Passive Navigation as a Pattern Recognition Problem

Cornelia Fermüller

Computational Vision and Active Perception Laboratory
Royal Institute of Technology
S-100 44 Stockholm, Sweden
and
Department for Pattern Recognition and Image Processing
Institute for Automation
Technical University Vienna
Treitlstraße 3
A-1040 Vienna, Austria
email : fer@bion.kth.se

Abstract

The most basic visual capabilities found in living organisms are based on motion. Machine Vision, of course, does not have to copy animal vision, but the existence of reliably functioning vision modules in nature gives us some reason to believe that it is possible for an artificial system to work in the same or a similar way. In this paper it is argued that many navigational capabilities can be formulated as pattern recognition problems. An appropriate retinotopic representation of the image allows to extract the information necessary to solve motion related tasks through the recognition of a set of locations on the retina. This position is demonstrated by introducing such a representation of image motion that allows to derive an observer's egomotion from information globally encoded in the image motion field. In the past the problem of determining a system's own motion from dynamic imagery has been considered as one of the classical visual reconstruction problems, where local constraints have been employed to compute from exact 2-D image measurements (correspondence, optical flow) the relative 3-D motion and structure of the scene in view. The approach introduced here is based on new global constraints defined on local normal flow measurements – the spatio-temporal derivatives of the image intensity function. Classifications are defined on normal flow vectors which allow to comprise vectors that form global patterns in the image

Please address all correspondence to: Cornelia Fermüller, Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD 20742-3275

plane. The position of these patterns is related to the 3-D motion of the observer, and their localization provides the axis of rotation and the direction of translation. The constraints introduced are utilized in algorithmic procedures formulated as search techniques. These procedures are very stable, since they are not affected by small perturbations in the image measurements. As a matter of fact, the solution to the two directions of translation and rotation is not affected, as long as the measurement of the sign of the normal flow is correct.

1. Introduction

Even simple visual systems that have the capability of navigating autonomously are faced with a number of tasks. Such tasks include the detection of independent motion by a moving observer, obstacle detection and avoidance, kinematic stabilization or the ability to determine the system's own motion, target tracking and pursuit , etc. In the literature these problems have been addressed as applications of the general principles related to the reconstruction of the visual world. Navigational problems were studied by first segmenting the image into areas of the same relative motion and then reconstructing the motion and shape of every point in the scene.

The computational theory behind this approach, known as "structure from motion" suggests to solve the problem in two stages. First, accurate image displacements between consecutive frames have to be computed, either in the form of point correspondences [Faugeras, 1992; Ullman, 1979] or as dense motion fields (optical flow fields) [Anandan and Weiss, 1985; Hildreth, 1983; Horn and Schunck, 1981]. Then, in a second step, the 3-D motion and structure are computed from the equations relating them to the 2-D image velocity [Bruss and Horn, 1983; Horn, 1990; Koenderink and van Doorn, 1976; Longuet-Higgins and Prazdny, 1980; Longuet-Higgins, 1981; Spetsakis and Aloimonos, 1988; Tsai and Huang, 1984].

This computational theory has been uncritically accepted, although it turned out that in both computational steps extreme difficulties are inherently involved. The computation of optical flow and correspondence in the general case is an ill-posed problem and additional assumptions must be made in order to solve it, and recovering 3-D motion from inexact or noisy flow fields has turned out to be a problem of extreme sensitivity. As a result navigational problems have basically been treated as numerical analysis problems where sophisticated techniques (such as singular value decomposition, simulated annealing, Kalman filtering, maximum likelihood estimation, etc.) have been employed in order to estimate the 3-D motion from the geometric and photometric constraints which relate local image motion to the 3-D world.

It is very doubtful, whether complete reconstruction in general is useful, since it provides us with an excess of information that might just postpone the extraction of information necessary for the solution to a certain task. If exact quantitative reconstruction were possible, the above mentioned navigational tasks could be solved with selected subsets of the computed parameters. However, it is not necessary to go through the cumbersome computations of reconstructing the 3-D world in order to solve these problems. We also know, that even very low animals such as arthropods, insects, and molluscs are solving these problems, and it is a fact that they are not performing reconstruction, since they don't posses adequate powerful enough computational machinery.

As we will show here, all the listed problems can be solved through the recognition of a set of locations on the retina (or rather on a retinotopic representation of the image) along with a set of attributes attached to these locations. To elaborate this point further, having collected some form of spatio-temporal information from the imagery, there exist spatial representations of this information that allow to directly extract the parameters necessary for solving specific tasks. A discussion of the problems mentioned above will help clarify this point.

The problem of independent motion detection by a moving observer amounts to recognizing a set of locations on the retina which have flow vectors that could not possibly originate from the rigid motion of the observer. For more advanced systems this recognition could take place on various representations of information using different detection mechanisms suited to well-defined situations [Nelson, 1991]. The problem of obstacle detection could be solved by recognizing a set of locations on the retina that represent the image of a part of the 3-D world being on a collision course with the observer. To perform this task it is not necessary to compute the exact motion between the observer and any object in the scene, but only to recognize that certain patterns of flow evolve in a way that signifies the collision of the corresponding scene points with the observer. Pursuing a target amounts to recognizing the target's location on the image plane along with a set of labels representing aspects of its relative motion sufficient for the observer to plan its actions. Motion measurements of this kind could be relative changes in the motion such as a turn to the left, right, above, down, further away, or closer. This paper addresses the problem of passive navigation, or estimating an observer's egomotion. We show that the axis of rotation and the axis of translation can be found by recognizing the intersection of these axes with the image plane. The fact that an observer's motion is rigid introduces constraints on the flow field that take the form of patterns in the image plane. Considering only the sign of image motion vectors (the spatio-temporal derivatives of the image intensity function) in certain directions, patterns are formed, whose positions encode the parameters of the translational and rotational axes. Finding the location of these patterns amounts to a simple search technique in a pattern space.

Instead of reconstructing the world, the problems described above are solved through the recognition of entities that are directly relevant to the task at hand. These entities are represented by only those parameters sufficient to solve the specific task. Since recognition amounts to comparing the information under consideration with prestored representations, the described approaches to solving these problems amount to matching patterns.

The motivation behind this approach of formulating visual modules as pattern recognition techniques, is that it leads to algorithms which are "qualitative" in nature. An approach towards qualitative methods is characterized by the fact that the visual modules have an output that does not depend on explicit quantitative models that could be unstable. Qualitativeness can be achieved for a number of reasons: The primitives might be expressible in qualitative terms, or their computation might be derived from inexact measurements and pattern recognition techniques, or the computational model itself might be proved stable and robust in all possible cases.

The approach to passive navigation introduced here has this set of attributes. First, the entities derived through pattern recognition techniques constitute only two directions - the axis of rotation and the direction of translation, which can be represented as two locations on the image plane. The redundancy with which the flow measurements are used – only the sign of the spatio-temporal derivatives is employed – allows us to speak of well-defined input.

Furthermore, the constraints developed are global and thus give rise to stable computations.

The paper is organized in the following way. Section 2 describes preliminaries, the input and the geometry relating 2-D image measurements to 3-D motion parameters. In Section 3 the new constraints, which are defined on classes of motion vectors are introduced. Section 4, using these constraints, describes algorithms to find through simple search techniques the rotation and translation axes of a general rigid motion. Section 5 concludes the paper with a discussion about the contribution of this work and possible applications of the constraints in the development of solutions to other navigational tasks.

