International Journal of Computer Vision, 15, 7–29 (1995)

Qualitative Egomotion*

CORNELIA FERMÜLLER
Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD 20742-3411 and Department for Pattern Recognition and Image Processing, Institute for Automation, Technical University of Vienna, Treitlstraße 3, A-1040 Vienna, Austria

YIANNIS ALOIMONOS
Computer Vision Laboratory, Center for Automation Research, Computer Science Department and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742-3411

Received October 12, 1992; Revised March 11, 1992; Accepted November 23, 1993

Abstract. Due to the aperture problem, the only motion measurement in images, whose computation does not require any assumptions about the scene in view, is normal flow—the projection of image motion on the gradient direction. In this paper we show how a monocular observer can estimate its 3D motion relative to the scene by using normal flow measurements in a global and qualitative way. The problem is addressed through a search technique. By checking constraints imposed by 3D motion parameters on the normal flow field, the possible space of solutions is gradually reduced. In the four modules that comprise the solution, constraints of increasing restriction are considered, culminating in testing every single normal flow value for its consistency with a set of motion parameters. The fact that motion is rigid defines geometric relations between certain values of the normal flow field. The selected values form patterns in the image plane that are dependent on only some of the motion parameters. These patterns, which are determined by the signs of the normal flow values, are searched for in order to find the axes of translation and rotation. The third rotational component is computed from normal flow vectors that are only due to rotational motion. Finally, by looking at the complete data set, all solutions that cannot give rise to the given normal flow field are discarded from the solution space.

1 Introduction

A system in order to navigate by visual means must possess a wide range of perceptual capabilities that can be classified hierarchically on the basis of their complexity. At the bottom of the hierarchy are low level tasks, such as obstacle avoidance, and the top is represented by high level abilities like homing or target pursuit. As a basic capability however, every visual navigation system must have an understanding of visual motion. It should be able to estimate the three-dimensional motions of objects in its environment and even more important, it should be able to determine its own motion. Usually the term "passive navigation" is used to describe the process by which a system can estimate its motion with respect to the environment. Passive navigation is a prerequisite for any other navigational ability. A system can be guided only if there is a way for it to acquire information about its own motion.

In its original formulation, passive navigation utilizes estimates of the scene points' projected motions. As a result, most algorithms that have appeared in the literature address the motion estimation problem in two steps. First the image displacements between consecutive image frames are computed; either discrete features in successive frames are corresponded or the vector field that represents the motion of every image point, the optical flow field, is computed (Barnard and Thompson 1980; Horn and Schunck 1981; Hildreth 1983). In general, the estimation of optical flow or of feature correspondence are ill-posed problems and additional assumptions must be made in order for a solution to be obtained. Therefore, their applicability to the problem of passive navigation must be reconsidered.

In the second step, under the assumption that optical flow or correspondence is known, the 3D motion is computed from the equations relating it to the 2D image velocity. These equations are determined by the specific geometric model of image formation which is used. Different geometric projection models have been employed. Orthographic projection (Ullman 1979) leads to linear equations, but in general is not realistic and

*Research supported in part by NSF (Grant IRI-90-57934), ONR (Contract N00014-93-1-0257) and ARPA (Order No. 8459).

should only be considered as an approximation in the case when lenses of very high focal length are used and the field of view is very small. A more adequate model is given by perspective projection. The image is projected either on a sphere (Nelson and Aloimonos 1988) or on a plane (Adiv 1985; Bruss and Horn 1983; Horn 1990; Longuet-Higgins 1981; Longuet-Higgins and Prazdny 1980; Waxman 1987; Spetsakis and Aloimonos 1988; Tsai and Huang 1984). The resulting equations relating the 3D motion described by three rotational parameters and two parameters for the direction of translation to image motion are nonlinear. Therefore the surface in view often is modeled as a smooth function (usually a polynomial) and nonlinear optimization techniques are applied to solve the 3D motion parameter estimation problem. Linear algorithms have also been developed, most of them based on a particular linearization technique, the intermediate computation of the "E" matrix (Longuet-Higgins 1981; Tsai and Huang 1984). But a critical investigation of the feasibility of addressing the motion estimation problem in this way, conducted by Spetsakis and Aloimonos (1988), shows that even the application of provably optimal algorithms to this problem cannot lead to methods that could be useful for applications in realistic domains.

Since most existence and uniqueness aspects of the problem (Tsai and Huang 1984; Horn 1987; Spetsakis and Aloimonos 1990; Aloimonos and Shulman 1989; Bergholm 1988; Faugeras and Maybank 1990; Liu and Huang 1988; Navab, Faugeras and Vieville 1993) are now well understood and initial attempts to construct algorithms that perform well in realistic domains have failed, motion estimation research has shifted its focus on the robustness issue. In order to overcome sensitivity, researchers started using redundant information. For the case of correspondence-based methods this meant the use of more features and more frames (Spetsakis and Aloimonos 1991). Several multi-frame approaches have been developed. In some of these works however an unrealistic assumption again is made—that of motion continuity over time (Chandrashekar and Chellappa 1991).

If our goal is to develop robust algorithms that can perform successfully in general environments, we should abandon all computational processes that are provably unstable. Any 3D motion estimation technique must make use of a representation for the image motion. Most existing algorithms rely, at this stage, on the computation of optic flow or correspondence, but the estimation of retinal correspondence is an ill-posed problem. Due to the "aperture problem," the only image motion that can in general be uniquely defined from a sequence of images is the normal flow— the projection of the optic flow on the gradient direction. It can be derived from local information assuming the conservation of some form of image information, such as the image intensity, which leads to the motion constraint equation and reduces the computation of normal flow to the computation of image gradients. The difficulty involved is only due to the discrete aspect of digital images. Computing normal flow in images is as difficult as detecting edges. The input to the 3D motion estimation algorithm introduced here is a normal flow field, even though it appears to contain less information than the optic flow field. Another reason for the sensitivity of most existing motion estimation techniques stems from the instability of the computations relating 2-D image measurements to 3-D motion and scene parameters. Especially, since most methods only employ local image information, a confusion between translation and rotation occurs. For example, in an area near the y-axis of the image plane, 3D rotation around the X-axis produces a flow field similar to the one of translation along the Y-axis (Adiv 1985).

Due to the aforementioned sensitivity issues, the fact that image measurements (also normal flow measurements) cannot be perfect and the fact that only the global use of data has the potential of leading to stable techniques has to be considered in the development of a robust algorithm. The problem is to estimate the motion parameters that describe the rotation and the direction of translation. In the chosen approach, we first employ only qualitative information to compute the direction of the rotation axis and the direction of translation. Motion rigidity introduces a number of constraints on the normal flow values. These constraints take the form of particular patterns in the image plane. In other words, for given positions of the translational and rotational axes, the normal flow values form certain patterns. The technique searches for these patterns. It uses data from different parts of the image plane and considers only the sign of the normal flow. For this reason the method for deriving the direction of the translation and rotation axes is of a qualitative and global character and can handle a considerable amount of error in the input. After having found the axis of rotation and the direction of translation further constraints are considered, and the complete set of motion parameters is obtained.

Previous qualitative approaches to 3D motion estimation are along the ideas suggested by Prazdny (1981), concentrating either on pure translation (Jain

1983) or general motion with restricted rotation (Burger and Bhanu 1990), where a global decomposition into rotation and translation is attempted by utilizing various experimentally inspired error norms. Another interesting qualitative approach (Verri and Poggio 1989; Verri, Straforini and Torre 1992) is motivated by a mathematical analysis in which first order spatial properties of the optical flow, such as singular points, are shown to be features that could be used in visual motion analysis.

Methods of estimating 3D-motion from only the normal flow field without going through the intermediate stage of computing optical flow have appeared in (Aloimonos and Brown 1984; Horn and Weldon 1987; Negahdaripour 1986). In (Aloimonos and Brown 1984) the case of purely rotational motion was studied, and linear equations relating the rotation parameters to the normal flow were derived. A similar result was reported by Horn and Weldon (1987), who presented several methods for the problem of motion and structure computation in addition to the purely rotational case, for only translation, for known rotation, and for known structure. The constraint of positive depth was used by Negahdaripour (1986) to estimate the focus of expansion for purely translational motion. In (White and Weldon 1988) translation and rotation were estimated for an observer rotating around the direction of translation, and in (Fermüller and Aloimonos 1992) the activity of dynamic tracking is used to compute the translational direction of a rigidly moving object.

The organization of this paper is as follows: In Section 2 we describe the geometry relating normal flow to three-dimensional motion. Then, we explain how to exploit these relations to develop a qualitative technique that searches for particular patterns of normal flow vectors in the image. The result of this search is a set of possible solutions for the direction of translation and the axis of rotation. Section 3 is devoted to the use of additional constraints to compute the value of the rotation and to further narrow down the possible space of solutions. If there is only one solution, the technique will find it uniquely. Section 4 is devoted to experimental results and the paper concludes with a discussion and outline of future work.

2 Geometric Constraints

To gain an insight into the problem and the difficulties involved in it we start with a brief summary of the equations relating the 3D-scene to the image measurements.

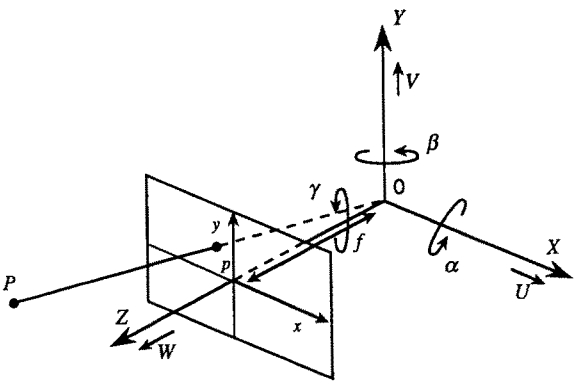


Fig. 1. Imaging geometry and motion representation.

2.1 *Formalization of the Problem*

The motion equations for a monocular observer moving in a static environment are defined by the following physical constraints: We assume that the coordinate system (X, Y, Z) is fixed to the observer with the origin 0 being the nodal point of the camera. If we denote by (U, V, W) the translational and by (α, β, γ) the rotational motion of the observer relative to the scene, then the velocity components of any point $P(X, Y, Z)$ in the image will be

$$\begin{aligned} \dot{X} &= -U - \beta Z + \gamma Y \\ \dot{Y} &= -V - \gamma X + \alpha Z \\ \dot{Z} &= -W - \alpha Y + \beta X \end{aligned} \tag{1}$$

As image formation model we use perspective projection on the plane. The image plane is parallel to the XY-plane and the viewing direction is along the positive Z axis (see Fig. 1). Under this projection the image position $p(x, y)$ of a 3D point $P(X, Y, Z)$ is defined through the relation

$$(x, y) = \left(\frac{fX}{Z}, \frac{fY}{Z}\right) \tag{2}$$

The constant f denotes the focal length of the imaging system. The equations relating the velocity (u, v) of an image point p to the 3D velocity can be derived by differentiating (1) and substituting from (2):

$$\begin{aligned} u &= \frac{(-Uf + xW)}{Z} + \alpha\frac{xy}{f} - \beta\left(\frac{x^2}{f} + f\right) + \gamma y \\ v &= \frac{(-Vf + yW)}{Z} + \alpha\left(\frac{y^2}{f} + f\right) - \beta\frac{xy}{f} - \gamma x \end{aligned} \tag{3}$$

The number of motion parameters that a monocular observer is able to compute under perspective

projection is limited to five: the three rotational parameters and the direction of translation. We therefore introduce coordinates for the direction of translation $(x_0, y_0) = (Uf/W, Vf/W)$, and rewrite the righthand side of Eq. (3) as sums of translational and rotational components.

$$\begin{aligned} u &= u_{\text{trans}} + u_{\text{rot}} \\ &= (-x_0 + x)\frac{W}{Z} + \alpha\frac{xy}{f} - \beta\left(\frac{x^2}{f} + f\right) + \gamma y \\ v &= v_{\text{trans}} + v_{\text{rot}} \\ &= (-y_0 + y)\frac{W}{Z} + \alpha\left(\frac{y^2}{f} + f\right) - \beta\frac{xy}{f} - \gamma x \end{aligned} \quad (4)$$

Since we can only compute the normal flow, the projection of the optical flow on the gradient direction (n_x, n_y), only one constraint on the actual flow can be derived at any given point. The value u_n of the normal flow vector along the gradient direction is given by

$$\begin{aligned} u_n &= un_x + vn_y \\ u_n &= \left((-x_0 + x)\frac{W}{Z} + \alpha\frac{xy}{f} \right. \\ &\quad \left. - \beta\left(\frac{x^2}{f} + f\right) + \gamma y\right)n_x \\ &\quad + \left((-y_0 + y)\frac{W}{Z} + \alpha\left(\frac{y^2}{f} + f\right)\right. \\ &\quad \left. - \beta\frac{xy}{f} - \gamma x\right)n_y \end{aligned} \quad (5)$$

This equation demonstrates the difficulties of motion computation. A monocular observer unable to measure depth is confronted with a motion field of five unknown motion parameters and one scaled depth component (W/Z) at every point. Since there is only one constraint at each point and since we do not want to make assumptions about depth, there is no straightforward way to compute the motion parameters analytically.

