COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 46, 303–326 (1989)

A Direct Method for Locating the Focus of Expansion

SHAHRIAR NEGAHDARIPOUR

Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, Hawaii 96822

AND

BERTHOLD K. P. HORN

MIT Artificial Intelligence Laboratory, Cambridge, Massachusetts 02139

Received December 10, 1986; revised October 18, 1988

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.

1. INTRODUCTION

One of the primary tasks of a computer vision system is to reconstruct, from 2-dimensional images, such 3-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 defined in terms 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.

Two types of approaches, discrete and continuous, have been pursued in most of the earlier work in motion vision. Discrete methods establish correspondences between images of points in the scene in a sequence of images in order to recover motion (see, for example, [21, 22, 14, 5, 15, 25, 27]). Among the shortcomings of the discrete methods are that they require the solution of the correspondence problem 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, [24, 26, 3]); however, the computation is still based on information in a relatively small portion of the image.

Jain [11] presents a method for the computation of the focus of expansion (for pure translational motion) that does not require determining the point correspondences in the two images, however, one still needs to identify feature points. Lawton [12] suggests a method based on simultaneous feature extraction and motion computation. In either case, the need for identifying feature points makes the method inattractive.

In the continuous approach, optical flow, an estimate of the velocities of the images of points in the scene, is used. Longuet-Higgins and Prazdny [13] show how image velocity as well as its first and second derivatives at a single image point can

0734-189X/89 $3.00

be used to recover motion and the local structure of the surface of the scene. This method, based on only local information, is sensitive to inherent ambiguities in the solution when data is noisy. In fact, since the method essentially works with a vanishingly small field of view, it is unable to estimate all components of motion [19]. Other methods have been suggested based on a least-squares approach (see [4, 6, 1]). Here, motion parameters are found that are most consistent with the estimated image velocity over the entire image. The least-squares methods are more robust, however, they make use of the unrealistic assumption that the computed optical flow is a good estimate of the true image motion. Also, the algorithms for estimating an optical flow field are computationally expensive. This has motivated investigation of methods that use brightness derivative information at every image point *directly* to recover 3-dimensional shape and motion. Several special cases of the motion vision problem have already been addressed using this notion.

When the motion is purely rotational, one only has to solve three linear equations in three unknowns (Aloimonos and Brown [2] apparently first reported a solution to this problem, while Horn and Weldon [8, 9] studied the 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 [19]. 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 the same distance from the viewer, that is, the scene lies in a frontal plane. In this case, Negahdaripour and Yu [19] show 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 and Horn [16] can be applied. This approach leads to both iterative and closed-form solutions.

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 do not compute optical flow nor do we use correspondences of image features. We use the information in the image brightness variations (both spatial and temporal) over the whole image, and 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. This physical constraint has been used in the past mainly to distinguish among multiple solutions to the problem of recovering scene structure from 2-dimensional images (for example, see [10]); that is, it has been used to eliminate those solutions for which the depth values are negative for some surface points. In our work, however, the constraint is used directly in the formulation of the problem to determine the best estimate of the motion. 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. We show how the constraint that depth is positive can be used to locate the focus of expansion, and consequently the direction of translation, given an image sequence. We explain how one can develop algorithms, based on our analysis and the proposed methods, that require simple computations. We will present examples based on experiments with real images. Finally, we explain how our results can be extended if the motion involves rotation as well as translation of the viewer. The

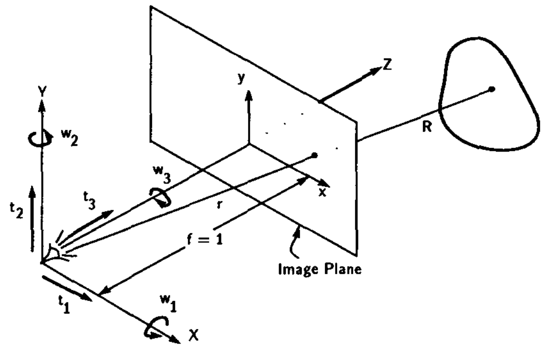

FIG. 1. Viewer-centered coordinate system and perspective projection.

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 rotation axis (see [17] for an example that demonstrates this fact, and [19] for a qualitative analysis).

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 assumed to be along the z-axis), and the focal length is taken to be the unit of length, without loss of generality (Fig. 1). Let $\mathbf{R} = (X, Y, Z)^T$ be the point in the 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 point $\mathbf{R}$.

Now, suppose the viewer moves with translational and rotational velocities $\mathbf{t}$ and $\boldsymbol{\omega}$ relative to a stationary scene. Then a point 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 [16]

$$\mathbf{r}_t = -\left(\hat{\mathbf{z}} \times \left(\mathbf{r} \times \left(\mathbf{r} \times \boldsymbol{\omega} - \frac{1}{\mathbf{R} \cdot \hat{\mathbf{z}}}\mathbf{t}\right)\right)\right),$$

which is the familiar result apparently first derived by Longuet-Higgins and Prazdny [13]. The velocities of all image points, given by the above equation, taken collectively, define a 2-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 [7] for a discussion of the distinction between optical flow and the motion field).

