Dept. for Pattern Recognition and Image Processing Institute for Automation Technical University of Vienna Treitlstr. 3/1832 A-1040 Vienna AUSTRIA Phone: +43 (222) 58801-8161 Fax: +43 (222) 569697 E-mail: fer@cfar.umd.edu

PRIP-TR-20

March 10,1993

QUALITATIVE EGOMOTION

Cornelia Fermüller¹ and Yiannis Aloimonos²

Abstract

Due to the aperture problem, the only general unambiguous motion measurement in images 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 mostly 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.

¹Computer 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

²Computer Vision Laboratory Center for Automation Research University of Maryland College Park, MD 20742-3411

1. Introduction

Visual navigation constitutes a problem which is of considerable practical as well as scientific interest. Navigation, in general, refers to the performance of sensory mediated movement, and visual navigation is defined as the process of motion control based on an analysis of images. A system with navigational capabilities interacts adaptively with its environment. The movement of the system is governed by sensory feedback which allows it to adapt to variations in the environment and does not have to be limited to a small set of predefined motions as it is the case, for instance, with cam-activated machinery.

Visual navigation encompasses a wide range of perceptual capabilities that can be classified hierarchically. 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 set of processes 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 motion and to control its parameters. Although it is possible to obtain the necessary information by using expensive inertial guidance systems, it remains a challenge to solve the task by visual means. In this paper, we address the problem of passive navigation on the basis of a sequence of images. For a monocular observer undergoing unrestricted rigid motion in the 3D world, we compute the parameters describing this motion. ¿From 2D images only five unknowns can be derived, three rotational parameters and two parameters describing the direction of translation.

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 [7, 15, 13]. This computation relies on smoothness assumptions that usually result in unrealistic assumptions about the 3D scene. Since both problems, the computation of optical flow and the correspondence of features, are ill-posed, 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 [26] leads to linear equations, but in general is not realistic and 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 [22] or on a plane [1, 9, 14, 19, 20, 28, 24, 25]. The resulting equations relating 3D motion 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 [19, 25]. But a critical

investigation of the feasibility of addressing the motion estimation problem in this way, conducted by Spetsakis and Aloimonos [24], shows that even application of provably optimal algorithms to this problem cannot lead to methods that are useful for applications in realistic domains.

In some methods the data used as input comes only from local parts of the motion field. Such limited use of data leads to inherently unstable methods, because completely different observer motions produce locally similar motion fields. 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 [2].

2. Motivation

In the past, research on motion estimation has concentrated mainly on issues of existence and uniqueness. It has been shown that two perspective views of at least seven points in a rigid configuration almost always guarantee uniqueness of the motion parameters [25]. A retinal motion field uniquely determines 3D motion, unless the surface in view belongs to a certain class of hyperboloids of one sheet [18]. Similar results have been obtained for features other than points, such as lines [23] or contours [4, 8]. Since most uniqueness aspects of the problem 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. We know that the motion problem as addressed in the last few years involves steps that are mathematically ill-defined, e.g the computation of optical flow and the correspondence problem. Unrealistic assumptions were made and the resulting algorithms were extremely sensitive to noise.

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 [30]. Several multi-frame approaches have been developed. In most of this work an unrealistic assumption again is made—that of motion continuity over time.

Recently a new concept has emerged-that of active [5], purposive [6] vision. One should not insist on recovering all motion and shape parameters in one module, but rather one should solve simpler problems and add additional information by making the observer active. An increasing number of researchers has been attracted by this idea, but a general theory that explains how particular observer activities facilitate specific perceptual tasks has not yet appeared.

If our goal is to develop robust algorithms that can perform successfully in general environments, we should abandon all computational processes which 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. The only image motion that can be uniquely defined from a sequence of images is the normal flow- the projection of the optic flow on the gradient direction. This is the well known "aperture problem". The normal flow can be computed from the image motion of edges (the edge flow0 or from the image gradients by employing the motion constraint equation. The difficulty in its computation is only due to the discrete aspect of digital images. Computing normal flow in images is as difficult as detecting edges. Since normal flow constitutes a uniquely definable image motion representation, we choose to use it as input to our 3D motion estimation algorithm, even though the normal flow field appears to contain less information than the optic flow field.