2.2 *Motion Field Interpretation*

A motion field is composed of a translational and a rotational component. Only the first of these is dependent on the distance from the observer. Therefore it seems reasonable to look for a way of determining the motion components by disregarding the depth components. The motion under consideration is rigid. Every point in 3D moves relative to the observer along a constrained trajectory. The rigidity constraint also imposes restrictions on the motion field in the image plane and these restrictions are reflected in the normal field as well. This is the motivation for investigating geometrical properties inherent in the normal flow field. The motion estimation problem then amounts to resolving the normal flow field into its rotational and its translational component.

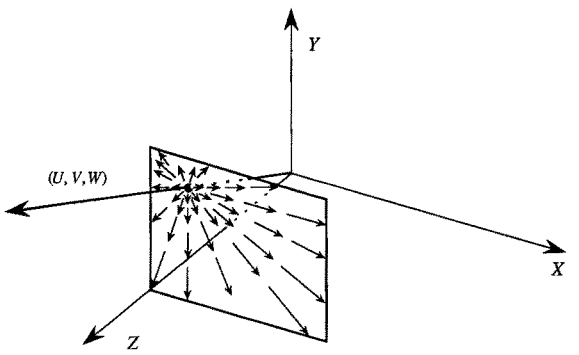


Fig. 2. Translational motion viewed under perspective projection: The observer is approaching the scene.

If the observer undergoes only translational motion, all points in the 3D scene move along parallel lines. Translational motion viewed under perspective results in a motion field in the image plane, in which every point moves along a line that passes through a vanishing point. This point is the intersection of the image plane with the translational trajectory passing through the nodal point. Its image coordinates are $x = Uf/W$ and $y = Vf/W$; the flow there has value zero. If the sensor is approaching the scene all the flow vectors emanate from the vanishing point, which is then called the Focus of Expansion (FOE) (Fig. 2). Otherwise the vectors point toward it, in which case we speak of the Focus of Contraction (FOC). The direction of every vector is determined by the location of the vanishing point; the lengths of the vectors depend on the 3D positions of the points in the scene. The vanishing point also constrains the direction of the normal flow vector at every point; it can only be in the half plane containing the optical flow vector.

In the case of purely rotational motion every point in 3D moves along a circle in a plane perpendicular to the axis of rotation. The perspective image of this circular path is the intersection of the image plane with the cone defined by the circle and the rotation axis (see Fig. 3). Depending on the relation between the aperture angle of the cone for a given image point and the angle that the image plane forms with the rotation axis, different second order curves are obtained for the intersection: ellipses, hyperbolas, parabolas, and even circles when

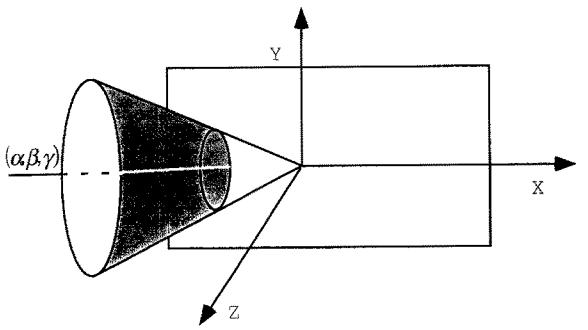


Fig. 3. The intersection of the image plane with the cone (determined by the circular path in 3D and the rotation axis) defines the projection of rotational motion on the image plane.

the rotation axis and the optical axis coincide. The specific conic sections due to rotational motion are defined by the axis of rotation. The rotation axis given by the two parameters $(\frac{\alpha}{\gamma})$ and $(\frac{\beta}{\gamma})$, defines the family $M(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma}; x, y)$ of conic sections:

$$
\begin{aligned}
&M\left(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma}; x, y\right) \\
&\quad = \left(\frac{\alpha^2}{\gamma^2}x^2 + 2xy\frac{\alpha}{\gamma}\frac{\beta}{\gamma} + y^2\frac{\beta^2}{\gamma^2} + 2xf\frac{\alpha}{\gamma}\right. \\
&\qquad \left. + 2yf\frac{\beta}{\gamma} + f^2\right) \Big/ (x^2 + y^2 + f^2) \\
&\quad = C \text{ with } C \text{ in } \left[0, \ldots, \left(1 + \frac{\alpha^2}{\gamma^2} + \frac{\beta^2}{\gamma^2}\right)\right] \qquad (6)
\end{aligned}
$$

Specifically, for a rotation around the Z-axis the second order curves are circles with center 0; we call them γ-circles (Fig. 4a). If the rotation axis is the X- or Y-axis the conic sections are hyperbolas whose axes coincide with the coordinate axes of the image plane. For the case of rotation around the X-axis the hyperbolas' major axis is the x-axis and they are called α-hyperbolas (Fig. 4b). For rotation around the Y-axis the major axis is the y-axis and we call the conic sections β-hyperbolas (Fig. 4c).

2.3 *Selection of Values*

A motion vector consists of a rotational component which can be parametrized by three unknowns and a translational vector which is everywhere directed away from (or towards) a point. However, the estimates we can compute at every point are only projections of the motion vector on the gradient direction.A general method of breaking up the normal flow vector at every point into its components does not seem to be possible, but there is a way of separating the components for vectors in certain directions.

The value of the normal flow at a point is the scalar product of the flow vector and the unit vector in the gradient direction. The right hand side of Eq. (5) can be written as a sum of scalar products by separating the translational components from the single rotational components around each of the coordinate axes:

$$
\begin{aligned}
u_n = &\frac{W}{Z}((-x_0 + x), (-y_0 + y))(n_x, n_y) \\
&+ \alpha\left(\frac{xy}{f}, \left(\frac{y^2}{f} + f\right)\right)(n_x, n_y) \\
&- \beta\left(\left(\frac{x^2}{f} + f\right), \frac{xy}{f}\right)(n_x, n_y) \\
&+ \gamma(y, -x)(n_x, n_y) \qquad (7)
\end{aligned}
$$

If two vectors are perpendicular to each other, their scalar product is zero. Thus, for normal flow vectors in particular directions one or more of the motion components may vanish. In particular, all the normal flow vector that form right angles with the γ-circles do not

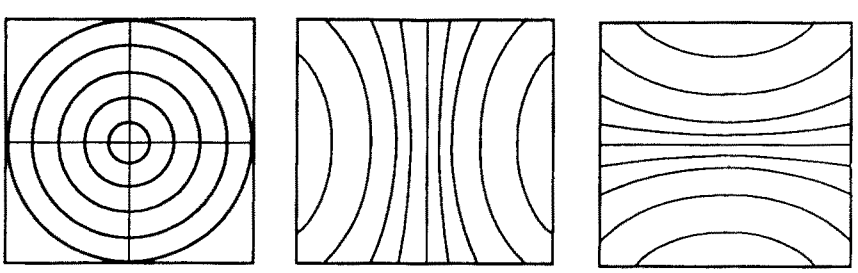


Fig. 4. Rotation around the Z-, X- or Y-axis gives rise in the image plane to γ-circles (a), α-hyperbolas (b), or β-hyperbolas (c).

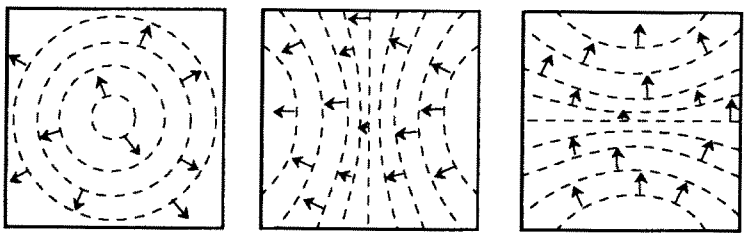


Fig. 5. Positive γ-, α-, and β-vectors.

contain a component due to rotation around the Z-axis. Similarly, there is no motion component due to rotation around the X-axis for the normal flow vectors perpendicular to the α-hyperbolas, and no component due to rotation around the Y-axis for vectors perpendicular to the β-hyperbolas. The motion estimation problem becomes easier when analyzing only the normal flow vectors perpendicular to one of these families of conic sections. It is reduced by one parameter for these subsets of the normal flow vectors.

We call these three subsets of the normal flow vectors the α-, β-, and γ-vectors. It is convenient to agree upon conventions for the vectors' orientations. A γ-vector at point (x, y) is said to have positive orientation if it is pointing in the direction (x, y); otherwise, its orientation is said to be negative. Similarly, we call an α-vector (or a β-vector) originating from a point (x, y) positive if it points in the direction $(-(f^2 + y^2), xy)$ (or $(xy, -(f^2 + x^2))$) (see Fig. 5).

2.4 *Properties of the Subsets*

Let us first concentrate on the γ-vectors. These vectors do not contain a component due to rotation around the Z-axis. Along the positive direction, the two remaining rotational components contribute

$$v_{\text{rot}}(r, \phi) = -\alpha(r^2/f + f)\sin\phi + \beta(r^2/f + f)\cos\phi,$$

where r is the distance from the image center and the angle ϕ is measured from the x-axis. Thus, the rotational component of the normal flow along a vector pointing away from the image center can be described by a trigonometric function with amplitude $\max(\alpha, \beta)$ and period 2π. Along the line which passes through the image center and makes an angle $\phi = \arctan(\beta/\alpha)$ with the x-axis the values are zero. This line divides the plane into two halves. In one half the vectors point in the positive direction, and in the other half they point in the negative direction; in the future we simply refer to them as positive and negative vectors (Fig. 6a).

The translational component of the motion field is characterized by the location of the FOE or FOC in the image plane. In (Fermüller and Aloimonos 1992) a qualitative method is described which can be applied to quickly distinguish whether an object is coming closer or moving away. This allows us to restrict our description to the approaching case; the extension to the opposite case is obvious.

The γ-vectors lie on lines passing through the image center and the optical flow values due to translation lie on lines passing through the FOE. These two lines are at right angles for all points on a circle which has the FOE and the image center as diametrical opposite points. This circle given as

$$k(x_0, y_0; x, y) = \left(x - \frac{x_0}{2}\right)^2 + \left(y - \frac{y_0}{2}\right)^2 - \left(\frac{x_0}{2}\right)^2 - \left(\frac{y_0}{2}\right)^2 = 0$$

defines the geometric locus of all points on the γ-circles where the γ vectors' translational components vanish. Thus, the diameter of this circle is the line segment connecting the image center and the FOE. At all points inside this circle the two lines enclose an angle greater than 90° and the normal flow along the γ-vector therefore has a negative value. The normal flow values outside the circle are positive (Fig. 6b).