As an object moves in the environment, the brightness of the image of a patch on the surface of the object may change for a variety of reasons including changes in illumination or shading. If the surfaces of the object have sufficient texture (or high contrast at large frequencies) and the lighting conditions vary rather slowly both spatially and with time, the image brightness changes due to changing surface orientation and changing illumination are negligible (relative to changes due to relative motion of the scene and the observer). Hence, we may assume that the brightness of a small patch on a surface in the scene remains essentially constant as it moves. This is the *brightness constancy assumption*.

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 $E_{\mathbf{r}} = (E_x, E_y, 0)^{\mathrm{T}}$ 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 [16],

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

where, for conciseness, we have defined

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

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

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

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

$$\mathbf{r} \cdot \mathbf{s} = 0, \qquad \mathbf{r} \cdot \mathbf{v} = 0, \qquad \mathbf{s} \cdot \mathbf{v} = 0.$$

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.

We need to mention that the brightness constancy assumption has been a controversial issue in recent years (for example, see [23]). While this assumption may be violated in some cases, for example, for surfaces with dominantly specular reflectance properties, it holds exactly for the case of many other surfaces including those with lambertian reflectance properties. More importantly, it is a good approximation when the surface texture has high contrast at large frequencies (see [9] for a more detailed explanation). We will later explain why our method is robust with respect to the errors in the constraint equation provided that the scene has strong texture. (Negahdaripour and Horn [18] show, through two selected examples with synthetic data, that the method is robust with respect to errors in the brightness change equation. We also show this through two selected examples from experiments we have performed using real images.)

3. POSITIVENESS OF DEPTH

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 temporal derivative 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 $\{\boldsymbol{\omega}', \mathbf{t}'\}$, depth values that satisfy the brightness change constraint equation can be determined using

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

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

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 $\{\boldsymbol{\omega}', \mathbf{t}'\}$ that gives rise to negative depth values cannot be the correct one. Thus, the problem is to determine the pair $\{\boldsymbol{\omega}, \mathbf{t}\}$ 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 $\{\boldsymbol{\omega}, \mathbf{t}\}$ and the surface $Z > 0$, is there another motion $\{\boldsymbol{\omega}', \mathbf{t}'\}$ 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 one, 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 $\tilde{c} = c + \mathbf{v} + \boldsymbol{\omega}$. For simplicity, we will from now on write c where $\tilde{c}$ should appear. The problem is still underconstrained 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(s) 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*. 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 ($E_\mathbf{r} \neq 0$). 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 and Weldon [9]). 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 nonparallel 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. So

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

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

This solution can be interpreted geometrically: At each stationary point, $\mathbf{t}$ is the normal to the vector $\mathbf{s}$. Two such vectors define the plane with normal vector $\mathbf{t}$. Another interpretation is in terms of quantities in the image plane: The brightness gradient at a stationary point is orthogonal to the line that connects the point to the FOE (see the appendix for more details); or, equivalently, the tangent of the isobrightness contour at a stationary point passes through the FOE. Intersecting the tangents of the isobrightness contours at two different stationary points allows us to determine the FOE.

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 all stationary points; that is, we minimize

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

(In the above we have used the identity $\mathbf{s} \cdot \mathbf{t} = \mathbf{s}^{\mathrm{T}}\mathbf{t}$.) Note that the resulting quadratic form cannot 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 obtain 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

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

The necessary conditions for stationary values of J are

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

Executing the indicated differentiations we arrive at

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

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

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

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

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

The least-squares method just described can be interpreted as follows. The normal of the plane defined by the vectors $\mathbf{s}$ at all stationary points defines the direction of $\mathbf{t}$, if there was no measurement error (note that, $\mathbf{s} \cdot \mathbf{t} = 0$ implies that $\mathbf{t}$ lies in the null space of all $\mathbf{s}$ vectors). In practice these vectors do not lie in the same plane due to noise. The best estimate of $\mathbf{t}$ is the direction for which the sum of the squared magnitude of the projections of all $\mathbf{s}$ vectors is minimum. We can alternatively choose another estimate, derived in the Appendix. At each stationary point, the tangent to the isobrightness 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.

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 [9]).

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

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

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 nonzero 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 the product $\mathbf{s} \cdot \mathbf{t}$ in the form

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

Similarly, we obtain

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

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

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

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

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

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.

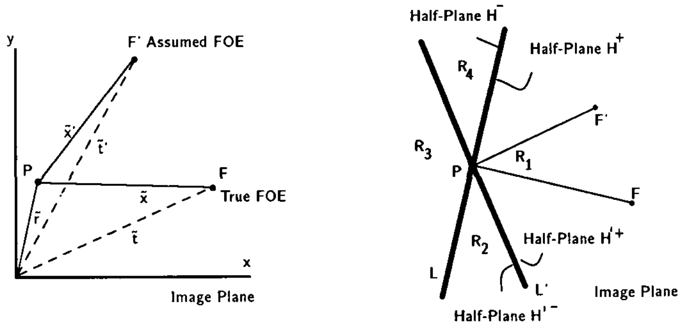

FIG. 2. True and spurious (assumed) FOEs define the positive and negative half planes.