One of the basic problems with many motion estimation algorithms, as mentioned before, is their sensitivity to slight perturbations of the input. The fact that normal flow measurements cannot be perfect 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 our approach, we first 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. Our technique searches for these patterns. It uses data from different parts of the image plane and considers only the sign of the normal flow. Our method for deriving the direction of the translation and rotation axes is of a qualitative and global character and thus can handle a considerable amount of error in the input.

Methods of estimating 3D-motion from only the normal flow field without going through the intermediate stage of computing optical flow have appeared in [3, 16, 21]. In [3] 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 [16], 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 [21] to estimate the focus of expansion for purely translational motion. In [29] translation and rotation were estimated for an observer rotating around the direction of translation, and in [12] the activity of tracking is used to compute the translational direction of a general rigidly moving object. Lately some techniques have appeared which are claimed to be of a qualitative character [10, 17, 27]. In these techniques solutions for noisy optical flow fields are proposed; but these techniques involve approaches different from ours, and thus they cannot be considered as related to our work.

The organization of this paper is as follows: In Section 3 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 4 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 5 is devoted to experimental results and the paper concludes with a discussion and outline of future work.

3. 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.

3.1. 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

$$\dot{X} = -U - \beta Z + \gamma Y
\dot{Y} = -V - \gamma X + \alpha Z
\dot{Z} = -W - \alpha Y + \beta X$$
(1)

Figure 1: Imaging geometry and motion representation

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 Figure 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):

$$u = \frac{(-Uf + xW)}{Z} + \alpha \frac{xy}{f} - \beta (\frac{x^2}{f} + f) + \gamma y$$
$$v = \frac{(-Vf + yW)}{Z} + \alpha (\frac{y^2}{f} + f) - \beta \frac{xy}{f} - \gamma x$$
(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 equation (3) as sums of translational and rotational components.

$$u = u_{\text{trans}} + u_{\text{rot}} = (-x_0 + xf)\frac{W}{Z} + \alpha \frac{xy}{f} - \beta(\frac{x^2}{f} + f) + \gamma y$$

$$v = v_{\text{trans}} + v_{\text{rot}} = (-y_0 + yf)\frac{W}{Z} + \alpha(\frac{y^2}{f} + f) - \beta \frac{xy}{f} - \gamma x$$
(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

$$u_{n} = un_{x} + vn_{y}$$

$$u_{n} = ((-x_{0} + xf)\frac{W}{Z} + \alpha \frac{xy}{f} - \beta(\frac{x^{2}}{f} + f) + \gamma y)n_{x}$$

$$+((-y_{0} + yf)\frac{W}{Z} + \alpha(\frac{y^{2}}{f} + f) - \beta \frac{xy}{f} - \gamma x)n_{y}$$
(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.

3.2. Motion field interpretation

A motion field is composed of a translational and a rotational component. Only the first of these is dependent on 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 in resolving the normal flow field into its rotational and its translational component.

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) (Figure 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.

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

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 Figure 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 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:

$$M(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma}; x, y) = (\frac{\alpha^2}{\gamma^2} x^2 + 2xy \frac{\alpha}{\gamma} \frac{\beta}{\gamma} + y^2 \frac{\beta^2}{\gamma^2} + 2xf \frac{\alpha}{\gamma} + 2yf \frac{\beta}{\gamma} + f^2)/(x^2 + y^2 + f^2) = C$$

with C in $[0, \dots, (1 + \frac{\alpha^2}{\gamma^2} + \frac{\beta^2}{\gamma^2})]$ (6)

