PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Computational Aspects of Motion Perception in Natural and Artificial Vision Systems

A. Verri, M. Straforini and V. Torre

Phil. Trans. R. Soc. Lond. B 1992 **337**, 429-443
doi: 10.1098/rstb.1992.0119

Computational aspects of motion perception in natural and artificial vision systems

A. VERRI, M. STRAFORINI AND V. TORRE

Dipartimento di Fisica dell' Università di Genova, Via Dodecaneso 33, 16146 Genova, Italy

CONTENTS

SUMMARY

In this paper a computational scheme for motion perception in artificial and natural vision systems is described. The scheme is motivated by a mathematical analysis in which first-order spatial properties of optical flow, such as singular points and elementary components of optical flow, are shown to be salient features for the computation and analysis of visual motion. The fact that different methods for the computation of optical flow produce similar results is explained in terms of the simple spatial structure of the image motion of rigid bodies. Singular points and elementary flow components are used to compute motion parameters, such as time-to-collision and angular velocity, and also to segment the visual field into areas which correspond to different motions. Then a number of biological implications are discussed. Electrophysiological findings suggest that the brain perceives visual motion by detecting and analysing optical flow components. However, the cortical neurons, which seem to detect elementary flow components, are not able to extract these components from more complex flows. A simple model for the organization of the receptive field of these cells, which is consistent with anatomical and electrophysiological data, is described at the end of the paper.

Phil. Trans. R. Soc. Lond. B (1992) **337,** 429–443
Printed in Great Britain

1. INTRODUCTION

The understanding of changing images is a remarkable feature of natural visual systems (Gibson 1950; Hassenstein & Reichardt 1956; Barlow & Levick 1965; Torre & Poggio 1978; Marr & Ullman 1981) and a major goal of machine vision is to build machines with similar capabilities. The classical notion of optical flow (Gibson 1950), or the motion of the image brightness pattern on the retina of the visual sensor, appears to be the bridge between research in motion perception of natural and artificial vision systems. In recent years, many aspects of motion understanding in artificial vision systems have been clarified (see, for example, Koenderink & van Doorn 1975; Horn & Schunck 1981; Longuet-Higgins & Prazdny 1981; Nagel 1983; Hildreth 1984), and it is now possible to compute optical flow almost in real time. However, recent electrophysiological recordings from cortical neurons of the monkey (Tanaka & Saito 1989; Anderson *et al.* 1990, 1991; Lagae *et al.* 1991; Duffy & Wurtz 1991*a*,*b*; Lagae 1991) have revealed the existence of units tuned to rotating or expanding flows, thus shedding new light on how the brain perceives motion. This paper outlines a theory of motion perception which has inspired a number of papers on the processing of image sequences in machine vision (Uras *et al.* 1988; Verri *et al.* 1989, 1990; Campani & Verri 1990, 1992; Rognone *et al.* 1992), and is also relevant for the understanding of motion mechanisms in the brain (Orban *et al.* 1992).

A general analysis of the many stages of motion perception is beyond the scope of the present research. The paper focuses on the segment of motion perception which, in the monkey, occurs along the visual pathway that goes from the visual area V1 (and V2) through the Middle Temporal area (MT) (Zeky 1974) to the Medial Superior Temporal area (MST) (Ungerleider & De Simone 1986), which is also referred as the visual area V5.

The paper is divided into two parts. The first part discusses basic mathematical facts about optical flow and methods for the computation and use of optical flow for motion analysis. Firstly, it is shown that the spatial structure of the optical flow produced by rigid, opaque objects is usually simple. Consequently, different methods for the computation of optical flow produce similar results which, equivalently, can be used for further analysis. In addition, the computation of optical flow is not critical. Secondly, optical flow, with the exception of surface boundaries, is often well described by its first-order spatial properties, such as singular points and elementary components. These spatial properties can be used to compute motion parameters, such as time-to-collision and angular velocity, and segment the visual field into the different moving objects.