In order to investigate the constraints associated with a general motion, the geometrical relations derived from rotation and from translation have to be combined. A circle separating the plane into positive and negative values and a line separating the plane into two half-planes of opposite sign always intersect (in two points or one point in case the line is tangential to the circle), because both the line and the circle pass through the

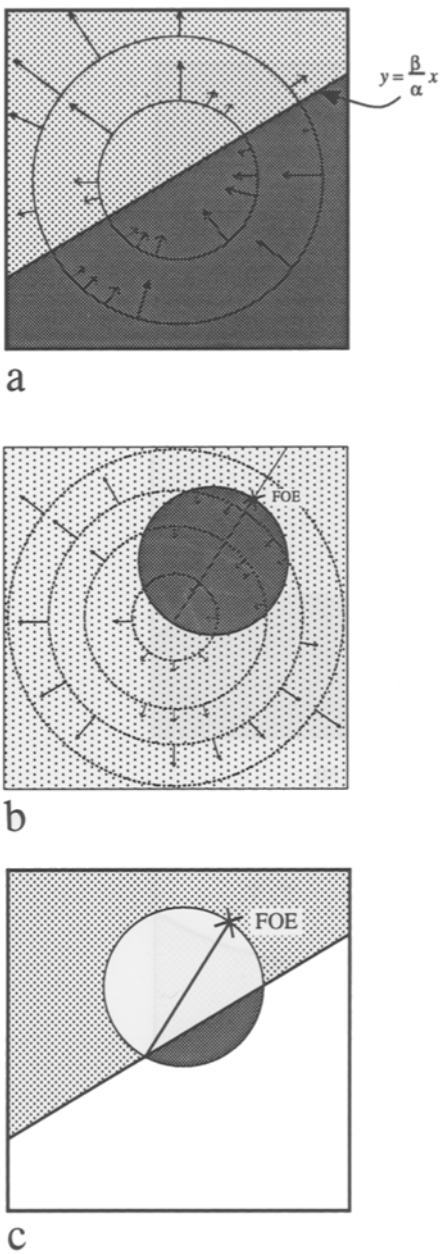


Fig. 6. (a) The γ-vectors due to rotation separate the image plane in a halfplane of positive values and a halfplane of negative values. (b) The γ-vectors due to translation are negative, if they lie within the circle defined by the FOE and the image center and are positive at all other locations. (c) A general rigid motion defines an area of positive γ-vectors and an area of negative γ-vectors. The rest of the image plane is not considered.

origin. This splits the plane into areas of only positive, of only negative γ-vectors, and into areas in which the rotational and translational flows have opposite signs. In the latter areas, unless we make depth assumptions, no information is derivable (Fig. 6c).

We thus obtain the following geometrical result for the case of general motion. Points in the image plane at which the gradient direction is perpendicular to circles around the image center can be separated into two classes. For a given FOE, and for a line through the image center which represents the quotient of two of the three rotational parameters, there are two geometrically defined areas in the plane, one containing positive and one containing negative values. We call this structure on the γ-values the γ-pattern. It depends on the three parameters x_0, y_0 and $\frac{\beta}{\alpha}$.

Similar relations can be derived when eliminating the motion components due to rotation around the X- and Y-axes.

The α- and β-vectors due to rotation are also separable into positive and negative vectors. In both cases the locus of zero normal flow which separates the two classes is a line. For the α-hyperbolas the line is parallel to the x-axis and is defined by the equation $y = \frac{\beta f}{\gamma}$ (Fig. 7a); for the β-hyperbolas it is parallel to the y-axis and is defined by $x = \frac{\alpha f}{\gamma}$ (see Fig. 7b).

The translational components of the α- and β-vectors are separated by hyperbolas. The α-vectors, which are perpendicular to lines through the FOE, and which therefore have zero normal flow lie on a hyperbola of the form

$$f(x_0, y_0; x, y) = x_0 y^2 - xyy_0 - xf^2 + x_0 f^2 = 0$$

When $f(x_0, y_0) > 0$, the normal flow values are positive; in the other part of the plane they are negative (Fig. 7c). Symmetrical relations hold for the β-vectors. The curve of zero normal flow is defined by

$$g(x_0, y_0; x, y) = x^2 y_0 - x_0 xy - yf^2 + y_0 f^2 = 0$$

and in areas of positive $g(x_0, y_0)$ the β-vectors are positive (see Fig. 7d).

The superposition of translational and rotational values again defines patterns in the plane each of which consists of a negative and a positive area. These patterns, called α- and β-patterns, are uniquely described by three parameters: x_0 and y_0, the coordinates of the FOE and the quotients $\frac{\beta}{\gamma}$ (and $\frac{\alpha}{\gamma}$) (Figs. 7e and 7f).

3 The Method

The estimation of motion for a rigid moving observer is performed by four modules. The strategy involves

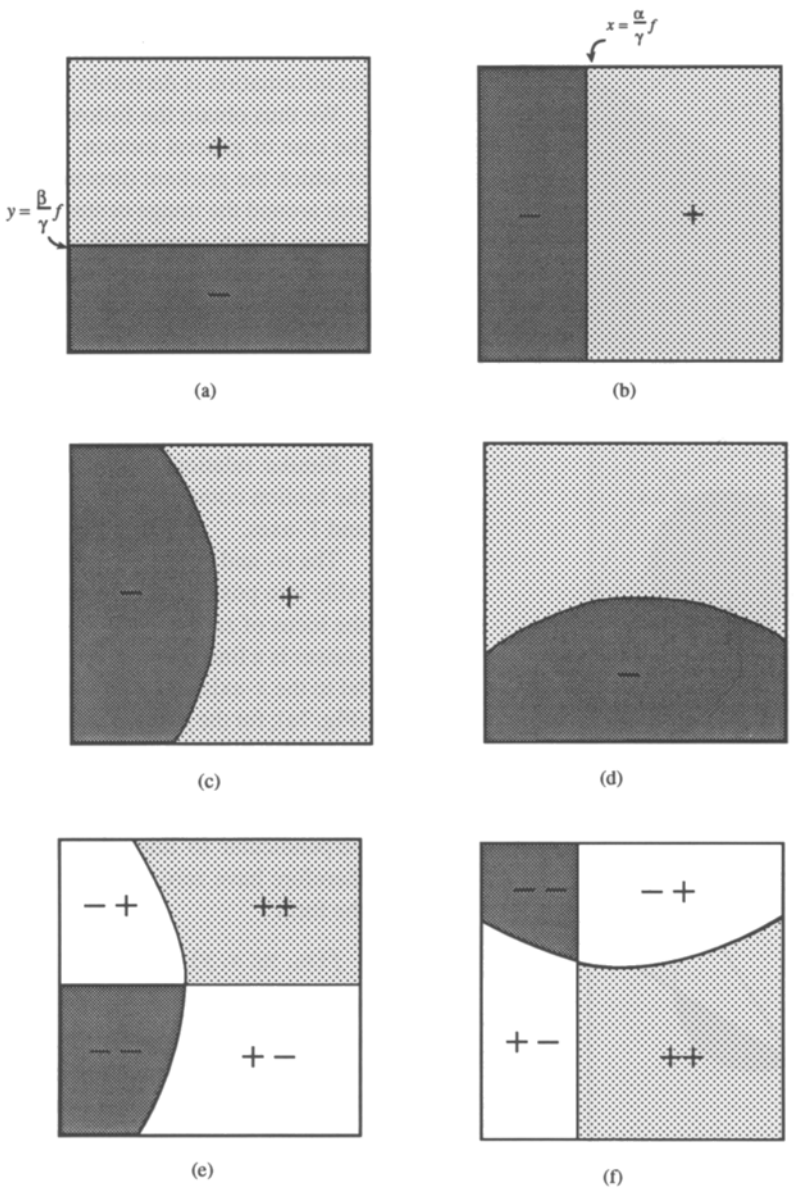


Fig. 7. (a) The α-vectors due to rotation are separated into positive and negative values by a straight line parallel to the x-axis. (b) The β-vectors due to rotation are separated by a straight line parallel to the y-axis. (c) (d) Hyperbolas separate the α- and β-vectors due to translation into areas of positive and negative values. (e) (f) α- and β-patterns for general rigid motion.

checking the constraints that a given solution would impose on the normal flow field and in this way discarding impossible solutions. From the first to the fourth module the constraints become more restrictive; hence the number of possible solutions computed by each module decreases. In the first module patterns are fitted to subsets of the normal flow field to search for the set S_1 of possible solutions for the direction of translation (FOE) and the direction of the axis of rotation. The number of these candidate solutions is reduced to a set S_2 in the second module by fitting another pattern to selected normal flow vectors which are not dependent on rotation. These pattern fitting processes use the input in a qualitative way; since only the sign of the normal flow is employed.[1] In the third module the third rotational parameter is computed from the normal flow

vectors that do not contain translational components and the space of solutions is further narrowed to a set S_3. Finally, the fourth module eliminates all impossible solutions by checking the validity of the motion parameters at every point and gives as output the set S_4.

3.1 *Pattern Fitting: Search in 3D Parameter Spaces*

The geometrical constraints developed in Section 3 are used in a search process to estimate the directions of the translation and the rotational axis. Finding these two directions is a four-dimensional problem, but through selection of values (the α-, β- and γ-vectors) and use of geometrical constraints the problem is reduced in dimensionality. With each subset of the normal flow vectors is associated a three-dimensional parameter space that spans the possible locations of the FOE and of a line defined by the quotient of two of the three rotational parameters. Instead of searching a four-dimensional space, three three-dimensional subspaces are searched for the solution.

The search in the three-dimensional subspaces is accomplished by checking the patterns which the subspaces' parameter triples define on selected values of the normal flow field. The α-patterns are fitted to the α-vectors; this provides possible solutions for the coordinates of the FOE: x_0, y_0, and the quotient $\frac{\beta}{\gamma}$. Similarly, the fitting of the β- or γ-patterns yields solutions for x_0, y_0 and $\frac{\alpha}{\gamma}$ or $\frac{\beta}{\alpha}$. The objective is to find the four parameters defining the directions of the translational and rotational axes which give rise to three successfully fitted patterns. Therefore the three subspaces' patterns are combined and the parameter quadruples which define possible solutions are determined. Since only subsets of the normal flow values are considered in the fitting process, the fitting does not uniquely define the motion, but just constitutes a necessary condition. Usually a number of parameter quadruples will be $\{x_0, y_0, \alpha/\gamma, \beta/\gamma\}$ that are selected as candidate solutions through pattern fitting.

In the general case none of the three translational and three rotational parameters is equal to zero. Then the FOE and the rotation center (the intersection of the rotation axis with the image plane) lie in a bounded area of the image plane and the three three-dimensional subspaces are also bounded.

The method can also deal with cases of one or more parameters of value zero but the search has to be extended by using additional patterns. If there is only translation then the α- β- and γ-patterns split the image plane into the insides and outsides of circles or hyperbolas which are of opposite sign. In cases of only rotation the pattern consists of an area of negative and an area of positive values separated by a line of zero rotational normal flow. If one or both of the translational parameters U and V are zero, then the FOE lies on the x- or y-axis; this case does not have to be considered separately. A translational value W of zero causes a degeneration of the γ-pattern's circle into a halfplane and of the α- and β-patterns' hyperbolas into simpler hyperbolas of the form

$$f(\infty, \infty; x, y) = y^2 - \frac{y_0}{x_0}xy + f^2 = 0$$

$$g(\infty, \infty; x, y) = x^2 - \frac{x_0}{y_0}xy + f^2 = 0$$

where $\frac{x_0}{y_0}$ is the direction of translation in the plane parallel to the image plane. If one or two of the rotational parameters vanish, this will result for the γ-pattern either in an nonexistent line of rotation or in a line which is parallel to the x- or y-axis. For the α- and β-patterns the rotation lines pass through the center or lie at infinity. For example, in case of zero α- and γ-values, the γ-pattern's rotation line is the y-axis and both the α- and β-patterns' rotation lines are at infinity.

In order to make the method work for any rigid motion, the above described patterns have to be searched for in addition to the patterns defined by the three three-dimensional subspaces.

3.2 *Partial Derotation*

Suppose, we want to test whether a quadruple $(x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$ given by the first module is a correct solution. Since we know the direction of the rotation axis $(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$, we can compute the field lines of the rotational vector field (i.e. the lines which have the property that at each point the rotational flow is tangential). As described in Section 2.2 the second order curves $M(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma}; x, y)$ are given through Eq. (6). The normal flow vectors perpendicular to $M(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma}; x, y)$ are only due to translation. We call these normal flow vectors the "rotation-axis vectors" and define a vector emanating from a point (x, y) to be of positive orientation if it is pointing in the direction $((-\frac{\alpha}{\gamma}(y^2+f^2)+\frac{\beta}{\gamma}xy+xf), (\frac{\alpha}{\gamma}xy-\frac{\beta}{\gamma}(x^2+f^2)+yf))$ (see Fig. 8). The signs of the rotation-axis vectors are defined by the location of the FOE.

