Signal Processing 16 (1989) 53–72
North-Holland

RECURSIVE AND/OR ITERATIVE ESTIMATION OF THE TWO-DIMENSIONAL VELOCITY FIELD AND RECONSTRUCTION OF THREE-DIMENSIONAL MOTION

Georges TZIRITAS

Laboratoire des Signaux et Systèmes, Ecole Supérieure d'Electricité, Plateau du Moulon, 91190 Gif-sur-Yvette, France

Received 4 February 1988
Revised 25 May 1988

Abstract. In this article an approach is presented for estimating the two-dimensional velocity field from a sequence of images as well as a method for reconstructing the 3-D motion and structure from the velocity field. We suppose that the 2-D velocity vector is the projection of the 3-D velocity vector on the image plane and that the surfaces of the 3-D objects are locally planar and rigid, to obtain spatial relations on the velocity field. If the projection is perspective, we use the "perspective" velocity vector. The spatial relations are used for a linear estimation of the velocity field using local measurements. If the velocity vector has to be estimated on a 2-D region, measurement of the spatial and temporal gradients of the intensity is needed; for edge-based estimation the normal velocity component on an edge is used. In the case of moving edges a Kalman filter is proposed, and in the case of two-dimensional regions a Gauss–Seidel relaxation algorithm, which is iterative and non-recursive, is proposed. We also present a recursive estimator which may be useful in predictive motion-compensating coding.

Concerning the reconstruction of the 3-D motion parameters, we demonstrate that, if the "perspective" velocity vector is known at three points, the 3-D motion parameters and the orientation of the plane defined by the three points can be reconstructed. We also demonstrate that, if the "perspective" velocity vector is known at four points, the 3-D motion parameters and the relative depths can be reconstructed by solving two systems of three linear equations and a system of four linear equations.

Zusammenfassung. In diesem Beitrag wird ein Ansatz zur Schätzung zweidimensionaler Geschwindigkeitsfelder aus Bildsequenzen beschrieben. Zusätzlich wird eine Methode vorgestellt, mit der aus den Geschwindigkeitsfeldern dreidimensionale Bewegung und Struktur rekonstruiert werden kann. Es wird angenommen, daß der zweidimensionale Bewegungsvektor die Projektion des dreidimensionalen auf die Bildebene darstellt, und daß die Oberflächen der dreidimensionalen Objekte lokal eben und starr sind, woraus die örtlichen Beziehungen des Geschwindigkeitsfeldes hergeleitet werden. Bei Annahme einer Perspektiv-Projektion wird der "Perspektiv-Vektor" benutzt. Die örtlichen Beziehungen werden benutzt, um aufgrund isolierter Messungen eine lineare Abschätzung des Bewegungsfeldes zu erhalten. Wenn für einen Bereich des zweidimensionalen Bildes ein Bewegungsvektor geschätzt werden soll, müssen örtliche und zeitliche Intensitäts-Gradienten gemessen werden. Für eine konturbasierte Bewegungsschätzung wird die Bewegungskomponente normal zur Konturrichtung verwendet. Für den Fall bewegter Konturen wird ein Kalman-Filter vorgeschlagen; für den Fall zweidimensionaler Gebiete wird ein Gauß–Seidel Relaxations-Algorithmus verwendet, der iterativ, aber nicht rekursiv arbeitet. Außerdem wird ein rekursiver Bewegungsschätzer vorgeschlagen, der bei prädiktiver bewegungskompensierter Codierung von Nutzen sein kann.

Es wird gezeigt, daß die Parameter der dreidimensionalen Bewegung genau dann rekonstruiert werden können, wenn der "Perspektiv-Vektor" an drei Punkten gegeben ist. Dann ist auch die Orientierung der durch die drei Punkte definierten Ebene gegeben. Es wird außerdem gezeigt, daß bei Vorliegen des "Perspektiv-Vektors" an vier Punkten die dreidimensionalen Bewegungsparameter sowie die relative Tiefeninformation rekonstruiert werden können durch Lösung zweier Systeme aus je drei linearen Gleichungen und eines Systems von vier linearen Gleichungen.

Résumé. Dans cet article nous présentons une approche pour estimer le champ des vitesses bidimensionnel à partir d'une séquence d'images, ainsi qu'une méthode pour reconstruire le mouvement et la structure 3-D à partir du mouvement apparent sur l'image. Nous supposons que le vecteur vitesse bidimensionnel est la projection du vecteur vitesse 3-D sur le plan image et que les surfaces des objets 3-D sont localement planes et rigides, afin d'obtenir de relations spatiales sur le champ des vitesses. Si la projection est perspective, nous utilisons sur le plan image le vecteur de vitesse "perspective". Les relations spatiales sont utilisées pour une estimation linéaire du champ des vitesses à l'aide de mesures locales, comme les gradients spatiaux ou la composante normale de la vitesse sur un contour. Si le vecteur vitesse doit être estimé sur une région 2-D, la mesure des gradients spatiaux et temporal de l'intensité sont nécessaires; pour une estimation sur des contours la composante perpendiculaire au contour est utilisée. Dans le cas du mouvement des contours un filtre de Kalman est proposé, et dans le cas de regions bidimensionnelles un algorithme de relaxation de Gauss-Seidel, qui est itératif et non-récursif, est proposé. Nous présentons aussi un estimateur récursif, qui peut étre utile au codage prédictif avec compensation du mouvement.

Concernant la reconstruction des paramètres du mouvement 3-D, nous démontrons que, si le vecteur de vitesse "perspective" est connu à trois points, on peut reconstruire les paramètres du mouvement 3-D et l'orientation du plan défini par les trois points. Nous démontrons aussi que, si le vecteur de vitesse "perspective" est connu à quatre points, les paramètres du mouvement 3-D et les profondeurs relatives peuvent être reconstruits en resolvant deux systèmes de trois équations linéaires et un système de quatre équations linéaires.

Keywords. Differential motion estimation, edge motion estimation, 3-D motion parameters, structure.

1. Introduction

The estimation of motion from a sequence of images is a fundamental domain of research with various applications. In image coding the prediction or the interpolation of the intensity based on motion estimation are used to reduce the quantity of data for transmission or storage [14]. In fact the source of information in a sequence of images is the movement of the objects in the scene and/or the motion of the camera. In scene analysis the task is the determination of the 3-D motion of the objects in the scene and/or the motion of the camera, and also the structure of the 3-D objects or the reconstruction of the depth map. Apart from a specific application there are many fundamental methods for motion analysis. In this paper we try to ilustrate certain fundamental approaches, like the geometrical model presented in Section 2, or the possible ambiguous velocity fields presented in Section 3. In the last sections we propose and discuss techniques for applying the models in a specific domain, such as image coding or scene analysis.

Our work is based on the differential approach. In this case the spatial and temporal derivatives of the image intensity $g(x, y; t)$ can be obtained and they are relevant for motion estimation. Thus the information extracted from the image sequence is the spatial and temporal gradient of the moving edges. The motion constraint equation gives a relationship between temporal and spatial gradients of image intensity $g(x, y; t)$ and retinal velocities [8, 18]

$$g_x u + g_y v + g_t = 0 \quad \text{or} \quad g(x+\Delta x, y+\Delta y; t+\Delta t) - g(x, y; t) = 0 \tag{1}$$

This equation assumes that the intensity of any pixel does not change significantly over a short time interval and that the perceived change in image irradiance must be entirely due to motion. Using this equation and some smoothness constraints, Horn and Schunck [8] proposed a method to estimate the two-dimensional field. Nagel and Enkelmann [15] proposed an improvement of this method using "oriented smoothness constraints".

Another gradient-based approach uses the normal component of the velocity at contour points. This method requires a stage of edge detection [4, 5]. To measure the normal component a displacement in the perpendicular direction from the first contour to the second is considered [7]. This measurement also introduces errors if the edge is locally not a straight line. To approximate the error let us consider a translational 2-D motion and let $v^{\perp}$ and v^{T} be the normal and tangential components of the velocity, respectively. A simple development shows that the error in the measurement of the normal component is approximately

$$|\Delta v^{\perp}| \cong \kappa (v^{\mathrm{T}})^2$$

where κ is the curvature of the contour at the point considered.

Hildreth [7] proposed a method using smoothness constraints and measurement of the normal component. Waxman and Wohn [26] proposed a method ("Velocity Functional Method") that uses the normal component of the velocity field of a moving contour to calculate u and v and their 1st and 2nd order derivatives. Murray and Buxton [13] proposed a method which uses the normal component and the hypothesis of planar surfaces.