The second part of the paper discusses the implications of this analysis on the design of artificial systems and on the understanding of how the brain analyses visual motion. Firstly, it is observed that both natural and machine vision systems overcome the intrinsic noise of visual motion estimation by using the regularity and density of motion estimates, and integrating visual information over large areas of the visual field. It is then argued that the brain does not necessarily have to compute optical flow. Instead, it may immediately extract the first-order spatial properties of optical flow from the changing image brightness. Finally, a model for the organization of the receptive field of cells in area MST or V5, which are likely candidates for the proposed analysis, is described. In agreement with experimental evidence, the model cells respond selectively to the stimuli presented and do so almost independently of the position of the stimuli on the receptive field; they do not extract elementary flow components from more complex flows.

The paper is organised as follows. In § 2, the spatial structure of optical flow is discussed. In § 3, four different methods for the computation of optical flow are briefly analysed. § 4 is dedicated to the first-order spatial properties of optical flow, that is, the singular points and the elementary components. Scope and goals of motion perception are then clarified in § 5. In § 6, the presented scheme is tested on real images. § 7 is devoted to the biological implications of the proposed analysis. A simple model of the organization of the receptive field of neurons sensitive to elementary flow components or their combinations is proposed in § 8. Finally, § 9 summarizes the results obtained and discusses similarities and differences of motion perception in natural and artificial visual systems.

2. MATHEMATICAL PROPERTIES OF THE OPTICAL FLOW OF RIGID OBJECTS

In this section, the notions of optical flow and motion are defined and the spatial structure of the optical flow of rigid objects is discussed. The important property of structural stability is briefly commented on. The mathematical properties of this section have severe implications for the understanding of motion perception in artificial and biological visual systems, implications which will be discussed in the following sections.

(*a*) *Preliminaries*

Let us define the motion field $\boldsymbol{v}$ as the perspective projection onto the image plane of the velocity field $\boldsymbol{V}$, the velocity of the moving surfaces in space. The optical flow $\boldsymbol{u}$, instead, can be defined as the apparent motion of the image brightness pattern over the image plane. These two vector fields $\boldsymbol{v}$ and $\boldsymbol{u}$ can be different because of the presence of shadows, specularities and changes in the illumination intensity. Because this motion does not necessarily coincide with the perspective projection of the velocity field, optical flow can only be considered to be an estimate of the motion field.

Let us now establish some notations. Let $\boldsymbol{X}$ be a point in space on one of these surfaces and $\boldsymbol{x}$ its perspective projection onto the image plane. In what follows, if $\boldsymbol{y}$ is a vector in space, let $y_i = \boldsymbol{y} \cdot \boldsymbol{e}_i$, $i = 1,2,3$, where the $\boldsymbol{e}_i$ are unit vectors parallel to the three mutually orthogonal axes (X_1, X_2 and X_3) of a system

of coordinates. The image plane is orthogonal to $\boldsymbol{e}_3$ (that is, the optical axis is parallel to $\boldsymbol{e}_3$) and the focus O and the origin of the system of coordinates coincide. Thus, we have

$$\boldsymbol{x} = f(\boldsymbol{X}/X_3), \tag{1}$$

where f is the focal length. Note that $x_3 = f$, for every $\boldsymbol{x}$, and that $X_3 > f$, for every $\boldsymbol{X}$, because otherwise $\boldsymbol{X}$ would not be visible. If $\boldsymbol{X}$ is moving with velocity $\boldsymbol{V}$, then the motion field $\boldsymbol{v}$ can also be obtained by differentiating equation (1) with respect to time; that is,

$$\boldsymbol{v} = f\frac{\boldsymbol{e}_3 \times (\boldsymbol{V} \times \boldsymbol{X})}{X_3^2}. \tag{2}$$

As v_3 is always zero, we write $\boldsymbol{v} = (v_1, v_2)$. Thus, the motion field $\boldsymbol{v}$ of a smooth surface, with the exception of the occluding boundaries, is a smooth two-dimensional (2D) vector field. In the presence of more than one moving surface the motion field is only piecewise smooth, and so a filtering step may be required to obtain a globally smooth motion field.