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 $((\mathbf{r} - \tilde{\mathbf{t}}) \cdot E_\mathbf{r})$ and $((\mathbf{r} - \tilde{\mathbf{t}}') \cdot E_\mathbf{r})$ 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.

4.3. *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 Fig. 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 Fig. 2(b)):

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

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

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

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

for some constants α and β. Then

$$(\mathbf{x} \cdot E_r)(\mathbf{x}' \cdot E_r) = -\alpha\beta|\mathbf{x} \times \mathbf{x}'|^2.$$

We see that the product denoted p is positive when E_r lies between **p** and $-\mathbf{p}'$ ($\alpha > 0$ and $\beta < 0$) and when E_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_r since it leads to positive depth values. Conversely, the product will be negative when E_r lies between **p** and $\mathbf{p}'$ ($\alpha > 0$ and $\beta > 0$) and when E_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_r since it leads to negative depth values.

Denoting the half planes separated by the line parallel to **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:

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

and

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

We see that $R_1 \cup R_3$ is the permissible range for E_r, because E_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_r since $Z' < 0$ when E_r lies in this region (see Fig. 3).

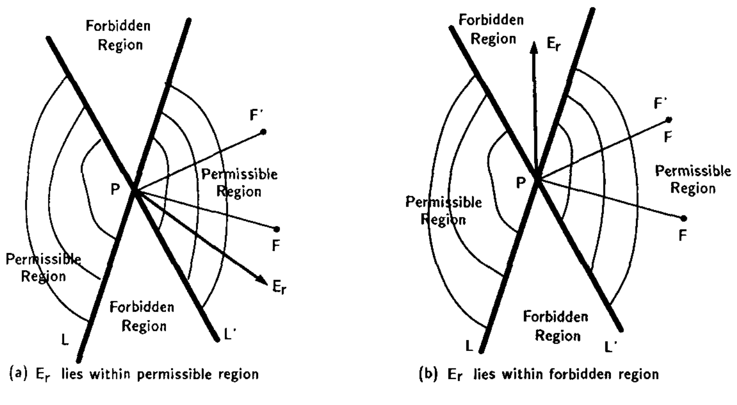

FIG. 3. Permissible and forbidden ranges for the brightness gradient direction.

(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 $\tilde{\mathbf{t}} \neq \tilde{\mathbf{t}}'$, 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 $\tilde{\mathbf{t}} = \tilde{\mathbf{t}}'$ 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.

4.4. *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

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

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 lies on the ray emanating from F' (segment $F'X'$). For points on the FOE constraint line, we have

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

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

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

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 $(\tilde{\mathbf{t}} - \tilde{\mathbf{t}}')$). 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.

When θ is small, the permissible range for $E_\mathbf{r}$ consists of a large set of allowed directions (see Fig. 4a). Therefore, the points where θ is small are likely to have

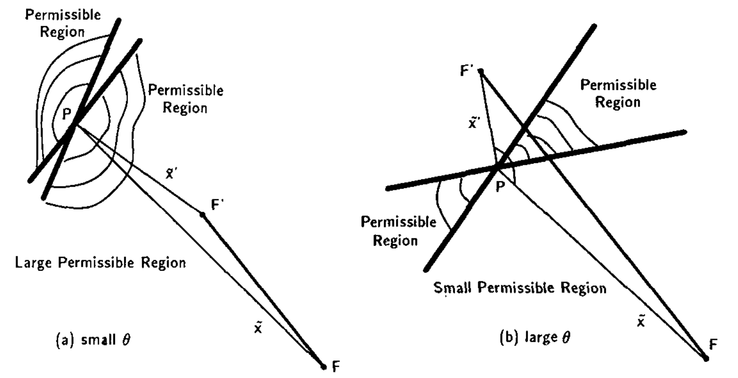

FIG. 4. Relationship between the size of the permissible range and the relative position of an image point with respect to the FOE constraint line.

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 Fig. 4b). Therefore, it is more likely that the brightness gradient lies outside this range, giving rise to a negative depth value. In the extreme case when $\theta = \pi$ (that is, the point lie along FF') the depth values are guaranteed to be negative (unless the point is a stationary point). The forbidden range for a point of 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 position-invariant and rotationally symmetric; that is, all directions of the brightness gradient are equally likely independent of the image position. It is not difficult to see that the probability that a point in the image plane gives rise to a negative depth value is given by

$$\mathrm{Prob}(Z' < 0) = \theta/\pi.$$

Since the 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 Fig. 5).

To summarize, 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 $(\mathbf{t} \cdot \hat{\mathbf{z}})(\mathbf{t}' \cdot \hat{\mathbf{z}})$ is negative, the situation is reversed. In this case, it is more likely that the points in the vicinity of FF' will give rise to positive depth values and the points along or in the vicinity of FX and $F'X'$ will give rise to negative depth values, but otherwise similar conclusions can be drawn.)