2. Preliminaries

We take the approach of considering the dynamic imagery as a three-dimensional function of two spatial and one temporal parameters. Considering only local information allows us to compute from the spatio-temporal derivatives of the intensity function partial information about the image velocity. If we assume that for a given scene point the intensity I at the corresponding image point (x, y) remains constant over a short time instant, we obtain the optical flow equation [Horn and Schunck, 1981],

$$I_x u + I_y v + I_t = 0.$$

which relates the flow (u, v) to the partial derivatives I_x, I_y, I_t of I. From this constraint alone, without making any additional assumptions, we can only compute the normal flow $u_n = -\frac{1}{\|\nabla I\|} I_t$, the projection of flow on the gradient direction. This is the well known aperture problem (see Figure 1). To give the reader some intuition Figures 2a and 2b show a synthetically created optical flow field and the corresponding normal flow field.

****
insert Figure 1 about here
****

****
insert Figure 2 about here
****

The normal flow is related to the 3-D measurements through the following constraints: We assume that the coordinate system is fixed to the observer with the center being the nodal point of the camera and f the focal length (see Figure 3).

****
insert Figure 3 about here
****

If we denote the 3-D translation by (U, V, W) and the 3-D rotation by (α, β, γ) with axis $(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$, and if we introduce the coordinates $(x_0, y_0) = (Uf/W, Vf/W)$ for the Focus of Expansion (FOE), the point that denotes the direction of translation, we obtain the following well-known equations relating the velocity $(u, v) = (u_{\text{trans}} + u_{\text{rot}}, v_{\text{trans}} + v_{\text{rot}})$ of an image point to the 3-D velocity and the depth Z of the corresponding scene point [Longuet-Higgins and Prazdny, 1980]:

$$
\begin{aligned}
u &= u_{\text{trans}} + u_{\text{rot}} = & (1)\\
&= (-x_0 + x)\frac{W}{Z} + \alpha\frac{xy}{f} - \beta(\frac{x^2}{f} + f) + \gamma y \\
v &= v_{\text{trans}} + v_{\text{rot}} = & (2)\\
&= (-y_0 + y)\frac{W}{Z} + \alpha(\frac{y^2}{f} + f) - \beta\frac{xy}{f} - \gamma x
\end{aligned}
$$

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

$$u_n = (u, v) \cdot (n_x, n_y),$$

where "$\cdot$" denotes the inner product, or

$$u_n = un_x + vn_y \text{ , or}$$

$$
\begin{aligned}
u_n = ((-x_0 + x)\frac{W}{Z} + \alpha\frac{xy}{f} - \beta(\frac{x^2}{f} + f) + \gamma y)n_x \\
+((-y_0 + y)\frac{W}{Z} + \alpha(\frac{y^2}{f} + f) - \beta\frac{xy}{f} - \gamma x)n_y \qquad (3)
\end{aligned}
$$

Using u_n we want to recover the parameters describing the direction of translation $(\frac{Uf}{W}, \frac{Vf}{W})$ and the axis of rotation $(\frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$.

3. Global motion constraints

In this section geometrical relations of normal flow vectors in selected directions are investigated. To be more precise, we study the sign of the normal flow vectors in certain directions and the locations of normal flow vectors of the same sign. The goal is to achieve a separation between translation and rotation. This leads us to group normal flow vectors according to their direction into two kinds of classes which are motivated by the concepts of rotation axis and FOE. For both kinds of classes we find that the FOE and the axis of rotation separate the normal flow values in the image according to their sign by a *second order curve* and a *straight line* .

3.1. Coaxis vectors

Consider an imaginary line given by an orientation vector (A, B, C), where $A^2+B^2+C^2 = 1$, passing through the nodal point. This line defines an infinite class of cones with axis (A, B, C) and apex at the origin. The intersections of the cones with the image plane give rise to a set of conic sections, hereafter called vector field lines, or field lines of the axis (A, B, C), or just (A, B, C) field lines. It is worth noting that the (A, B, C) field lines are the lines along which the image points would move if the observer rotated around axis (A, B, C). Let us now combine normal flow vectors into a single class if they are perpendicular to the vector field lines of the same axis (A, B, C). At a point (x, y) the orientation perpendicular to the (A, B, C) vector field lines is given by a vector $\vec{M} = (M_x, M_y)$:

$$
\begin{aligned}
(M_x, M_y) &= ((-A(y^2 + f^2) + Bxy + Cxf), \\
&\quad (Axy - B(x^2 + f^2) + Cyf))
\end{aligned}
$$

and its unit vector $\vec{m} = (m_x, m_y)$ is thus $\vec{m} = \frac{\vec{M}}{||\vec{M}||}$. We call the vectors of the class corresponding to the axis (A, B, C) the (A, B, C) coaxis vectors. These are the normal flow vectors where the gradient (n_x, n_y) is equal to (m_x, m_y). In order to establish conventions about the vector's orientation, a vector will be said to be of positive orientation if it is pointing in direction (m_x, m_y). Otherwise, if it is pointing in direction $(-m_x, -m_y)$, its orientation will be said to be negative[a] (see Figure 4).

****

insert Figure 4 about here

****

Next we evaluate the translational components of the normal flow vectors. The value t_n of any translational vector component at point (x, y) in direction (n_x, n_y) is given by

$$
t_n = \frac{W}{Z}(x - x_0, y - y_0) \cdot (n_x, n_y). \tag{4}
$$

Since $\frac{W}{Z}$ is positive in case the observer is approaching the scene, the sign of u_n is equal to the sign of

$$
\begin{aligned}
h(A, B, C, x_0, y_0; x, y) &= (x - x_0, y - y_0) \cdot (n_x, n_y) \\
&= x^2(Cf + By_0) + y^2(Cf + Ax_0) - xy(Ay_0 + Bx_0) \\
&\quad -xf(Af + Cx_0) - yf(Bf + Cy_0) + f^2(Ax_0 + By_0). \qquad (5)
\end{aligned}
$$

The equation $h = 0$ describes a curve which separates the positive from the negative translational components of the (A, B, C) coaxis vectors. The curve passes through the FOE and

[a]Obviously, the proposed classification is not based on an equivalence relation, since the intersection of the sets of normal flow vectors belonging to different axes is not empty. However, for our purpose this is not of importance.

through the point $(\frac{Af}{C}, \frac{Bf}{C})$. Where $h(x,y) > 0$, the normal flow values are positive; where $h(x,y) < 0$, they are negative; and where $h(x,y) = 0$, the normal flow values are zero. For any selected class of coaxis vectors there exists a curve $h = 0$ which is uniquely defined by the two coordinates x_0, y_0 of the FOE (see Figure 5a); furthermore it is linear in x_0 and y_0.

