient vectors is independent of the image position and is rotationally symmetric; that is, all directions of the brightness gradients are equally likely. It is not difficult to see that the probability that a point in the image plane gives rise to a negative depth value is then given by

$$\text{Prob}(Z' < 0) = \frac{\theta}{\pi}.$$

A chord of a circle subtends a constant angle. It follows that the constant probability loci are circles that pass through F and F', and that there is symmetry about the FOE constraint line (see Figure 5a).

This can be shown algebraically as follows: Let Q be the projection of an image point P on XX', and let O be the midpoint of the FOE constraint line FF' (see Figure 5b). Further, let θ_1 be the angle between PF and PQ, while θ_2 is the angle between PF' and PQ, and define

$$f = \frac{1}{2}|FF'|, \quad h = |PQ|, \quad \text{and} \quad s = |OQ|.$$

Then, we have

$$\tan\theta_1 = \frac{f-s}{h} \quad \text{and} \quad \tan\theta_2 = \frac{f+s}{h}.$$

Using the identity

$$\tan\theta = \tan(\theta_1+\theta_2) = \frac{\tan\theta_1 + \tan\theta_2}{1-\tan\theta_1\tan\theta_2},$$

we arrive at

$$\tan\theta = \frac{2hf}{s^2+h^2-f^2}.$$

The locus of points with constant θ (and, equivalently, constant $\tan\theta$) is thus determined by the equation

$$s^2+h^2-f^2 = 2khf,$$

for some constant k. This can be written in the form

$$s^2+(h-fk)^2 = (1+k^2)f^2,$$

which is the equation of a circle centered at $(s,h) = (0,kf)$ that passes through $(s,h) = (f,0)$ and $(s,h) = (-f,0)$. Solving for θ we obtain

$$\theta = \tan^{-1}\frac{2hf}{s^2+h^2-f^2},$$

and, therefore,

$$\text{Prob}(Z'<0) = \frac{1}{\pi}\left(\tan^{-1}\frac{2hf}{s^2+h^2-f^2}\right).$$

For constant s (where $s < f$), this function has a maximum of 1 for $h = 0$; that is, on the line segment FF'.

To summarize, we have shown that there are points in the image that give rise to a negative depth value if an incorrect translation vector ($\mathbf{t}'$) is assumed. These points are more likely to be found in the vicinity of the line segment that connects the incorrect focus of expansion to the true one (later this is exploited to locate the true focus of expansion). As F' approaches F, the region around FF' that is likely to contain points with negative depth values shrinks in size. In the limit when F' coincides with F, all depth values become positive. (When the product $(\mathbf{t}\cdot\hat{\mathbf{z}})(\mathbf{t}'\cdot\hat{z})$ is negative, the situation is reversed. In this case, it is more likely that the points in the vicinity of FF' will give

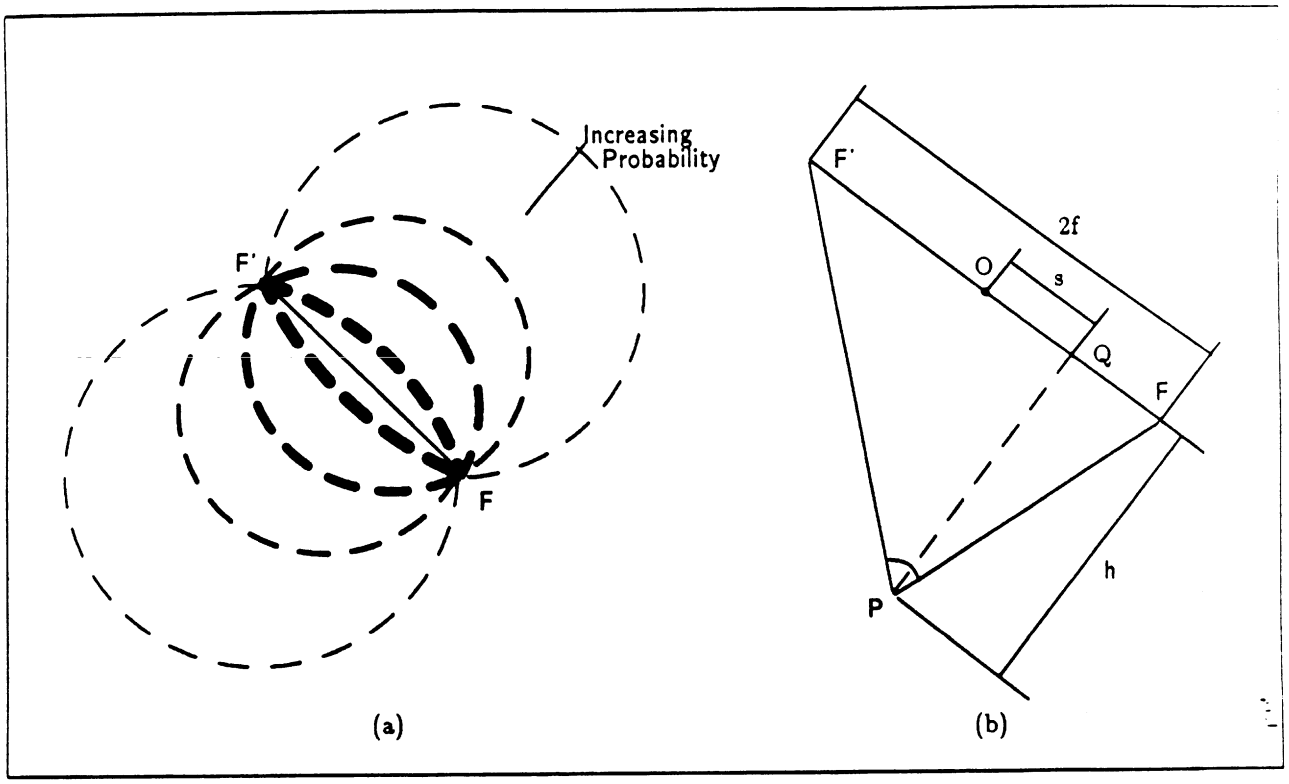

Figure 5. Constant probability loci for $Z' < 0$.

rise to positive depth values and the points along or in the vicinity of FX and $F'X'$ will give rise to positive depth values, but otherwise similar conclusions can be drawn.)