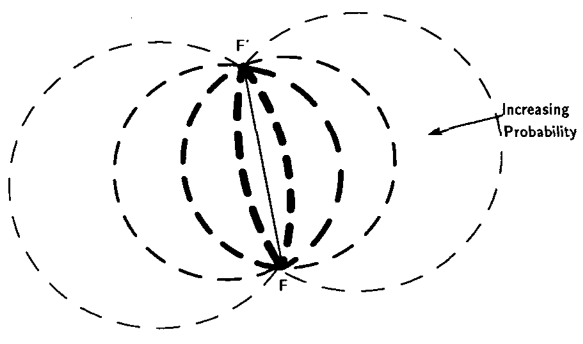

FIG. 5. Constant probability loci for $Z' < 0$.

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.

4.5. *Interpretation of Negative and Positive Depth Clusters*

We now give another interpretation of the negative and positive depth clusters using motion field vectors. For a purely translational motion, the motion field vectors emanate radially from the FOE for an approaching motion (or point toward the focus of contraction, for a departing motion). Let us assume an approaching motion, without loss of generality. For the sake of argument, let two points, one in the top left region and the other in the bottom right region of the image, denote the true and some assumed FOEs, respectively. Figure 6 shows, at each image point, the resulting motion field vectors based on the two FOEs. We see that the largest discrepancy between the two motion fields is observed in the region around the line from one FOE to the other. Along this direction, the true motion field vectors point away from the true toward the assumed FOE, while the assumed motion field vectors point in the opposite direction. Hence, we need to reverse, along this line, the direction of the assumed motion field vectors to make them consistent with the true motion field vectors. This is equivalent to multiplying either the translational vector or the depth values by the constant scale factor -1. Since the relative

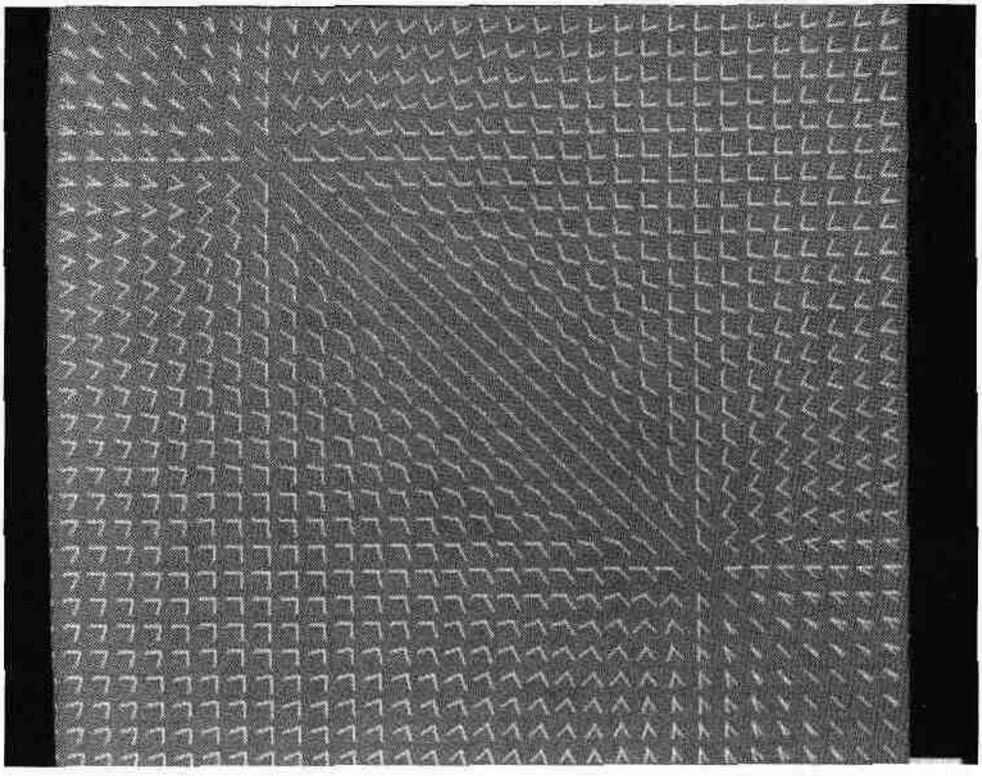

FIG. 6. Motion field vectors emanating from the true and assumed FOEs.

translational motion between the camera and all scene points is the same, we should reverse the sign of the depth values. The true depth values are positive, though. Therefore, the depth values for the assumed FOE will come out negative for points along this line. The same situation holds for points that are close to this line. This gives rise to the negative depth cluster we expect to observe around the so-called FOE constraint line. The motion field vectors, both the true and assumed ones, are consistent along the FOE constraint line away from the two FOEs. This implies that the true depth values and those obtained based on the assumed FOE will have the same sign (positive). This gives rise to the two positive depth clusters we expect along the extensions of the constraint line.

In other regions of the image, the degree of discrepancy between the two motion field vectors at each point is related to the probability that the resulting depth value will come out negative (for the particular assumed FOE). This is consistent with the constant probability loci shown in Fig. 5.