In a similar way we analyze the components of the (A, B, C) coaxis vectors due to rotation. The rotational components of the flow vectors are defined only by the three rotational parameters α, β, and γ. Along the positive direction of the coaxis vectors, the value r_n of the rotational component is

$$\begin{aligned} r_n &= ((\alpha\frac{xy}{f} - \beta(\frac{x^2}{f} + f) + \gamma y), (\alpha(\frac{y^2}{f} + f) - \beta\frac{xy}{f} - \gamma x)) \cdot (n_x, n_y) \text{ or} \\ r_n &= (y(\alpha C - \gamma A) - x(\beta C - \gamma B) + \beta A f - \alpha B f) \cdot (x^2 + y^2 + f^2) \qquad (6) \end{aligned}$$

The sign of r_n is equal to the sign of

$$g(A, B, C, \alpha, \beta, \gamma; x, y) = y(\alpha C - \gamma A) - x(\beta C - \gamma B) + \beta A f - \alpha B f. \qquad (7)$$

Thus considering only the sign of the rotational component along the (A, B, C) coaxis vectors, the image plane is separated by a straight line $g(A, B, C, \alpha, \beta, \gamma) = 0$ into two halves containing values of opposite sign. The straight line passes through the point where the rotation axis pierces the image plane. We call this point, whose coordinates are $(\frac{\alpha f}{\gamma}, \frac{\beta f}{\gamma})$, the Axis of Rotation point (AOR). Since the line also passes through the point $(\frac{Af}{C}, \frac{Bf}{C})$, it is defined by only one unknown parameter (see Figure 5b).

In order to investigate constraints for general motion the geometrical relations due to rotation and due to translation have to be combined. A second order curve separating the plane into positive and negative values and a line separating the plane into two half-planes of opposite sign intersect. This splits the plane into areas of only positive coaxis vectors, areas of only negative coaxis vectors, and areas in which the rotational and translational flow have opposite signs. In these last areas, no information is derivable without making depth assumptions (Figure 5c).

****

insert Figure 5 about here

****

We thus obtain the following geometrical result for the case of general rigid motion. For any class of (A, B, C) coaxis vectors, there exist two areas on the image plane with one containing only positive vectors and the other containing only negative vectors. The boundaries of these areas, a conic and a straigt line, are defined by the FOE and the rotation axis.

We call these structures on the coaxis vectors the coaxis patterns. The family of coaxis patterns depends on the four parameters x_0, y_0, $\frac{\alpha}{\gamma}$, and $\frac{\beta}{\gamma}$, but every single (A, B, C) coaxis pattern is defined by only three parameters.

3.2. Copoint vectors

For the second kind of class of normal flow vectors, namely the ones "perpendicular to the lines emanating from a defined point", similar patterns are obtained. In this case, the rotational components are separated by a second order curve into positive and negative values and the translational components are separated by a straight line. We call the vectors perpendicular to straight lines passing through a point (r, s), the (r, s) copoint vectors[b] (see Fig. 6).

****

insert Figure 6 about here

****

At point (x, y) an (r, s) copoint vector of unit length in the positive direction is defined as

$$\frac{(-y+s, x-r)}{\sqrt{(x-r)^2+(y-s)^2}}.$$

The functions which define the curves are obtained by substituting the above unit vector for (n_x, n_y) in equations 4 and 6 and are as follows: The straight line $k(r, s, x_0, y_0; x, y)$, passing through (r, s) and separating the translational components is (see Fig. 7a)

$$k(r, s, x_0, y_0; x, y) = y(x_0 - r) - x(y_0 - s) - x_0 s + y_0 r = 0 \tag{8}$$

and the second-order curve $l(r, s, \alpha, \beta, \gamma; x, y)$ separating the rotational components is (similar to $h(A, B, C, x_0, y_0; x, y)$) defined as (see Fig. 7b)

$$\begin{aligned} &l(r, s, \alpha, \beta, \gamma; x, y) = \\ &\quad -x^2(\beta s + \gamma f) - y^2(\alpha r + \gamma f) + xy(\alpha s + \beta r) + xf(\alpha f + \gamma r) \\ &\quad +yf(\beta f + \gamma s) - f^2(\alpha r + \beta s) = 0 \end{aligned}$$

The superposition of translational and rotational values again defines patterns in the plane which consist of a negative and a positive area (see Fig. 7c). These patterns, called copoint patterns, are described by the same four coordinates of the axes of translation and rotation, which characterize the coaxis patterns. Every single copoint pattern is defined by only three parameters describing the coordinates of the AOR and a line passing through the FOE.

****

insert Figure 7 about here

****

[b] The copoint and coaxis vectors are dual to each other.

4. Motion recognition

The direction of translation and the axis of rotation define patterns on subsets of the normal flow vectors. For a general rigid motion four independent variables describe these patterns. As shown in Section 3, knowledge of the patterns' position provides information about the rigid motion. In order to derive the motion parameters, we localize the boundaries of different patterns by performing a search in the appropriate parameter spaces. On the basis of the constraints described before, one can envision a large number of algorithms for estimating the motion parameters by choosing different coaxis and copoint vectors and by following different search techniques in relevant subspaces of the whole pattern space.

The way we tackle the problem is to concentrate in an initial search only on a small number of normal flow vectors. For example, one could choose three classes of coaxis vectors, such as the ones corresponding to the three orthogonal coordinate axes. The (A, B, C) coaxis vectors when the orientation vector (A, B, C) is the Z- axis are perpendicular to circles whose center is the origin of the image plane; let us call them γ-vectors. Similarly, when (A, B, C) is the X- or Y- axis, we call the (A, B, C) coaxis vectors the α-vectors and β-vectors. The corresponding field lines are hyperbolas whose major axes are the image plane's x- and y-axes, respectively. Figure 8 depicts these sets of vector field lines and the corresponding γ-, α-, and β-vectors in positive orientation. The patterns defined on these vectors are shown in Fig. 9. In the case of the γ-vectors the conic becomes a circle, which has the FOE and the image center as two diametrically opposite points. The straight line passes through the image center and has slope $\frac{\alpha}{\gamma}$. In the case of the α- and β-vectors the conics become hyperbolas and the lines are horizontal and vertical respectively. To obtain the motion parameters, we search in the appropriate three-dimensional parameter spaces[c] of the α, β, γ vectors for the patterns that fit the data. This provides a set of solutions for the FOE and AOR.

****
insert Figure 8 about here
****

Another strategy to start localizing the motion parameters would be to search for the class of normal flow vectors which are only translational or only rotational. For any rigid motion there exists one class of coaxis vectors which does not contain any rotational components. This set is defined by the actual rotation axis $\frac{A}{C} = \frac{\alpha}{\gamma}$ and $\frac{B}{C} = \frac{\beta}{\gamma}$. Coaxis vectors of this kind are due only to translation and the pattern of these vectors is solely defined by the two-parameter second order curve $h(\alpha, \beta, \gamma, x_0, y_0; x, y) = 0$. There is only one curve separating the positive from the negative values and thus the pattern is defined on the whole image plane. Since $h(\alpha, \beta, \gamma, x_0, y_0; x, y) = 0$ is linear in x_0 and y_0, the problem of finding the FOE from the normal vectors due only to translation reduces to estimating the linear

[c]The range of values for the coordinates of the FOE and for $\frac{\beta}{\gamma}$ and $\frac{\alpha}{\gamma}$ is $[-\infty, +\infty]$. If a wide-angle lens or a logarithmic retina [Tistarelli and Sandini, 1992] is employed, most of the directions representing the FOE lie in a bounded area of the image plane. Alternatively, in order to cover all possible cases, a coordinate transformation on the sphere can be performed, in which case the coordinates are expressed by two angles.

discriminant function [d] separating two classes of values (labeled positive and negative). The pattern is due to only two parameters. In order to find the axis of rotation a search in the two-dimensional parameter space of $\frac{\alpha}{\gamma}$ and $\frac{\beta}{\gamma}$ is performed. For every possible rotation axis the data is checked for linear discrimination. If a second-order curve can be found that separates the positive from the negative values the quadruple $(x_0, y_0, \frac{\alpha}{\gamma}, \frac{\beta}{\gamma})$ will be added to the set of possible solutions. Similarly, there is only one class of copoint vectors which does not contain any translational components, the (x_0, y_0) copoint vectors which can be used to search in a two-dimensional space for the coordinates of the FOE.