The most approaches cited here do not use any geometrical model for the velocity field. Optical flow often results from the presence in the scene of moving 3-D objects and/or from the motion of the camera in a static environment. Some geometrical aspects have thus to be considered as a guide in order to evaluate the spatial coherence of the 2-D velocity field. We present these geometrical considerations in Section 2, particularly the case of locally planar and rigid surfaces. Both orthographic and perspective projections are considered. In the case of perspective projection, the 3-D "perspective" velocity field, introduced by Scott [19], is used. The advantage of this vector is that, under the hypotheses considered, the velocity vector is linearly dependent on 2-D coordinates.

In Section 3 we present some results concerning the observability of the 2-D displacement field on a contour from the single normal component of the field. Equations of curves under ambiguous 3-D motion are given. In Section 4 we present a method for estimating the velocity field using only the geometrical considerations of Section 2. A Kalman filter is proposed as a recursive estimator of the 2-D velocity field on contour points and a relaxation method for an iterative estimation in homogeneous regions. In Section 5 we present a method which can be applied in non-homogeneous regions. In this case the spatial coherence in the velocity field does not only depend on geometry but also on image intensity gradients. The approach of Section 4 may be more useful for scene analysis in robotics whereas that of Section 5 may be used in image coding for movement compensation.

The last stage of the analysis consists of determining the 3-D motion parameters and the structure or depth characteristics from the 2-D velocity field (Section 6). Two approaches for reconstructing the structure and motion of 3-D objects can be distinguished. The first assumes that the velocity vector is given, in fact estimated, at some sparse image points, and the second assumes that the set of points is dense, for example as dense as the spatial sampling. In the second approach one can dispose of the first and second derivatives of the velocity field. We consider here the first case and we give a system of linear equations for determining the 3-D motion parameters and structure. For the same problem Bruss and Horn [3] applied a least-squares method in nonlinear equations, as did Zacharias et al. [27]. In Section 7 applications of the algorithms of Sections 4 and 5 are given.

2. The two-dimensional velocity field

In this section we will present and use a geometrical model of the 3-D motion in order to obtain some spatial relations on the two-dimensional velocity field. We wish to establish relations, e.g. linear differential or difference equations, which will permit us to determine linear estimators in association with motion measurements from the sequence of images.

Let us consider a 3-D coordinate system $O(X, Y, Z)$ (Fig. 1) that is fixed with respect to the retina, OZ being the line of sight [10]. The focal length of the camera is normalized ($f=1$). Let $\boldsymbol{V}_t=(V_X, V_Y, V_Z)$ be the translational velocity and $\boldsymbol{\Omega}=(\Omega_X, \Omega_Y, \Omega_Z)$ the angular velocity of a point $P(X, Y, Z)$ in an observer-relative decomposition. If the observer is moving through a static environment, the velocity of P is $-\boldsymbol{V}_t$ and $-\boldsymbol{\Omega}$. For a point $P(X, Y, Z)$ the velocity components are given by

$$X' = V_X + Z\Omega_Y - Y\Omega_Z \quad Y' = V_Y + X\Omega_Z - Z\Omega_X \quad Z' = V_Z + Y\Omega_X - X\Omega_Y \tag{2}$$

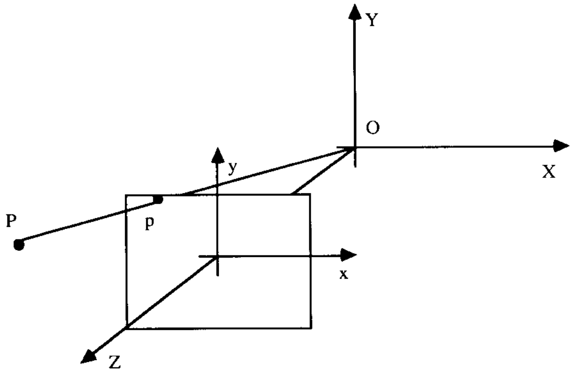

Fig. 1. The 3-D coordinate system.

We will consider orthographic as well as perspective projection.

2.1. Perspective projection and "perspective" velocity field

For a perspective projection a point P(X, Y, Z) is projected at p(x, y) with

$$x = X/Z \quad \text{and} \quad y = Y/Z \tag{3}$$

The "perspective" velocity vector is obtained by dividing the 3-D velocity vector of equation (2) by Z [19]:

$$\begin{aligned} u &= V_X/Z + \Omega_Y - y\Omega_Z \\ v &= V_Y/Z - \Omega_X + x\Omega_Z \\ w &= V_z/Z + y\Omega_X - x\Omega_Y \end{aligned} \tag{4}$$

It is useful to note that the depth in the equations (4) is not distinguishable from the velocity components. If the projection is perspective, the depth appears in a constant ratio with the translational velocity components. This means that objects far away and moving fast are perceived exactly as objects that are closer and moving more slowly. Thus the motion and depth parameters which can be determined using the two-dimensional velocity field are V_X/Z, V_Y/Z and V_Z/Z. No information concerning the depth can be obtained if the motion has only an angular component.

Let (φ, ψ) be the retinal velocity field. It is related to the "perspective" velocity field by

$$\varphi = u - xw \quad \text{and} \quad \psi = v - yw \tag{5}$$

Equations (4) show that the "perspective" velocity components depend on the depth Z of P. If we want to obtain expressions depending only on two-dimensional coordinates, we must make a hypothesis concerning the shape of the surface. A simple and plausible hypothesis is that the surface is locally plane. In other words we take into consideration only the orientation of the surface. Let us define a plane (excluding the degenerate case in which the plane contains the nodal point) $n_X X + n_Y Y + n_Z Z = 1$. For a perspective projection we have $1/Z = n_X x + n_Y y + n_Z$ and we obtain from equations (4):

$$\begin{aligned} u &= (n_Z V_X + \Omega_Y) + n_X V_X x + (n_Y V_X - \Omega_Z) y \\ v &= (n_Z Y_Y - \Omega_X) + (n_X V_Y + \Omega_Z) x + n_Y V_Y y \\ w &= n_Z V_Z + (n_X V_Z - \Omega_Y) x + (\Omega_X + n_Y V_Z) y \end{aligned} \tag{6}$$

Scott [19] indicates that if there exists a first-order "perspective" velocity field, like that of equations (6), there exists a one-parameter family of such fields, consistent with the same motion: $(u-cx,\ v-cy,\ w-c)$.

The expressions (6) for the "perspective" velocity field suggest spatial relations which are locally valid, depending on the validity of the planarity and rigidity assumption. We shall develop the discrete case in detail. For the continuous case we give only the general spatial relations. For the u-component we have

$$\frac{\partial^2 u}{\partial x^2}=\frac{\partial^2 u}{\partial y^2}=\frac{\partial^2 u}{\partial x \partial y}=0 \tag{7}$$

All second-order derivatives also vanish for the other components, v and w. This is the differential form of the spatial relations for the velocity field.

Let us now consider the case of discrete points in order to obtain the equivalent difference equations. We put the expressions (6) in matrix form

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where the identification of $\{(a_i, b_i, c_i);\ i=1,2,3\}$ from equations (6) is obvious.

We consider a series of points (x_k, y_k). Considering the first line of the above equation for four points, we can write,

$$\begin{bmatrix} u_{k+1}-u_k & x_{k+1}-x_k & y_{k+1}-y_k \\ u_k-u_{k-1} & x_k-x_{k-1} & y_k-y_{k-1} \\ u_{k-1}-u_{k-2} & x_{k-1}-x_{k-2} & y_{k-1}-y_{k-2} \end{bmatrix} \begin{bmatrix} -1 \\ a_1 \\ b_1 \end{bmatrix} = 0 \tag{8}$$

We can write the same equations for the components v and w. A consequence of equation (8) is that the determinant of the matrix must be zero,

$$\begin{vmatrix} u_{k+1}-u_k & x_{k+1}-x_k & y_{k+1}-y_k \\ u_k-u_{k-1} & x_k-x_{k-1} & y_k-y_{k-1} \\ u_{k-1}-u_{k-2} & x_{k-1}-x_{k-2} & y_{k-1}-y_{k-2} \end{vmatrix} = 0 \tag{9}$$

and the same for v and w. We can then write

$$D_{1,k}(u_{k+1}-u_k)-D_{2,k+1}(u_k-u_{k-1})+D_{1,k+1}(u_{k-1}-u_{k-2})=0$$

with

$$D_{1,k}=\begin{vmatrix} x_k-x_{k-1} & y_k-y_{k-1} \\ x_{k-1}-x_{k-2} & y_{k-1}-y_{k-2} \end{vmatrix} \quad \text{and} \quad D_{2,k+1}=\begin{vmatrix} x_{k+1}-x_k & y_{k+1}-y_k \\ x_{k-1}-x_{k-2} & y_{k-1}-y_{k-2} \end{vmatrix}$$