Specifically, for a rotation around the Z-axis the second order curves are circles with center 0; we call them γ -circles (Figure 4a). If the rotation axis is the X- or Y- axis the rotation 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 (Figure 4b). For rotation around the Y-axis the major axis is the y-axis and we call the conic sections β -hyperbolas (Figure 4c). Figure 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

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

3.3. Selection of values

A motion vector consists of a rotational component which can be parameterized 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 equation (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:

$$u_{n} = \frac{W}{Z}((-x_{0} + xf), (-y_{0} + yf))(n_{x}, n_{y}) + \alpha(\frac{xy}{f}, (\frac{y^{2}}{f} + f))(n_{x}, n_{y}) -\beta((\frac{x^{2}}{f} + f), \frac{xy}{f})(n_{x}, n_{y}) + \gamma(y, -x)(n_{x}, n_{y})$$

$$(7)$$

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 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 thee 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 Figure 5).

3.4. 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

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

where r is 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(α , β) and period 2π . Along the line

Figure 5: Positive γ -, α - and β -vectors

which passes through the image center and makes 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 (Figure 6a).

The translational component of the motion field is characterized by the location of the FOE or FOC in the image plane. In [11] 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 the FOE and the image center as diametrical opposite points. At these points the γ vectors' translational components vanish. Thus, the geometric locus of all points on the γ -hyperbolas where there is zero translational normal flow, is a circle. 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 (Figure 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 halfplanes 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 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 (Figure 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

Figure 6:

6a: The γ -vectors due to rotation separate the image plane in a halfplane of positive values and a halfplane of negative values.

6b: 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.

6c: 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. Figure 7: Rotational positive and negative values are separated by a straight line parallel to the x-axis in case of the α -vectors and by a line parallel to the y-axis in case of the β -vectors.

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 $y = \frac{\beta f}{\gamma}$ (Figure 7a); for the β -hyperbolas it is parallel to the y-axis and is defined by $x = \frac{\alpha f}{\gamma}$ (see Figure 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 (Figures 8a, 8c). Symmetrical relations hold for the β -vectors. The curve of zero normal flow is defined by

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

and in areas of positive $g(x_0, y_0)$ the β -vectors are positive (see Figures 8b, 8d).

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 $\frac{\alpha}{\gamma}$) (Figure 9).

Figure 8: 8a, 8c: Hyperbolas separate the α - and β -vectors due to rotation into areas of positive and negative values. 8b, 8d: When using lenses of conventional focal lengths only one arm of the hyperbola is within the image plane. Figure 9: α - and β -patterns for general rigid motion.

4. The method

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

4.1. 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 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 $\frac{\beta}{\alpha}$. Similarly, the fitting of the β - or γ -patterns yields solutions for X_0 , Y_0 and $\frac{\beta}{\gamma}$ or $\frac{\alpha}{\gamma}$. 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 rotation only the pattern consists of an area of negative and an area of positive value 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 nonexistant 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 in 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.

4.2. 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 3.2, the second order curves $M(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma}; x, y)$ are given through equation (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 $((-\frac{\alpha}{\gamma}(y^2 + f^2) + \frac{\beta}{\gamma}xy + y^2)$

Figure 10: Field lines of a rotational vector field and positive rotation axis vectors.

xf, $(\frac{\alpha}{\gamma}xy - \frac{\beta}{\gamma}(x^2 + f^2) + yf)$ (see Figure 10). 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 (see Figure 11). The curve $h(x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$ which contains rotation axis vectors of value zero is given by the equation

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

4.3. 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 Figure 11: Rotation axis pattern.

normal flow vectors perpendicular to these lines do not contain translational components; thus have only rotational components. This can be seen from equation (7). If the selected gradient direction at a point (x, y) is $((y_0 - yf), (-x_0 + xf))$ 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 Figure 12). 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

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

4.4. 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

Figure 12: Normal flow vectors perpendicular to lines passing through the FOE are only due to rotation.

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:

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

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 Figure 13). The technique checks all points for this property and eliminates solutions that cannot give rise to the given normal flow field.

