Motion Field and Optical Flow: Qualitative Properties

ALESSANDRO VERRI AND TOMASO POGGIO, MEMBER, IEEE

***Abstract*—This paper shows that the *motion field*, the 2-D vector field which is the perspective projection on the image plane of the 3-D velocity field of a moving scene, and the *optical flow*, defined as the estimate of the motion field which can be derived from first order variation of the image brightness pattern, are in general different, unless special conditions are satisfied. Therefore, dense optical flow is often ill-suited for computing *structure from motion* and for reconstructing the 3-D velocity field by means of algorithms which require a locally accurate estimate of the motion field. A different use of the optical flow is suggested. It is shown that the (smoothed) optical flow and the motion field can be interpreted as vector fields tangent to flows of planar dynamical systems. Stable qualitative properties of the motion field, which give useful information about the 3-D velocity field and the 3-D structure of the scene, can be usually obtained from the optical flow. The idea is supported by results from the theory of structural stability of dynamical systems.**

***Index Terms*—Motion computation, optical flow.**

I. Introduction

A KEY task for many vision systems is to extract information from a sequence of images. This information can be useful for solving important problems such as recovering the 3-D velocity field, segmenting the image into parts corresponding to different moving objects, or reconstructing the 3-D structure of surfaces in the viewed scene. The recovery of the *motion field*, which is the perspective projection onto the image plane of the true 3-D velocity field of moving surfaces in space, is thought to be an essential step in the solution of these problems. The data available, however, are only the spatial and temporal variations in the image brightness pattern E. From these variations it is possible to derive an estimate of the motion field, called *optical flow* [1]–[3]. The assumption that the motion field and the optical flow coincide has often been made, the intuitive rationale being that this is true when spatial variations in E correspond to physical features on the visible 3-D surfaces [4]–[6]. Horn [7], however, has pointed out examples in which this assumption does not hold. Algorithms which deal with the recovery of the motion field from dense optical flow data have been proposed, with the more or less implicit assumption that the two fields are the same [8]–[10].

In this paper we show that the optical flow and the motion field are in general different, unless special conditions are satisfied. In particular, even the hypothesis of a Lambertian reflectance function of the viewed surfaces is not sufficient by itself to guarantee that the two vector fields are the same. A rigorous derivation of this result is provided. Indeed, where sharp changes in intensity over time are due to physical events on the moving surface (e.g., texture and surface markings), the estimates of the component of the motion field along the direction of the spatial gradient of the image brightness pattern—estimates which can be obtained by means of first order derivative of the image brightness pattern—are accurate. These estimates, therefore, are unlikely to be useful for methods which rely upon a very precise, local reconstruction of the motion field. One may then ask, what is the optical flow for? In the final part of the paper it is suggested that meaningful information about the 3-D velocity field and the 3-D structure of the viewed scene can be obtained from qualitative properties of the motion field. At any fixed time, the motion field can be seen as the flow associated with some dynamical system and useful motion motion information can be retrieved from its qualitative properties, e.g., from its singular points. A thorough analysis of this approach—proposed first by [11]—is presented in [12]. Then, results from the theory of structural stability of dynamical systems suggest that, if the motion field and the optical flow are sufficiently similar, they also have the same qualitative properties. Therefore, the qualitative properties of the optical flow might be very useful in recovering motion information (see [13], for example).

The paper is organized as follows. Section II defines the problem and considers in detail how image irradiance can be related to scene radiance in the case of a scene consisting on non-Lambertian surfaces. Section III describes the method used to show that the optical flow and the motion field are almost always different. We consider the Lambertian model of reflectance and a more realistic model for arbitrary rigid motion of a generic smooth surface. Section IV shows that at any given time the motion field and the optical flow can be processed to become smooth vector fields tangent to flows of dynamical systems. Results from the theory of structural stability of dy-

Manuscript received June 1, 1987; revised July 20, 1988. Recommended for acceptance by W. B. Thompson. This paper describes research done within the Center for Biological Information Processing, in the Department of Brain and Cognitive Sciences, and at the Artificial Intelligence Laboratory. This work was supported by a grant from the Office of Naval Research (ONR), Cognitive and Neural Sciences Division; by the Artificial Intelligence Center of Hughes Aircraft Corporation; by the Alfred P. Sloan Foundation; by the National Science Foundation; by the Artificial Intelligence Center of Hughes Aircraft Corporation (S1-801534-2); and by the NATO Scientific Affairs Division (0403/87). Support for the A. I. Laboratory's artificial intelligence research is provided by the Advanced Research Projects Agency of the Department of Defense under Army contract DACA76-85-C-0010, and in part by ONR contract N00014-85-K-0124. T. Poggio was supported by the Uncas and Helen Whitaker Chair at the Massachusetts Institute of Technology, Whitaker College. A. Verri is supported by a Fairchild Fellowship.

A. Verri was with the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139. He is now with the Department of Physics, University of Genova, Genova, Italy.

T. Poggio is with the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139.

IEEE Log Number 8926695.

0162-8828/89/0500-0490$01.00

namical systems, then, are used to suggest that qualitative, stable properties of the motion field hold for smooth optical flows, provided the two fields are sufficiently similar. Finally, in Section V, possible biological connections are discussed. Details on the geometry of perspective projection and a brief review of the theory of planar dynamical systems are summarized in the Appendixes.

II. Preliminaries

In this section notations, definitions, and assumptions which have been used throughout the paper are stated. The motion field and the optical flow are defined and image irradiance is related to scene radiance in the case of a scene consisting of non-Lambertian surfaces.