4.6. *Brightness Change Equation Revisited*

The brightness change equation was derived under the assumption of brightness constancy. For pure translation, we have

$$c + \frac{1}{Z}\mathbf{s} \cdot \mathbf{t} = 0.$$

Now suppose this equation does not hold exactly, and let ε denote the error; that is,

$$c + \frac{1}{Z}\mathbf{s} \cdot \mathbf{t} = \varepsilon.$$

Solving for Z, we obtain

$$Z = -\frac{1}{c + \varepsilon}(\mathbf{s} \cdot \mathbf{t}).$$

Our earlier analysis is based on the information in the sign of depth values, Z, and not their magnitudes. If the scene has strong texture, the magnitude of $c = E_t$ is expected to be larger than ε for most image points. Therefore c and $c + \varepsilon$ will have the same sign, and consequently Z will preserve its correct sign. The errors in the brightness change equation affect mostly those points with a small c value. The information at these points is usually less reliable anyway, due to noise in the images and quantization error. Therefore, we expect any error in the brightness change equation (if it exists) due to noise in the images, quantization error, etc. to affect mostly those points with small c values. This, unfortunately, includes the critical points we identified earlier, which, at least in theory, provide strong constraints for locating the focus of expansion.

4.7. *Locating the Focus of Expansion Using Gradient Vectors*

We have shown that the clusters of positive and negative depth values, obtained by choosing an arbitrary point as the estimate of the FOE, form particular patterns in the image. These clusters can be used to determine the location of the FOE. Hence, the problem of locating the FOE simplifies to that of identifying clusters of negative and positive depth values, and the corresponding FOE constraint lines.

4.7.1. *A Recursive Method*

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, $\tilde{\mathbf{t}}'$, in the brightness change constraint equation. We then determine the cluster of negative depth values. Finally, we use the fact that the centroid of this cluster is expected to lie halfway 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

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

Then the position of the FOE can be estimated using:

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

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:

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

where $(\bar{\mathbf{t}})^n$ is the centroid of the cluster of points with negative depth values obtained using the estimate $(\tilde{\mathbf{t}}')^n$ 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.

4.7.2. *Intersection of FOE Constraint Lines*

We can alternatively consider other methods which 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. (In this case, there is no problem 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. The FOE constraint lines simply intersect outside the image plane.)

4.7.2.1. *Determining the FOE 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 direction of 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.

Consider a circle of radius r centered at the assumed FOE, which does not include the true FOE. We can construct a histogram of the number of image points, within the circle, with negative depth values for every possible direction $0 \le \phi < 2\pi$ (In practice, we cannot have every angle because of quantization.) We expect the peak of the histogram to correspond to the direction toward the FOE; this gives the direction of the line passing through the assumed FOE that we choose as the FOE constraint line. Generally, the solution may not be unique because the histogram does not have a distinct peak; that is, we may obtain a range of possible angles around the correct solution, denoted by ϕ_f. We can obtain a better estimate of ϕ_f using the fact that $\phi_f \pm \pi$ is expected to correspond to a minimum in the histogram (this is the direction with the most number of positive depth values). Let us denote the number of negative (positive) depth values in a particular direction ϕ by N^- (N^+), and let $D = N^- - N^+$ denote the difference between the number of negative and positive depth values in that direction. We can instead construct the histogram of $D(\phi) - D(\phi + \pi)$ for all directions $0 \le \phi < \pi$. This new histogram is expected to have a more distinct peak, corresponding to the direction of the FOE constraint line.

The size of r, the radius of the circle, can be an important factor. If r is too small, the histogram may not have a distinct peak due to noise or quantization errors. If r is too large, the circle may include the FOE. In this case, the histogram may be flattened particularly around ϕ_f due to the effect of points within the positive cluster beyond the FOE.

Alternatively, we can compute the center of mass of the negative points and the positive points within the circle. For a symmetric cluster, these points are expected to lie on the FOE constraint line. We then determine the line that best fits the assumed FOE as well as the two centers of mass.

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.

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 the 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 encountered when 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 from an approaching motion). Unfortunately, these vectors do not intersect at the FOE when the rotational component is nonzero. Similarly, we expect that the FOE constraint lines will not intersect at a common point when the rotational component is nonzero (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 [21] 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 $\boldsymbol{\omega}_{ijk}$.