If the three points (x_k, y_k), (x_{k-1}, y_{k-1}) and (x_{k-2}, y_{k-2}) are not aligned, then $D_{1,k} \neq 0$ and we can write

$$u_{k+1}=\left(1+\frac{D_{2,k+1}}{D_{1,k}}\right)u_k-\frac{D_{1,k+1}+D_{2,k+1}}{D_{1,k}}u_{k-1}+\frac{D_{1,k+1}}{D_{1,k}}u_{k-2} \tag{10}$$

which is an autoregressive relation on the velocity. A similar relation is valid for the other components. If the points are aligned, we can assume, without loss of generality, that $x_k-x_{k-1} \neq 0$. From equation (8) we obtain the following autoregressive relation

$$u_k=\left(1+\frac{x_k-x_{k-1}}{x_{k-1}-x_{k-2}}\right)u_{k-1}-\frac{x_k-x_{k-1}}{x_{k-1}-x_{k-2}}u_{k-2} \tag{11}$$

Let us determine a relation on the image plane. We consider four points (i,j), $(i-1,j)$, $(i,j-1)$ and $(i-1,j-1)$. Applying expression (10) we obtain

$$u_{i,j}-u_{i-1,j}-u_{i,j-1}+u_{i-1,j-1}=0 \tag{12a}$$

From equation (11) we obtain two expressions along the vertical and the horizontal axes

$$u_{i-1,j}-2u_{i,j}+u_{i+1,j}=0 \tag{12b}$$

$$u_{i,j-1}-2u_{i,j}+u_{i,j+1}=0 \tag{12c}$$

The same expressions are valid for the other components, v and w. These difference equations constitute the discrete version of the spatial relations on the velocity field, and they are the principal result of this section.

2.2. Orthographic projection

For an orthographic projection we have $x=X$ and $y=Y$. The velocity of $p(x,y)$ is then given by

$$u=V_X+Z\Omega_Y-y\Omega_Z \quad \text{and} \quad v=V_Y-Z\Omega_X+x\Omega_Z \tag{13}$$

If the projection is orthographic, the depth is not distinguishable from the components (Ω_X, Ω_Y) and if these components are vanishing, no depth information is obtained from the image flow.

Equations (13) show that the 2-D velocity components depend on the depth Z of P. If we want to obtain expressions depending only on two-dimensional coordinates, we must assume a hypothesis concerning the shape of the surface. A simple and plausible hypothesis, as in the case of the "perspective" velocity vector, is that the surface is locally plane. In other words we take into consideration only the surface orientation.

For a plane, as defined above (with $n_Z \neq 0$), we have $Z=(1-n_Xx-n_Yy)/n_Z$ and we obtain:

$$\begin{aligned} u &= \left(V_X+\frac{\Omega_Y}{n_Z}\right)-\frac{n_X}{n_Z}\Omega_Y x-\left(\Omega_Z+\frac{n_Y}{n_Z}\Omega_Y\right)y \\ v &= \left(V_Y-\frac{\Omega_X}{n_Z}\right)+\left(\Omega_Z+\frac{n_X}{n_Z}\Omega_X\right)x+\frac{n_Y}{n_Z}\Omega_X y \end{aligned} \tag{14}$$

The expressions (14) for the 2-D velocity field suggest spatial relations, which are locally valid, depending on the validity of the planarity and rigidity assumption. The spatial relations on the velocity components are the same as in the case of the "perspective" velocity vector.

3. Ambiguity in the estimation of the 2-D velocity field from the normal component

An important task in computer vision is to recover the motion and structure of 3-D objects from the changing intensities in an image. A crucial question which arises is the unicity of the interpretation of the image flow due to the motion. In this paper we discuss this question. We assume that the measurement of motion takes place along contours in the image. The existence of ambiguous interpretations of 3-D motion is known and discussed by Hildreth [7].

There are two reasons for possible ambiguous interpretations. The first is that the 3-D world is projected onto 2-D images. The result of this is that different surfaces which undergo different motions can have the same 2-D motion. The second reason is that the contours in the image are due to changing intensities and there may be ambiguous or incorrect interpretations resulting from the unknown relations between the shape of the moving surface and the changing intensities of the image. In this paper we investigate only ambiguities due to the first reason.

Hildreth [7] gives some examples of human motion perception from which illusions result, such as the rotating ellipse, the rotating spiral and the barberpole illusion. We will give here a geometrical interpretation of these cases. More generally we will give the differential equation of all curves for which there are multiple interpretations of motion. We also give the solution of this equation in the case of planar surfaces under orthographic projection and certain cases of non-planar surfaces.

This study is particularly interesting in computer vision. The existence of ambiguities in recovering 3-D motion and structure from image flow means that using an exact method does not result in a unique solution. In this paper we propose a method giving the exact velocity vector for planar surfaces (Section 4.1); Waxman and Wohn [26] proposed another exact method under the same hyothesis. However, if constraints are introduced, such as the velocity smoothing constraints [7], a unique solution may be obtained.

In Ref. [23] we have demonstrated that curves of second degree do not admit a unique solution to the problem of reconstruction of the velocity field from the normal component of velocity. This formulation is equivalent to the present formulation of the problem but the method used in Ref. [23] is applicable only to curves in polynomial form. The same result concerning the second-degree curves is obtained by Waxman and Wohn [26] using another method. The result given below permits the classification of all plane curves under orthographique projection.

Let us define a moving curve by the set of points

$$C_2 = \{(x(\tau), y(\tau)) \colon \tau \in \mathbb{R}\}$$

where τ is a curve parameter. An ambiguous 3-D motion results if for every point on the curve the tangential vector may be proportional to a velocity vector

$$\mathrm{d}y/\mathrm{d}x = v/u \tag{15}$$

Such a motion is not apparent.

We will examine below different cases of curves in orthographic or perspective projection and give some solutions and examples.

3.1. Planar curves under orthographic projection

Using equation (8) we obtain the following plane differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{a_2 x + b_2 y + c_2}{a_1 x + b_1 y + c_1}$$

Let us define the matrix

$$\boldsymbol{A} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$$

and assume that det $\boldsymbol{A} \neq 0$. We can perform a coordinate translation $x' = x - x_0$ and $y' = y - y_0$ and therefore obtain

$$\frac{\mathrm{d}y'}{\mathrm{d}x'} = \frac{a_2 x' + b_2 y'}{a_1 x' + b_1 y'}$$

where x_0 and y_0 are the solutions of the following equation

$$\boldsymbol{A} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

The solution of the above plane differential equation is known [16].

We will give only the results here. For simplicity and without loss of generality we use for the curve equation the initial coordinates (x, y), and not (x', y').

*3.1.1. The matrix **A** has two distinct real eigenvalues*

The curve equation is

$$|y| = c|x|^{\lambda_2/\lambda_1} \tag{16}$$

where λ_1 and λ_2 are real eigenvalues of $\boldsymbol{A}$ $(\lambda_1 \neq \lambda_2)$. Any curve obtained by a reversible linear transformation from equation (16) also admits a velocity vector as tangential vector. As examples we can consider the parabola $(\lambda_2/\lambda_1 = 2)$ and the hyperbole $(\lambda_2/\lambda_1 = -1)$. Let us consider the parabola $y = cx^2$ and the velocity vector $u = 1$, $v = 2cx$. This vector is tangential to the curve. In Fig. 2 we illustrate the type of curves corresponding to equation (16) (from Ref. [16]).

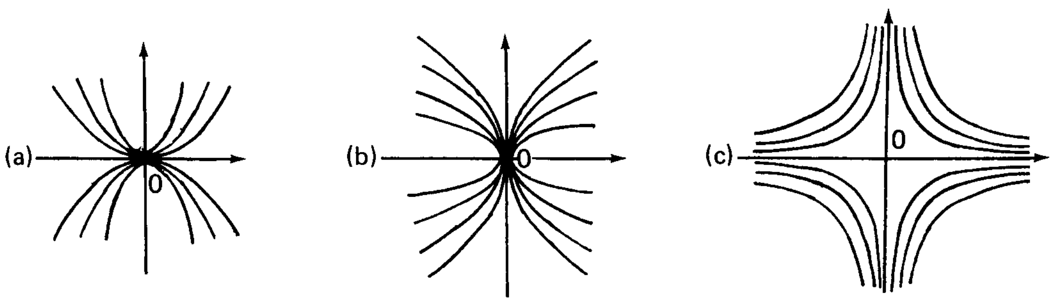

Fig. 2. Curves for the case where the matrix $\boldsymbol{A}$ has two real eigenvalues. (a) $\lambda_2/\lambda_1 > 1$; (b) $0 < \lambda_2/\lambda_1 < 1$; (c) $\lambda_2/\lambda_1 < 0$ (from Ref. [16]).

*3.1.2. The eigenvalues of **A** are complex*

The curve equation is given in polar coordinates

$$\rho = c \exp(a\theta/b)$$

In general this is a logarithmic spiral. If $a = 0$, it is a circle. A reversible linear transformation gives an ellipse. Let us consider three examples: the circle, the ellipse and a logarithmic spiral.