A. Definitions

Let us define notations and summarize concepts which will be useful in what follows. For more details on the geometry of perspective projection see Appendix A. Let

$$\vec{x}_p = \frac{f}{f + \vec{x} \cdot \vec{n}} [\vec{x} - (\vec{x} \cdot \vec{n})\, \vec{n}] \qquad (2.1.1)$$

be the equation which defines the perspective projection of a generic point on the image plane, where $\vec{x}_p = (x_p, y_p, 0)$ is the position vector of the projected point, $\vec{x}$ is the position vector of the point, $\vec{n} = (0, 0, 1)$ is the unit vector normal to the image plane (projection plane), and f is the focal length in a suitable systems of coordinates (see Fig. 1). Notice that the origin O is on the image plane, the focus of projection F is located at $(0, 0, -f)$, and $f\vec{n} + \vec{x}$ is the vector pointing from F to the point. The equation for $\vec{x}_p$, in the case of orthographic projection, can be derived simply by taking the limit of the right-hand side of (2.1.1) for $f \to \infty$, which yields

$$\vec{x}_p = \vec{x} - (\vec{x} \cdot \vec{n})\, \vec{n}.$$

The motion field $\vec{v}_p$ can be obtained differentiating (2.1.1) with respect to time. If $\vec{v} = d\vec{x}/dt$, then[1]

$$\vec{v}_p = \frac{f}{f + \vec{x} \cdot \vec{n}} \left\{ \vec{v} - (\vec{v} \cdot \vec{n})\, \vec{n} - \frac{\vec{v} \cdot \vec{n}}{f + \vec{x} \cdot \vec{n}} [\vec{x} - (\vec{x} \cdot \vec{n})\, \vec{n}] \right\}. \qquad (2.1.2)$$

Notice that in (2.1.2) $\vec{v}_p$ is given in terms of $\vec{x}$ and $\vec{v}$, position and velocity of the moving points in the scene respectively, which are not known. In what follows, $\vec{x}_p$ and $\vec{v}_p$ will be considered as 2-D vectors defined on the image plane, since their third component vanishes identically.

Let $E = E(x_p, y_p, t)$ be the image brightness pattern that is, the intensity of light at the point (x_p, y_p) of the image plane at time t. If $\vec{\nabla}_p$ is the gradient with respect to the image plane coordinates, then

$$\frac{dE}{dt} = \frac{\partial E}{\partial t} + \vec{\nabla}_p E \cdot \vec{v}_p \qquad (2.1.3)$$

[1]It can be easily shown that the perspective projection of the 3-D velocity error $\vec{v}$ is equal to the velocity $\vec{v}_p$ of the projected point on the image plane, since both vectors are defined in terms of infinitesimal.

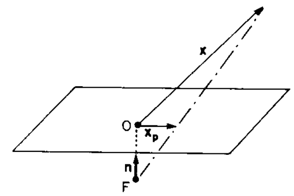

Fig. 1. The geometry of perspective projection.

where $\partial E/\partial t$ is the partial derivative with respect to the time at the point (x_p, y_p) and dE/dt, also called *total temporal derivative*, can be thought of as the temporal derivative along the trajectory of the point (x_p, y_p) onto the image plane. Equation (2.1.3) can be rewritten as

$$\frac{dE}{dt} = \frac{\partial E}{\partial t} + \|\vec{\nabla}_p E\|\, v_\perp$$

where $v_\perp$ is the norm of $\vec{v}_\perp$, component of the motion field $\vec{v}$ along the direction of $\vec{\nabla} E$. Now if

$$\frac{dE}{dt} = 0 \qquad (2.1.4)$$

and $\|\vec{\nabla}_p E\| \neq 0$, then

$$\vec{v}_\perp = -\frac{\partial E/\partial t}{\|\vec{\nabla}_p E\|} \frac{\vec{\nabla}_p E}{\|\vec{\nabla}_p E\|}. \qquad (2.1.5)$$

Therefore, if (2.1.4) holds, the component of the motion field along the direction of the gradient of the image brightness $\vec{v}_\perp$ can be written in terms of derivatives of E (which can be computed). Equation (2.1.5) can be interpreted as an instance of the well-known *aperture* problem [2], [3], [14], [15] for the unknown $\vec{v}_p$: that is, the information available at each point of a sequence of frames is only the component of the motion field along the direction of $\vec{\nabla}_p E$. In order to compute the full 2-D optical flow, some other constraints are needed: Horn and Schunck [3] for example, look for the smoothest 2-D vector field whose component along $\vec{\nabla}_p E$ coincides with the right-hand side of (2.1.5). Examples for which (2.1.5) is not true are well-known [3]. Consider, for instance, a rotating sphere with no texture on it (i.e., with uniform albedo) under arbitrary, fixed illumination. Since the image brightness at each image location does not change with time, the left-hand side of (2.1.5) is identically equal to zero, while the right-hand side is different from zero almost everywhere. Notice that keeping the sphere fixed and moving the light source, (2.1.5) is again wrong. In this case, however, the right-hand side is different from zero while $\vec{v}_\perp$ is zero everywhere. Since (2.1.4) is often assumed as a starting point for computing the optical flow, it is interesting to calculate explicitly the total temporal derivative of E with respect to time. The calculations require both an accurate definition of image formation properties and an analytical model of the reflectance function.