As in the case of the α-, β- and γ-vectors a second order curve separates the plane into an area of positive and an area of negative rotation-axis vectors and therefore defines another pattern in the image plane, the "rotation-axis pattern" pattern (see Fig. 9). The curve

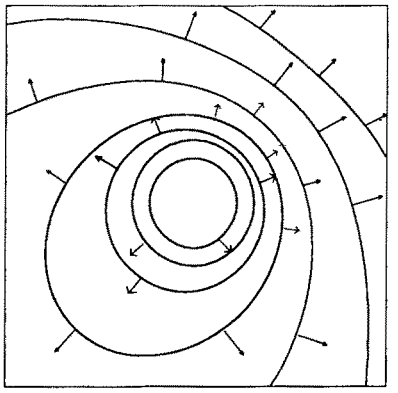


Fig. 8. Field lines of a rotational vector field and positive rotation axis vectors.

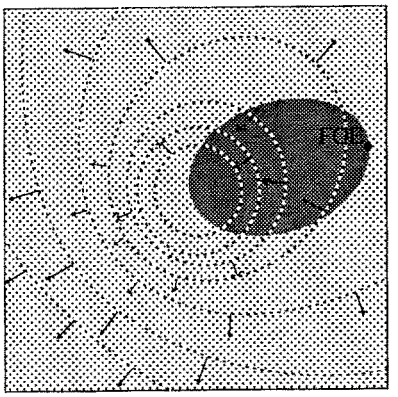


Fig. 9. Rotation-axis pattern.

$h(x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$ which contains rotation-axis vectors of value zero is given by the equation

$$h\left(x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma}\right) = x^2\left(f + \frac{\beta}{\gamma}y_0\right) + y^2\left(f + \frac{\alpha}{\gamma}x_0\right) - xy\left(\frac{\alpha}{\gamma}y_0 + \frac{\beta}{\gamma}x_0\right) - xf\left(\frac{\alpha}{\gamma}f + x_0\right) - yf\left(\frac{\beta}{\gamma}f + y_0\right) + f^2\left(\frac{\alpha}{\gamma}x_0 + \frac{\beta}{\gamma}y_0\right) = 0 \tag{8}$$

By considering only the rotation-axis vectors, we achieve derotation for a subset of the normal flow vectors without actually subtracting rotational values. Thus a fourth set of normal flow vectors can be used for further reducing the set of candidate solutions for the axes of translation and rotation computed in the first module. For every quadruple of the set S_1 we find the rotation-axis vectors defined by $(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$ and test if each vector's sign is consistent with the sign defined by the rotation-axis pattern due to $(x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$. All quadruples that lead to a successful fitting of the corresponding rotation-axis pattern are kept as possible solutions in the set S_2.

3.3 Detranslation

Proper selection of normal flow vectors also makes it possible to eliminate the normal flow's translational components. If the location of the FOE is given the directions of the translational motion components are also known. The optical flow vectors lie on lines passing through the FOE. The normal flow vectors perpendicular to these lines do not contain translational components; thus they have only rotational components. This can be seen from Eq. (7). If the selected gradient direction at a point (x, y) is $((y_0 - y), (-x_0 + x))$ the scalar product of the translational motion component and a vector in the gradient direction is zero. In the third module, this method of eliminating the translational component, in the future referred to as "detranslation", is used to compute the third rotational component and to further reduce the possible number of solutions.

For each of the possible solutions computed in the second module the normal flow vectors perpendicular to the lines passing through the FOE have to be tested to determine if they are really only due to rotation (see Fig. 10). This results in solving an overdetermined system of linear equations. Since two of the rotational parameters are already computed, there is only one unknown, the value γ. Every point supplies an equation of the form

$$\gamma = u_n \Big/ \left(\frac{\alpha}{\gamma}\left(\frac{xy}{f}n_x + \left(\frac{y^2}{f} + f\right)n_y\right) - \frac{\beta}{\gamma}\left(\left(\frac{x^2}{f} + f\right)n_x + \frac{xy}{f}n_y\right) + (yn_x - xn_y)\right) \tag{9}$$

If the chosen normal flow vectors are due only to rotation then the solution to the overdetermined system gives the γ value. In a practical application a threshold has to be chosen to discriminate between possible and impossible solutions. The value of the residual is used to confirm the presumption that the selected normal

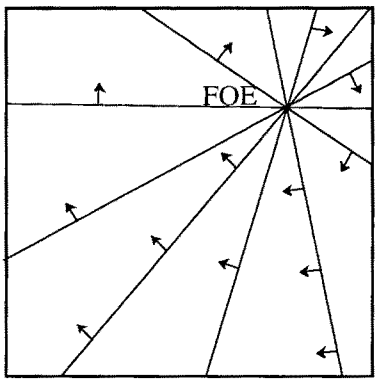


Fig. 10. Normal flow vectors perpendicular to lines passing through the FOE are only due to rotation.

flow values are purely rotational. Usually "detranslation" will not result in only one solution, but will provide a set S_3 of possible parameter quintuples.

3.4 *Complete Derotation*

In the fourth module the elements of the set S_3 are examined for further constraints. The modules described so far considered only subsets of the normal flow vectors. In order to eliminate all motion parameters which are not consistent with the given normal flow field, every normal flow vector has to be checked.

This check is performed using a "derotation" technique. For every parameter quintuple of S_4 a possible FOE and a rotation is defined. The three rotational parameters are used to derotate the normal flow vectors by subtracting the rotational component. At every point the flow vector $(u_{\text{der}}, v_{\text{der}})$ is computed:

$$\begin{aligned} u_{\text{der}} &= u_n n_x - u_{\text{rot}} n_x \\ v_{\text{der}} &= u_n n_y - v_{\text{rot}} n_y \end{aligned} \tag{10}$$

If the parameter quintuple defines the correct solution, the remaining normal flow is purely translational. Thus the corresponding optic flow field consists of vectors that all point away from one point, the FOE. Since the direction of optical flow for a given FOE is known, the possible directions of the normal flow vectors can be determined. The normal flow vector at every point is confined to lie in a half plane (see Fig. 11). The technique checks all points for this property and eliminates solutions that cannot give rise to the given normal flow field.

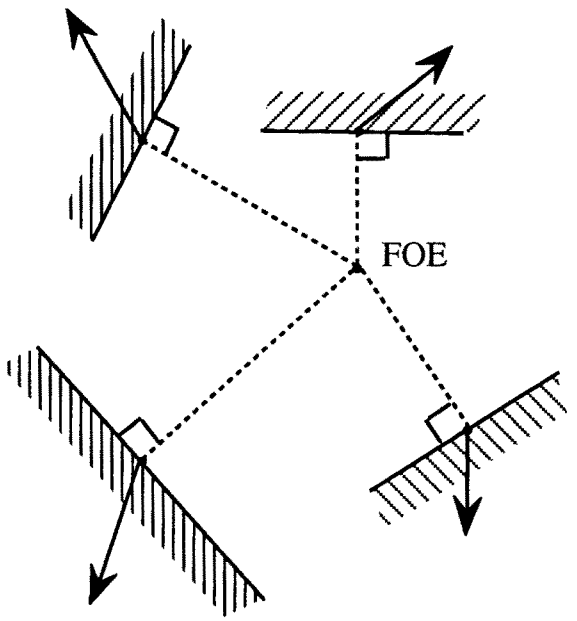


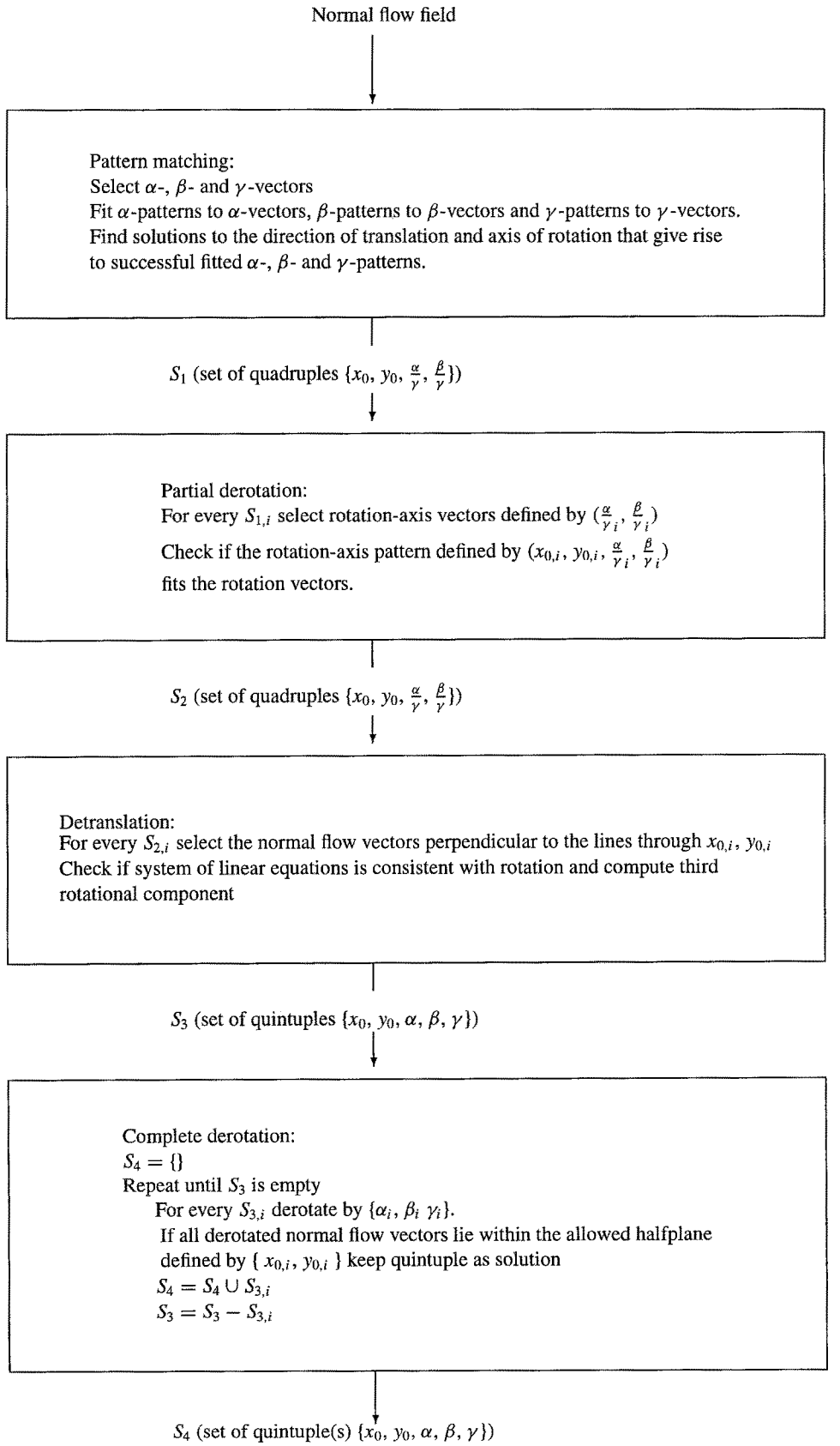
Fig. 11. Normal flow vectors due to translation are constrained to lie in halfplanes.

3.5 *The Algorithm*

In this section we summarize the complete algorithm in form of a block diagram (see the next page). The sets of candidate solutions which are determined in the four modules are called S_1, S_2, S_3, and S_4. To denote single solutions or single parameters, subscripts are used: $S_{1,i}, S_{2,i}, x_{0,i}, y_{0,i}$, etc. The input to the algorithm is a normal flow field and the outputs are all possible solutions (direction of translation and rotation) which could give rise to this normal flow field.

The complexity of the algorithm is $O(n^3)$ in the size of the image because of the complexity of pattern matching. As explained later, in the implementation search is carried out in a straightforward uninformed manner. However, one could envision a massively parallel implementation with a polynomial number of processors that reduces the complexity of the algorithm to a constant time.