The true FOE is at infinity when $\mathbf{t} \cdot \hat{\mathbf{z}} = 0$. First consider the situation where $\mathbf{t}' \cdot \hat{\mathbf{z}} = 0$ for a spurious solution. Then we have

$$\mathbf{s} \cdot \mathbf{t} = -\mathbf{t} \cdot E_{\mathbf{r}} \quad \text{and} \quad \mathbf{s} \cdot \mathbf{t}' = -\mathbf{t}' \cdot E_{\mathbf{r}}.$$

Using these, we obtain

$$(\mathbf{s} \cdot \mathbf{t})(\mathbf{s} \cdot \mathbf{t}') = (\mathbf{t} \cdot E_{\mathbf{r}})(\mathbf{t}' \cdot E_{\mathbf{r}}).$$

The half-planes $\{H^+, H^-\}$ and $\{H'^+, H'^-\}$ are now defined by the vector $\mathbf{t}$ and $\mathbf{t}'$, instead of $\mathbf{x}$ and $\mathbf{x}'$ for the case $\mathbf{t} \cdot \hat{\mathbf{z}} \neq 0$ (that is, we need to replace $\mathbf{x}$ and $\mathbf{x}'$ by $\mathbf{t}$ and $\mathbf{t}'$, respectively, in our earlier analysis). Since these vectors are constants, we conclude that θ (in this case, this becomes the angle between the two vectors $\mathbf{t}$ and $\mathbf{t}'$) is the same for every image point. If the distribution of brightness gradient vectors is rotationally symmetric and independent of the image position, each image point can give rise to a negative depth value with probability equal to θ/π. We conclude that the depth values will be negative for some image points unless $\mathbf{t} = \mathbf{t}'$. Similar arguments can be made when only one of the FOEs lies at infinity.

4.2.3 Locating the Focus of Expansion using Gradient Vectors

It is somewhat easier to locate the FOE when it lies within the field of view than when it lies outside. We first compute the sign of the depth values using an initial estimate of the solution, $\tilde{t}'$, in the brightness change constraint equation. We then determine the cluster of negative depth values. The first method to be presented here uses the fact that the centroid of this cluster is expected to lie half-way between the true FOE and the assumed FOE. That is, because of the symmetry of the probability distribution, we have for the expected position of the centroid

$$\bar{t} = \frac{1}{2}(\tilde{t} + \tilde{t}').$$

Then the position of the FOE can be estimated using:

$$\tilde{t} = 2\bar{t} - \tilde{t}'.$$

This estimate will be biased if the border of the image cuts off a significant portion of the cluster. Nevertheless, a simple iterative scheme can be based on the above approximation that updates the estimate as follows:

$$(\tilde{t}')^{n+1} = 2(\bar{t})^n - (\tilde{t}')^n,$$

where $(\bar{t})^n$ is the centroid of the cluster of points with negative depth values obtained using the estimate $(\tilde{t}')^n$ for the FOE. The cluster will shrink at each iteration, so in subsequent computations we may restrict attention to the image region containing the major portion of the previous cluster rather than the whole of the initial image region under consideration.

Other methods we have investigated work even when the FOE is outside the field of view. Suppose that we identify at least two FOE constraint lines corresponding to two assumed FOEs. The intersection of these lines will be the estimated FOE. In practice more than two FOE constraint lines are used to reduce the effects of measurement error. These lines will no longer all intersect in a common point because of noise in the images, quantization error, and error in the estimate of brightness derivatives. It makes sense then to choose as the estimate of the true FOE the point with the least sum of squares of distances from the constraint lines.

The axis of symmetry or axis of least inertia of the clusters of positive and negative depth values for a particular assumed FOE can be chosen as the FOE constraint line. Alternatively, we may employ a direction histogram method. In this case, we need to determine the line through the assumed FOE along which the largest number of negative depth values are found on one side of the assumed FOE, and the largest number of positive depth values on the other side (Negahdaripour [1986]).

To summarize, we first choose arbitrary points in the image as estimates of the FOE. For each assumed FOE, we determine the signs of the depth values at each image point using the brightness change equation, and then the FOE constraint line using either a clustering technique or a histogram method. Finally, we use the best estimate of the common intersection of the constraint lines corresponding to the assumed FOEs as the best estimate of the FOE.

Even when the FOE is outside the field of view (including the case $\mathbf{t} \cdot \hat{\mathbf{z}} = 0$ where the FOE is at infinity) we should choose the estimates of the FOE to be in the image plane. Otherwise, the whole FOE constraint line (FF') lies outside the field of view and the clusters of negative or positive depth values cannot be properly identified. The FOE is still determined from the best estimate of the common intersection of the FOE constraint lines; however, the intersection will be outside the image plane.

The accuracy of the estimate of the location of the FOE will depend on the choice of the assumed FOEs and the resulting shape and size of the clusters of negative and positive depth values. These, in turn, depend on the distribution of the directions of the brightness gradient, that is, the "richness of texture" in the images.

5 Unknown Rotation

The problem of locating the FOE from gradient vectors has similar properties to those of the problem of estimating the location of the FOE from optical flow vectors, in the following sense: When the motion is pure translation, the FOE can be determined rather easily from the intersection of the optical flow vectors (using the fact that these vectors point toward the FOE for a departing motion and emanate from the FOE for an approaching motion). Unfortunately, these vectors do not intersect at the FOE when the rotational component is non-zero. Similarly, we expect that the FOE constraint lines will not intersect at a common point when the rotational component is non-zero (and is unknown).