B. Scene Radiance and Image Irradiance

Let us review some definitions of photometry and make explicit the constraints under which the image irradiance is related to the scene radiance. The image irradiance can be thought of as the image brightness pattern $E = E(x_p, y_p)$, since it is the power of light per unit area at each point (x_p, y_p) of the image plane. The scene radiance L is the power of light per unit area which can be thought of as emitted by each point of a surface S in the scene in a particular direction. This surface can be fictitious, or may be the actual radiating surface of a light surface of a light source, or the illuminated surface of a solid. The scene radiance can be thought of as a function of the point of the surface and of the direction in space. If (a, b) is a point on the surface S in intrinsic coordinates of the surface and (α, β) polar coordinates determining a direction in space with respect to the normal vector to the surface, then $L = L(a, b, \alpha, \beta)$ gives the scene radiance at the point (a, b) in the direction (α, β). Given the scene radiance, in principle, it is possible to compute the expected image irradiance. For example in the case of pinhole camera approximation, i.e., assuming that the camera has an infinitesimally small aperture, the image irradiance at a point (x_p, y_p) is proportional to the scene radiance at the point (a, b) on the surface in the direction of the pinhole, say (α_0, β_0), where the projected point, the original point, and the pinhole lie on the same line (see Fig. 2). Therefore,

$$E(x_p(a, b), y_p(a, b)) = L(a, b, \alpha_0, \beta_0) \quad (2.2.1)$$

if $(x_p(a, b), y_p(a, b))$ is the image point which lies on the line through (a, b) and the pinhole. In practice, however, the aperture of any real optical device is finite and not very small and (2.2.1) does not necessarily hold. Assuming that the surface is Lambertian i.e., $L(a, b, \alpha, \beta) = L(a, b)$, that there are not losses within the system and that the angular aperture (on the image side) is small, it can be proved [16] that

$$E(x_p(a, b), y_p(a, b)) = L(a, b)\, \Omega \cos^4 \varphi$$

where Ω is the solid angle corresponding to the angular aperture and φ is the angle between the principal ray (that is, the ray through the center of the aperture) and the optical axis. With the further assumption that the aperture is much smaller than the distance of the viewed surface, the Lambertian hypothesis can be relaxed to give [17]

$$E(x_p(a, b), y_p(a, b)) = L(a, b, \alpha_0, \beta_0)\, \Omega \cos^4 \varphi \quad (2.2.2)$$

where α_0 and β_0 are the polar coordinates of the direction of the principal ray. It must be pointed out that (2.2.2) holds if L is continuous with respect to α and β. In what follows, it will be assumed that the optical system has been calibrated so that (2.2.2) can be rewritten as (2.2.1). Finally, notice that

$$\vec{\nabla}_p E \cdot \vec{v}_p = \vec{\nabla}_S L \cdot \left(\frac{da}{dt}, \frac{db}{dt}\right) \quad (2.2.3)$$

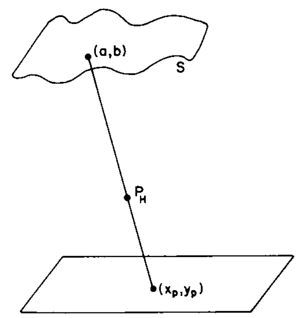

Fig. 2. Scene radiance and image irradiance in the pinhole approximation. The image irradiance at the point (x_p, y_p) is given by the scene radiance at the point (a, b) on the surface in the direction of the line connecting the two points through the pinhole P_H.

where $\vec{\nabla}_S$ is the gradient with respect to the surface coordinates, since differentiating (2.2.1) yields

$$\vec{\nabla}_p E \cdot (dx_p, dy_p) = \vec{\nabla}_S L \cdot (da, db).$$

III. Computing Derivatives of the Image Brightness in Terms of Scene Radiance

The general method which is used to show that optical flow data almost always give inaccurate estimates of the component of the motion field along the gradient of the image brightness is presented. The Lambertian model of reflectance and a more realistic model are assumed for pure rotation, pure rotation, and general rigid motion of a generic surface. The motion field and the optical flow are exactly the same only for Lambertian objects which translate under uniform, fixed illumination.

A. The Method

Consider a rigid surface S moving in space. From (2.2.1), the image irradiance E at time t at the point (x_p, y_p) is equal to the scene radiance L at the point (a, b) on S, i.e., $E(x_p, y_p, t) = L(a, b)$. The image irradiance at time $t + \Delta t$ is given by the scene radiance from S at time $t + \Delta t$. As shown in Fig. 3, the point on S which radiates toward (x_p, y_p) at time $t + \Delta t$ is the point $(a - \Delta a, b - \Delta b)$.[2] The unit normal vector $\vec{N}$ to S at the point $(a - \Delta a, b - \Delta b)$ at time $t + \Delta t$ is

$$\vec{N}_{t+\Delta t}(a - \Delta a, b - \Delta b) = \vec{N}_t(a - \Delta a, b - \Delta b) + \Delta\vec{N}$$

where $\Delta\vec{N}$ is the first order variation of $\vec{N}$ due to the motion of S during the time interval Δt. Now in the case of translation

$$\Delta\vec{N} = 0$$

while in the case of rotation with angular velocity $\vec{\omega}$

$$\Delta\vec{N} = \vec{\omega} \times \vec{N}\Delta t. \quad (3.1.1)$$

[2]The surface is assumed to correspond to a moving convex body to avoid self-occlusions. In fact, the computations which follow hold for any convex surface patch.

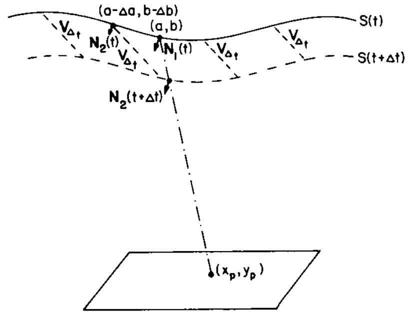

Fig. 3. Computing $\partial E/\partial t$. The point (a, b) on S radiates toward (x_p, y_p) at time t. The point $(a - \Delta a, b - \Delta b)$ radiates toward the same point at time $t + \Delta t$. The vectors $\vec{N}_1$ and $\vec{N}_2$ are the unit normal vectors to the surface S at the point (a, b) and $(a - \Delta a, b - \Delta b)$, respectively.