4.5. The algorithm

In this section we summarize of the complete algorithm in form of a block diagram. 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.

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.

$$S_1 \text{ (set of quadruples } \{x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma}\})$$

Partial derotation: For every $S_{1,i}$ select rotation axis vectors defined by $\left(\frac{\alpha}{\gamma_i}, \frac{\beta}{\gamma_i}\right)$ 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.

 $S_2 \text{ (set of quadruples } \{x_0, y_0, rac{lpha}{\gamma}, rac{eta}{\gamma} \})$

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 $\{A_i, B_i \ C_i\}$. 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) { $x_0, y_0, \alpha, \beta, \gamma$ })

Figure 13: Normal flow vectors due to translation are constrained to lie in halfplanes.

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 vector which is compatible with both motions (Figure 14).

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 a matching of prominent features to eliminate the incorrect motion parameters.

The computed 3D motion parameters and the normal flow values supply two linear equations in u and v at every point from which the optical flow field can be determined:

$$\frac{u - u_{\text{rot}}}{v - v_{\text{rot}}} = \frac{u_t}{v_t}$$

$$u_n = un_x + vn_y \tag{11}$$

The unique solution is then derived by checking whether prominent feature points in the first frame exist in the second frame at the locations computed by the optical flow values.

5. Experiments

Our approach to motion estimation is of a geometric nature. Sophisticated implementations of our theory could be developed by taking into account statistics of the input error; but we constructed a very simple implementation to demonstrate our approach. The elimination of impossible parameters from the space of solutions involves discrimination on the basis of quantitative values. We have

Figure 14: The intersection of two circles defines the normal flow vector which corresponds to two different optical flow vectors.

implemented in the following way: Normal flow values in certain directions are selected, if they are within a tolerance interval of 10°. This relatively large degree of freedom, of course, will introduce some error, but there is a tradeoff between accuracy and the amount of data used by the algorithm. In the first, second, and fourth modules counting is applied to discriminate between possible and impossible solutions. 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 third 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 and second module no quantitative use of values is made, since only the sign of the normal flow is considered. Such a limited use of data makes the module 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, our "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 within an interval and the gradient direction were chosen randomly. The parameter combinations leading to a success rate of more than 98% in the final module are considered as candidate solutions. In all experiments on noiseless data the correct solution was found as the best one. Figure 15 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 Figure 16 the fitting of the circle and the hyperbolas to the α -, β -, and γ -vectors is displayed. 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.

As an example of a real scene the NASA–Ames sequence 1 was chosen. The camera undergoes

¹This is a calibrated motion sequence made public for the Workshop on Visual Motion, 1991

Normal flow field

Optical flow field

Figure 15: Flow fields for synthetic data

only translational motion, and we added different amounts of rotation: 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 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 Figure 18). All solutions with higher success rates were very close to the correct one with the FOE deviating by at most 6% of the focal length from the correct positions (x_0, y_0) . In Figure 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.

6. 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 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



а

d

Figure 16: a,b,c: Positive and negative α -, β - and γ -vectors for synthetic data. d,e,f: Fitting of α -, β - and γ -patterns.

Figure 17: Real scene: Normal flow field and fitting of α - β - and γ -vectors.

Figure 18: The FOE's of all solutions with a success rate of 99% and higher are marked by black squares. The biggest square denotes the correct solution

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. 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 the understanding of this method and is part of our future planned research.

References

- [1] G. Adiv. Determining 3d motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:384 401, 1985.
- [2] G. Adiv. Inherent ambiguities in recovering 3d motion and structure from a noisy flow field. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pages 70 - 77, 1985.