(a) Let us consider the circle $x = \cos \tau$, $y = \sin \tau$ and the velocity vector $u = -y$, $v = x$. The 3-D motion parameters are given by

$$n_X \Omega_Y = 0 \qquad n_Y \Omega_X = 0$$

$$V_X + \Omega_Y / n_Z = 0 \qquad V_Y - \Omega_X / n_Z = 0$$

$$\Omega_Z + n_X \Omega_X / n_Z = 1 \qquad \Omega_Z + n_Y \Omega_Y / n_Z = 1$$

There are two families of solutions:

$$(\forall n_X, \forall n_Y; \Omega_X = 0, \Omega_Y = 0, \Omega_Z = 1; V_X = 0, V_Y = 0)$$

$$(n_X = 0, n_Y = 0; \Omega_Z = 1, V_X + \Omega_Y / n_Z = 0, V_Y - \Omega_X / n_Z = 0)$$

The first family is an orientation-independent solution, and gives an interpretation for a permanent rigid motion. This is the natural interpretation in human vision.

(b) Let us consider the ellipse $x = a \cos \tau$, $y = b \sin \tau$ and the velocity vector $u = -ay/b$, $v = bx/a$. There are two families of solutions

$$(n_X = 0, \Omega_X = 0, n_Y \Omega_Y / n_Z = a/b - b/a, \Omega_Z = b/a, V_X + \Omega_Y / n_Z = 0, V_Y = 0)$$

$$(n_Y = 0, n_X \Omega_X / n_Z = b/a - a/b, \Omega_Y = 0, \Omega_Z = a/b, V_X = 0, V_Y - \Omega_X / n_Z = 0)$$

Both families of solutions provide an instantaneous interpretation of 3-D motion. This means that in a long-range observation the perceived motion is a curve deformation. It is known in human motion perception that an ellipse does not appear rigid when it rotates, but appears to be deformed continuously [7].

(c) Let us consider the logarithmic spiral $x = \mathrm{e}^{\tau} \cos \tau$, $y = \mathrm{e}^{\tau} \sin \tau$ and the velocity vector $u = x - y$, $v = x + y$. The parameters of the 3-D motion are given by the following family

$$(V_X + \Omega_Y / n_Z = 0, V_Y - \Omega_X / n_Z = 0, n_X \Omega_Y / n_Z = -1,$$
$$n_Y \Omega_X / n_Z = 1, \Omega_Z + n_X \Omega_X / n_Z = 1, \Omega_Z + n_Y \Omega_Y / n_Z = 1)$$

There are no solutions to the plane orientation and the 3-D motion parameters. This means that such a tangential vector is always interpreted as a curve deformation. This is very well illustrated by Hildreth [7] who gives an example of a logarithmic spiral undergoing a rotation around the line of sight; the perceived motion is an expansion of the curve.

3.1.3. *The matrix A has two equal eigenvalues*

The curve is given by

$$|y| = (\gamma + \delta \ln |x|)|x|$$

This family of curves is illustrated in Fig. 3 (from Ref. [16]).

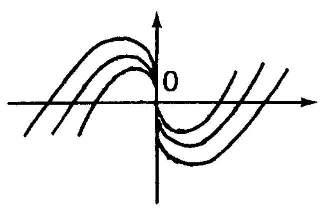

Fig. 3. Curves for the case where the two eigenvalues of $\boldsymbol{A}$ are equal (from Ref. [16]).

3.2. *Arbitrary 3-D curve in orthographic projection*

Let us consider the 3-D curve

$$C_3 = \{(X(\tau),\ Y(\tau), Z(\tau));\ \tau \in \mathbb{R}\}$$

The 3-D motion is ambiguous if the tangent vector is given by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{aX(\tau) + b_2 Z(\tau) + c_2}{b_1 Z(\tau) - aY(\tau) + c_1}$$

As an example let us consider the barberpole illusion. The parameter equation of the curve is given by: $(X(\tau) = \tau,\ Y(\tau) = \cos\tau,\ Z(\tau) = \sin\tau)$. We can verify that for $a = b_1 = c_2 = 0$, $c_1 = 1$, and $b_2 = -1$ the plane differential equation is valid. This corresponds to the motion

$$V_X = 1, \quad V_Y = 0, \quad \Omega_X = 1, \quad \Omega_Y = 0, \quad \Omega_Z = 0$$

We discuss below the consequence of the ambiguous interpretation of 3-D motion in computer vision. Although the ambiguities are inherent in the motion projection onto 2-D images, algorithms exist which give a unique solution. This results from certain constraints. However, exact methods, such as that proposed here or that of Waxman and Wohn [26], do not give a unique solution. We examine the constraining algorithms and exact methods separately.

Let us first consider the algorithm proposed by Hildreth [7]. The smoothing constraint means that the velocity vector is assumed constant, the residual considered as a noise. Under this constraint only the straight lines may admit an ambiguous interpretation. The exact methods are more vulnerable on ambiguities. But as the number of independent coefficients is limited (for example six for a planar surface under orthographic projection) it is expected that the noise introduced by edge detection and particularly sampling will be sufficient to stabilize the algorithm. Otherwise we can test the fitting of any contour in the image and the plane differential equation and apply the algorithm only to contours which will give a unique solution, and then infer the computed velocity field in all points. There are two ways of applying the test, either directly to a system of linear equations of unknown coefficients, or by applying the algorithm computing the velocity field with different initial conditions and verifying if it is converging to the same solution.

4. Estimation of the velocity field using the geometric model

4.1. *Estimation at contour points*

We propose to use the autoregressive relation given in Section 2.1 for estimating the "perspective" velocity vector is assumed constant, the residual considered as a noise. Under this constraint only the procedure. Let us consider equation (10) and write the autoregressive relation for the three velocity components

$$u_{k+1} = \beta_k u_k + \beta_{k-1} u_{k-1} + \beta_{k-2} u_{k-2}$$

$$v_{k+1} = \beta_k v_k + \beta_{k-1} v_{k-1} + \beta_{k-2} v_{k-2}$$

$$w_{k+1} = \beta_k w_k + \beta_{k-1} w_{k-1} + \beta_{k-2} w_{k-2}$$

We designate $\boldsymbol{\xi}_k$ the state vector. It is given by

$$\boldsymbol{\xi}_k = [u_k \quad u_{k-1} \quad u_{k-2} \quad v_k \quad v_{k-1} \quad v_{k-2} \quad w_k \quad w_{k-1} \quad w_{k-2}]^{\mathrm{T}}$$

The state equation is given below

$$\boldsymbol{\xi}_{k+1} = \begin{bmatrix} \boldsymbol{\Phi}_{k+1|k} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{k+1|k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi}_{k+1|k} \end{bmatrix} \boldsymbol{\xi}_k + \boldsymbol{\omega}_k$$

where the noise vector $\boldsymbol{\omega}_k$ is zero-mean with covariance $\boldsymbol{Q}_k$ and the transition matrix $\boldsymbol{\Phi}_{k+1|k}$ is

$$\boldsymbol{\Phi}_{k+1|k} = \begin{bmatrix} \beta_k & \beta_{k-1} & \beta_{k-2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

The equation of measurement is given by

$$y_k = [c_{1k} \quad 0 \quad 0 \quad c_{2k} \quad 0 \quad 0 \quad c_{3k} \quad 0 \quad]\boldsymbol{\xi}_k + z_k$$

where y_k is the measured projection of the velocity on the normal vector of the contour

$$(c_{1k}, c_{2k}, c_{3k}) = \left(f_x(x_k, y_k), f_y(x_k, y_k), -\frac{x_k f_k(x_k, y_k) + y_k f_y(x_k, y_k)}{\sqrt{f_x^2(x_k, y_k) + f_y^2(x_k, y_k)}} \right)$$

and z_k is a measurement noise which is assumed zero-mean with variance R_k ($f(x, y) = 0$ is the equation of the contour). The system and measurement noise are assumed to be independent and independent between different points. This is a classical linear estimation problem and the solution is the well-known Kalman filter for the discrete case [9, 11]. We discuss here how to apply the filter on a contour. If the contour is closed, we can choose any point as initial point and develop the filter around the contour. It is obvious that the velocity at the first points will be less well estimated than that at the last points. In all cases a second application of the filter around the contour will be necessary. If the contour is not closed, the same technique can be applied inverting the sense of direction of the filter at the final point.