Notice that (3.1.1) can be considered as the expression of $\Delta\vec{N}$ for any kind of motion. Similarly, for each argument A of the scene radiance, we can write

$$A_{t+\Delta t}(a - \Delta a, b - \Delta b) = A_t(a - \Delta a, b - \Delta b) + \Delta A. \quad (3.1.2)$$

To compute ΔA, let us distinguish between arguments of L which are intrinsic function of the surface coordinates, such as texture and albedo, and those which can be thought of as function of the surface coordinates, but, in fact, are function of 3-D space coordinates, such as the illumination and the point of view. If A is an intrinsic function of the surface coordinates, it follows easily that

$$\Delta A = 0$$

while if A is a function of 3-D space coordinates, from Taylor's expansion we have

$$\Delta A = \vec{\nabla} A \cdot \vec{v}\Delta t \quad (3.1.3)$$

where $\vec{\nabla}$ is the gradient operator with respect to the 3-D space coordinates. Let us assume that L can be written as a function of m arguments A^i, $i = 1, \cdots, m$, and of $\vec{N}$. Then, taking into account (3.1.1) and (3.1.2), (2.2.1) becomes

$$E(x_p, y_p, t + \Delta t) = L\big(A_t^i(a - \Delta a, b - \Delta b) + \Delta A^i, \vec{N}_t(a - \Delta a, b - \Delta b) + \Delta\vec{N}\big) \quad (3.1.4)$$

at time $t + \Delta t$ and

$$E(x_p, y_p, t) = L\big(A_t^i(a, b), \vec{N}_t(a, b)\big) \quad (3.1.5)$$

at time t. Therefore, using (3.1.4) and (3.1.5) and dropping the subscript t we have

$$\frac{\partial E}{\partial t} = -\vec{\nabla}_S L \cdot \left(\frac{da}{dt}, \frac{db}{dt}\right) + \sum_{i=1}^{p} \frac{\partial L}{\partial A^i} \vec{\nabla} A^i \cdot \vec{v} + \frac{\partial L}{\partial \vec{N}} \cdot \vec{\omega} \times \vec{N} \quad (3.1.6)$$

if p, $p \le m$, of the A^i, $i = 1, \cdots, m$, require the use of (3.1.3) to compute ΔA, $\partial L/\partial\vec{N} = (\partial L/\partial N_x, \partial L/\partial N_y, \partial L/\partial N_z)$, and $\vec{N} = (N_x, N_y, N_z)$.

From (2.2.3) and the definition of $\vec{v}_\perp$, (3.1.6) can be rewritten as

$$\frac{\partial E}{\partial t} = -\|\vec{\nabla}_p E\| \, v_\perp + \sum_{i=1}^{p} \frac{\partial L}{\partial A^i} \vec{\nabla} A^i \cdot \vec{v} + \frac{\partial L}{\partial \vec{N}} \cdot \vec{\omega} \times \vec{N}. \quad (3.1.7)$$

Finally, if $\Delta v_\perp$ is the norm of the difference between $\vec{v}_\perp$—the true component of the motion field $\vec{v}$ along $\vec{\nabla}_p E$—and $\|\vec{\nabla}_p E\|^{-1} \partial E/\partial t$—the estimate of $\|\vec{v}_\perp\|$ which can be obtained from (2.1.5)—from (3.1.7) we have

$$\Delta v_\perp = \frac{1}{\|\vec{\nabla}_p E\|} \left| \sum_{i=1}^{p} \frac{\partial L}{\partial A^i} \vec{\nabla} A^i \cdot \vec{v} + \frac{\partial L}{\partial \vec{N}} \cdot \vec{\omega} \times \vec{N} \right|. \quad (3.1.8)$$

Thus, $\Delta v_\perp$ vanishes, or the motion field and the optical flow are the same, if the reflectance function does not depend upon 3-D space coordinates and the surface undergoes pure translation.

Let us consider now some interesting examples in detail.

B. Translation of a Lambertian Surface

Consider a Lambertian surface S. In the hypothesis of uniform illumination, the scene radiance due to S is

$$L = \rho \vec{I} \cdot \vec{N} \quad (3.2.1)$$

where ρ is the albedo of the surface, $\vec{I}$ the unit vector which gives the direction of illumination, and $\vec{N}$ the unit normal vector to S. If the surface is translating, substitution of (3.2.1) in (3.1.8) yields

$$\Delta v_\perp = 0 \quad (3.2.2)$$

since $\vec{\omega} = 0$ and none of the arguments in L depends upon 3-D space coordinates. Therefore, in the case of a Lambertian surface which is translating under uniform illumination it is possible to estimate correctly the component of the motion field in the direction of $\vec{\nabla}_p E$ from optical flow data.

In the case of nonuniform illumination the right-hand side of (3.2.2) contains an extra term due to ΔI. A rigorous analysis of the relevance of this term in (3.2.2) would require a realistic model of illumination. In practice, if the object is not moving very slowly, (3.2.2) is expected to be satisfied almost everywhere a part from locations where the gradient of illumination cannot be neglected (e.g., shadow boundaries).

Let us consider now the case of a rotating Lambertian surface.

C. Rotation of a Lambertian Surface

Let S be a Lambertian surface rotating in space with angular velocity $\vec{v}$ around an arbitrary positioned axis. Applying the same argument of the previous section but taking into account the constraint (3.1.1) for $\vec{\Delta} N$, (3.1.8) yields

$$\Delta v_\perp = \frac{|\rho \vec{N} \cdot \vec{I} \times \vec{\omega}|}{\|\vec{\nabla}_p E\|}. \tag{3.3.1}$$

In the case of rotation, therefore, even under uniform illumination, the estimate of the motion field by means of optical flow data is corrupted by a systematic error. Notice that the error vanishes if the surface is rotating around an axis parallel to the direction of illumination. In the case of nonuniform illumination again a corresponding extra term must be added to the right-hand side of (3.3.1).