An intuitively appealing approach is one that assumes some rotation vector in order to discount the contribution of the rotational component before we apply the method given for the case of a purely translational motion (Prazdny [1981] suggested this procedure to decouple the rotational and translational components of the motion field). Obviously, the estimate we obtain for the FOE is likely to be very poor if the rotational component is not chosen accurately. This, however, is exactly the behavior we want if our method is to work in the general case. That is, in order to have a distinct peak in the measure we use as a criterion for selecting the best estimate of the FOE, we should have a large error when we assume a rotation far from the correct one. The measure of "badness," denoted $e(\boldsymbol{\omega})$, can be the total square distance of the estimated FOE from the constraint lines. Then the best estimate of the motion parameters is the one that minimizes this error. It is not possible to compute this function for every possible rotation. An approach for dealing with this problem follows.

Suppose, an upper bound for each component of the rotational vector $\boldsymbol{\omega}$ is available; for example, it is known that $|\omega_i| \leq \omega_i^{\max}$. If each interval from $-\omega_i^{\max}$ to $\omega_i^{\max}$ is divided into n smaller intervals, we can restrict the search to the n^3 discrete points in $\boldsymbol{\omega}$-space. Let us denote a point in this space by $\boldsymbol{\omega}_{ijk}$ for $i, j, k = 1, 2, \ldots, n$. For each possible point in this space (that is, for each $\boldsymbol{\omega}_{ijk}$) we estimate the location of the FOE using the method given earlier. We store the value of the error $e(\boldsymbol{\omega}_{ijk})$ for the best FOE in each case. The best estimate of the rotation corresponds to a minimum of the error function. To obtain an even more accurate result we may perform a local search in the neighborhood of $\boldsymbol{\omega}_{ijk}$.

6 Selected Examples

Through examples, we show that it is possible to determine the location of the true FOE from the distribution of the clusters of positive and negative depth values around the assumed FOE. In these examples, we have used synthetic data so that the underlying motion is known exactly. The focal length is assumed to be unity and the image plane is a unit square divided into 64 rows of 64 picture cells. The half-angle of the field of view of is thus $\tan^{-1} 0.5 \approx 27°$. The positive x-axis points towards the right and the positive y-axis points downward. Positive depth values and the spatial brightness derivatives were chosen randomly. The depth values vary in a range of one to nine units. The time derivative $E_t = c$ of image brightness was computed using the brightness change constraint equation,

$$c = -\big(\mathbf{v} \cdot \boldsymbol{\omega} + \frac{1}{Z}(\mathbf{s} \cdot \mathbf{t})\big).$$

To simulate the effect of noise, random noise was added to both $E_\mathbf{r}$ and $c = E_t$.

6.1 Example One: Focus of Expansion in the Image

In this example, we consider an observer approaching a scene; the motion parameters are $\boldsymbol{\omega} = (0, 0, 0)^T$ and $\mathbf{t} = (0, 0, 1)^T$; that is, there is no rotation and the focus of expansion is at the center of the image plane. Figure 6 shows the regions of negative (white) and positive (black) depth values for several assumed FOEs. The diagrams in columns one through four show the results when the added noise has a mean of about 20%, 40%, 60%, and 80%, respectively. These plots show that the negative and positive depth values form symmetric clusters with respect to the line from the assumed FOE to the true FOE. Using these maps, it is possible to estimate the location of the true FOE with good accuracy even when there is as much as 80% noise in the data.

6.2 Example Two: Focus of Expansion at Infinity

In this example, $\boldsymbol{\omega} = (.1, 0, 0)^T$ and $\mathbf{t} = (0, -1, 0)^T$, that is, the focus of expansion is at infinity along the negative y-axis. Here the rotational component is non-zero but assumed