4.2. Estimation on the image plane

We can use the spatial relations of the "perspective" velocity field and the motion constraint equation (1) for estimating the retinal motion. We have three spatial relations for each velocity component at each point. The more natural way to use them is the minimization of a quadratic functional, which has the following form

$$\lambda^2(\|\boldsymbol{A}_1\boldsymbol{U}\|^2 + \|\boldsymbol{A}_2\boldsymbol{U}\|^2 + \|\boldsymbol{A}_3\boldsymbol{U}\|^2 + \|\boldsymbol{A}_1\boldsymbol{V}\|^2 + \|\boldsymbol{A}_2\boldsymbol{V}\|^2 + \|\boldsymbol{A}_3\boldsymbol{V}\|^2 + \|\boldsymbol{A}_1\boldsymbol{W}\|^2 + \|\boldsymbol{A}_2\boldsymbol{W}\|^2 + \|\boldsymbol{A}_3\boldsymbol{W}\|^2) + \|\boldsymbol{G}_x\boldsymbol{U} + \boldsymbol{G}_y\boldsymbol{V} + \boldsymbol{G}_z\boldsymbol{W} + \boldsymbol{G}_t\|^2$$

where $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{W}$ are the complete velocity components, $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_3$, $\boldsymbol{G}_x$, $\boldsymbol{G}_y$, $\boldsymbol{G}_z$ and $\boldsymbol{G}_t$ are linear operators on the image plane, and $\| \ \|^2$ is the euclidean norm. The operators $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ are determined from the spatial relations. The operators $\boldsymbol{G}_x$, $\boldsymbol{G}_y$, $\boldsymbol{G}_z$, with

$$g_{z;i,j} = -x_{i,j} g_{x;i,j} - y_{i,j} g_{y;i,j},$$

contain the spatial gradient and $\boldsymbol{G}_t$ contains the temporal gradient of the image intensity. The minimization of the above quadratic functional gives a system of linear equations, given below

$$\begin{aligned} &\lambda^2[(\boldsymbol{A}_1)^{\mathrm{T}}\boldsymbol{A}_1 + (\boldsymbol{A}_2)^{\mathrm{T}}\boldsymbol{A}_2 + (\boldsymbol{A}_3)^{\mathrm{T}}\boldsymbol{A}_3]\boldsymbol{U} + \boldsymbol{G}_x^{\mathrm{T}}(\boldsymbol{G}_x\boldsymbol{U} + \boldsymbol{G}_y\boldsymbol{V} + \boldsymbol{G}_z\boldsymbol{W} + \boldsymbol{G}_t) = 0 \\ &\lambda^2[(\boldsymbol{A}_1)^{\mathrm{T}}\boldsymbol{A}_1 + (\boldsymbol{A}_2)^{\mathrm{T}}\boldsymbol{A}_2 + (\boldsymbol{A}_3)^{\mathrm{T}}\boldsymbol{A}_3]\boldsymbol{V} + \boldsymbol{G}_y^{\mathrm{T}}(\boldsymbol{G}_x\boldsymbol{U} + \boldsymbol{G}_y\boldsymbol{V} + \boldsymbol{G}_z\boldsymbol{W} + \boldsymbol{G}_t) = 0 \\ &\lambda^2[(\boldsymbol{A}_1)^{\mathrm{T}}\boldsymbol{A}_1 + (\boldsymbol{A}_2)^{\mathrm{T}}\boldsymbol{A}_2 + (\boldsymbol{A}_3)^{\mathrm{T}}\boldsymbol{A}_3]\boldsymbol{W} + \boldsymbol{G}_z^{\mathrm{T}}(\boldsymbol{G}_x\boldsymbol{U} + \boldsymbol{G}_y\boldsymbol{V} + \boldsymbol{G}_z\boldsymbol{W} + \boldsymbol{G}_t) = 0 \end{aligned} \tag{17}$$

where superscript T signifies the adjoint operator. We specify in the following the operators A_1, A_2, A_3, and give a method for solving the above system of equations.

The quadratic functional for minimizing is obtained from the spatial relations (10); a natural way to use equations (10) is to minimize the mean of the sum of traces like

$$\operatorname{tr}\left\{\begin{bmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{bmatrix}^2\right\}.$$

We thus obtain the following functional

$$\iint [\lambda^2[(u_{xx})^2+2(u_{xy})^2+(u_{yy})^2+(v_{xx})^2+2(v_{xy})^2+(v_{yy})^2+(w_{xx})^2+2(w_{xy})^2+(w_{yy})^2] \\ +(g_x u+g_y v+g_z w+g_t)^2]\,dx\,dy$$

This approach is related to the regularization approach [17, 22]. The smoothness constraint of this functional is to be close to the surface reconstruction constraints by a thin plate model [21]. This smoothness constraint is also used by Anandan and Weiss [1] as a heuristic for the two-dimensional velocity field. The smoothness constraint of Anandan and Weiss [1] thus corresponds to a 2-D velocity field from an orthographic projection under the local planarity assumption.

The minimization leads to a system of linear equations of the following form

$$u_{i,j}=u_{i,j}-g_{x;i,j}\gamma_{i,j} \quad v_{i,j}=v_{i,j}-g_{y;i,j}\gamma_{i,j} \quad w_{i,j}=w_{i,j}-g_{z:i,j}\gamma_{i,j}$$

with

$$\gamma_{i,j}=\frac{(g_{x;i,j}\bar{u}_{i,j}+g_{y;i,j}\bar{v}_{i,j}+g_{z;i,j}\bar{w}_{i,j}+g_{t;i,j})}{[\lambda^2+(g_{x;i,j})^2+(g_{y;i,j})^2+(g_{z;i,j})^2]}$$

The averaging is given below for the u-component (without loss of generality we use the same weighting factor λ^2). It is also represented in Fig. 4.

$$\bar{u}_{i,j}=0.4(u_{i+1,j}+u_{i-1,j}+u_{i,j+1}+u_{i,j-1})-0.1(u_{i+1,j+1}+u_{i+1,j-1}+u_{i-1,j+1}+u_{i-1,j-1}) \\ -0.05(u_{i+2,j}+u_{i,j-2}+u_{i-2,j}+u_{i,j+2}) \tag{18}$$

The numerical coefficients in equation (18) are obtained using the discrete form of the spatial relations (12) to determine the linear operators of the expression (17).

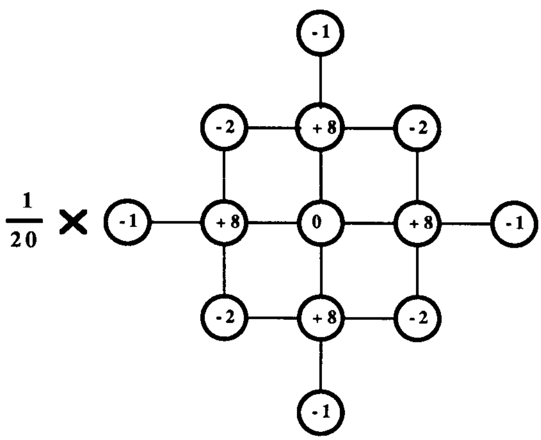

Fig. 4. The nodal molecule corresponding to the averaging of equation (18).

We can use these equations in an iterative method, such as the Gauss–Seidel method, for determining the solution of the system of equations (n indicates the number of iteration)

$$u_{i,j}^{(n+1)} = \bar{u}_{i,j}^{(n)} - g_{x;i,j}\gamma_{i,j}^{(n)} \quad v_{i,j}^{(n+1)} = \bar{v}_{i,j}^{(n)} - g_{y;i,j}\gamma_{i,j}^{(n)} \quad w_{i,j}^{(n+1)} = \bar{w}_{i,j}^{(n)} - g_{z;i,j}\gamma_{i,j}^{(n)}$$

This is a classical relaxation method [6] for solving an optimization problem; it is also used by Horn and Schunck [8] to determine the optical flow using other smoothness constraints. The difference lies in the choice of local averaging.

Remark: In the case of an orthographic projection the same methods can be applied to the two components of the velocity vector.

5. Recursive–iterative estimation of the velocity vector in non-homogeneous regions

In this section we propose a recursive estimator of the displacement field which can be used for predictive coding with movement compensation. The estimator is composed of three parts. The first part is an a priori estimator, or a recursive predictor, of the velocity vector, using the estimated values of velocity vectors in the causal neighborhood of the considered element. The second part is a detector of movement discontinuities, using the displacement frame difference (1). The third part is an a posteriori estimator, which operates in an iterative method in order to minimize the displacement frame difference. In the following we shall present the three parts of the estimator separately. The structure of this motion estimator is classical and is also used by Biemond et al. [2], by Walker and Rao [24] and by Moorhead and Rajala [12].