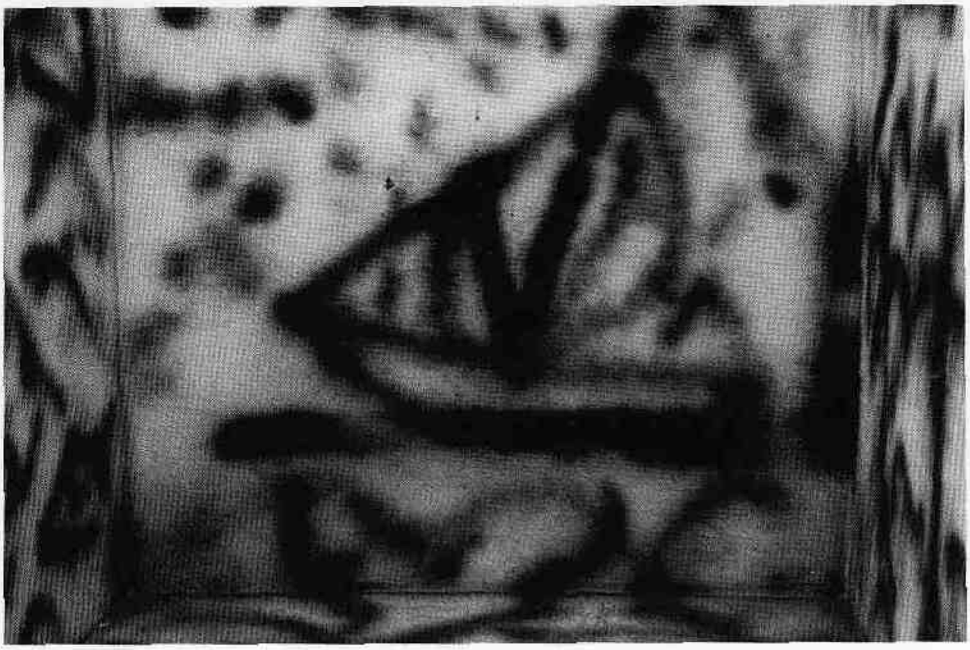

FIG. 7. A band-limited scene.

6. EXPERIMENTAL RESULTS

Negahdaipour and Horn [18] present two examples, using synthetic data, to 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 FOE constraint lines. Synthetic data was used so that the underlying motion is known exactly. In addition, one can evaluate the performance, in terms of stability and robustness, of the proposed methods in the presence of artificial noise added to the data. We will present two examples using 64×61 real images of a band-limited scene (see Fig. 7).

6.1. *Example One*: *Focus of Expansion in the Image*

In this experiment, the camera was moved 0.5 in toward the scene; that is, the focus of expansion is at the origin of the image plane. The corresponding image motion varies from 0, at the FOE, to about 0.3 pixels on the image boundary. Figure 8 shows the regions of negative (gray) and positive (black) depth values for several assumed FOEs, which are shown as white points. We have obtained essentially the same clusters for motions as large as 5.0 in (the image motion is as large as about 3 pixels for boundary points). These results show that we can estimate the location of the FOE with very good accuracy using the information from the negative and positive depth clusters.

6.2. *Example Two*: *Focus of Expansion Outside the Image*

In this experiment, the camera was moved 0.25 in forward and 0.5 in to the right. The FOE is out of the image, along the x-axis and about 4 pixels to the right of the image boundary. The corresponding image motion varies from about 0, on the right

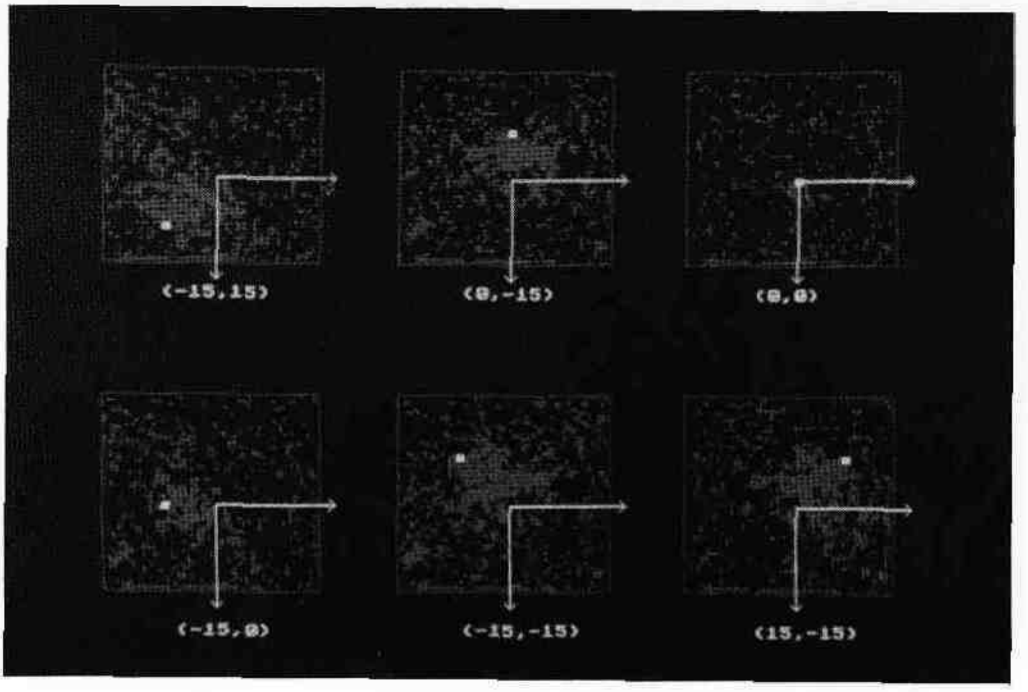

FIG. 8. The regions of negative (gray) and positive (black) depth values for several assumed FOEs; true FOE is approximately at the image origin.

boundary (along the x-axis) to about 0.5 pixel on the left boundary (along the x-axis). The regions of negative (black) and positive(gray) depth values shown, for several assumed FOEs, are shown in Fig. 9. In this example, the boundary to the right of the image cuts out a large portion of the negative depth cluster; however, there is still sufficient information to estimate the location of the FOE using the negative depth regions. Again, very similar results have been obtained, for the same FOE, for an order of magnitude larger camera motion, see Fig. 8. The regions of negative depth regions. Again, very similar results have been obtained, for the same FOE, for an order of magnitude larger camera motion.