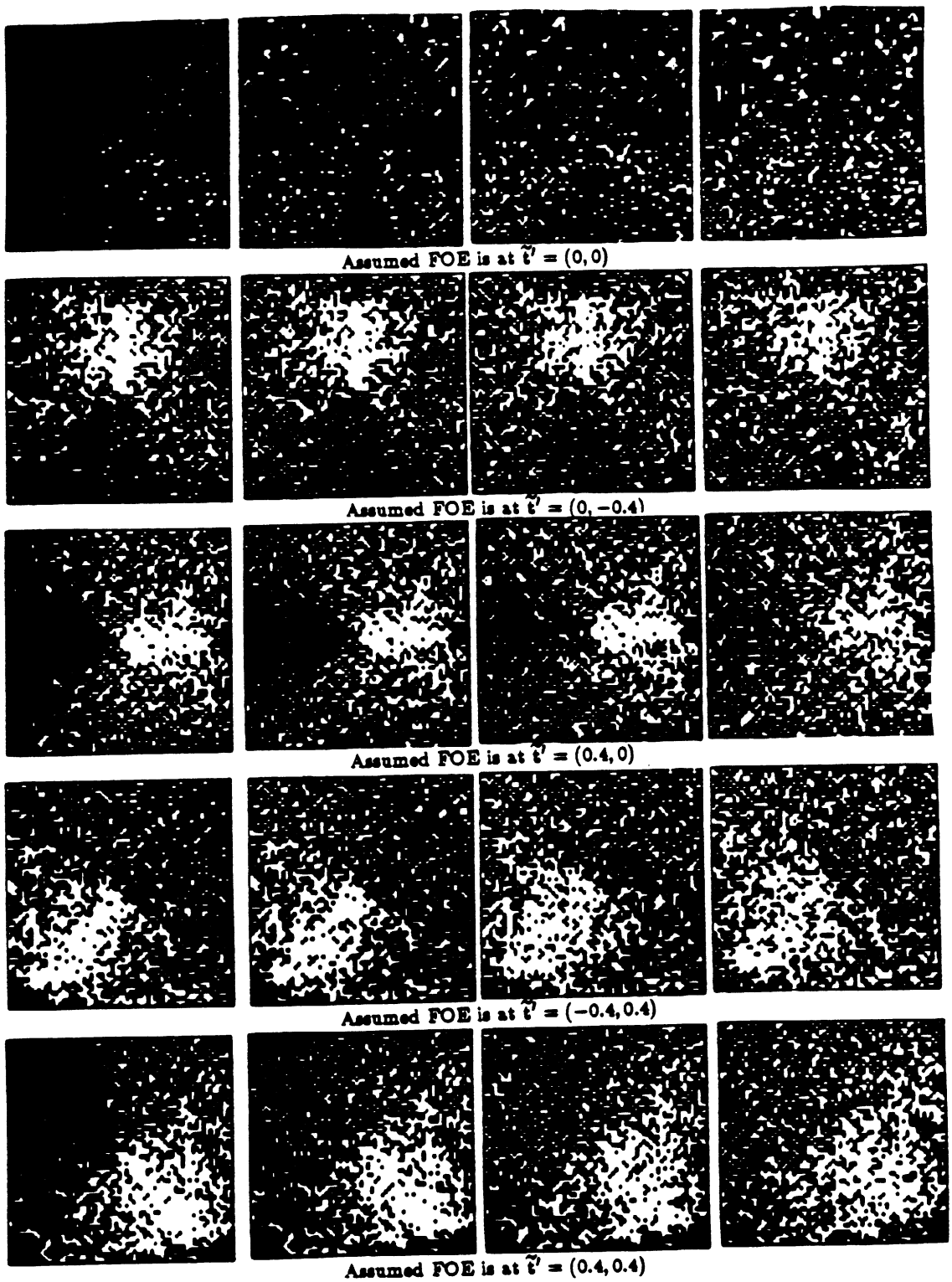

Figure 6. Regions with positive (black) and negative (white) depth values when noise is added to brightness derivatives; Example one, focus of expansion in the field of view. See text for explanation.

known. Figure 7 shows the regions of negative (white) and positive (black) depth values for several assumed FOEs with random noise added to the brightness derivatives. The diagrams in columns one through four show the results when the added noise has a mean of about 20%, 40%, 60%, and 80%, respectively. We can determine the direction toward the true FOE at infinity with good accuracy with as much as 40% noise in the data. The results deteriorate to some extend with 60% noise for some of the assumed FOEs. With 80% noise in the data, it is hard to define the clusters of positive and negative depth values and so the FOE cannot be located accurately.

6.3 Example Three: Unknown Rotation

In this example, we investigate the sensitivity of the solution to local variations due to non-zero rotational parameters. The motion is toward the scene with no rotation (as in example one) so that the true FOE is at the origin of the image plane. The depth values vary in a range from one to nine units with an average of about five units. To study the dependency of the solution on the choice of the rotational vector, the procedure given in the previous examples was repeated for six values of the rotational vector with 40% noise in the data. The results are shown in Figure 8. Again, the regions with negative depth are shown in white and the regions with positive depth are in black.

The first column shows the results for an assumed rotation of $\boldsymbol{\omega}' = (.05, 0, 0)^T$. These results show that the estimated FOE is located on the positive y-axis around $y = 0.25$ (since the axis of symmetry of the clusters of negative depth values for the assumed FOEs intersects the y-axis around $y = 0.25$). Interestingly, this is consistent with a translation of $\mathbf{t}' = (0, .25, 1)^T$. Therefore, we have overestimated the rotation about the x-axis by 0.05 radians and the translation along y-axis by about 0.25 units. As explained earlier, with noisy data, it is possible to interpret a rotation about the positive x-axis as a translation in the direction of the negative y-axis, scaled by the distance of the object from the viewer (note that the average of depth values is about five units). In this case, we need to add a translation in the positive y direction to offset the rotation about the positive x-axis.

The second column shows the results for an assumed rotation of $\boldsymbol{\omega}' = (-.05, 0, 0)^T$. In this case, the estimated FOE is along the negative y-axis at a distance of about 0.25 units from the origin. This is consistent with a translation vector $\mathbf{t}' = (0, -0.25, 1)^T$. The same conclusion as in the previous case can be made; we need to add a translation in the negative y direction to offset the rotation about the negative x-axis.

The third column shows the results for an assumed rotation of $\boldsymbol{\omega}' = (0, 0.05, 0)^T$. In this case, the FOE constraint lines do not seem to have a common intersection point. This is good news since the conclusion is that the assumed rotation cannot be correct. The situation is different for $\boldsymbol{\omega}' = (0, -0.05, 0)^T$ (the results are shown in the fourth column). In this case, the axes of symmetry of the negative depth clusters seem to intersect around the point $\tilde{\mathbf{t}}' = (0.25, 0)^T$. This is consistent with a translation of $\mathbf{t}' = (0.25, 0, 1)^T$.