5.1. *Prediction of the velocity vector*

Let us consider the problem of estimating the vector $\boldsymbol{\eta}_{i,j} = [u_{i,j} \;\; v_{i,j}]^{\mathrm{T}}$ using the estimates of the velocity at the points $\{(i-m, j): m = 1, \ldots, i\} \cup \{(l, j-n): \forall l;\ n = 1, \ldots, j\}$. We limit ourselves to the neighborhood $\{(i-1, j), (i, j-1), (i-1, j-1)\}$. Let us consider the following quadratic form

$$Q_0(\boldsymbol{\eta}) = \operatorname{tr}\left\{[\boldsymbol{\eta} - \boldsymbol{\eta}_{i-1,j} \quad \boldsymbol{\eta} - \boldsymbol{\eta}_{i,j-1} \quad \boldsymbol{\eta} - \boldsymbol{\eta}_{i-1,j-1}]\boldsymbol{K}^{-1}\begin{bmatrix} (\boldsymbol{\eta} - \boldsymbol{\eta}_{i-1,j})^{\mathrm{T}} \\ (\boldsymbol{\eta} - \boldsymbol{\eta}_{i,j-1})^{\mathrm{T}} \\ (\boldsymbol{\eta} - \boldsymbol{\eta}_{i-1,j-1})^{\mathrm{T}} \end{bmatrix}\right\} \tag{19}$$

where $\boldsymbol{\eta}_{i-1,j}$ is the estimate of the velocity vector at $(i-1, j)$, $\boldsymbol{\eta}_{i,j-1}$ at $(i, j-1)$ and $\boldsymbol{\eta}_{i-1,j-1}$ at $(i-1, j-1)$. Matrix $\boldsymbol{K}$ can be interpreted as the covariance of the vectors $[u - u_{i-1,j} \;\; u - u_{i,j-1} \;\; u - u_{i-1,j-1}]$ and $[v - v_{i-1,j} \;\; v - v_{i,j-1} \;\; v - v_{i-1,j-1}]^{\mathrm{T}}$, which are also considered uncorrelated. The estimator of $\boldsymbol{\eta}$ minimizes the above quadratic form and is given by

$$\boldsymbol{\eta} = \frac{[\boldsymbol{\eta}_{i-1,j} \quad \boldsymbol{\eta}_{i,j-1} \quad \boldsymbol{\eta}_{i-1,j-1}]\boldsymbol{K}^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\boldsymbol{K}^{-1}\mathbf{1}} \tag{20}$$

with $\mathbf{1}^{\mathrm{T}} = [1 \;\; 1 \;\; 1]$.

Matrix $\boldsymbol{K}$ is unknown, however. We propose an intensity-dependent expression of $\boldsymbol{K}$. We postulate that the velocity vector field should vary more smoothly in the direction of small intensity variations, than in the direction of prominent transitions. Adopting this conservative standpoint in non-homogeneous regions, we have not to detect occluding edge lines. We attribute to the covariance a value proportional to the inner product of the intensity gradients in the two respective directions and to the product of the distances of the points in the difference of velocity. This assumes that u and v are unknown deterministic parameyers; $\boldsymbol{\eta}_{i-1,j}$, $\boldsymbol{\eta}_{i,j-1}$ and $\boldsymbol{\eta}_{i-1,j-1}$ are supposed random with the same mean $\boldsymbol{\eta}$. It is then assumed that

$$\boldsymbol{\eta}_{i-1,j} = \boldsymbol{\eta} + \boldsymbol{\zeta}_1 + g_x \boldsymbol{\varepsilon}_1$$

$$\boldsymbol{\eta}_{i,j-1} = \boldsymbol{\eta} + \boldsymbol{\zeta}_2 + g_y \boldsymbol{\varepsilon}_2$$

$$\boldsymbol{\eta}_{i-1,j-1} = \boldsymbol{\eta} + \boldsymbol{\zeta}_3 + g_x \boldsymbol{\varepsilon}_1 + g_y \boldsymbol{\varepsilon}_2$$

where the covariance of the variables $\boldsymbol{\zeta}_i$ $(i=1,2,3)$ contains only diagonal elements equal to $\mu\sigma^2$, and the covariance matrix of the variables $\boldsymbol{\varepsilon}_i$ $(i=1,2,3)$ contains only diagonal elements equal to σ^2; $\boldsymbol{\zeta}_i$ $(i=1,2,3)$ and $\boldsymbol{\varepsilon}_i$ $(i=1,2,3)$ are assumed uncorrelated. We thus propose the following expression for the covariance matrix

$$\boldsymbol{K} = \begin{bmatrix} \mu + g_x^2 & 0 & g_x^2 \\ 0 & \mu + g_y^2 & g_y^2 \\ g_x^2 & g_y^2 & \mu + g_x^2 + g_y^2 \end{bmatrix} \tag{21}$$

After some approximations we obtain from equations (20) and (21) the following a priori estimation of $\boldsymbol{\eta}$,

$$\boldsymbol{\eta}^0 = \alpha_x \boldsymbol{\eta}_{i-1,j} + \alpha_y \boldsymbol{\eta}_{i,j-1} - \alpha_x \alpha_y \boldsymbol{\eta}_{i-1,j-1} \tag{22}$$

where the coefficients α_x and α_y depend on the spatial gradients g_x, g_y and on $\mu (0 < \alpha_x, \alpha_y < 1)$

$$\alpha_x = \frac{\mu + g_y^2}{\mu + g_x^2 + g_y^2} \quad \text{and} \quad \alpha_y = \frac{\mu + g_x^2}{\mu + g_x^2 + g_y^2}$$

If $g_x^2 + g_y^2$ is small with respect to μ, then $\alpha_x = \alpha_y = 1$, and we have the same spatial relation to the velocity field as that given in equation (12a) for an orthographic projection under a local planarity and rigidity assumption.

5.2. Detection of movement estimation discontinuities

The motion estimation in non-homogeneous regions necessitates the detection of discontuities. The detection must be based on the prediction of the velocity, because the latter is an expression of the spatial coherence of the velocity field. The measurement on which the detection can be based is the prediction error of the intensity, that is, the displacement frame difference

$$e^0(i,j) = g(i,j;k) - g(i-u^0, j-v^0; k-1).$$

The discontinuities detected using the displacement frame difference may be caused by the fact that the motion model is not valid or by the existence of ambiguities in the measurements (spatio-temporal gradients).

If we take into account the constraints of transmission, two approaches, which are described below, can be considered.

(1) We compare the prediction error with the frame difference. If a discontinuity is detected, the frame difference is transmitted in place of the displacement frame difference, as well as a message of discontinuity detection. The a priori estimation is put back to zero ($\boldsymbol{\eta}^0 = 0$).

(2) We consider a causal neighborhood and apply the predicted displacement. This operation gives the following prediction error

$$e^0(i,j)=\sum_{m,n}|g(i-n,j-m;\,k)-g(i-n-u^0,j-m-v^0;\,k-1)|$$

We compare the prediction error with the frame difference at the same neighborhood. If a discontinuity is detected, the a priori estimation is again put back to zero, But now no message of discontinuity detection is transmitted, because the decoder can also operate the detection.

5.3. *A posteriori estimation of the velocity vector*

The a posteriori estimation is realized by an iterative method. The estimation error is given by

$$e(i,j)=g(i,j;\,k)-g(i-u,j-v;k-1)$$

Let $\boldsymbol{\eta}^n$ be the velocity estimation after n iterations. The optimization is obtained using a quadratic criterion

$$Q(\boldsymbol{\eta})=e^2+\lambda\,\|\boldsymbol{\eta}-\boldsymbol{\eta}^n\|^2 \tag{23}$$

The second part of this equation can be considered as a regularization constraint for the underdetermined problem of estimating the displacement vector by minimizing the displacement frame difference. Expression (23) is not quadratic on the unknown variable $\boldsymbol{\eta}$. A linearization of the estimation error can be obtained using a first-order Taylor development, and we obtain

$$e(i,j)=e^{n-1}(i,j)+g_x(i-u^{n-1},j-v^{n-1};\,k-1)(u-u^{n-1})+g_y(i-u^{n-1},j-v^{n-1};\,k-1)(v-v^{n-1})$$

With this approximation we obtain the following solution

$$\boldsymbol{\eta}^n=\boldsymbol{\eta}^{n-1}-\frac{e^{n-1}(i,j)[g_x(i-u^{n-1},j-v^{n-1};\,k-1)\quad g_y(i-u^{n-1},j-v^{n-1};\,k-1]^{\mathrm{T}}}{\lambda+g_x^2(i-u^{n-1},j-v^{n-1};\,k-1)+g_y^2(i-u^{n-1},j-v^{n-1};\,k-1)} \tag{24}$$

The number of iterations is either fixed a priori or determined using a test on estimation error.

6. 3-D motion parameters and structure estimation

We now consider that the retinal velocity field is estimated and we will determine the 3-D motion parameters and structure. If the 2-D velocity field is dense, Longuet-Higgins and Prazdny [10] and Waxman and Ullman [25] give a method to calculate at each point the 3-D motion parameters and the surface orientation and curvature scaled by the depth. They use φ, ψ and their 1st and 2nd order spatial derivatives, and reconstitute all information contained in the optical flow, but we can note here that derivation amplifies the errors of 2-D velocity field estimation. Nevertheless Waxman and Ullman [25] give the velocity field and its spatial derivatives directly as initial observables.