It can easily be shown that normal flow fields, in general, are not unique. In fact, for any two flow fields a common normal flow field can be constructed. Consider two different normal flow fields that arise from different scenes and different observer motions. At every point in the image plane there exist two motion vectors. A normal flow vector, which is defined as the projection of a flow vector, is constrained to lie on a circle. The intersection of the two circles defines a


Normal flow field
Pattern matching:
Select α-, β- and γ-vectors
Fit α-patterns to α-vectors, β-patterns to β-vectors and γ-patterns to γ-vectors.
Find solutions to the direction of translation and axis of rotation that give rise to successful fitted α-, β- and γ-patterns.
S_1 (set of quadruples $\{x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma}\}$)
Partial derotation:
For every $S_{1,i}$ select rotation-axis vectors defined by $(\frac{\alpha}{\gamma_i}, \frac{\beta}{\gamma_i})$
Check if the rotation-axis pattern defined by $(x_{0,i}, y_{0,i}, \frac{\alpha}{\gamma_i}, \frac{\beta}{\gamma_i})$
fits the rotation vectors.
S_2 (set of quadruples $\{x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma}\}$)
Detranslation:
For every $S_{2,i}$ select the normal flow vectors perpendicular to the lines through $x_{0,i}$, $y_{0,i}$
Check if system of linear equations is consistent with rotation and compute third rotational component
S_3 (set of quintuples $\{x_0, y_0, \alpha, \beta, \gamma\}$)
Complete derotation:
$S_4 = \{\}$
Repeat until S_3 is empty
For every $S_{3,i}$ derotate by $\{\alpha_i, \beta_i\ \gamma_i\}$.
If all derotated normal flow vectors lie within the allowed halfplane defined by $\{ x_{0,i}, y_{0,i} \}$ keep quintuple as solution
$S_4 = S_4 \cup S_{3,i}$
$S_3 = S_3 - S_{3,i}$
S_4 (set of quintuple(s) $\{x_0, y_0, \alpha, \beta, \gamma\}$)

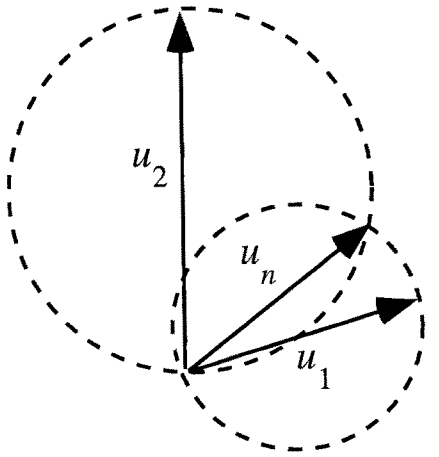


Fig. 12. The intersection of two circles defines the possible location of the normal flow vector which corresponds to two different optical flow vectors.

normal flow vector which is compatible with both motions (Fig. 12).

The algorithm determines the complete set of solutions. If for a given normal flow field the algorithm finds more than one solution, then from that normal flow field alone the 3D motion cannot be determined uniquely. In this case one can use matching of prominent features to eliminate the incorrect motion parameters, in the following manner:

If the quintuple $(x_{0,i}, y_{0,i}, \alpha_i, \beta_i, \gamma_i)$ for $i = 1 \cdots n$ is a candidate solution from the set of possible solutions, then from these parameters and the normal flow values the optical flow values can be computed. If (u, v) is the optical flow value at point (x, y) and $u_{\text{trans}}, v_{\text{trans}}, u_{\text{rot}}, v_{\text{rot}}$ the translational and rotational parts of the flow respectively, the following linear equations determine the optical flow (u, v) completely.

$$\frac{u - u_{\text{rot}}}{v - v_{\text{rot}}} = \frac{u_{\text{trans}}}{v_{\text{trans}}}$$
$$u_n = un_x + vn_y \qquad (11)$$

The matching of prominent features can then be used to verify the correctness of the optical flow and thus whether $(x_{0,i}, y_{0,i}, \alpha_i, \beta_i, \gamma_i)$ should be rejected.

4 Experiments

In a series of experiments the four modules of the egomotion recovery strategy were tested. In the implementation of these modules the following approach was taken: The elimination of impossible parameters from the space of solutions involves discrimination on the basis of quantitative values. This was realized in the following way: Normal flow values in certain directions are selected, if they are within a tolerance interval. In the pattern matching-, partial derotation-, and complete derotation-modules pattern fitting is basically implemented as a template matching, where the quality of the fitting, the "success rate", is measured by the number of values with correct signs normalized by the total number of selected values. The amount of rotation in the detranslation module is computed as an average of the values derived at every point and the discrimination between accepted and rejected motion parameters is based on the value of the standard deviation.

In the first (pattern matching) and second (partial derotation) module no quantitative use of values is made, since only the sign of the normal flow is considered. This limited use of data makes the modules very robust, and the correct solutions for the axes of translation and rotation are usually found even in the presence of high amounts of noise. To give some quantitative justification of this we define the error in the normal flow at a point as a percentage of the correct vector's length. Since the sign of the vector is not affected as long as the error does not exceed the correct vector in value, the "pattern fitting" will find the correct solution in all cases of up to 100% error.

Several experiments have been performed on synthetic data. For different 3D motion parameters normal flow fields were generated; the depth value (in an interval) and the gradient direction were chosen randomly. In all experiments on noiseless data the correct solution was found as the best one. Figure 13 shows the optical flow field and the normal flow field for one of the generated data sets: The image size is 100×100, the focal length is 150, the image coordinates of the FOE are $(-5, +30)$, and the relationship of the rotational components is $\alpha{:}\beta{:}\gamma = 10{:}11{:}150$. In Fig. 14 the fitting of the circle and the hyperbolas to the α-, β-, and γ-vectors and the rotation-axis vectors is displayed. The tolerance interval for the direction of normal flow vectors was taken to be 10°. Points with positive normal flow values are rendered in a light color and points with negative values are dark. Perturbation of the normal flow vectors' lengths by up to 50% did not prevent the method from finding the correct solution.

Figures 15 and 16 show in a graphical form the intermediate and final results of the algorithm demonstrating how uncertainty decreases through the successive application of the four modules. Since not all five dimensions of the algorithm's output can be displayed,

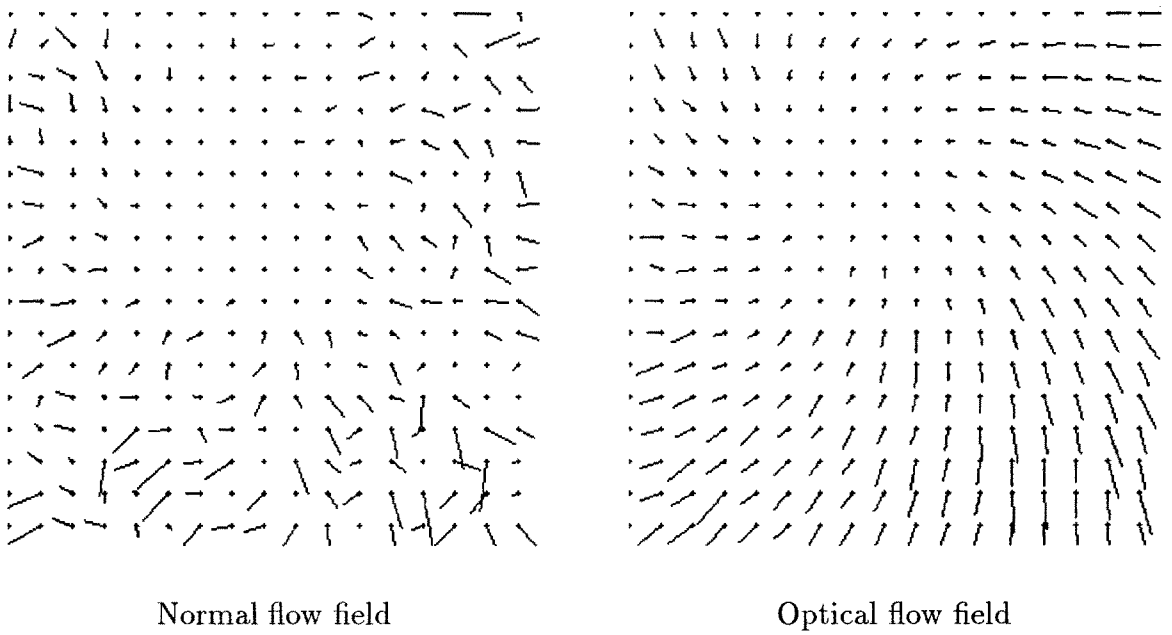


Fig. 13. Flow fields of synthetic data.

only values for the FOE are presented. The figures show two different areas of the plane where the coordinates of the FOE have been searched for, with the success rate in the three modules where pattern matching is performed displayed as a grey value—white denoting the highest and black the lowest value. For the case of detranslation, the standard deviation derived from the computation of the third rotational component is rendered with the smallest value displayed in the lightest grey shade. (Although the whole half sphere of all possible directions has been searched, for the sake of clarity of the pictorial description only a part of the search space on the plane is displayed). For every candidate location of the FOE the rotation leading to the highest success rate (smallest standard deviation) has been chosen to derive the grey value at the location considered. In Figs. 15a,b,c, and d the results of the synthetic experiment described above are displayed for a search area equal to the size of the image ($[-50, 50] \times [-50, +50]$). Figure 15a shows the result of pattern fitting, 15b the result of partial derotation, 15c the result of detranslation, and finally Fig. 15d shows the result of total derotation and thus the output of the complete algorithm. In this presentation every fifth pixel is displayed. Figures 16a–d display results for the same data, but in this case the search area is $[-155, +145] \times [-120, -180]$ while the image is again of size $[-50, 50] \times [-50, +50]$. (The image is displayed as a black rectangle in Fig. 16d.) In this presentation every fifteenth pixel is displayed.

As a first example of a real scene the NASA-Ames sequence[2] was chosen. The camera undergoes only translational motion; thus different amounts of rotation were added. For all points at which translational motion can be found the rotational normal flow is computed, and the new position of each pixel is evaluated. The "rotated" image is then generated by computing the new greylevels through bilinear interpolation. The images were convolved with a Gaussian of kernel size 5×5 and standard deviation $\sigma = 1.4$. The normal flow was computed by using 3×3 Sobel operators to estimate the spatial derivatives in the x- and y-directions and by subtracting the 3×3 box-filtered values of consecutive images to estimate the temporal derivatives.