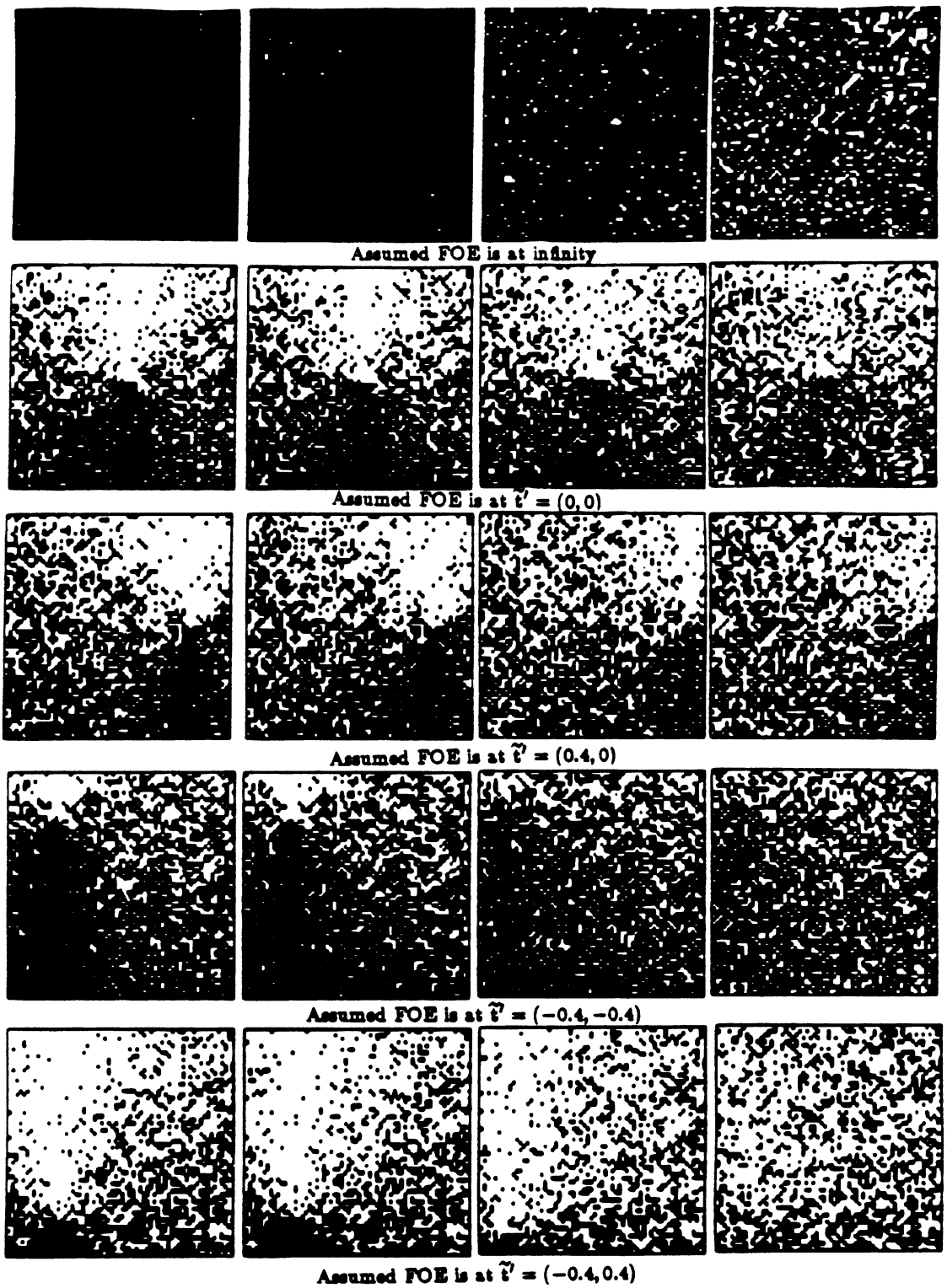

Figure 7. Regions with positive (black) and negative (white) depth values when noise is added to brightness derivatives; Example two: Focus of expansion at infinity. See text for explanation.