For a sphere rotating around an axis through its center, due to rotational symmetry we have

$$\vec{N}(a - \Delta a, b - \Delta b) + \vec{\omega} \times \vec{N}(a - \Delta a, b - \Delta b)\, \Delta t = \vec{N}(a, b)$$

for every point (a, b) on the sphere. Therefore, if ρ is uniform, since the displacement in space equals the displacement on the sphere, (3.1.6) gives

$$\frac{\partial E}{\partial t} = 0$$

which means that

$$\Delta v_\perp = \|\vec{v}_\perp\|$$

or, in other words, that all the motion information in the image brightness pattern is lost.

D. Translation of a Specular Surface

Let us consider, now, a more realistic reflectance model. The scene radiance can be thought of as a suitable linear combination of a Lambertian and a specular term [18], i.e.,

$$L = L_{\text{lamb}} + L_{\text{spec}} \tag{3.4.1}$$

where the Lambertian contribution is the same of (3.2.1) and the specular term is

$$L = s\left(\frac{\vec{D} \cdot \vec{R}}{D}\right)^n \tag{3.4.2}$$

where s is the fraction of light reflected by the surface, $\vec{D} = f\vec{n} + \vec{x}$ is the vector pointing from the focus of projection to the radiating point and

$$\vec{R} = \vec{I} - 2(\vec{I} \cdot \vec{N})\vec{N} \tag{3.4.3}$$

is the unit vector which gives the direction of perfect specular reflection. Assuming that s is not a function of the direction of the incident light and that is constant on the surface, the specular term is proportional to the nth power of the cosine of the angle between the direction of specular reflection and the line of sight. It is clear that the contribution to $\Delta v_\perp$ due to the Lambertian term has already been computed. In the case of translation the contribution of the specular term can be obtained substituting (3.4.2) in (3.1.8) which gives

$$\Delta v_\perp = \frac{1}{\|\vec{\nabla}_p E\|}\left|\frac{sn}{D^3}\left(\frac{\vec{D} \cdot \vec{R}}{D}\right)^{n-1} (\vec{v} \times \vec{D}) \cdot (\vec{R} \times \vec{D})\right| \tag{3.4.4}$$

since

$$\frac{d\vec{D}}{dt} = \vec{v}$$

where $\vec{v}$ is the velocity of the translating surface. Thus, if the reflectance function has a specular component, the motion field and the optical flow are different even in the case of translation. In the case of orthographic projection, i.e., for $f \to \infty$, the right-hand side of (3.4.4) vanishes since $\vec{D} \to \infty$. Therefore, in the approximation of orthographic projection, the estimate of $\vec{v}_\perp$ is correct for any translating surface whose reflectance function satisfies (3.4.1).

E. Rotation of a Specular Surface

Let us compute the contribution of the specular term of (3.4.1) to $\Delta v_\perp$ in the case of a rotating surface. Substituting (3.4.2) in (3.1.8) and taking into account the constraint (3.1.1), we have

$$\Delta v_\perp = \frac{1}{\|\vec{\nabla}_p E\|}\left|\frac{sn}{D^3}\left(\frac{\vec{D} \cdot \vec{R}}{D}\right)^{n-1} \Big\{(\vec{v} \times \vec{D}) \cdot (\vec{R} \times \vec{D}) - 2D^2\,[(\vec{D} \cdot \vec{N})(\vec{I} \cdot \vec{\omega} \times \vec{N}) + (\vec{I} \cdot \vec{N})(\vec{D} \cdot \vec{\omega} \times \vec{N})]\Big\}\right| \tag{3.5.1}$$

which, in general, is different from zero. Since for $f \to \infty$ (3.5.1) yields

$$\Delta v_\perp = \frac{1}{\|\vec{\nabla}_p E\|}\left|2sn(\vec{R} \cdot \vec{n})^{n-1}\,[(\vec{n} \cdot \vec{N})(\vec{I} \cdot \vec{\omega} \times \vec{N}) + (\vec{I} \cdot \vec{N})(\vec{n} \cdot \vec{\omega} \times \vec{N})]\right| \tag{3.5.2}$$

the estimate of $\vec{v}_\perp$, in this case, is affected by error even in orthographic approximation, unless $\vec{I}$, $\vec{\omega}$, and $\vec{n}$ are parallel.

F. General Case

Let us assume now that a generic object whose reflectance function is described by (3.4.1) undergoes a given arbitrary rigid motion (composition of a translation and rotation in space). Adding together contributions (3.3.1), (3.4.4), and (3.5.1), for the difference $\Delta v_\perp$ between the expected and the computed component of the motion field by means of optical flow data we obtain

$$\Delta v_\perp = \frac{1}{\|\vec{\nabla}_p E\|} \left| \rho \vec{N} \cdot \vec{I} \times \vec{\omega} + \frac{sn}{D^3} \left(\frac{\vec{D} \cdot \vec{R}}{D} \right)^{n-1} \cdot \left\{ (\vec{v} \times \vec{D}) \cdot (\vec{R} \times \vec{D}) - 2D^2 \left[(\vec{D} \cdot \vec{N})(\vec{I} \cdot \vec{\omega} \times \vec{N}) + (\vec{I} \cdot \vec{N})(\vec{D} \cdot \vec{\omega} \times \vec{N}) \right] \right\} \right| \tag{3.6.1}$$

where $\vec{v}$ is the velocity (with rotational and translational components) of the moving surface. The right-hand side of (3.6.1) is generally different from zero. However, as pointed out in Section III-A, it is different from zero whenever arguments of the reflectance function depend upon 3-D space coordinates. It is then clear [see (3.1.7) and (3.1.8)] that the $\Delta v_\perp / \| v_\perp \|$ is lower if the variation of the image brightness pattern over time at a given location (measured by $\partial E / \partial t$) is due to physical events on the moving surface (e.g., texture and surface markings); conversely, it is larger the more the change over time in intensity is due to lightness condition, abrupt changes in the reflectance properties of the moving surface at the corresponding location in space, or highlight boundaries of poorly textured surfaces. It must be noted that in this analysis shadows and self-shadow effects have not been considered. They also give rise to sharp changes in intensity which do not correspond to features in the scene. Furthermore, the Phong model of reflectance does not include explicitly sharp intensity changes due to highlights.