(*b*) *The spatial structure of the motion field*

The motion field $\boldsymbol{v} = (v_1, v_2)$ of a moving planar surface at the point $\boldsymbol{x} = (x_1, x_2, f)$ of the image plane can always be written as a quadratic polynomial of the image coordinates

$$v_1(x_1,x_2) = a_1x_1^2 + a_2x_1x_2 + a_3x_1f + a_4x_2f + a_5f^2$$
$$v_2(x_1,x_2) = a_1x_1x_2 + a_2x_2^2 + a_6x_1f + a_7x_2f + a_8f^2, \tag{3}$$

where the a_i, $i = 1, \ldots, 8$ depend on the motion and structure parameters (Longuet-Higgins 1984; Negahdaripour & Horn 1987). From the qualitative point of view, the spatial structure of the vector field of equation (3) is fairly simple as there are three singular points at most (Verri *et al.* 1989) and there cannot be limit cycles (Verri & Aicardi 1990). Most importantly, as f, the focal length, is usually much larger than x_1 and x_2, the motion field $\boldsymbol{v}$, over relatively large patches of the image plane, and apart from special values of the a_i, $i = 1, \ldots, 8$, is well approximated by a linear vector field (Campani & Verri 1992). This fact is illustrated in figure 1. Figure 1*a*,*b* shows the motion fields of a translating and a rotating planar surface, respectively. In figure 1*c*,*d* the piecewise linear vector fields which best approximate the vector fields of figure 1*a*,*b* are reproduced. The vector fields of figure 1*c*,*d* are obtained by computing the best linear approximation of the vector fields of figure 1*a*,*b* over 8×8 non-overlapping patches of the image plane (each patch contains 32×32 flow vectors). It is evident that the piecewise linear approximations are almost indistinguishable from the actual vector fields.

Theoretical and experimental evidence shows that, with the exception of occluding boundaries, the same argument holds true for smooth non-planar surfaces (Campani & Verri 1992). In this case, the size of the patches of the image plane on which the motion field can be approximated by a linear vector field changes depending on the motion parameters and the apparent curvature of the viewed surface.

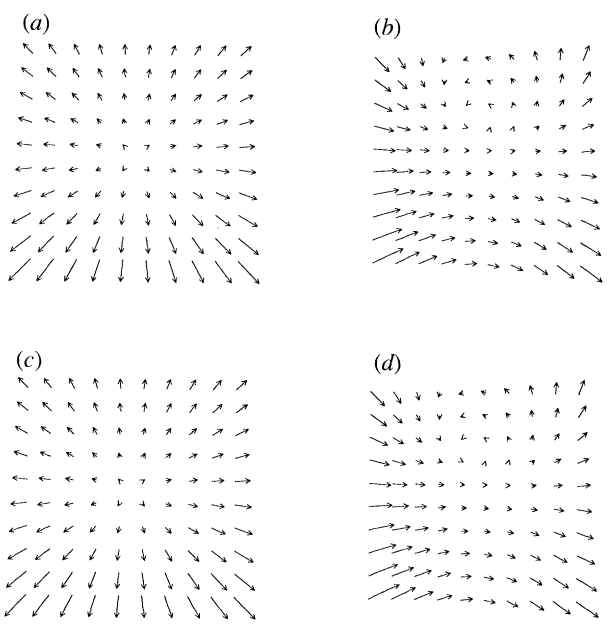


Figure 1. Linear approximation of the motion field of a moving plane. (*a*) The motion field of the plane of equation $2X_2 + X_3 = 100\,f$, where f is the focal length. The plane is translating at unit speed along the optical axis (the vector field has been subsampled for better visualization). (*b*) The motion field of the plane of equation $0.06X_1 - 0.1X_2 + X_3 = 100\,f$. The plane is rotating at unit angular speed around an axis parallel to the X_2 axis. (*c*) The vector field which is obtained by computing the linear approximation of the motion field of (*a*) over 8×8 non-overlapping patches of 32×32 flow vectors each. (*d*) The vector field which is obtained by computing the linear approximation of the motion field of (*b*) over 8×8 non-overlapping patches of 32×32 flow vectors each.