Our approach is different, but it can be considered as complementary. We use the "perspective" velocity field at distinct image points. If the retinal velocity field is sparse, this is the single method for 3-D motion and depth reconstruction. Information on the structure in the case of a sparse velocity field is the relative depth, and in the case of a dense field it is the surface orientation and eventually the curvature.

We study the cases of planar surfaces and curved surfaces separately.

6.1. Planar surface

If the "perspective" velocity vector is known at three non-aligned points from the expressions [6] we can obtain nine values, called the essential parameters.

$$n_X V_X + c = a_1, \quad n_Y V_X - \Omega_Z = b_1, \quad n_Z V_X + \Omega_Y = c_1$$

$$n_X V_Y + \Omega_Z = a_2, \quad n_Y V_Y + c = b_2, \quad n_Z V_T - \Omega_X = c_2$$

$$n_X V_Z - \Omega_Y = a_3, \quad \Omega_X + n_Y V_Z = b_3, \quad n_Z V_Z + c = c_3$$

This system of non-linear equations was solved by Subbarao and Waxman [20] to obtain: n_X/n_Z, $n_Y/n_Z (n_Z \neq 0)$, $n_Z V_X$, $n_Z V_Y$, $n_Z V_Z$, Ω_X, Ω_Y and Ω_Z. (It can easily be demonstrated that the condition $n_Z \neq 0$ is equivalent to: $(a_3 + c_1)^2(c_3 - b_2) + (b_3 + c_2)^2(c_3 - a_1) + (b_3 + c_2)(a_3 + c_1)(a_2 + b_1) \neq 0$).

In general two solutions exist. The duality can be resolved using spatial or temporal coherence [20].

6.2. Non-planar surface

Let us assume that the three components of the "perspective" velocity vector are known at four image points. From equation (3) we can write

$$Z_i = \frac{V_X}{u_i - \Omega_Y + y_i\Omega_Z + x_i c} = \frac{V_Y}{v_i + \Omega_X + y_i c - x_i \Omega_Z} = \frac{V_Z}{w_i + c - y_i\Omega_X + x_i\Omega_Y} \tag{25}$$

for $i = 1, 2, 3, 4$. We assume that $V_Z \neq 0$ and we put $e_1 = V_X/V_Z$ and $e_2 = V_Y/V_Z$. Let us consider two points. We have

$$e_1(w_i - w_j) - (\Omega_Z + e_1\Omega_X)(y_i - y_j) + (e_1\Omega_Y - c)(x_i - x_j) = u_i - u_j \tag{26}$$

$$e_2(w_i - w_j) - (e_2\Omega_X + c)(y_i - y_j) + (\Omega_Z + e_2\Omega_Y)(x_i - x_j) = v_i - v_j \tag{27}$$

Knowing the "perspective" velocity vector at four non-aligned points we obtain two systems of three linear equations (the first from equation (26) and the second from equation (27)). Solving the two linear systems we obtain:

$$e_1, \quad \Omega_Z + e_1\Omega_X = e_3, \quad e_1\Omega_Y - c = e_4 \tag{28}$$

$$e_2, \quad \Omega_Z + e_2\Omega_Y = e_5, \quad e_2\Omega_X + c = e_6 \tag{29}$$

From the two last equations of (28) and (29) we obtain a system of four linear equations

$$\begin{bmatrix} e_1 & 0 & 1 & 0 \\ 0 & e_1 & 0 & -1 \\ 0 & e_2 & 1 & 0 \\ e_2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Omega_X \\ \Omega_Y \\ \Omega_Z \\ c \end{bmatrix} = \begin{bmatrix} e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

If $e_1^2 + e_2^2 \neq 0$, the solution is:

$$\Omega_X = [e_1(e_3 - e_5) + e_2(e_4 + e_6)]/(e_1^2 + e_2^2)$$

$$\Omega_Y = [e_1(e_4 + e_6) - e_2(e_3 - e_5)]/(e_1^2 + e_2^2)$$

$$\Omega_Z = [e_1^2 e_5 + e_2^2 e_3 - e_1 e_2(e_4 + e_6)]/(e_1^2 + e_2^2)$$

$$c = [e_1^2 e_6 - e_2^2 e_4 - e_1 e_2(e_3 - e_5)]/(e_1^2 + e_2^2)$$

From equation (25) we then obtain Z_i/V_Z. If $e_1^2+e_2^2=0$, equation (28) gives

$$\Omega_Z = e_3 = e_5 \quad \text{and} \quad c = e_6 = -e_4$$

Ω_X is obtained from: $v_i + \Omega_X + y_i c - x_i \Omega_Z = 0$, and Ω_Y from: $u_i - \Omega_Y + y_i \Omega_Z + x_i c = 0$.

The solution has been obtained assuming that $V_Z \neq 0$. This is equivalent to having

$$\begin{vmatrix} w_1 - w_2 & y_1 - y_2 & x_1 - x_2 \\ w_2 - w_3 & y_2 - y_3 & x_2 - x_3 \\ w_3 - w_4 & y_3 - y_4 & x_3 - x_4 \end{vmatrix} \neq 0$$

The above derivation assumes that the velocity vectors are noiseless. In reality, these vectors are estimated and therefore noisy. The above equations have then to be solved in a least-squares sense; but we do not develop this point here.

We note that the translational and the rotational component of the 2-D velocity field can be separated. It is therefore possible to determine the focus of expansion, that is, the point $(V_X/V_Z, V_Y/V_Z)$ where the translational velocity component vanishes.

We can also determine an instantaneous time-of-collision, which is an interesting parameter in passive navigation

$$T_{\text{col}} = -\frac{Z}{Z'} = \frac{1}{x\Omega_X - y\Omega_Y - V_Z/Z}$$

7. Applications

We have applied the method presented here to a simulated motion of a quartic curve, whose equation on the image plane is given by $x^4 + y^4 = 1$.

The 3-D parameters were: $n_X = n_Y = 0$, $n_Z V_Z = 0.1$, $V_X = V_Y = 0$, $\Omega_X = \Omega_Z = 0$, $\Omega_Y = 0.1$. Four iterations along the contour were necessary for the exact estimation of the retinal velocity field, when the measurement noise is zero. In Fig. 5 we give the results obtained, that is, the true 2-D velocity field and the estimated one. These coincide.

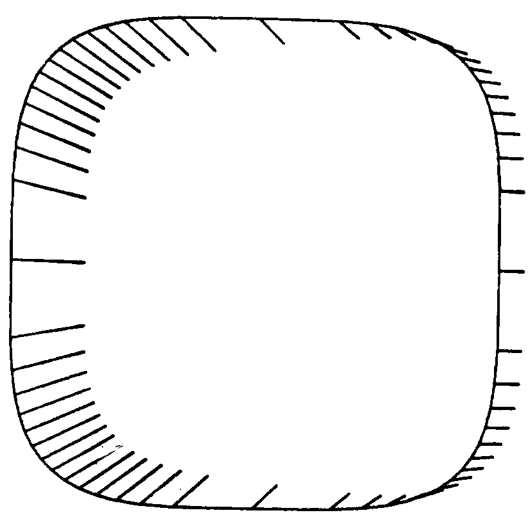

Fig. 5. The true and the estimated 2-D velocity field for a quartic curve lying on a plane.

We have also applied the algorithm of Section 6 in the case of a standardized TV sequence of images (COST 211bis) with a moving car, and a moving camera. There are 512×512 pixels of 8 bits in each frame. We give the results of the application of the algorithm for the two first frames.

The horizontal spatial gradient is obtained using a finite impulse response filter given below

$$F_x = \frac{1}{80} \begin{bmatrix} -3 & -5 & 0 & 5 & 3 \\ -5 & -8 & 0 & 8 & 5 \\ -3 & -5 & 0 & 5 & 3 \end{bmatrix}$$

The transposed matrix gives the vertical gradient.

The average of the absolute difference between the first two images is equal to 18.472. The average absolute prediction error is equal to 11.358 and the average absolute estimation error after two iterations is 6.926. Figure 6 gives the difference between the two frames, Figure 7 gives the prediction error e^0 (§ 5.2) and Fig. 8 the displacement frame difference after two iterations of the a posteriori estimation (equation (24)).