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). There is no need to compute optical flow, detect image features, or establish feature correspondences.

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 values around the line that connects the assumed FOE to the true FOE; that is, the line that we referred to as an FOE constraint line. These clusters can be used to determine the direction toward the true FOE; that is, the orientation of the FOE constraint line. By finding the

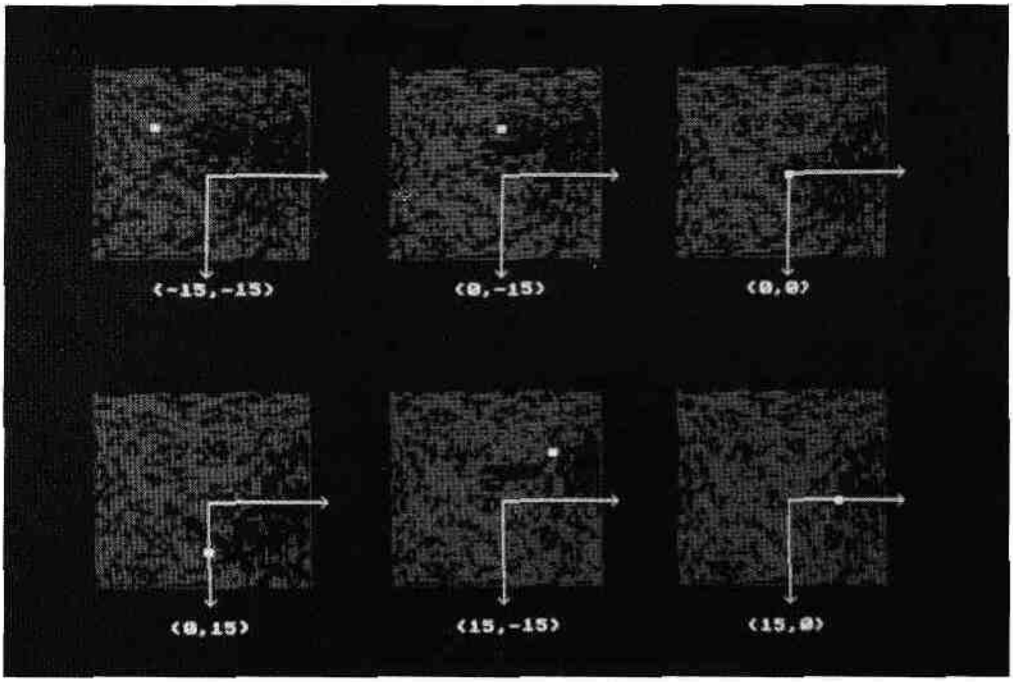

FIG. 9. The regions of negative (black) and positive (gray) depth values for several assumed FOEs; true FOE is outside the image along the x-axis.

common intersection of several such constraint lines, it is possible to obtain a reasonable estimate of the true FOE.

We have described several possible methods for developing algorithms for recovering the translational motion, which require simple computations. We have also shown, through two selected examples using real images, that these methods can give a good estimate of the location of the FOE.

When the rotational component is not known (and is nonzero), the FOE 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 a 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, initial studies using synthetic data, have shown some difficulties in estimating 3D motion accurately due to ambiguities in distinguishing rotations about some axis in the image plane from image plane translations along the axis perpendicular to the rotation axis (appropriately scaled by the average distance of the viewer from the scene). These interpretations, however, are consistent with those obtained from the corresponding noisy 2-dimensional optical flow estimate by other means.

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 $\tilde{\mathbf{t}} = (x_0, y_0, 1)^T$. At a stationary point, $\mathbf{s} \cdot \mathbf{t} = 0$, and so $\mathbf{s} \cdot \tilde{\mathbf{t}} = 0$ (unless $\mathbf{t} \cdot \hat{\mathbf{z}} = 0$). This in turn can be expanded to yield

$$(x - x_0)E_x + (y - y_0)E_y = 0;$$

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

$$x_0\left(E_{x_1}E_{y_2} - E_{x_2}E_{y_1}\right) = \left(x_1 E_{x_1} + y_1 E_{y_1}\right)E_{y_2} - \left(x_2 E_{x_2} + y_2 E_{y_2}\right)E_{y_1},$$
$$y_0\left(E_{x_1}E_{y_2} - E_{x_2}E_{y_1}\right) = \left(x_2 E_{x_2} + y_2 E_{y_2}\right)E_{x_1} - \left(x_1 E_{x_1} + y_1 E_{y_1}\right)E_{x_2}.$$

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

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

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

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

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

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

a set of equations that can be solved for the location of the FOE in a way similar 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 $(\tilde{\mathbf{t}} \cdot \hat{\mathbf{z}}) = 1$ rather than $(\mathbf{t} \cdot \mathbf{t}) = 1$.)