Clearly, since only restricted use of data is made, the above two techniques will not result in a unique solution in general. In order to further narrow the space of solutions, additional coaxis and copoint patterns have to be employed. Since the first searches already result in a number of candidate solutions, additional algorithmic procedures can be formulated as verification techniques that merely test for the existence of further patterns to either eliminate or verify correct solutions.

The global constraints allow to derive the parameters for the axes of rotation and translation by only employing the sign of the normal flow. If someone is interested in obtaining the third rotational parameter, quantitative information has to be used. It can can be computed from the normal flow vectors that do not contain translational components. These computations are accomplished in the "detranslation" module described in [Fermüller, 1993]. Furthermore, after having found the complete rotation another module, called "complete derotation" [Fermüller, 1993] allows to check every normal flow vector for its consistency with the correct solution. The idea behind "detranslation" is the following: The translational components of optic flow vectors lie on lines passing through the FOE. Thus, if for a computed FOE we select normal flow vectors perpendicular to these lines, we have selected vectors which are only due to rotation. The third rotational component can be computed from these vectors by solving an overdetermined system of linear equations. In "complete derotation" the computed rotational values are subtracted from the given normal flow vectors. If the parameter quintuple defines the correct solution, the remaining normal flow is purely translational. Therefore it is confined to lie in the half plane which is on the opposite side of the grey-level edge from the FOE.

5. Experiments

In a series of experiments for synthetic and real data the algorithms described for finding the axis of rotation and the direction of translation as well as the two other modules to obtain all the five motion parameters ("detranslation" and "partial derotation") were tested. A detailed description of the preprocessing steps, the implementation of the algorithms, and pictorial descriptions of the algorithms' intermediate results are given elsewhere [Fermüller and Aloimonos, 1992]. Here only the fitting of patterns for one of the synthetic and one of

[d] Concerning the computational aspect of solving the discrimination problem, different algorithms from the pattern recognition literature can be applied. For example, the Ho-Kashyap algorithm [Ho and Kashyap, 1965] decides whether a data set is linearly discriminable and will also find the best discrimination.

the real scenes is displayed.

In Figure 10 results of the fitting of patterns to various classes of vectors of the synthetic normal flow field shown in Figure 2 is demonstrated. The image size was $[-50,+50] \times [-50,+50]$, the FOE was at (–40, –40) and the ratio of the rotational components was $\alpha : \beta : \gamma = 10 : 11 : 150$. Figures 10a,c,e show the vectors which have been found to be α-, β-, or γ-vectors. Points with positive normal flow values are rendered in light color and points with negative values are dark. In Figures 10b,d,f the fitting of the α-, β-, and γ-patterns found by the algorithms is displayed. The dark and grey areas denote the parts of the image plane which contain either only negative or only positive vectors, and the white areas are those about which no statement can be made. Figures 10g and 10h show the results of the fitting of the (α, β, γ) coaxis vectors and the fitting of the (x_0, y_0) copoint vectors.

****
insert Figure 9 about here
****

In Figures 11, 12, 13 the results for a real sequence are displayed. The motion of the camera relative to the scene contains all three translational and rotational components. 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 $\times$ 652, center of the image: (332, 305) (measuring from the bottom left corner). The ground truth is FOE = (–74, + 176) (measured from the image center); $\alpha : \beta : \gamma = 1 : 1 : 8$. Figure 11 shows the computed normal flow field. In Figures 12a,b,c the positive and negative α-, β-, and γ-vectors and in Figures 12d,e,f the corresponding patterns are displayed. For the clarity of the pictorial description, all the points corresponding to the chosen vectors were enlarged by a factor of four. The fact that the focal length is different in X- and Y-direction causes a distortion of the patterns, most obvious, the circle in the γ-patterns becomes an ellipse. Figure 13 shows, overlaid on the original image, in white, gray, and black the curves and lines separating the positive and negative α-, β-, and γ-vectors. At the intersection of the curve lies the FOE, and the intersection of the straight lines denotes the AOR.

****
insert Figure 10 about here
****

****
insert Figure 11 about here
****

****
insert Figure 12 about here
****

****

insert Figure 13 about here

****

6. Conclusions

The traditional view that regards "seeing" and "thinking" as two separate processes has started fading away. It becomes clear that Vision is not a process that just recovers descriptions of the visible word but is part of a system that perceives and acts [Zeki, 1992]. Vision, therefore, must be studied in connection with the behavior of the organism or system that "sees", and it is visual capabilities that will constitute the building blocks of an intelligent visual system. Thus, in order to develop a sound theory of Vision we must provide a systematic way of generating and developing visual capabilities which will serve the system with the information necessary to achieve its goals.

Two questions of immediate concern are what capabilities should we develop and how. The approach of addressing capabilities as recognition procedures attempts to give answers to both questions. What capabilities to develop is determined by the system's goals and the quantities that can possibly be recognized. The output of each of the system's modules has to be derived through a recognition procedure, where the amount of memory available naturally is limited. How a recognition procedure is carried out, of course, depends on the problem at hand and the available memory, but a guiding principle should be robustness which relates to qualitative techniques and global computations whenever possible.

We argued here that a number of capabilities related to navigation could be developed through the recognition of a set of locations on a retinotopic image representation and we addressed in detail the problem of egomotion recognition. It was shown that a moving observer can recognize its motion (direction of translation and axis of rotation) by matching a representation of image motion which only employs the sign of normal flow against a set of prestored patterns.

The major contribution of this paper is the introduction of a set of geometric constraints that reveal a global structure hidden in normal flow fields. We restricted the discussions about the applicability of the geometrical constraints to the most general capability related to egomotion, the computation of rigid motion for an observer moving rigidly with all six degrees of freedom. Of course, many more capabilities that deal in some form with an observer's motion can be designed on the basis of these constraints. For example, in most practical applications (e.g. a car going on a street) we are dealing with only restricted motion capabilities, and particular algorithms have to be developed that use the constraints in some modified form. Also, various techniques based on modifications are required for active observers that supply additional information. In [Fermüller and Aloimonos, 1993] it was shown how the motion estimation problem becomes easier if the observer is able to perform tracking. Furthermore, in [Fermüller, 1993] it was demonstrated that in order to solve various navigational tasks, such as servoing, for example, it is not necessary to go through the intermediate computation of the exact motion parameters, but control strategies might be developed that work on the basis of minimizing the distance of prestored patterns with the data.