IV. Motion Fields, Optical Flows, and Dynamical Systems

In the previous section, we have shown that the estimates of the motion field which can be given in terms of first order derivatives of the image brightness pattern, are often inaccurate. Moreover, uncertainty of these estimates cannot be measured unless further information about the nature of viewed objects is known (or provided by some other vision module). This result seems to cast a shadow on the use of dense optical flow data for motion computations, such as recovering 3-D motion parameters and 3-D structure from motion. In what follows, however, we will argue that useful information can be extracted from dense optical flows. We introduced a theoretical framework in which the motion field can be compared with several plausible optical flows. Smooth planar vector fields can be seen as flows associated with some 2-D dynamical systems and the theory of dynamical systems can be used to confront them. An optical flow can be thought of as *close* to the true motion field, if the topological description of the two vector fields, in terms of the theory of dynamical systems, is the same at any fixed time.

Let us first show in what sense the motion field and optical flows can be associated with dynamical systems. As mentioned in Section I, a thorough analysis—outlined in [11]—is presented in [12]. Let a smooth convex surface undergo a given rigid motion in space. At any given time t, the motion field produced by the moving surface onto the image plane—provided it has the appropriate degree of smoothness that is, if it is continuous with first order partial (spatial) derivatives continuous—can be thought of as the vector field tangent to the solution to a planar system of ordinary autonomous differential equations (see [19] or Appendix B for mathematical details). Therefore, the qualitative theory of planar differential equations seems to be a natural tool for studying properties of the motion field. The analogy is between *phase portraits* of dynamical systems and motion flows. The motion field is considered at a fixed time: the physical meaning of the underlying dynamical system is irrelevant. Clearly, the same argument applies to other smooth planar vector fields, e.g., the optical flow. Since most of the relevant information about a dynamical system can be extracted from its *singular points*—that is, the points where the vector field vanishes (see Appendix B for a list of the main singular points of planar flows)—a natural criterion to study and confront motion field and optical flow seems to be that they have the same *number* and *kind* of singular points at about the same *position*, or, in other words, that they are qualitatively the same. From this perspective, quantitative difference between the motion field and optical flow might no longer be relevant. It is worth noticing that where singular points lie *close* to discontinuities of the motion field, a smoothing step is expected to change the qualitative properties of the field and of the corresponding optical flows (how *close* is determined by the size of the smoothing filter). Consider, for example, an arbitrary object which is translating *against* the viewer. The main feature of the motion field is an isolated singular point which is a *focus of expansion*. Note that the qualitative structure of the field neither depends on the shape of the object nor on its speed. If the focus of expansion lies *close* to the boundary of the moving surface, the qualitative properties of the *smoothed* motion field might change and a detailed, quantitative analysis might be required.

V. Discussion

If our point of view is correct, the only critical feature of the optical flow is that it must be topologically equivalent to the motion field. This requirement also satisfies two important uses of optical flow, namely to *detect discontinuities* and *help long-range matching* of the stereo type, which are needed for computing *structure-from-motion*. Quantitative equivalence, which is unlikely in general, is irrelevant for this use. As a consequence, *many different "optical flows" can be defined*. Different definitions could be chosen on the basis of criteria such as computability (from image data) or ease of implementation (for given hardware constraints). This point of view has clear implications for biological visual systems: movement detecting calls (say, in the retina) do not have to compute a specific optical flow, because simple estimates of the motion field which preserve its qualitative properties are equally good candidates (e.g., correlation-like algorithms). This argument may explain why the models proposed to explain motion dependent behavior in insects [20], motion perception in humans [21], and physiology of cells [22], [23], are all implementing quite different computations of optical flows. A basic question to answer is, of course, whether these biological models are in fact sufficiently *close* to the motion field to be topologically equivalent to it. Indeed, we conjecture that they are usually similar enough to preserve the qualitative properties of the motion field. The conjecture is based on results [24] showing that most of the biological models proposed so far can be considered as special instances or approximations of a general class of nonlinear models (characterized as Volterra systems of the second order).

Appendix A
Perspective and Orthographic Projections

In this section we explain in more detail the geometry of perspective projection used in the paper. In order to obtain orthographic projection as the simple limit of perspective projection for $f \to \infty$, where f is the focal length, the focus of projection cannot be located at the origin of the system of coordinates. To simplify the geometry without losing in generality, let the origin lie on the projection plane. The vector pointing from the focus to a point $\vec{x} = (x, y, z)$ is now $f\vec{n} + \vec{x}$. To obtain the expression of the projected point $\vec{x}_p$, from Fig. 1 notice that

$$\frac{f\vec{n} + \vec{x}}{(f\vec{n} + \vec{x}) \cdot \vec{n}} = \frac{f\vec{n} + \vec{x}_p}{f}. \tag{A.1}$$