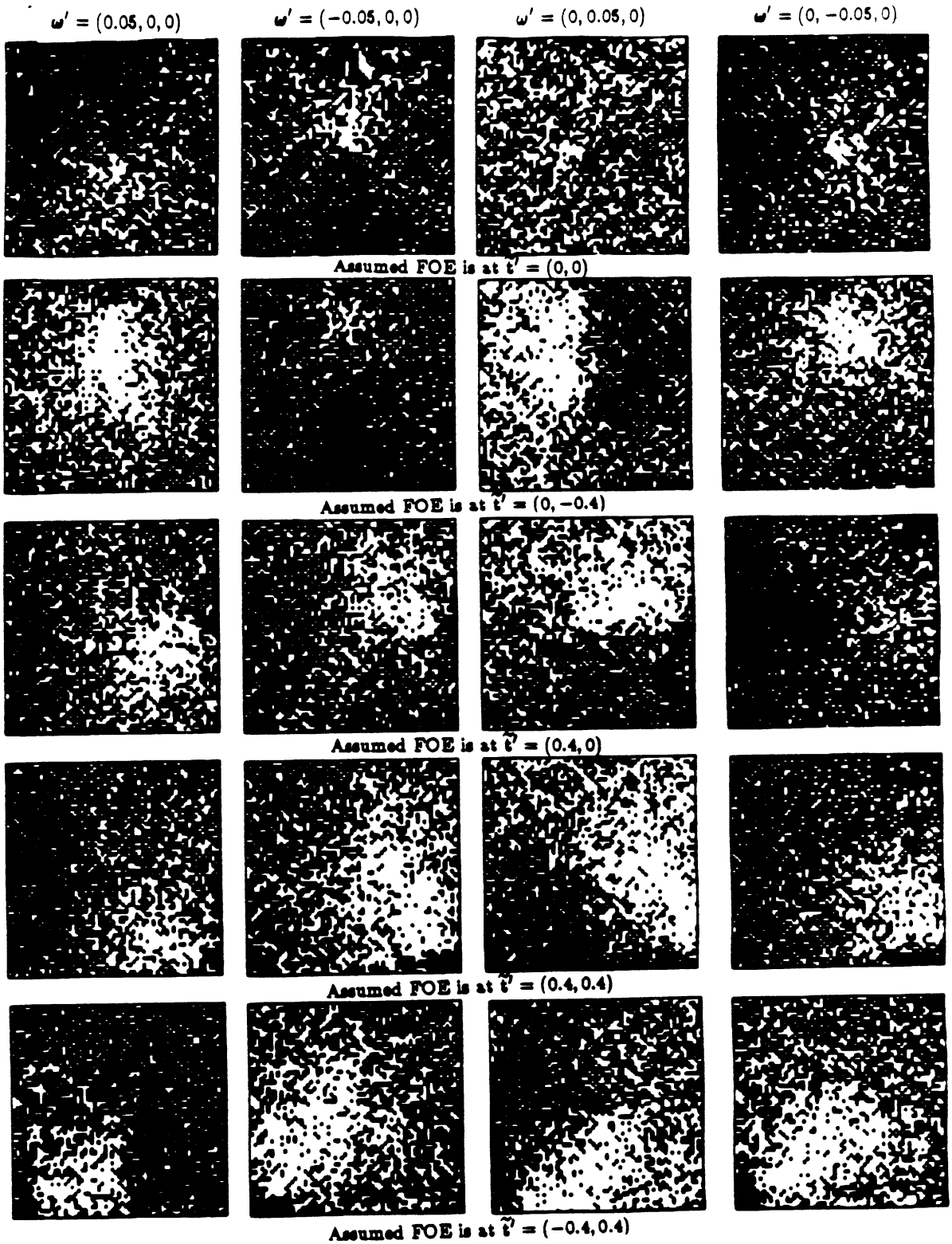

Figure 8. Positive (black) and negative (white) depth regions with noise added to brightness derivatives; Example three: Unknown Rotation. See text for explanation.

Again, with noisy data, it is possible to interpret a rotation about the negative y-axis as a translation in the negative x direction scaled by the distance of the object from the viewer. In this case, we need to add a translation in the positive x direction to offset the rotation about the negative y-axis.

The remaining plots from the leftmost column to the rightmost column (Figure 8, continued) are for an assumed rotation of $\boldsymbol{\omega} = (0, 0, 0.05)^T$, $\boldsymbol{\omega} = (0, 0, -0.05)^T$, $\boldsymbol{\omega} = (0, 0, 0.1)^T$, and $\boldsymbol{\omega} = (0, 0, -0.1)^T$, respectively.

A careful review of these plots reveals that, for each assumed rotation, the FOE constraint lines do not intersect at a common point, but seem to intersect in points lying on a circle centered at the origin with radius proportional to the assumed rotation rate about the optical axis. To explain this, we need to remember that a rotation about the optical axis generates motion field vectors that are tangent to concentric circles with center at the FOE (the origin in this case). For a rotation of the viewer about the positive z-axis (the optical axis) the motion field vectors travel counterclockwise. Conversely, they are clockwise for a rotation about the negative z-axis. Take rotation about the positive z-axis, for example (results shown in the first and third columns). Along the negative y-axis (remember this points upward) the motion field vectors point from right to left (and increase in magnitude linearly with y). This is indicated by the shift in the negative depth cluster toward the negative x-direction (second row in the first and third columns). Along the positive x direction, these vectors point upward (and increase in magnitude linearly with x). This appears as an upward shift in the negative depth cluster (third row in the first and third columns).

The same behavior is observed in the plots in the last two rows of the first and third columns. In each case, the negative depth cluster is shifted somewhat in the direction consistent with a rotation about z-axis. This implies, as mentioned earlier, that the axes of symmetry of these clusters do not intersect at a common point (the origin) because of the shifts, but rather intersect at several points that are located approximately on a circle with center at the origin and radius proportional to the magnitude of the assumed rotation. The plots in the second and fourth columns for a rotation about the negative z-axis show a similar behavior except that the shifts are now in the opposite directions. Therefore, we expect that the axes of symmetry of the clusters intersect almost at a common point when the magnitude of rotation about z-axis tends toward zero; that is, the correct rotation is assumed (see also the plots given in the first example).

7 Summary

In this paper we have shown that one can exploit the positiveness of depth as a constraint in order to estimate the location of the focus of expansion when the motion is either purely translational or the rotational component is known. The approach is based on the fact that when an arbitrary point in the image is chosen as the FOE, the depth values that are computed based on the assumed FOE tend to form clusters of positive and negative