Recent results from the neurophysiological literature [Tanaka and Saito, 1989; Duffy and Wurtz, 1991] suggest that there exist areas in the brain containing neurons that respond to particular patterns in changing images. The analysis described here shows a way of extracting globally encoded information from normal motion fields. The particular global structure of patterns that has been investigated depends only on the 3-D motion and not on the shape of the scene in view. One could envision that several other problems involving multiple views of a scene could be addressed in a similar manner. Any form of representation encoding changes in images, such as optical flow fields or stereo disparity fields, deserves investigation for global constraints that might allow us to solve various tasks related to the computation of extrinsic and intrinsic parameters (self calibration of active vision systems) through pattern matching techniques.

References

[Anandan and Weiss, 1985] P. Anandan and R. Weiss. Introducing a smoothness constraint in a matching approach for the computation of optical flow fields. In *Proc. 3rd Workshop on Computer Vision: Representation and Control*, pages 186–194, 1985.

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

[Duffy and Wurtz, 1991] C.J. Duffy and R.H. Wurtz. Sensitivity of MST neurons to optical flow stimuli I: a continuum of response selectivity to large field stimuli. *Journal of Neurophysiology*, 65:1329–1345, 1991.

[Faugeras, 1992] O. Faugeras. *Three Dimensional Computer Vision*. MIT Press, Cambridge, MA, 1992.

[Fermüller and Aloimonos, 1992] C. Fermüller and Y. Aloimonos. Qualitative egomotion. Technical Report, Center for Automation Research, University of Maryland, 1992.

[Fermüller and Aloimonos, 1993] C. Fermüller and Y. Aloimonos. The role of fixation in visual motion analysis. *International Journal of Computer Vision*, 11(2):165–186, 1993.

[Fermüller, 1993] C. Fermüller. Navigational preliminaries. In Y. Aloimonos, editor, *Active Perception*, Advances in Computer Vision, pages 103–150. Lawrence Erlbaum, Hillsdale, NJ, 1993.

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

[Ho and Kashyap, 1965] Y-C. Ho and R.L. Kashyap. An algorithm for linear inequalities and its applications. *IEEE Transactions on Electronic Computers*, 14:683–688, 1965.

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

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

[Koenderink and van Doorn, 1976] J.J. Koenderink and A.J. van Doorn. Local structure of movement parallax of the plane. *Journal of the Optical Society of America*, 66:717–723, 1976.

[Longuet-Higgins and Prazdny, 1980] 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.

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

[Nelson, 1991] R.C. Nelson. Qualitative detection of motion by a moving observer. *International Journal of Computer Vision*, 7:33–46, 1991.

[Spetsakis and Aloimonos, 1988] 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.

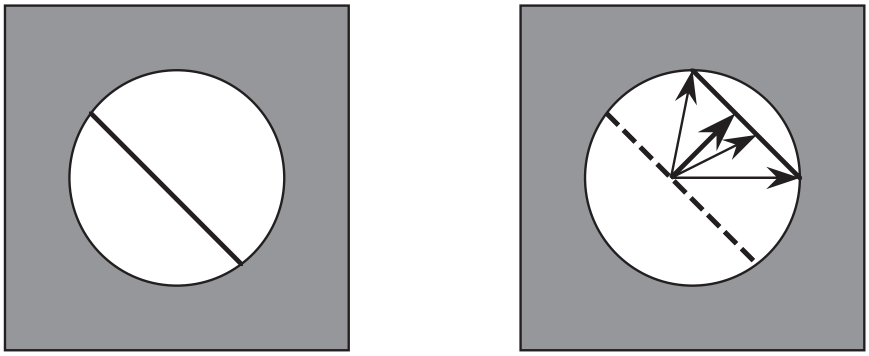
[Tanaka and Saito, 1989] K. Tanaka and H.A. Saito. Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells illustrated in the dorsal part of the Medial Superior Temporal area of the macaque monkey. *Journal of Neurophysiology*, 62:626–641, 1989.

[Tistarelli and Sandini, 1992] M. Tistarelli and G. Sandini. Dynamic aspects in active vision. *CVGIP: Image Understanding: Special Issue on Purposive, Qualitative, Active Vision, Y. Aloimonos (Ed.)*, 56:108–129, 1992.

[Tsai and Huang, 1984] 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.

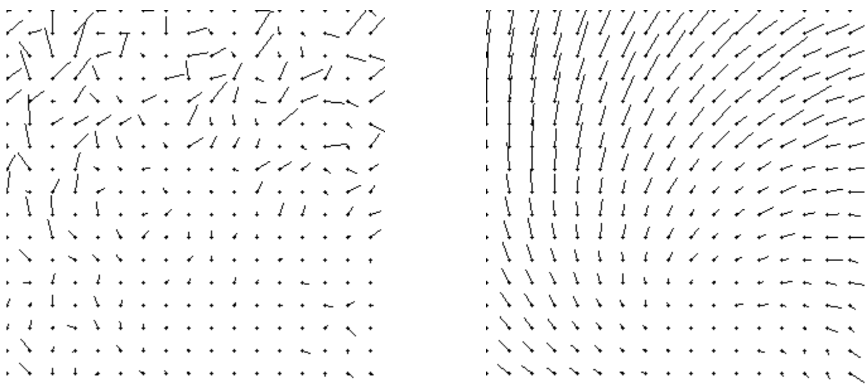
[Ullman, 1979] S. Ullman. *The Interpretation of Visual Motion.* MIT Press, 1979.

[Zeki, 1992] S. Zeki. The visual image in mind and brain. *Scientific American*, 267(3):69–76, 9 1992.



(a) (b)

Figure 1: The aperture problem: (a): Line feature observed through a small aperture at time t. (b): At time $t + \delta t$ the feature has moved to a new position. It is not possible to determine exactly where each point has moved to. From local measurements only the flow component perpendicular to the line feature can be computed.



Normal flow field　　　　Optical flow field

Figure 2: Flow fields of synthetic data.

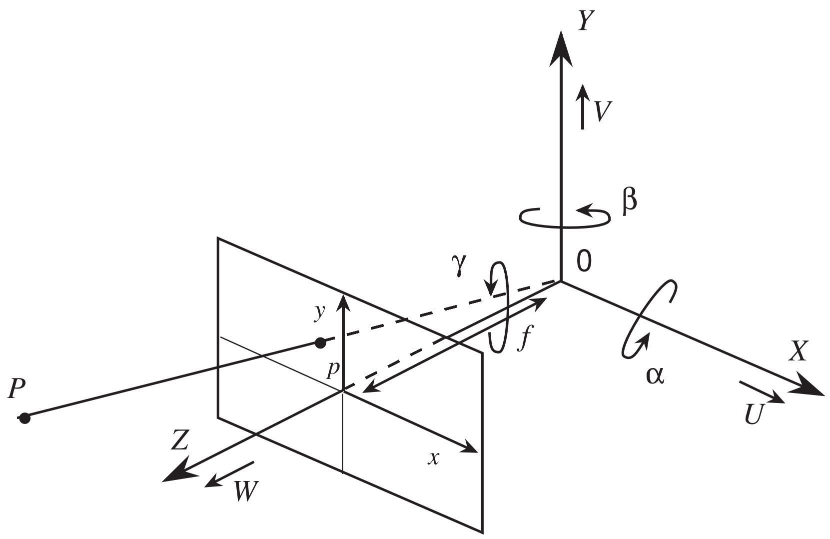


Figure 3: Imaging geometry and motion representation.