From (A.1), we have

$$\vec{x}_p = f\frac{f\vec{n} + \vec{x}}{f + \vec{x} \cdot \vec{n}} - f\vec{n},$$

and finally

$$\vec{x}_p = \frac{f}{f + \vec{x} \cdot \vec{n}}(\vec{x} - (\vec{x} \cdot \vec{n})\,\vec{n})$$

or

$$\vec{x}_p = \frac{f}{f + \vec{x} \cdot \vec{n}}(\vec{n} \times (\vec{x} \times \vec{n})). \tag{A.2}$$

In the limit of orthographic projection (i.e., $f \to \infty$), $\vec{x}_p \to \vec{x}_0$, and (A.2) yields

$$\vec{x}_0 = \vec{n} \times (\vec{x} \times \vec{n}). \tag{A.3}$$

Combining (A.2) and (A.3), we obtain the general relationship between perspective ($\vec{x}_p$) and orthographic ($\vec{x}_0$) projection that is,

$$\vec{x}_p = \frac{f}{f + \vec{x} \cdot \vec{n}}\,\vec{x}_0.$$

Appendix B
Planar Dynamical Systems

A planar *dynamical system* is a C^1 map $\phi: R \times A \to A$, where A is an open subset of R^2 and writing $\phi(t, x) = \phi_t(x)$, the map $\phi_t: A \to A$ satisfies:

a) $\phi_0: A \to A$ is the identity;

b) the composition $\phi_t(\phi_s(x)) = \phi_{t+s}$ for each $t, s \in R$.

A planar dynamical system ϕ_t on A gives rise to a planar systems of differential equations on A that is, a vector field $y: A \to R^2$ defined as follows:

$$y(x) = \frac{d}{dt}\phi_t(x)\bigg|_{t=0}. \tag{B.1}$$

Thus, for every x, $y(x)$ is the tangent vector to the curve $t \to \phi_t(x)$ at $t = 0$. Equation (B.1) can be rewritten as

$$\frac{dx}{dt} = y(x). \tag{B.2}$$

If $y(x)$ is a C^1 vector field, (B.2) defines a planar dynamical system which can be thought of as a one-parameter family of transformation $\phi_t: A \to A$ describing the motion of the points in A as the time passes. The trajectories of the points are given by the solution curves to (B.2). Since (B.2) is *autonomous* (that is, the right-hand side does not depend explicitly on t), if $y(x^0) = 0$, then $x = x^0$ is a solution to it. Solutions like $x = x^0$ are called *equilibrium points* or *singular points*. In the case of linear systems, useful qualitative information about the behavior of the solution to (B.2) can be obtained from the eigenvalues of the matrix $\boldsymbol{M}$ of the coefficients of the differential equation computed at x_0. The restriction to planar systems reduces the classification of singular points to four fundamental cases:

1) $\boldsymbol{M}$ has real eigenvalues of opposite signs. In this case the singular point is a *saddle:* the saddle is *unstable* (a singular point is *stable* if any nearby solutions to it stays nearby for all the future time. It is *unstable* otherwise).

2) The eigenvalues have negative real parts. The singular point is a *sink* which is stable. The main property of a sink is that

$$\lim_{t \to \infty} x(t) = x_0$$

for any nearby solution $x(t)$. Qualitatively, the *phase portrait* of the solutions—that is, the family of the solution curves as a subset of R^2—looks like Fig. 4, where only some tangent vectors of some of the solution curves have been drawn.

Sinks can be classified depending on further characteristics of the eigenvalues. A *focus* (Fig. 4), for example, represents the case of coincident eigenvalues ($\boldsymbol{M}$ is a mul-

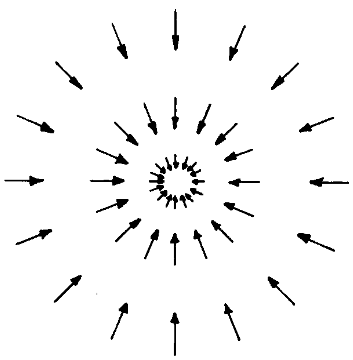

Fig. 4. Vector field tangent to a planar *sink*; all the solutions curves are pointing toward the singular point.

tiple of the identity matrix); a *node*, the case of different real eigenvalues; a *spiral*, the case of complex conjugate eigenvalues. A *sink-increasing* rotational component corresponds to each different case.

3) The eigenvalues have positive real parts. The singular point is a *source*. The main property of a source is that

$$\lim_{t \to \infty} |x(t)| = \infty$$

and

$$\lim_{t \to -\infty} x(t) = x_0$$

for any nearby solution $x(t)$. A source can be considered as the dual case of a sink: the phase portrait of a source and of the corresponding sink are the same except that for the direction of the vectors which must be reversed. Reversing the arrows in Fig. 4, for example, the phase portrait of a system with coincident real positive eigenvalues would be obtained. Obviously, sources are unstable.

4) The eigenvalues are pure imaginary. The singular point is a *center*. All the nearby solutions are *periodic* with the same period. A center is a stable equilibrium. For a reason that will be made clear soon, this last case is of little practical interest, since even a small perturbation of the field will make the orbits spiral inward toward (or outward from) the singular point, changing the qualitative properties of the solution curves. In other words, a center is not *structurally stable*.

The crucial point is that this classification is *exhaustive*. Every singular point (in the linear case) looks like a saddle, a sink, a source, or a center. The same classification holds for the nonlinear case with respect to the eigenvalues of the derivative of the right-hand side of (B.2), considered as a linear operator. This can be seen considering the best linear approximation of the system in the neighbor of the singular point. Assuming that the real part of the eigenvalues of the matrix representative of the linear approximation at the singular point does not vanish, the phase portrait of the system in the neighbor of the singular point looks like in the corrisponding linear approximation. A dynamical system for which the real part of the eigenvalues of the linear approximation at each singular point does not vanish—with the additional conditions that there are no trajectories joining saddles and all the *limit cycles* are either periodic *attractor* or periodic *repeller*—is said to be *structurally stable*. It is clear that a dynamical system with a singular point whose linearization is a center is not structurally stable. Intuitively, a dynamical system is structurally stable if all the dynamical systems sufficiently *close* to it, share the same qualitative properties. A very important result of the theory of planar dynamical systems [25] says that the set S of planar dynamical systems which are structurally stable is dense in the set T of all the planar dynamical systems. Since S is also open in T, this result implies that almost all the planar dynamical systems are structurally stable.