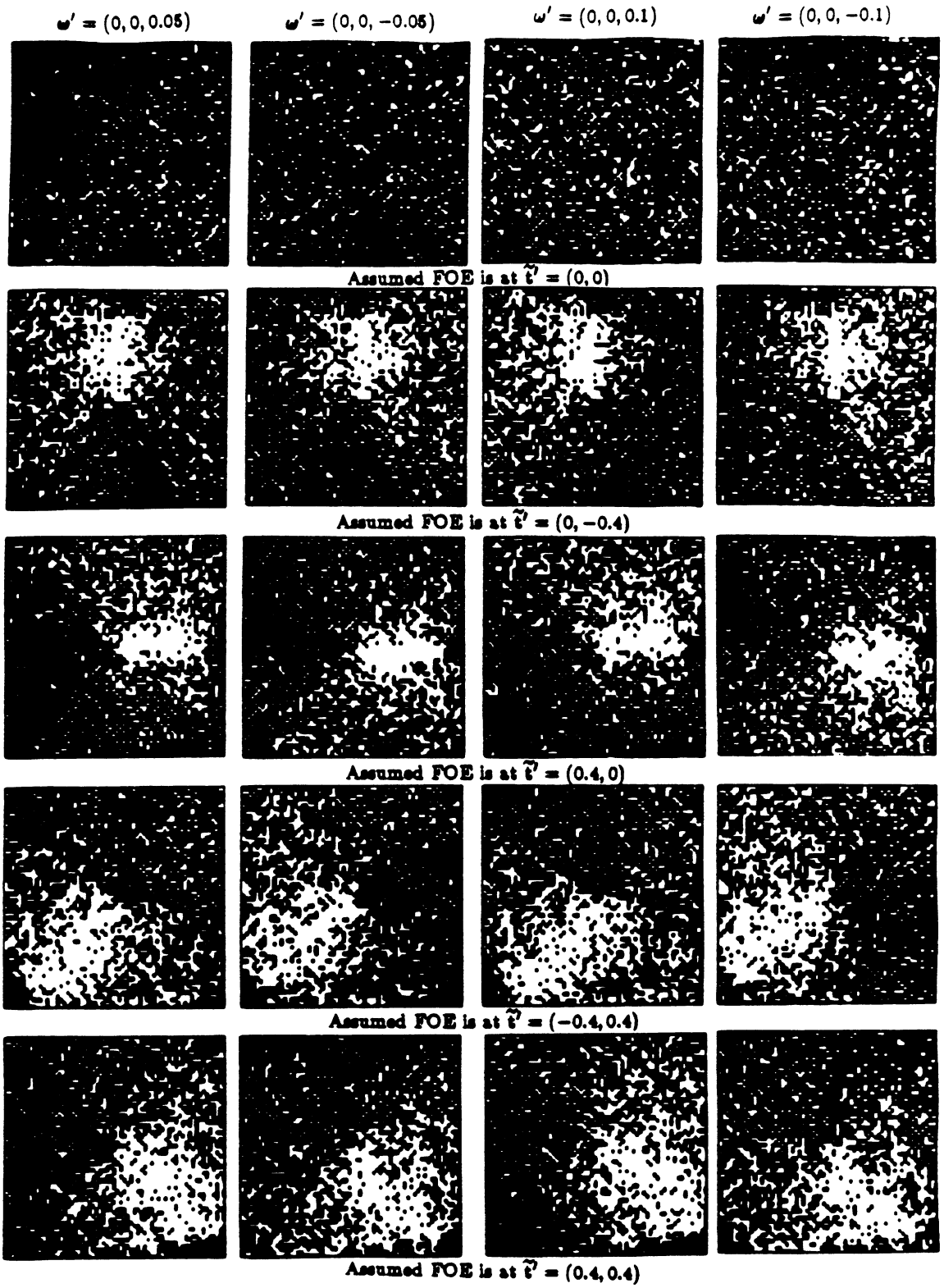

Figure 8. Continued.

values around the line that connects the assumed FOE to the true FOE; that is, the line that we referred to as the FOE constraint line. These clusters are symmetrical with respect to the FOE constraint line and can be used to determine the direction toward the true FOE; that is, the orientation of the FOE constraint line. By finding the common intersection of several such constraint lines, it is possible to obtain a reasonable estimate of the true FOE. In two selected examples, we showed that when the rotation is known, the method we suggested can give a good estimate of the location of the FOE in the presence of noisy data (with noise of as much as 60%).

When the rotational component is not known (and is non-zero), these constraint lines do not have a common intersection point. This is reminiscent of the fact that motion field vectors do not intersect at a common point when the viewer rotates about some axis through the viewing point as well as translating in an arbitrary direction. In this case, we proposed a method based on discounting the component due to rotation (by assuming some arbitrary rotation) before we apply the method developed for the case of pure translation. Ideally, a reasonable estimate of the FOE is obtained only when the correct rotation is assumed; this corresponds to a distinct optimum solution. We have not implemented the method to evaluate the accuracy of the solution; however, we presented an example to demonstrate the behavior of the solution, with noisy data, where the rotation vector was varied locally. The results showed some of the difficulties we have to deal with in estimating 3-D motion when the rotational component of motion is unknown. For example, several interpretations were possible (based on a qualitative analysis) related to the ambiguity in distinguishing rotation from translation (appropriately scaled by the average distance of the viewer from the scene). These interpretations, however, are consistent with those obtained from the corresponding noisy two-dimensional optical flow estimate by other means.

8 Appendix—Image Plane Formulae for the FOE

Some of the results presented above have been expressed concisely using vector notation. It is occasionally helpful to develop corresponding results in terms of the components of these vectors. Consider, for example, the methods for recovering the FOE from the brightness gradient at stationary points (where $c = 0$). Let the FOE be at $\widetilde{\mathbf{t}} = (x_0, y_0, 1)^T$. At a stationary point, $\mathbf{s} \cdot \mathbf{t} = 0$, and so $\mathbf{s} \cdot \widetilde{\mathbf{t}} = 0$ (unless $\mathbf{t} \cdot \mathbf{z} = 0$). This in turn can be expanded to yield

$$x_0 E_x + y_0 E_y = x E_x + y E_y$$

that is, the brightness gradient is perpendicular to the line from the stationary point to the FOE.

Now suppose that we have the brightness gradient at two stationary points, (x_1, y_1) and (x_2, y_2) say. Then

$$x_0 E_{x_1} + y_0 E_{y_1} = x_1 E_{x_1} + y_1 E_{y_1},$$

$$x_0 E_{x_2} + y_0 E_{y_2} = x_2 E_{x_2} + y_2 E_{y_2},$$

which gives us

$$x_0(E_{x_1}E_{y_2} - E_{x_2}E_{y_1}) = (x_1 E_{x_1} + y_1 E_{y_1})E_{y_2} - (x_2 E_{x_2} + y_2 E_{y_2})E_{y_1},$$

$$y_0(E_{x_1}E_{y_2} - E_{x_2}E_{y_1}) = (x_1 E_{x_1} + y_1 E_{y_1})E_{x_2} - (x_2 E_{x_2} + y_2 E_{y_2})E_{x_1}.$$

This in turn yields the location of the FOE, (x_0, y_0), provided that the brightness gradients at the two stationary points are not parallel. This result corresponds exactly to

$$\widetilde{\mathbf{t}} = \frac{\mathbf{s}_1 \times \mathbf{s}_2}{(\mathbf{s}_1 \times \mathbf{s}_2) \cdot \hat{\mathbf{z}}}.$$

Next, consider the case were many stationary points are known. Suppose there are n such points. Then we may wish to minimize

$$\sum_{i=1}^{n} \left((x_0 E_{x_i} + y_0 E_{y_i}) - (x_i E_{x_i} + y_i E_{y_i})\right)^2.$$

Differentiating with respect to x_0 and y_0 and setting the results equal to zero yields,

$$x_0 \sum E_{x_i}^2 + y_0 \sum E_{x_i} E_{y_i} = \sum (x_i E_{x_i} + y_i E_{y_i}) E_{x_i}$$

$$x_0 \sum E_{y_i} E_{x_i} + y_0 \sum E_{y_i}^2 = \sum (x_i E_{x_i} + y_i E_{y_i}) E_{y_i}$$

a set of equations that can be solved for the location of the FOE in a similar way to that used to solve the set of equations above. (This produces a result that, in the presence of noise, will be slightly different from the one given in vector form earlier, since we are here enforcing the condition $(\widetilde{\mathbf{t}} \cdot \hat{\mathbf{z}}) = 1$ rather than $(\mathbf{t} \cdot \mathbf{t}) = 1$.)