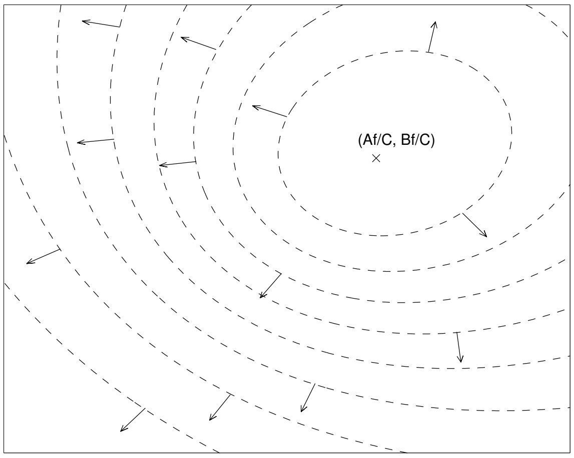
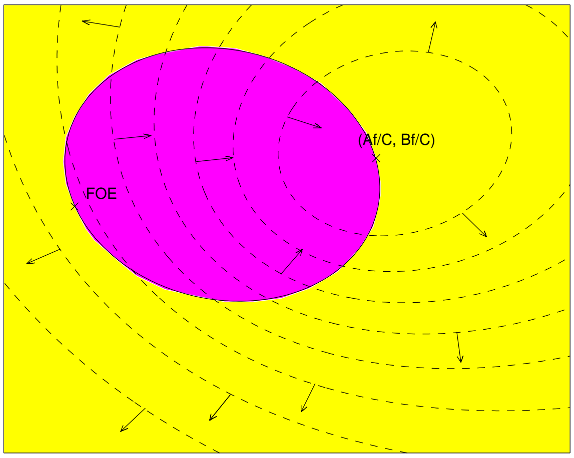
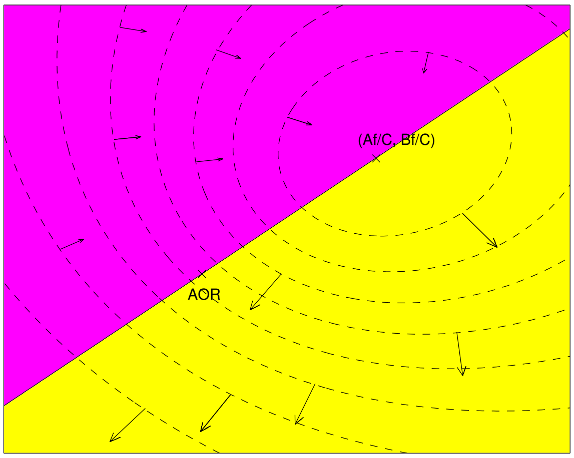


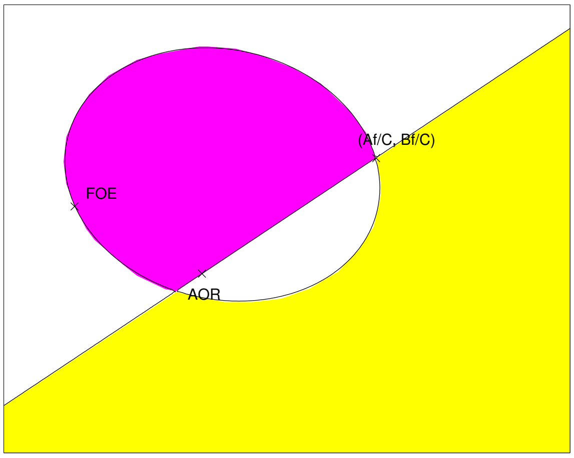
Figure 4: Field lines corresponding to an axis (A, B, C) and positive (A, B, C) coaxis vectors.



(a)



(b)



(c)

Figure 5: (a) The (A, B, C) coaxis vectors due to translation are negative if they lie within a second-order curve defined by the FOE, and are positive at all other locations. (b) The coaxis vectors due to rotation separate the image plane into a half-plane of positive values and a half-plane of negative values. (c) A general rigid motion defines an area of positive coaxis vectors and an area of negative coaxis vectors. The rest of the image plane is not considered.

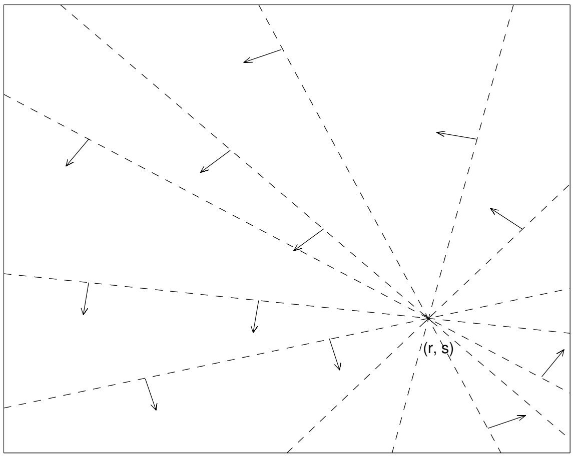
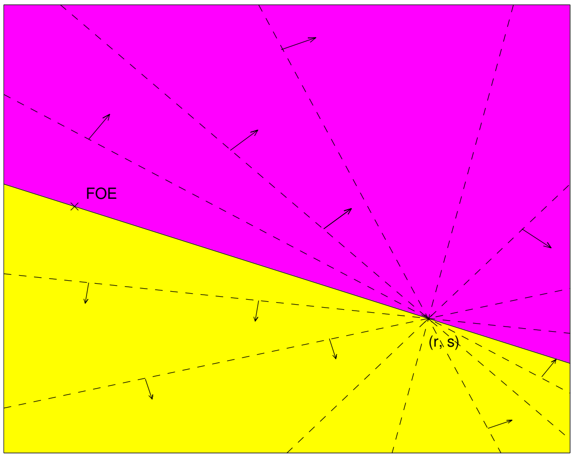
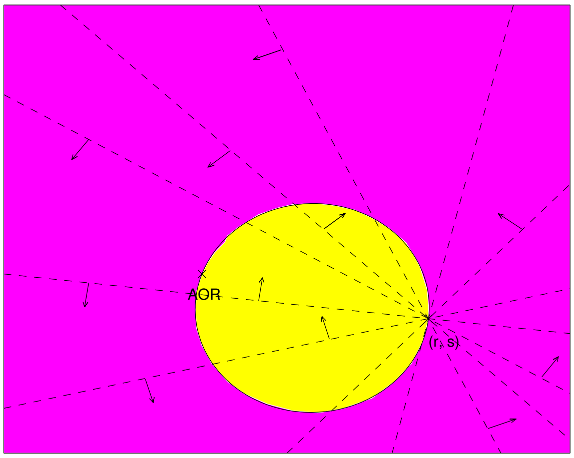


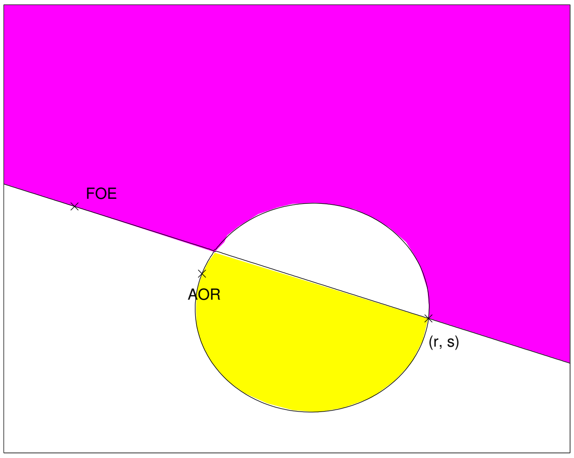
Figure 6: Positive (r, s) copoint vectors.