(*c*) *Motion field, optical flow, and structural stability*

Motion field and optical flow are usually different (Verri & Poggio 1989), but the notion of structural stability provides a natural framework within which they can be considered equivalent (Verri *et al.* 1989). Intuitively, a sufficiently smooth vector field is structurally stable if its qualitative properties (like the number and the topological structure of its singular points and limit cycles) remain unchanged for small (and smooth) perturbations. The emphasis on structural stability is due to a fundamental theorem of the theory of dynamical systems, which says that, among planar dynamical systems, the property of structural stability is generic (for an elementary account of the theory of dynamical systems and structural stability see Hirsch & Smale (1974)). This result suggests that certain qualitative and, to some extent, quantitative properties of motion field and optical flow are the same.

3. THE COMPUTATION OF OPTICAL FLOW

One important consequence of the simple spatial structure of optical flow described in § 2 is that

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

different methods for the computation of optical flow produce similar results. In this section, four methods are briefly described and commented on in the light of this observation. Other methods such as the energy models proposed by Adelson & Bergen (1985) are almost identical to the techniques here discussed.

(*a*) *The spatial gradient constancy method*

Let us assume that the spatial gradient of the image brightness, ∇E, is stationary over time, *i.e.*

$$\frac{d}{dt}\nabla E = 0. \tag{4}$$

Equation (4) can be rewritten as a pair of linear algebraic equations for the optical flow components (u_1,u_2), or

$$E_{xx}u_1 + E_{xy}u_2 + E_{xt} = 0$$
$$E_{xy}u_1 + E_{yy}u_2 + E_{yt} = 0,$$

where the subscripts x, y, and t denote partial derivatives with respect to the spatial and temporal coordinates respectively. The solution to equation (4) can be written as

$$\boldsymbol{u} = -\mathbf{H}^{-1}\frac{\partial}{\partial t}\nabla E,$$

where

$$\mathbf{H} = \begin{pmatrix} E_{xx} & E_{xy} \\ E_{yx} & E_{yy} \end{pmatrix}$$

is the Hessian matrix. Equation (4), which can be derived through an analogy with the theory of deformable objects (Helmholtz 1858), has been proposed by several authors (Haralick & Lee 1983; Nagel 1983; Tretiack & Pastor 1984; Uras *et al.* 1988). The optical flow $\boldsymbol{u}$ is uniquely determined when $\mathbf{H}$ is invertible and can be written in terms of the true motion field $\boldsymbol{v}$ as

$$\boldsymbol{u} = \boldsymbol{v} + \mathbf{H}^{-1}\left(\mathbf{M}^T\nabla E - \nabla\frac{dE}{dt}\right), \tag{5}$$

where $\mathbf{M}^T$ is the transpose of $\mathbf{M}$ and $dE/dt = \nabla E \cdot \boldsymbol{v} + E_t$. The matrix $\mathbf{M}$ is the jacobian of the motion field $\boldsymbol{v}$ and is defined as:

$$\mathbf{M}_{ij} = \partial v_i/\partial x_j \qquad i = 1,2.$$

From equation (5) the true motion field and the optical flow of equation (4) coincide if the image brightness pattern is stationary ($dE/dt = 0$) and the motion field is spatially constant ($\mathbf{M} = 0$). Strictly speaking, both of the conditions above do not usually hold, but it can be shown (Verri *et al.* 1990) that, provided dE/dt and the entries of $\mathbf{M}$ are small, then the larger the eigenvalues of the Hessian matrix $\mathbf{H}$, the smaller is the relative difference between $\boldsymbol{u}$ and $\boldsymbol{v}$.

Figure 2 shows four frames of a sequence in which the viewing camera is translating toward a picture posted on the wall. Figure 3*a* shows the optical flow obtained from equation (4) associated with the third frame of figure 2. It is evident that the computed optical flow is qualitatively consistent with the true motion.