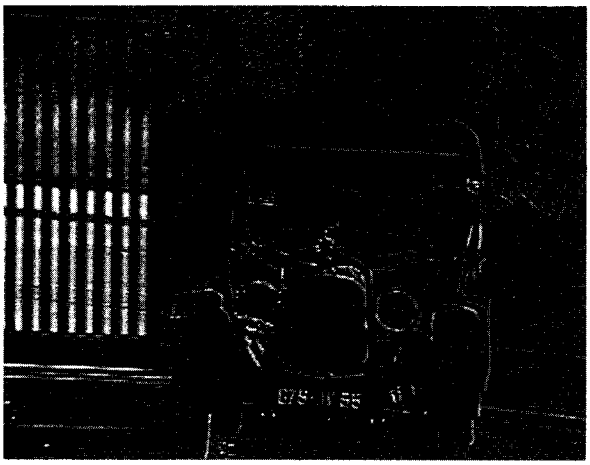

Fig. 6. Frame difference.

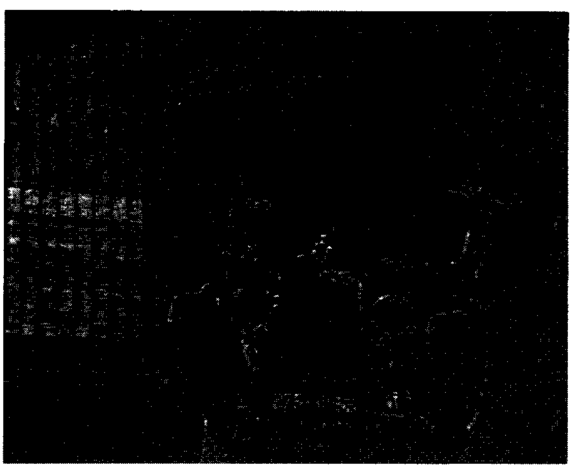

Fig. 7. Displaced frame difference with the a priori estimation.

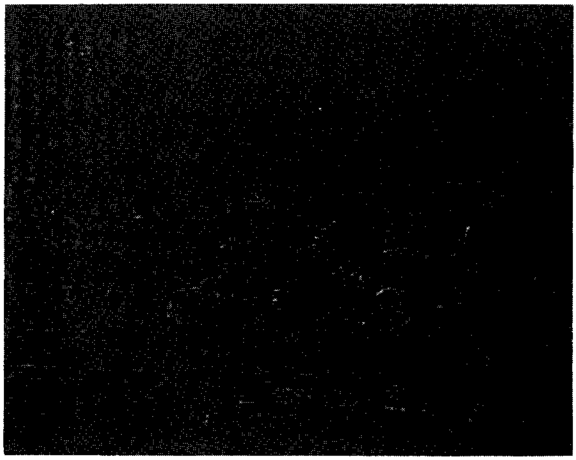

Fig. 8. Displaced frame difference after two iterations.

8. Conclusion

Using the assumption of local 3-D rigidity and local planarity we have given spatial relations on the "perspective" velocity field, using the fact that the "perspective" velocity vector is linear-dependent on the 2-D coordinates (equation (6)). For discrete points an autoregressive relation of order 3 for each velocity component is given (equation (10)). In the whole image plane three spatial relations are given (12) for each velocity component.

The spatial relations on the 2-D velocity field are used to estimate it, using for a contour the perpendicular component of the velocity and for the whole image the motion constraint equation. In the case of a contour we propose a recursive method using a Kalman filter as estimator. In the case of the whole image plane we propose a classical relaxation method to solve the system of linear equations. A recursive/iterative estimator operating on the whole image plane is also proposed, which may be useful for predictive coding using motion compensation. The iterative part of this estimator improves considerably the displacement frame difference, as indicated by the numerical results obtained. We think that an adaptive identification of the recursive part would further improve these results.

We give the differential plane equation of curves presenting ambiguities on the determination of the 2-D velocity field from the single normal component of the field, and illustrate it with several examples (Section 3).

A system of linear equations is used to determine the motion parameters and the relative depths. Four points and their velocities are considered. A least-squares approach may be used to reduce the effect of the retinal velocity field estimation noise on the estimated 2-D velocity field. We also show that the translational and rotational components of the velocity can be separated, and therefore the focus of expansion may be determined. We also give the instantaneous time-of-collision.

Finally we would like comment on certain aspects which merit more attention. The stage of initial local measurements from image intensities is very important. We have seen that a good estimation of the retinal velocity field can be obtained from a single component and that we can reconstruct the 3-D motion and structure exactly from the retinal velocity field. The stage which eventually introduces important errors is the initial measurement. We think that great efforts should be made to improve this measurement stage, or that eventually a method matching contours globally should be determined.

References

[1] P. Anandan and R. Weiss, "Introducing a smoothness constraint in a matching approach for the computation of displacement fields", *Proc. 3rd IEEE Workshop on Computer Vision: Representation and Control*, Oct. 1985, pp. 186-194.
[2] J. Biemond, L. Looijenga and D.E. Boekee, "A pel-recursive Wiener-based displacement estimation algorithm", *Signal Processing*, Vol. 13(4), Dec. 1987, pp. 399-412.
[3] A. Bruss and B. Horn, "Passive navigation", *Comput. Vision, Graphics, Image Process.*, Vol. 21, 1983, pp. 3-20.
[4] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Anal. Machine Intelligence*, Vol. PAMI-8 (6), Nov. 1986, pp. 679-698.
[5] R. Deriche, "Using Canny's criteria to derive an optimal edge detector recursively implemented", *Int. J. Comput. Vision*, Vol. 1 (2), 1987.
[6] T. Henderson and O. Faugeras, *Relaxation Techniques in Computer Vision*, Oxford University Press, U.K., 1986.
[7] E. Hildreth, *The Measurement of Visual Motion*, MIT Press, Cambridge, MA, 1983.
[8] B. Horn and B. Schunck, "Determining optical flow", *Artificial Intelligence*, Vol. 7, 1981, pp. 185-203.
[9] A. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, NY, 1970.
[10] H. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image", *Proc. R. Soc., London*, Vol. B208, 1980, pp. 385-397.
[11] P. Maybeck, *Stochastic Models, Estimation, and Control*, Vol. 1, Academic Press, New York, NY, 1979.
[12] R. Moorhead, S. Rajala, "Motion-compensated interframe coding", *Proc. ICASSP*, 1985, pp. 347-350.
[13] D. Murray and B. Buxton, "Reconstructing the optic flow field from edge motion: an examination of two different approaches", *1st Conf. on AI Applications*, Denver, CO, 1984.
[14] H.G. Musmann, P. Pirsch, and H.-J. Grallert, "Advances in picture coding", *Proc. IEEE*, Vol. 73 (4), Apr. 1985, pp. 523-548.
[15] H.-H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences", *IEEE Trans. Pattern Anal. Machine Intelligence*, Vol. PAMI-8 (5), Sept. 1986, pp. 679-698.
[16] I. Petrovski, *Ordinary Differential Equations*, Prentice-Hall, London, 1966.
[17] T. Poggio, "Early vision: from computational structure to algorithms and parallel hardware", *Comput. Vision, Graphics, Image Process.*, Vol. 31 (2), Aug. 1985, pp. 139-155.
[18] B. Schunck, "The image flow constraint equation", *Comput. Vision, Graphics, Image Process.*, Vol. 35, 1986, pp. 20-46.
[19] G.L. Scott, "Smoothing the optic flow field under perspective projection", *IEEE Conf. on Computer Vision and Pattern Recognition*, Miami Beach, FL, June 22-26, 1986, pp. 504-509.
[20] M. Subbarao and A. Waxman, "On the uniqueness of image flow solutions for planar surfaces in motion", *Proc. 3rd IEEE Workshop on Computer Vision: Representation and Control*, Oct. 1985, pp. 129-140.
[21] D. Terzopoulos, "Multiresolution computation of visible-surface reconstruction", Ph.D. Thesis, MIT, Cambridge, MA, Jan. 1984.
[22] A. Tikhonov and V. Arsenine, *Méthodes de résolution de problèmes mal posés*, Editions MIR, 1974.
[23] G. Tziritas, "Estimation du flux optique basée sur le gradient" *Traitement du Signal*, Vol. 3 (1), 1986, pp. 3-11.
[24] D. Walker, K. Rao, "Improved pel-recursive motion compensation", *IEEE Trans. Commun.*, Vol. COM-32 (10), 1984, pp. 1128-1134.
[25] A. Waxman and S. Ullman, "Surface structure and 3-D motion from image flow: a kinematic analysis", *Robotics Res.*, Vol. 4 (3), 1985, pp. 72-94.
[26] A. Waxman and K. Wohn, "Contour evolution, neighborhood deformation and global image flow: planar surfaces in motion", *Robotics Res.*, Vol. 4 (3), 1985, pp. 95-108.
[27] G. Zacharias, A. Caglayan and J. Sinacori, "A model for visual flow-field cueing and self-motion estimation", *IEEE Trans. Syst. Man Cybern.*, Vol. SMC-15 (3), May/June 1985, pp. 385-389.