(a)



(b)



(c)

Figure 7: (a) Separation of the translational (r, s) copoint vectors. (b) Separation of the rotational copoint (r, s) vectors. (c) Separation of the (r, s) copoint vectors due to general rigid motion into an area of negative vectors, an area of positive vectors, and an area that may contain vectors of any value.

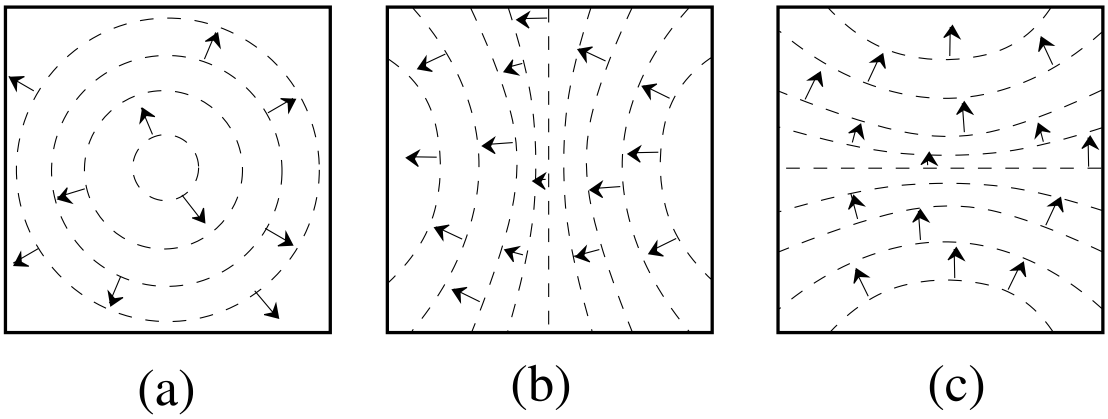
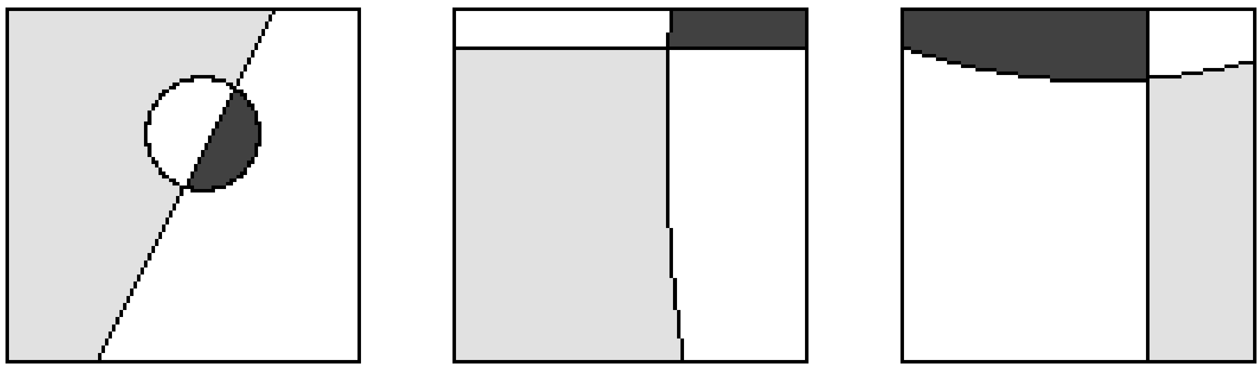


Figure 8: If the (A, B, C) axis is the Z-, X-, or Y-axis, the corresponding vector field lines are circles with center O (a), or hyperbolas whose axes coincide with the coordinate axes of the image plane (b and c). Normal flow vectors perpendicular to these field lines are called γ-, α-, and β-vectors.



(a) (b) (c)

Figure 9: α-, β-, and γ-patterns for a general rigid motion.

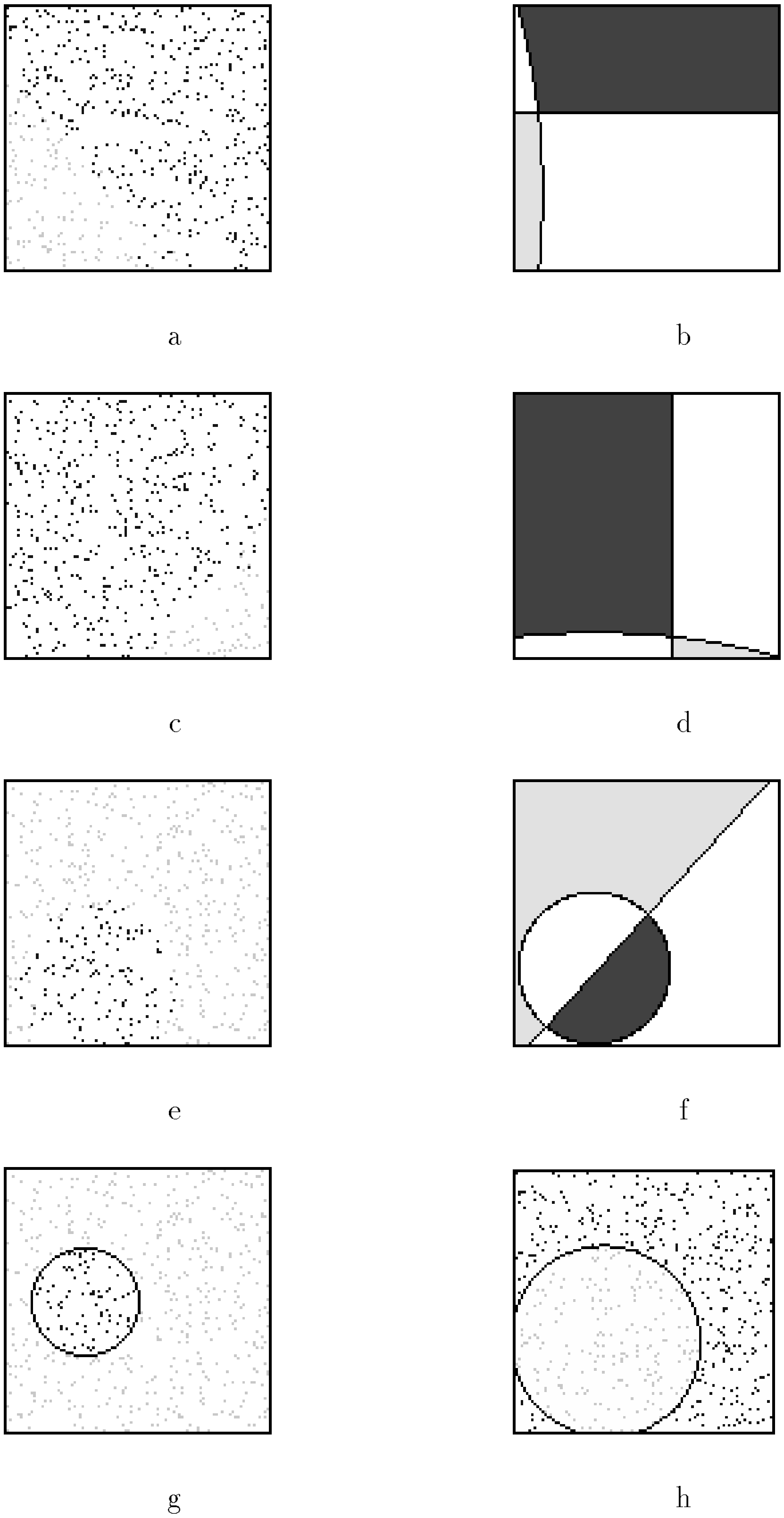


Figure 10: (a), (c), (e): Positive and negative α-, β-, and γ-vectors. (b), (d), (f) Fitting of α-, β-, and γ-patterns. (g): Separation of the (α, β, γ) coaxis pattern. (h): Separation of the (x_0, y_0) copoint vectors.