When adding rotational normal flow on the order of a third to three times the amount of the translational flow, the exact solution was always found among the best fitted parameter sets. The solution could not clearly be derived as a unique point in the five-dimensional parameter space; rather we obtained a number of solutions that form a "fuzzy blob" in the solution space (see Fig. 18). All solutions with success rates higher than 99% were very close to the correct one with the FOE deviating by at most 6% of the focal length from the

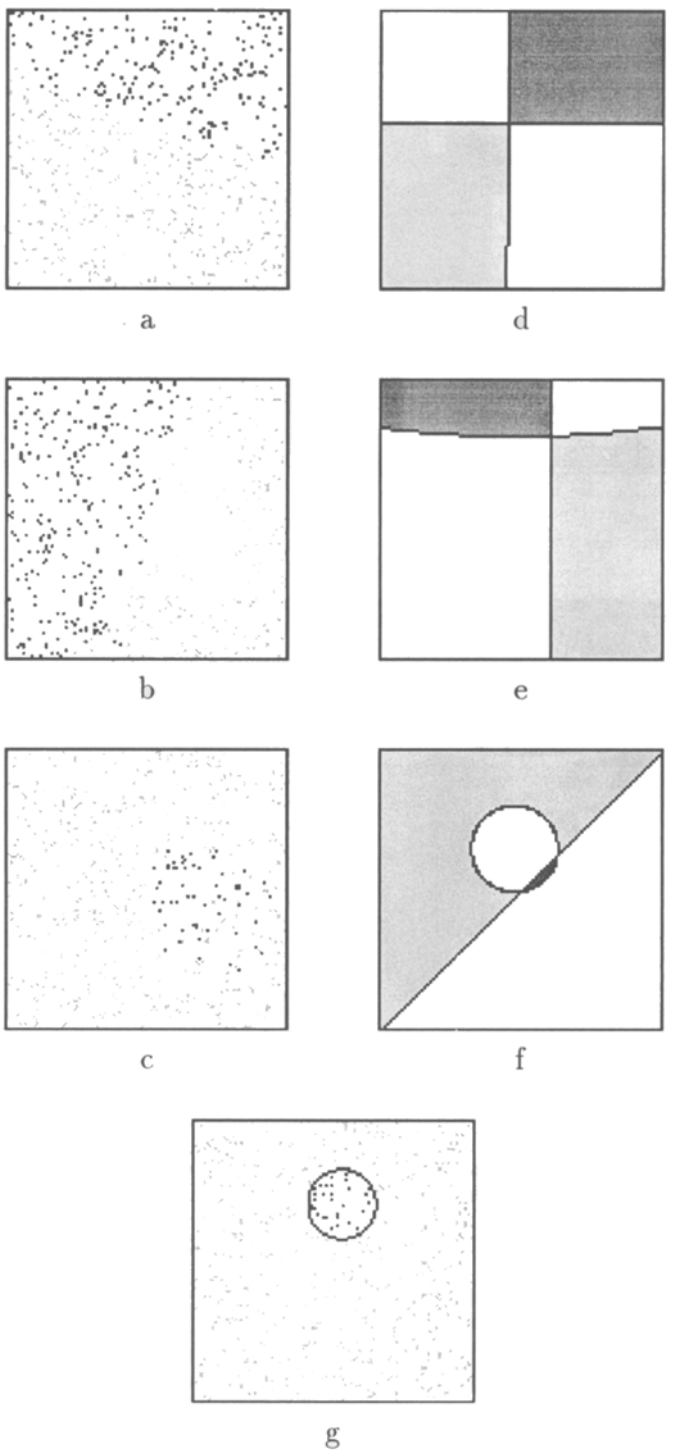


Fig. 14. (a), (b), (c): Positive and negative α-, β-, and γ-vectors. (d), (e), (f) Fitting of α-, β-, and γ-patterns. (g): Fitting of rotation axis pattern.

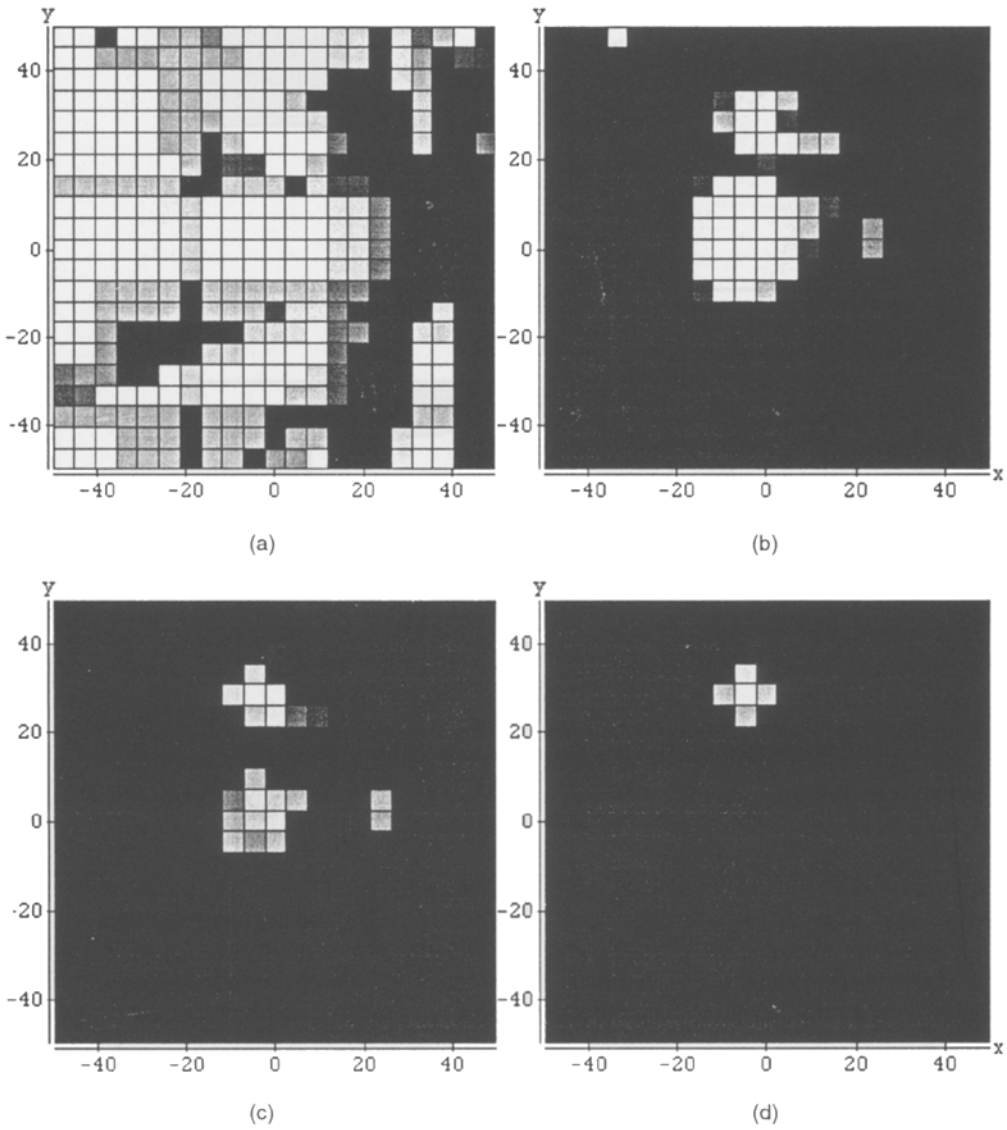


Fig. 15. Synthetic data: Intermediate and final results of the algorithm displayed for a search are for the FOE equal to the size of the image. (a) Results after pattern matching. (b) Results after partial derotation. (c) Results after detranslation. (d) Final result.

correct positions (x_0, y_0). It should be explained here that the algorithm in its first two modules is a search procedure and it is carried out in a discrete manner. It searches for a quadruple (axis of translation, axis of rotation). Then, in the "detranslation" module, the complete rotation is computed quantitatively. As opposed to the previous steps, during this step small errors can occur. What is displayed in Fig. 18 is the projection of the cluster of solutions in the five-dimensional motion space on the image plane. If we chose the "success rate" threshold to be lower than 99%, all solutions would be connected. If we chose it higher, only the cluster in the center would appear. In Fig. 17 the computed normal flow vectors and the fitting of the α-, β- and γ-vectors for one of the "rotated" images is shown. Areas of negative normal flow vectors are marked by horizontal lines and areas of positive values with vertical lines. The ground truth for the FOE is $(-5, -8)$, the focal length is 599 pixels, and the rotation between the two image frames is $\alpha = 0.0006$, $\beta = 0.0006$, and $\gamma = 0.004$. The algorithm computed the solution exactly. Figure 19 shows, overlaid

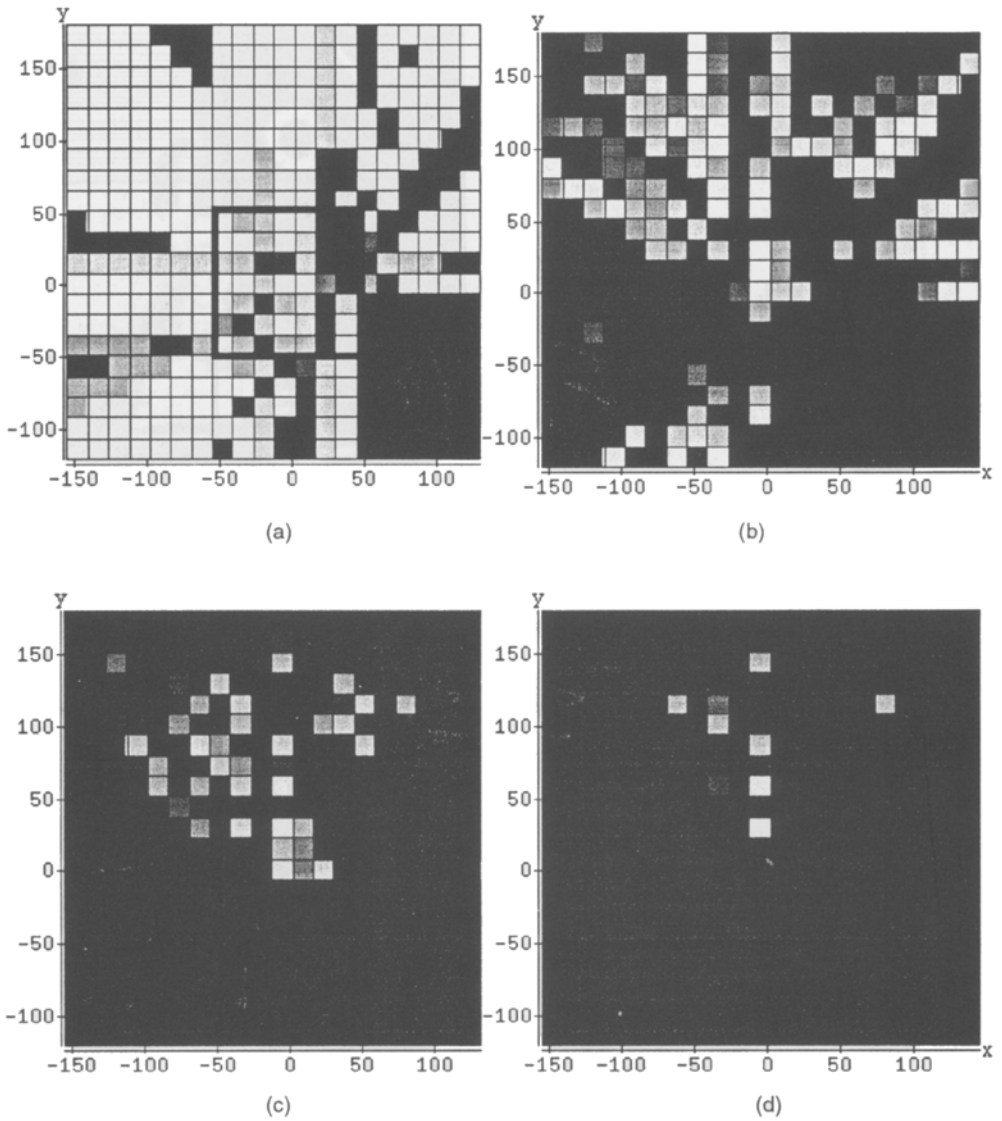


Fig. 16. Synthetic data: Intermediate and final results of the algorithm displayed for a search area for the FOE three times the size of the image. (a) Results after pattern matching. (b) Results after partial derotation. (c) Results after detranslation. (d) Final result.

on the original image in black, the curves and lines separating positive and negative α-, β-, and γ-vectors. (Due to the large focal length the parts of the α- and β-hyperbolas which appear in the image plane are close to straight lines.) The curve separating the rotation-axis vectors is rendered in white. This is to visualize the fact that the intersection of the second order curves gives the FOE, and the intersection of the straight lines and the curve separating the rotation-axis vectors gives the point where the axis of rotation pierces the image plane.

A second series of experiments was performed on a series of images containing translation and rotation around the X- and Y-axes (table top scene). The camera translated along a straight line, while at the same time rotating around the X- and Y-axes. The parameters of the camera were as follows: focal length in X-direction: 1163 pixels, focal length in Y-direction: 1316 pixels, image size: 574×652, center of the image: (332, 305) (measuring from the bottom left corner).