9 Acknowledgments

We would like to thank E. J. Weldon for pointing out the importance of points where the time derivative of brightness is small.

10 References

Adiv, G. (1985) "Determining 3-D Motion and Structure from Optical Flow Generated by Several Moving Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume PAMI-7, No 4, July.

Aloimonos, J., & C. Brown, (1984) "Direct Processing of Curvilinear Sensor Motion from a Sequence of Perspective Images," *Proceedings of Workshop on Computer Vision: Representation and Control*, Annapolis, Maryland.

Aloimonos, J., & A. Basu (1986) "Determining the Translation of a Rigidly Moving Surface, without Correspondence," *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, Miami, Florida, June.

Ballard, D.H., & O.A. Kimball (1983) "Rigid Body Motion from Depth and Optical Flow," *Computer Vision, Graphics, and Image Processing*, Vol. 22, No 1, April.

Barnard, S.T., & W.B. Thompson (1980) "Disparity Analysis of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, No 4, July.

Bruss, A.R., & B.K.P. Horn (1983) "Passive Navigation," *Computer Vision, Graphics, and Image Processing*, Vol. 21, No 1, January.

Horn, B.K.P. (1986) *Robot Vision*, MIT Press, Cambridge and McGraw-Hill, New York.

Horn, B.K.P., & E.J. Weldon (1986) "Robust Direct Methods for Recovering Motion," submitted to *International Journal of Computer Vision.*

Jerian, C., & R. Jain (1984) "Determining Motion Parameters for Scenes with Translation and Rotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No 4, July.

Longuet-Higgins, H.C., & K. Prazdny (1980) "The Interpretation of a Moving Retinal Image," *Proceedings of the Royal Society of London*, Series B, Vol. 208.

Longuet-Higgins, H.C. (1981) "A Computer Algorithm for Reconstructing a Scene from Two Projections," *Nature*, Vol. 293.

Mitiche, A. (1984) "Computation of Optical Flow and Rigid Motion," *Proceedings of Workshop on Computer Vision: Representation and Control*, Annapolis, Maryland.

Negahdaripour, S., & B.K.P. Horn (1985) "Determining 3-D Motion of Planar Objects from Image Brightness Patterns," *Proceedings of the Ninth International Joint Conference in Artificial Intelligence*, Los Angeles, CA, August.

Negahdaripour, S., & B.K.P. Horn (1987) "Direct Passive Navigation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No 1, January.

Negahdaripour, S., "Direct Passive Navigation," Ph.D. Thesis, Department of Mechanical Engineering, MIT, Cambridge, MA, November.

Prazdny, K. (1979) "Motion and Structure from Optical Flow," *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, Tokyo, Japan, August.

Prazdny, K. (1981) "Determining the Instantaneous Direction of Motion from Optical Flow Generated by a Curvilinearly Moving Observer," *Computer Graphics and Image Processing*, Vol. 17, No 3, November.

Roach, J.W., & J.K. Aggarwal (1980) "Determining the Movement of Objects from a Sequence of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, No 6, November.

Tsai, R. (1983) "Estimating 3-D Motion Parameters and Object Surface Structures from the Image Motion of Curved Edges," *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, Washington, D.C., June.

Tsai, R.Y., T.S. Huang (1984) "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No 1, January.

Waxman, A.M., & S. Ullman (1983) "Surface Structure and 3-D Motion from Image Flow: A Kinematic Analysis," CAR-TR-24, Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD, October.

Yen, B.L., & T.S. Huang (1983) "Determining 3-D Motion/Structure of a Rigid Body Over 3 Frames Using Straight Line Correspondence," *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, Washington, D.C., June.

This blank page was inserted to preserve pagination.

CS-TR Scanning Project
Document Control Form

Date : 10/12/95

Report # AIM-939

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☐ Technical Report (TR) ☒ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 29(34-IMAGES)
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☒ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:

☒ DOD Form ☐ Funding Agent Form ☐ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages(by page number):_____

Photographs/Tonal Material (by page number):_____

Other (note description/page number):

Description :	Page Number:
(A) IMAGE MAP:	(1-29) UNNUMBERED TITLE PAGE, 1-28 (30-34) SCANCONTROL, DOD, TRGT'S (3)
(B) CUTOUTS FIGS ON PAGES	3, 10, 11, 15, 19, 21-22, 24.

Scanning Agent Signoff:

Date Received: 10/12/95 Date Scanned: 10/17/95 Date Returned: 10/19/95

Scanning Agent Signature: Michael W. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER A.I. Memo 939	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A185840
4. TITLE *(and Subtitle)* A Direct Method for Locating the Focus of Expansion		5. TYPE OF REPORT & PERIOD COVERED memorandum 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Shariar Negahdaripour Berthold K.P. Horn		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0643
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE January 1987 13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* Office of Naval Research Information Systems Arlington, Virginia 22217		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Distribution of this document is unlimited.

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

18. SUPPLEMENTARY NOTES

None

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

focus of expansion optical flow motion field
structure from motion correspondence problem

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

We address the problem of recovering the motion of a monocular observer relative to a rigid scene. We do not make any assumptions about the shapes of the surfaces in the scene, nor do we use estimates of the optical flow or point correspondences. Instead, we exploit the spatial gradient and the time rate of change of brightness over the whole image and explicitly impose the constraint that the surface of an object in the scene must be in front of the camera for it to be imaged.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 10/17/1995

M.I.T. Libraries
Document Services