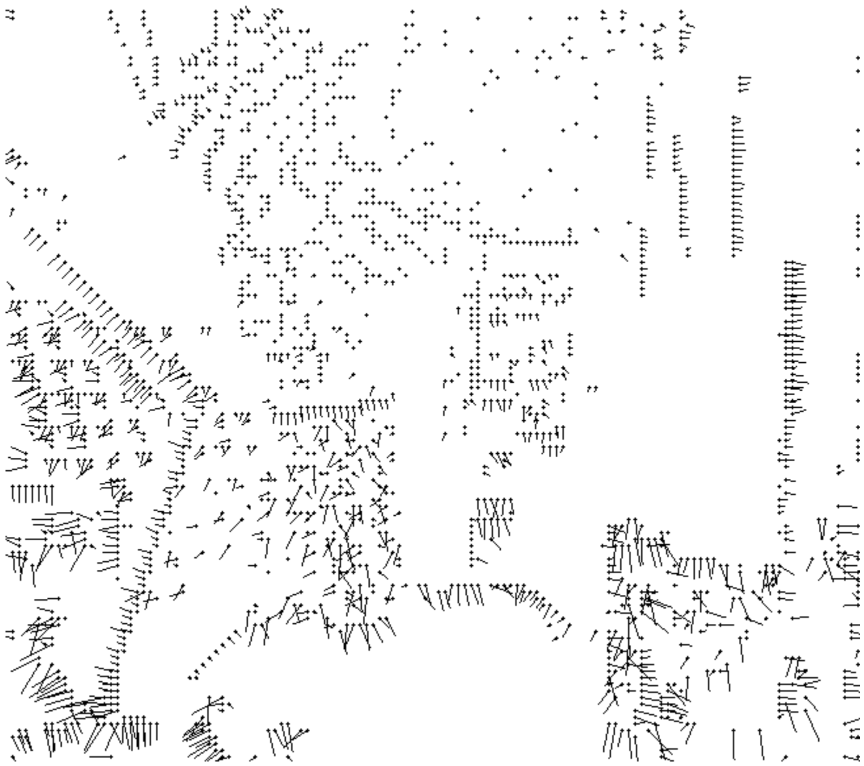


Figure 11: Normal flow field : Real scene.

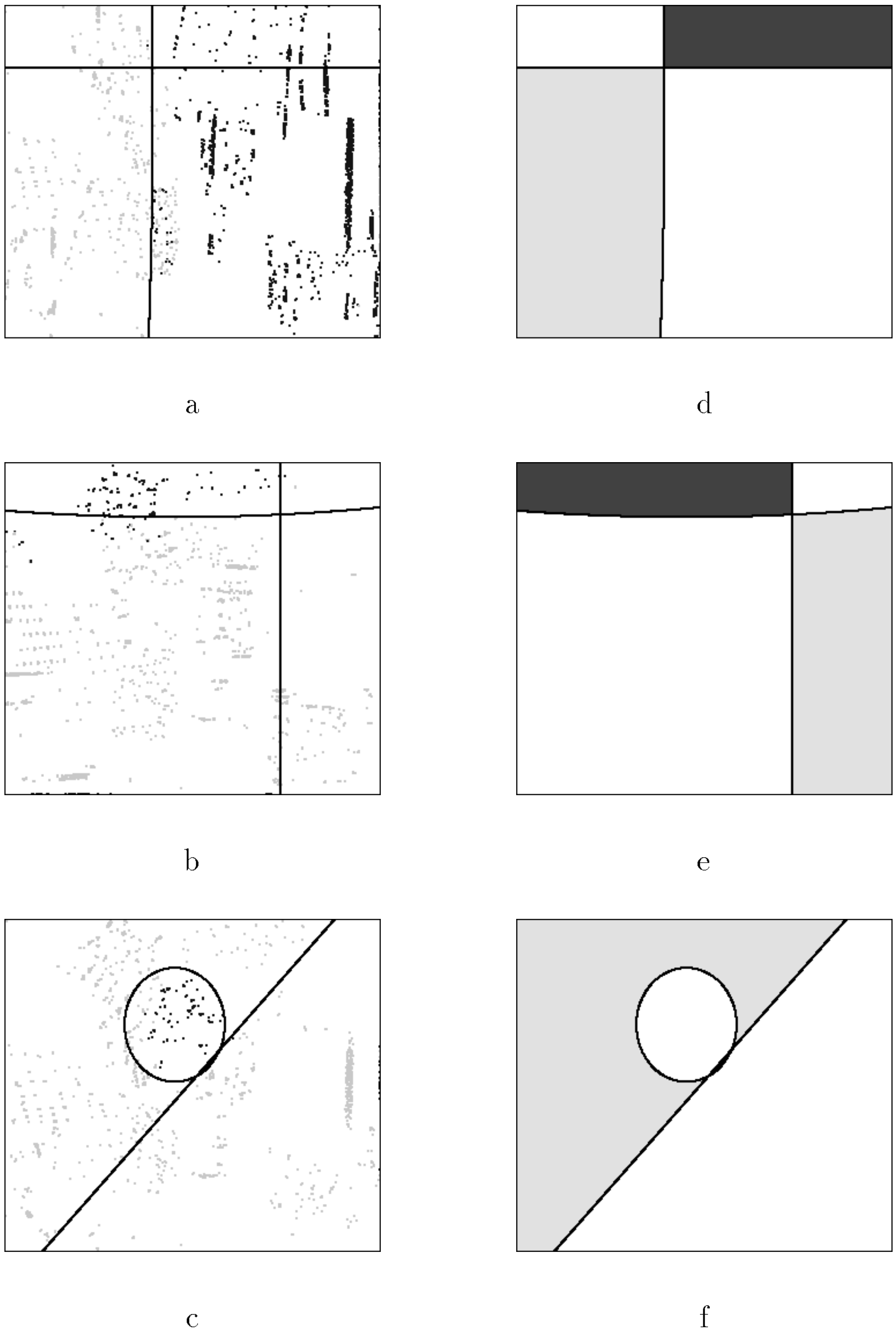


Figure 12: Real scene: (a), (b), (c): Positive and negative α-, β-, and γ-vectors. (d), (e), (f) Fitting of α-, β-, and γ-patterns.

Figure 13: Real scene: Superimposed on the image are the curves and lines separating the α- vectors (grey), β-vectors (black) and γ-vectors (white). At the intersection of the second order curves lies the FOE, and at the intersection of the straight lines lies the AOR.