ACKNOWLEDGMENTS

The first author would like to thank Roy E. Shaw, Gordon T. Sasamori, and Sean Richardson for performing the experiments with real images and preparing the output images. This paper describes research done at the MIT Artificial Intelligence Laboratory and the *Waikiki Surfing and Computer Vision Society* of the Department of Electrical Engineering at the University of Hawaii. Support for the computer vision research at the MIT Artificial Intelligence Laboratory is provided in part by the National Science Foundation under Grant DMC85-11966 and in part by the Defense Advanced Research Projects Agency under the Office of Naval Research Contract N00014-75-C-0643. Support for the research at the Waikiki Surfing and Computer Vision Society is provided by the Pacific International Center for High Technology Research.

REFERENCES

1. G. Adiv, Determining 3-D motion and structure from optical flow generated by several moving objects, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-7**, No. 4, 1985.
2. J. Aloimonos and C. Brown, Direct processing of curvilinear sensor motion from a sequence of perspective images, *Proceedings Workshop on Computer Vision: Representation and Control, Annapolis, Maryland, 1984.*
3. J. Aloimonos and A. Basu, Determining the translation of a rigidly moving surface, without correspondence, in *Proc. IEEE Computer Vision and Pattern Recognition Conference, Miami, Florida, June 1986.*
4. D. H. Ballard and O. A. Kimball, Rigid body motion from depth and optical flow, *Comput. Vision Graphics Image Process.* **22**, No. 1, 1983.
5. S. T. Barnard and W. B. Thompson, Disparity analysis of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-2**, No. 4, 1980.
6. A. R. Bruss and B. K. P. Horn, Passive navigation, *Comput. Vision Graphics Image Process.* **21**, No. 1, 1983.
7. B. K. P. Horn, *Robot Vision*, MIT Press McGraw–Hill, New York, 1986.
8. B. K. P. Horn and E. J. Weldon, Robust direct methods for recovering motion, in *Proceedings First Int. Conference on Computer Vision, London, June 1987.*
9. B. K. P. Horn and E. J. Weldon, Robust direct methods for recovering motion, *Int. J. Comput. Vision.* **2**, 1988.
10. T. S. Huang, *Image Sequence Analysis*, Chap. 1, Springer-Verlag, New York, 1981.
11. R. Jain, Direct computation of the focus of expansion, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-5**, No. 1, 1983.
12. D. Lawton, Processing translational motion sequences, *Comput. Vision Graphics Image Process.* **22**, 1983.
13. H. C. Longuet-Higgins and K. Prazdny, The interpretation of a moving retinal image, *Proc. Roy. Soc. London Ser. B* **208**, 1980.
14. H. C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature* **293**, 1981.
15. A. Mitiche, Computation of optical flow and rigid motion, in *Proceedings Workshop on Computer Vision: Representation and Control, Annapolis, MD, 1984.*
16. S. Negahdaripour and B. K. P. Horn, Direct passive navigation, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-9**, No. 1, 1987.
17. S. Negahdaripour and B. K. P. Horn, *A Direct Method for Locating the FOE*, A.I. Memo 939, MIT Artificial Intelligence Lab, MIT, Cambridge, MA, January 1987.
18. S. Negahdaripour and B. K. P. Horn, Using depth-is-positive constraint to recover translational motion, in *IEEE Workshop on Computer Vision, Miami Beach, FL, November 30–December 2, 1987.*
19. S. Negahdaripour and C. H. Yu, Robust recovery of motion: Effects of field of view and surface orientation, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Ann Arbor, MI, June 1988.*
20. K. Prazdny, Motion and structure from optical flow, in *Proceedings Sixth International Joint Conference on Artificial Intelligence, Tokyo, Japan, August 1979.*
21. K. Prazdny, Determining the instantaneous direction of motion from optical flow generated by a curvilinearly moving observer, *Comput. Graphics and Image Process.* **17**, No. 3, 1981.

22. J. W. Roach and J. K. Aggarwal, Determining the movement of objects from a sequence of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-2**, No. 6, 1980.

23. B. G. Schunck, Image flow continuity for motion and density, in *Proceedings, Workshop on Motion: Representation & Control, South Carolina, 1986.*

24. R. Tsai, Estimating 3-D motion parameters and object surface structures from the image motion of curved edges, in *Proceedings IEEE Computer Vision and Pattern Recognition Conference, Washington, DC, June 1983.*

25. R. Y. Tsai and T. S. Huang, Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**, No. 1, 1984.

26. B. L. Yen and T. S. Huang, Determining 3-D motion/structure of a rigid body over 3 frames using straight line correspondence, in *Proceedings IEEE Computer Vision and Pattern Recognition Conference, Washington, DC, June 1983.*

27. X. Zhuang and R. Haralick, Rigid body motion and the optical flow image, in *Proceedings IEEE First Conference on AI Applications, Denver, CO, December 1984.*