- [3] J. Aloimonos and C. Brown. Direct processing of curvilinear sensor motion from a sequence of perspective images. In Proc. Workshop on Computer Vision: Representation and Control, pages 72 - 77, 1984.
- [4] J. Aloimonos and D. Shulman. Integration of Visual Modules: An Extension of the Marr Paradigm. Academic Press, Boston, 1989.
- [5] J. Aloimonos, I. Weiss, and A. Bandopadhay. Active vision. International Journal of Computer Vision, 2:333-356, 1988.
- [6] J. Y. Aloimonos. Purposive and qualitative active vision. In Proc. DARPA Image Understanding Workshop, pages 816 - 828, 1990.
- S. Barnard and W. Thompson. Disparity analysis of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2:333 - 340, 1980.
- [8] F. Bergholm. Motion from flow along contours: A note on robustness and ambiguous cases. International Journal of Computer Vision, 3:395 - 415, 1988.
- [9] A. Bruss and B. Horn. Passive navigation. Computer Vision, Graphics, and Image Processing, 21:3-20, 1983.
- [10] W. Burger and B. Bhanu. On computing a fuzzy focus of expansion for autonomos navigation. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pages 563 - 564, 1989.
- [11] C. Fermüller and Y. Aloimonos. Estimating time to collision. Technical Report CAR-TR, Center for Automation Research, University of Maryland, 1992.
- [12] C. Fermüller and Y. Aloimonos. Tracking facilitates 3-d motion estimation. Biological Cybernetics, 67:259 - 268, 1992.
- [13] E. Hildreth. The Measurement of Visual Motion. MIT Press, Cambridge, Massachusets, 1983.
- [14] B. Horn. Relative orientation. International Journal of Computer Vision, 4:59-78, 1990.
- [15] B. Horn and B. Schunck. Determining optical flow. Artificial Intelligence, 17:185 203, 1981.
- [16] B. K. P. Horn and J. Weldon. Computationally efficient methods for recovering translational motion. In Proc. International Conference on Computer Vision, pages 2 – 11, 1987.
- [17] R. Jain. Direct computation of the focus of expansion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 5:58–64, 1983.
- [18] B. K.P.Horn. Motion fields are hardly ever ambiguous. International Journal of Computer Vision, 1:259-274, 1987.
- [19] H. C. Longuet-Higgins. A computer algorithm for reconstruction of a scene from two projections. Nature, 293:133 - 135, 1981.
- [20] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. Proceedings of the Royal Society, London B, 208:385-397, 1980.
- [21] S. Negahdaripour. Direct Passive Navigation. PhD thesis, Department of Mechanical Engineering, MIT, 1986.

- [22] R. Nelsons and J. Aloimonos. Finding motion parameters from spherical flow fields (or the advantage of having eyes in the back of your head). Biological Cybernetics, 58:261-273, 1988.
- [23] M. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. International Journal of Computer Vision, 1:171-183, 1990.
- [24] M. E. Spetsakis and J. Aloimonos. Optimal computing of structure from motion using point correspondence. In Proc. International Conference on Computer Vision, pages 449-453, 1988.
- [25] R. Y. Tsai and T. S. Huang. 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, 1984.
- [26] S. Ullman. The interpretation of structure from motion. Proceedings of the Royal Society, London, B 203:405 - 426, 1979.
- [27] A. Verri, F. Girosi, and V. Torre. Mathematical properties of the 2d motion field: From singular points to motion parameters. In Proc. IEEE Workshop on Visual Motion, pages 190 - 200, 1989.
- [28] A. Waxman. Image flow theory: A framework for 3-d inference from time-varying imagery. In C. Brown, editor, Advances in Computer Vision. Erlbaum, Hillsdale, NJ, 1987.
- [29] G. White and E. Weldon. Utilizing gradient vector distributions to recover motion parameters. In Proc. International Conference on Computer Vision, pages 64 – 73, Tampa, Florida, 1988.
- [30] G.-S. J. Young and R. Chellappa. 3-d motion estimation using a sequence of noisy stereo images: Models, estimation, and uniqueness results. *IEEE Transactions on Pattern Analysis* and Machine Intelligence, 12:735-759, 1990.