The algorithm was run on three different sequences. In all cases it computed the axes of rotation and

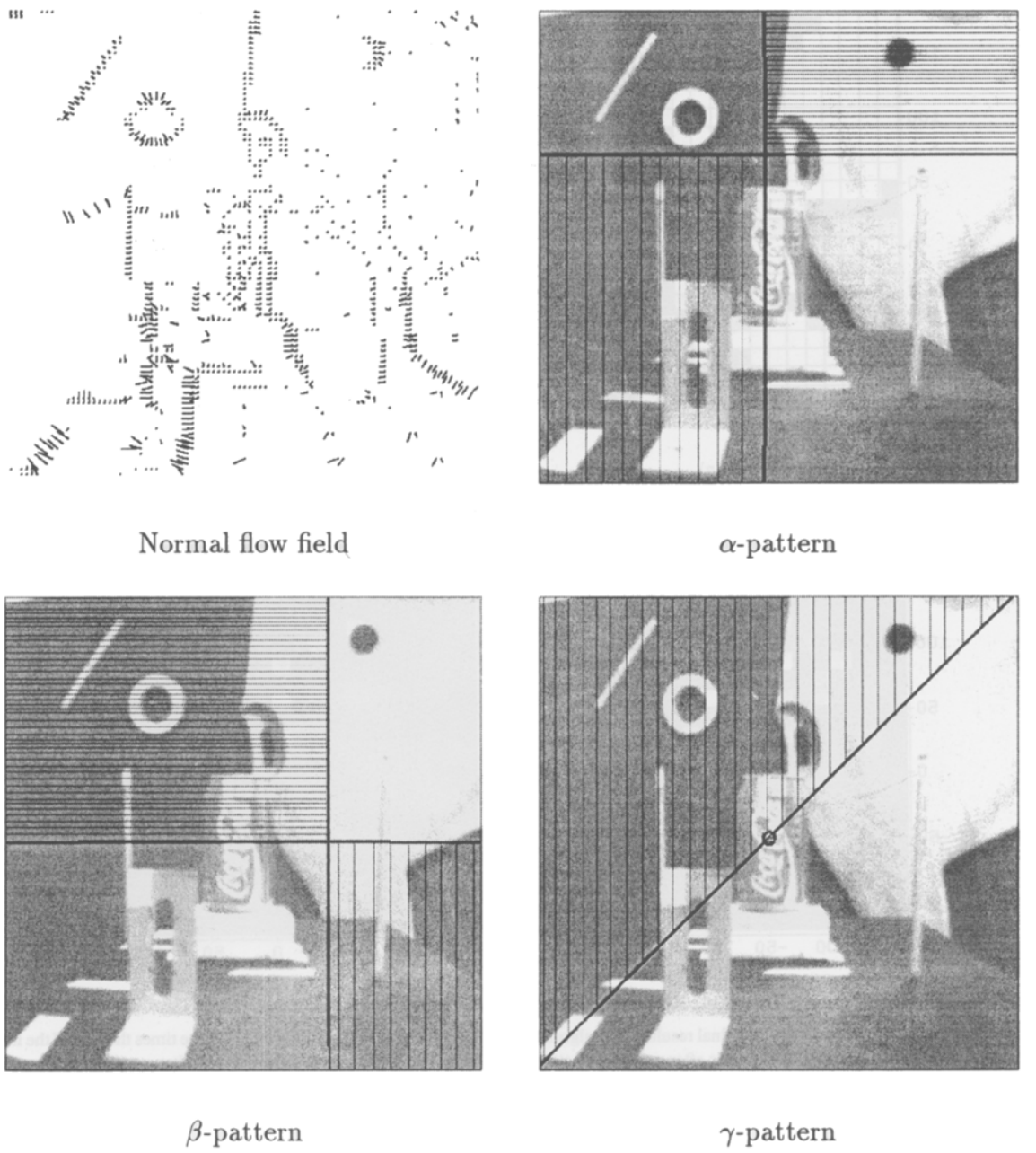


Fig. 17. NASA scene: Normal flow field and fitting of α-, β-, and γ-vectors.

translation correctly, but due to the high value of the focal length a number of other solutions (in the range ± 0.08 times the focal length away from the true solution) had acceptable confidence. Furthermore, since the rotational components were very small with regard to the translational ones (the absolute values were five to ten times smaller) the amount of rotation was not computed correctly in all cases. Figure 20 shows the scene. For one of the settings the results are displayed: the ground truth is FOE$=$ $(-129, +146)$ (measured from the image center); $\alpha = 0.000125663$ rad; $\beta = 0.000251327$ rad, $\gamma = 0.0$ rad. Notice, because $\gamma = 0$, the lines separating the rotational components of the α- and β-vectors lie in infinity and thus do

Fig. 18. NASA scene: The FOE's of all solutions with a success rate of 99% and higher are marked by black squares. The actual solution is at the center of the blob.

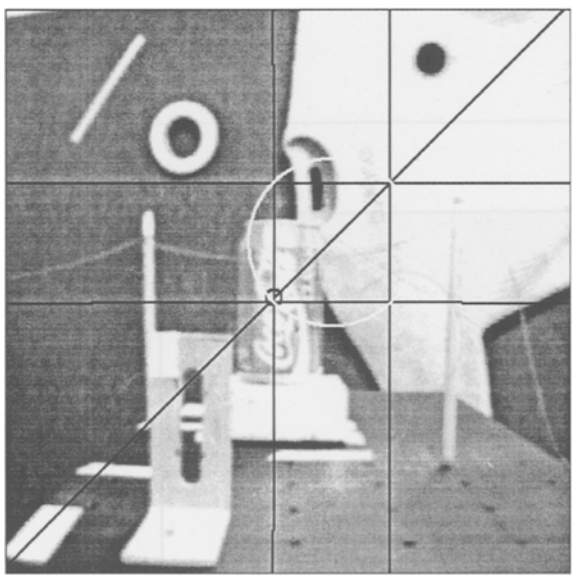


Fig. 19. NASA scene with overlaid curves and lines separating positive and negative α-, β-, and γ-vectors and rotation-axis vectors. At the intersection of the second order curves is the FOE (at the center of the coke can). The intersection of the straight lines denotes the point where the rotation axis pierces the image plane (over the pullover).

not appear in the image plane. Since the focal length is different in X- and Y-direction the circle in the γ-pattern is distorted to an ellipse. For this case the algorithm computed the correct FOE and the correct ratio $\frac{\beta}{\alpha}$,

Fig. 20. First image of series containing both translation and rotation (table top scene).

however the estimated value for the rotation was equal to 90 percent of the actual one. In Fig. 21 the positive and negative α-, β-, and γ-vectors and the corresponding patterns are displayed. For these experiments a tolerance interval of 6° has been allowed in the choice of normal flow vectors for the α-, β-, and γ-vectors. For the clarity of the pictorial description, all the points corresponding to the chosen vectors were enlarged by a factor of four. Figure 22 shows the rotation-axis pattern (partial derotation). Since the rotation is very small in relation to the translation, a smaller tolerance interval, namely only 3° was chosen. Figure 23 shows the computed normal flow field, and Fig. 24 shows the conic sections separating the translation vectors of the α-, β-, γ-patterns and the rotation-axis pattern overlaid on the image. At the intersection of these curves lies the FOE.

In Fig. 25a and b final and intermediate results of the algorithm's output for this scene are displayed. Figure 25a shows the results of pattern fitting and 25b the final result of total derotation.

A final set of experiments was performed imaging the same last scene (table top scene) but with translation along and rotation around all three axes. The ground truth is FOE = (−129, +146) (measured from the image center); α = 0.000125663 rad; β = 0.000251327 rad, γ = 0.001005304 rad, and thus the intersection point of the rotation axis with the image plane is outside the image at (+145, +329). The algorithm computed the correct solution for the axes of translation and rotation, but the estimated value for the rotation was 105% of the actual

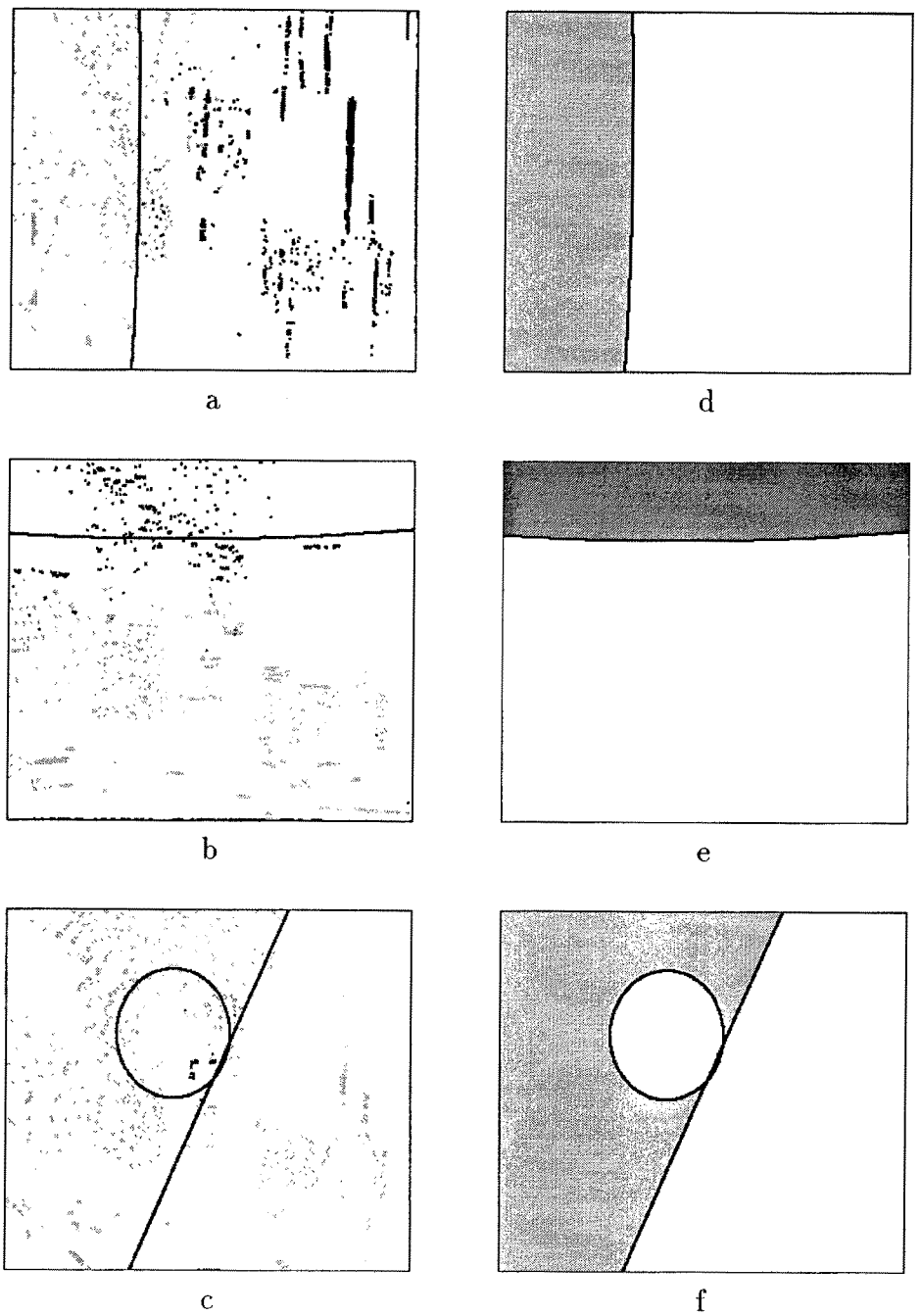


Fig. 21. Table top scene: (a), (b), (c): Positive and negative α-, β-, and γ-vectors. (d), (e), (f) Fitting of α-, β-, and γ-patterns.

one. Figure 26a show the intermediate results for pattern fitting and the final results computed in the derotation module. Comparing Figs. 25a and 26, we observe that in the absence of γ-rotation, the result of pattern fitting provided a much larger set of candidate solutions.

5 Conclusions

We have described several geometric relations that are characteristic of a normal flow field due to rigid motion. These relations were exploited to solve the problem of computing the 3D motion of an observer relative to a

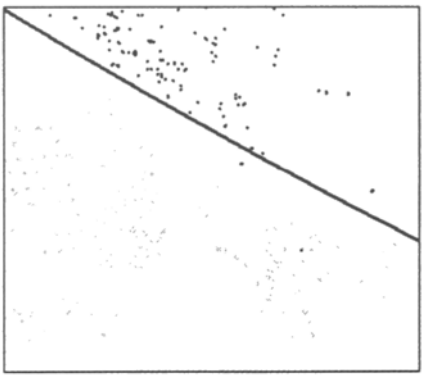


Fig. 22. Table top scene: Positive and negative vectors containing only translation $(\alpha, \beta, 0)$ and fitting of rotation-axis pattern.