ACKNOWLEDGMENT

We would like to thank V. Torre, D. Geiger, B. Caprile, and B. K. P. Horn for useful discussion and especially B. Saxberg for important comments and suggestions.

REFERENCES

[1] J. J. Gibson, *The Perception of the Visual World*. Boston: Houghton Mifflin, 1950.
[2] C. L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," *Comput. Graphics Image Processing*, vol. 9, pp. 301-315, 1979.
[3] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185-203, 1981.
[4] E. C. Hildreth, *The Measurement of Visual Motion*. Cambridge; MA: MIT Press, 1984.
[5] —, "The Computation of the velocity field," *Proc. Roy. Soc. London*, vol. B221, pp. 189-220, 1984.
[6] A. M. Waxman, "Image flow theory: A framework for 3-D inference from time-varying imagery," in *Advances in Computer Vision*, C. Brown, Ed. Norwood, NJ: Erlbaum, to be published.
[7] B. K. P. Horn, *Robot Vision*. Cambridge, MA, MIT Press/McGraw-Hill, 1986.
[8] K. Kanatani, "Structure from motion without correspondence: General principles," in *Proc. Image Understanding Workshop*, Miami, FL, 1985, pp. 107-116.
[9] H. H. Nagel, "Recent advances in image sequence analysis," in *Proc. Premier Colloque Image—Traitment, Synthese, Technologie et Applications*, Biarritz, France, 1984, pp. 545-558.
[10] 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 Intell.*, vol. PAMI-8, pp. 565-593, 1986.
[11] A. Verri and T. Poggio, "Against quantitative optical flow," in *Proc. 1st ICCV*, London, June 1987.
[12] A. Verri, F. Girosi, and V. Torre, "Mathematical properties of the 2-D motion field: From singular points to motion parameters," *J. Opt. Soc. Amer.*, in press, 1989.
[13] S. Uras, F. Girosi, A. Verri, and V. Torre, "Computational approach to motion perception," *Biol. Cybern.*, vol. 60, pp. 79-87, 1988.
[14] P. Burt and G. Sperling, "Time, distance, and feature trade-offs in visual apparent motion," *Psych. Rev.*, vol. 88, pp. 171-195, 1981.
[15] D. Marr and S. Ullman, "Directional selectivity and its use in early vision processing," *Proc. Roy. Soc. London*, vol. B211, pp. 151-180, 1981.
[16] M. Born and E. Wolf, *Principle of Optics*. New York: Pergamon, 1959.

[17] B. K. P. Horn and R. W. Sjoberg, "Calculating the reflectance map; *Appl. Opt.*, vol. 18, pp. 1770-1779, 1979.
[18] B. T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, pp. 311-317, 1975.
[19] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*. New York: Academic, 1974.
[20] B. Hassenstein and W. Reichardt, "Systemtheoretische Analyse der Zeit-, Reihenfolgen und Vorzeichenauswertung bei der Bewgungsperzeption der Russelkafers Chlorophanus," *Z. Naturforsch*, vol. IIb, pp. 513-524, 1956.
[21] J. P. H. Van Santen and G. Sperling, "A temporal covariance model of motion perception," *J. Opt. Soc. Amer.*, vol. A1, pp. 451-473, 1984.
[22] H. B. Barlow and R. W. Levick, "The mechanism of directional selectivity in the rabbit's retina," *J. Physiol.*, vol. 173, pp. 477-504, 1965.
[23] V. Torre and T. Poggio, "A synaptic mechanism possibly underlying directional selectivity to motion," *Proc. Roy. Soc. London*, vol. B202, pp. 409-416, 1978.
[24] T. Poggio and W. Reichardt, "Considerations on models of movement detection," *Kybernetik*, vol. 13, pp. 223-227, 1973.
[25] M. Peixoto, "Structural stability on 2-dimensional manifolds," *Topology*, vol. 1, pp. 101-120, 1962.

Alessandro Verri received the Ph.D. degree in physics from the University of Genova, Italy, in 1984.

From December 1985 to November 1986 and from May 1988 to July 1988 he was a Visiting Scientist at the Massachusetts Institute of Technology Center for Biological Information Processing and at the M.I.T. Artificial Intelligence Laboratory, Cambridge, MA. Currently, he is a postdoctoral fellow at the Department of Physics of the University of Genova. He has published research papers in computational vision and image understanding. His current research interests are in applied mathematics and computer vision.

Tomaso Poggio (A'86) was born in Genoa, Italy, on September 11, 1947. He received the Ph.D. degree in theoretical physics from the University of Genoa in 1970.

From 1971 to 1982, he was Wissenschaftlicher Assistant at the MaxPlanckInstitut für Biologische Kybernetik in Tübingen, West Germany. Since 1982, he has been a Profesor at the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology in Cambridge. In 1984, he was additionally appointed Professor at M.I.T.'s Whitaker College of Health Sciences and Technology, and was named the first director of the Center for Biological Information Processing. In 1988, he was named to the Uncas and Helen Whitaker Professorship. He has authored over 100 papers in areas ranging from psychophysics and biophysics to information processing in man and machine, artificial intelligence, and machine vision.

Prof. Poggio is on the editorial boards of *Biological Cybernetics, Advances in Applied Mathematics*, and several other interdisciplinary journals.