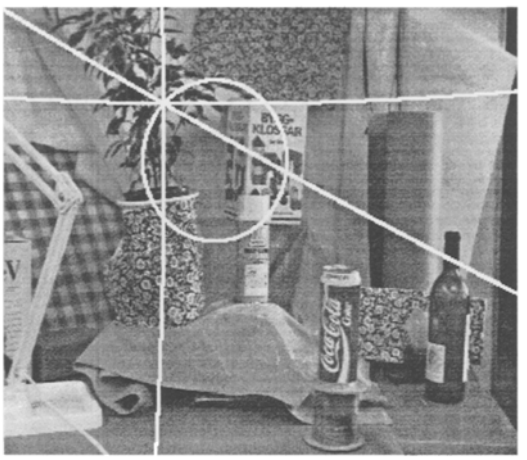


Fig. 24. Table top scene: Curves separating the translational components of the α-, β-, γ-vectors and the rotation-axis vectors. At their intersection lies the FOE.

scene in a robust way. Robustness is achieved by using the data in a global and mostly qualitative manner. The algorithm is qualitative, because for estimation of the translational and rotational axes only the sign of the normal flow vectors is used; and it is global, because values in all parts of the image are considered. The algorithm can be regarded as a search technique in a parameter space, where appropriate selection of normal flow values is used in different ways to reduce the dimensionality of the motion estimation problem. In order to compute the axes of translation and rotation, three different subsets of the vector field are examined for patterns defined by only three of the five parameters.

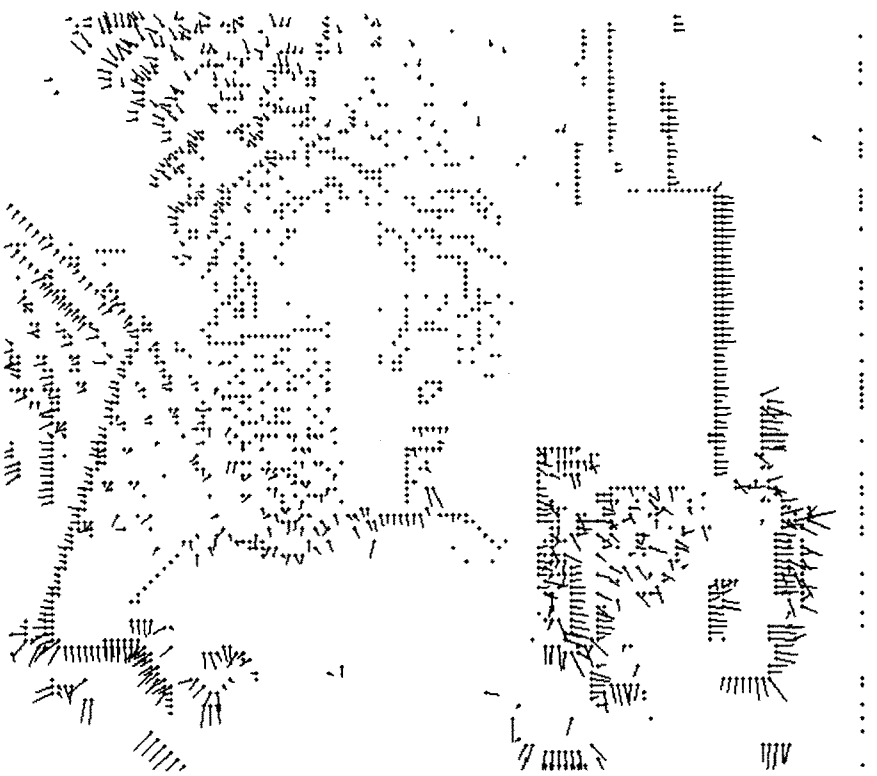


Fig. 23. Normal flow field: Table top scene.

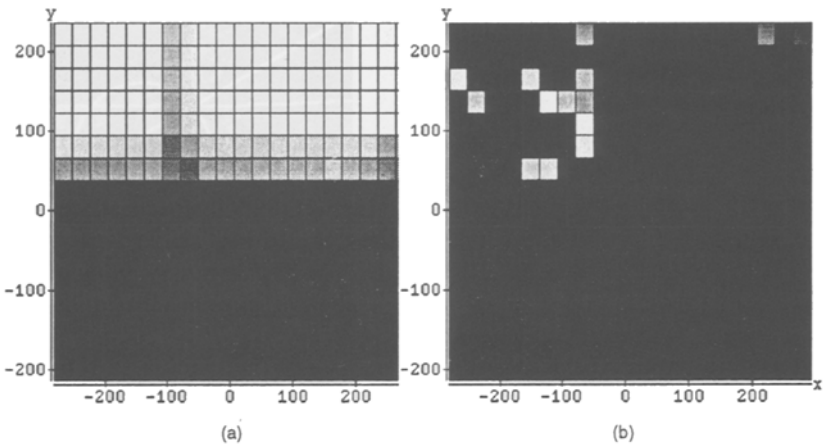


Fig. 25. Table top scene: (a) Results after pattern matching. (b) Final result.

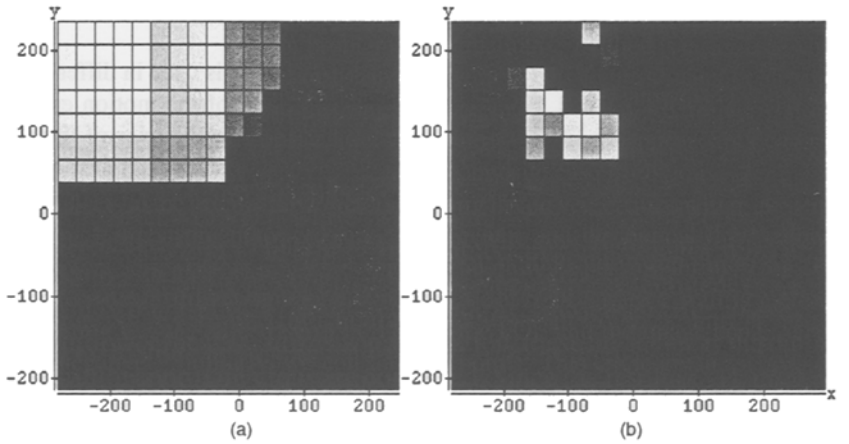


Fig. 26. Scene containing translation along and rotation around all three axes: (a) Results after pattern matching, (b) Final result.

A fourth set of values, which does not contain any rotational components, can be used to further reduce the set of candidate solutions for the two axes. By selecting values which are only due to rotation, the complete rotation is computed, and in the last phase of the algorithm every normal flow vector is tested for consistency with the computed motion parameters.

Normal flow measurements alone do not always define a unique 3D-motion interpretation. Our algorithm might be used as a front end tool in combination with other methods that use correspondence or optical flow. A study of uniqueness aspects of normal flow will be a very valuable theoretical contribution to this method and is part of our future research. Also, an interesting research problem would be the study of heuristics or other geometric constraints for pruning the search tree and reducing the complexity of the algorithm.

Notes

1. If the image intensity pattern in the neighborhood of a point is such that it allows reliable computation of the sign of the projection of the flow on directions other than the image gradient, this additional information can be utilized in the pattern matching procedure (Fermüller & Aloimonos, 1994).
2. This is a calibrated motion sequence made public for the Workshop on Visual Motion, 1991.

References

Adiv, G. 1985. Determining 3d motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:384–401.

Adiv, G. 1985. Inherent ambiguities in recovering 3d motion and structure from a noisy flow field. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 70–77.

Aloimonos, J. and Brown, C. 1984. Direct processing of curvilinear sensor motion from a sequence of perspective images. In *Proc. Workshop on Computer Vision: Representation and Control*, pp. 72–77.

Aloimonos, J. and Shulman, D. 1989. *Integration of Visual Modules: An Extension of the Marr Paradigm*. Academic Press, Boston.

Barnard, S. and Thompson, W. 1980. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:333–340.

Bergholm, F. 1988. Motion from flow along contours: A note on robustness and ambiguous cases. *International Journal of Computer Vision*, 3:395–415.

Bruss, A. and Horn, B. 1983. Passive navigation. *Computer Vision, Graphics, and Image Processing*, 21:3–20.

Burger, W. and Bhanu, B. 1990. Estimating 3-d egomotion from perspective image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:1040–1058.

Chandrashekar, S. and Chellappa, R. 1991. Passive navigation in a partially known environment. In *Proc. IEEE Workshop on Visual Motion*, pp. 2–7.

Faugeras, O. and Maybank, S. 1990. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 4:225–246.

Fermüller, C. and Aloimonos, Y. 1992. Estimating time to collision. Technical Report CAR-TR, Center for Automation Research, University of Maryland.

Fermüller, C. and Aloimonos, Y. 1992. Tracking facilitates 3-d motion estimation. *Biological Cybernetics*, 67:259–268.

Fermüller, C. and Aloimonos, Y. 1994. Vision and action. Technical Report CAR-TR-722, Center for Automation Research, University of Maryland (also *Image and Vision Computing Journal*, in press).

Hildreth, E. 1983. *The Measurement of Visual Motion*. MIT Press, Cambridge, Massachusets.

Horn, B. 1987. Motion fields are hardly ever ambiguous. *International Journal of Computer Vision*, 1:259–274.

Horn, B. 1990. Relative orientation. *International Journal of Computer Vision*, 4:59–78.

B. Horn and B. Schunck. 1981. Determining optical flow. *Artificial Intelligence*, 17:185–203.

Horn, B. and Weldon, E. 1987. Computationally efficient methods for recovering translational motion. In *Proc. International Conference on Computer Vision*, pp. 2–11.

Jain, R. 1983. Direct computation of the focus of expansion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:58–64.

Liu, Y. and Huang, T. 1988. Estimation of rigid body motion using straight line correspondences. *Computer Vision, Graphics, and Image Processing*, 43:37–52.

Longuet-Higgins, H. 1981. A computer algorithm for reconstruction of a scene from two projections. *Nature*, 293:133–135.

Longuet-Higgins, H.C. and Prazdny, K. 1980. The interpretation of a moving retinal image. *Proceedings of the Royal Society, London B*, 208:385–397.

Navab, N., Faugeras, O., and Vieville, T. 1993. The critical sets of lines for camera displacement estimation: A mixed euclidian-projective and constructive approach. In *Proc. International Conference on Computer Vision*.

Negahdaripour, S. 1986. *Direct Passive Navigation*. PhD thesis, Department of Mechanical Engineering, MIT, Cambridge, MA.

Nelson, R. and Aloimonos, J. 1988. Finding motion parameters from spherical flow fields (or the advantage of having eyes in the back of your head). *Biological Cybernetics*, 58:261–273.

Prazdny, K. 1981. Determining instantaneous direction of motion from optical flow generated by a curvilinear moving observer. *Computer Vision, Graphics, and Image Processing*, 17:238–248.

Spetsakis, M. and Aloimonos, J. 1988. Optimal computing of structure from motion using point correspondence. In *Proc. International Conference on Computer Vision*, pp. 449–453.

Spetsakis, M. and Aloimonos, J. 1990. Structure from motion using line correspondences. *International Journal of Computer Vision*, 1:171–183.

Spetsakis, M. and Aloimonos, J. 1991. A multi-frame approach to visual motion perception. *International Journal of Computer Vision*, 6:245–255.

Tsai, R. and Huang, T. 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27.

Ullman, S. 1979. The interpretation of structure from motion. *Proceedings of the Royal Society, London B*, 203:405–426.

Verri, A. and Poggio, T. 1989. Motion field and optical flow: Qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:490–498.

Verri, A., Straforini, M., and Torre, V. 1992. Computational aspects of motion perception in natural and artificial vision systems. *Philosophical Transactions of the Royal Society London, B*, 337:429–443.

Waxman, A. 1987. Image flow theory: A framework for 3-d inference from time-varying imagery. In C. Brown, editor, *Advances in Computer Vision*. Erlbaum, Hillsdale, NJ.

White, G. and Weldon, E. 1988. Utilizing gradient vector distributions to recover motion parameters. In *Proc. International Conference on Computer Vision*, pp. 64–73.