(*b*) *The image brightness constancy method*

Many approaches to the computation of optical flow are based on the view that the image brightness E is stationary over time (Fenmena & Thompson 1978; Horn & Schunck 1981; Hildreth 1984),

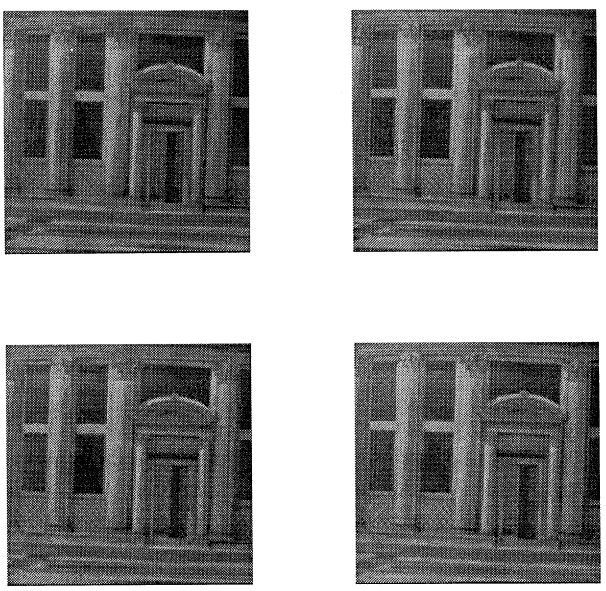


Figure 2. Four frames of a sequence (from upper left to lower right) taken while the viewing camera is translating along the optical axis toward a picture posted on the wall. Images were grabbed by means of a PULNIX TM46 camera and an Imaging Technology board FG100. Each image consists of 256×256 pixels.

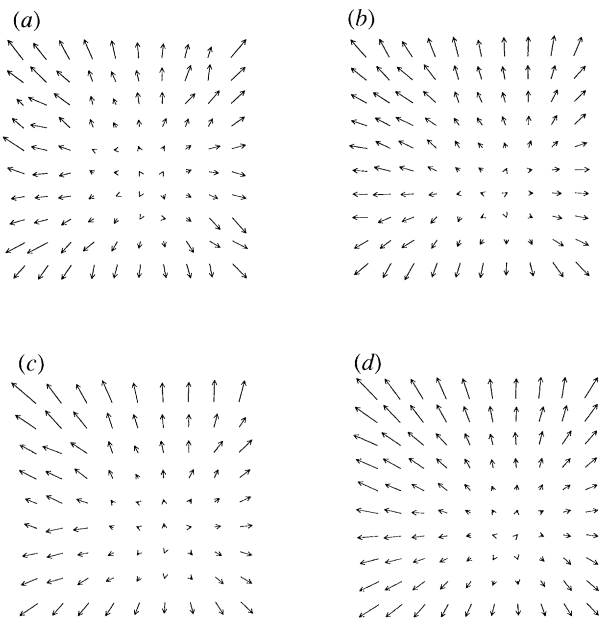


Figure 3. The optical flow of the third frame of the sequence of figure 2 computed through (*a*) the spatial gradient constancy method, (*b*) the image brightness constancy method, (*c*) the matching method, and (*d*) the shape constraint method described in § 3. The vector fields have been subsampled for better visualization.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

$$\frac{d}{dt}E = 0. \tag{6}$$

Equation (6) can be rewritten as

$$\nabla E \cdot \boldsymbol{u} + E_t = 0,$$

from which it is clear that the vector field $\boldsymbol{u}$ cannot be uniquely determined from equation (6). In the search for additional constraints the most common approach (Horn & Schunck 1981) looks for the vector field $\boldsymbol{u}$ which minimizes the functional

$$\boldsymbol{\Phi}(\boldsymbol{u}) = \iint \left(\frac{dE}{dt}\right)^2 dxdy + \alpha \iint (|\nabla u_1|^2 + |\nabla u_2|^2) dxdy$$

where the double integral extends over the whole image plane and α is a positive parameter. Intuitively, the vector field which minimizes $\boldsymbol{\Phi}$ is the smoothest vector field that nearly satisfies equation (6). Quite interestingly, the conditions for which $\boldsymbol{\Phi}$ is 'small' are similar to those which ensure that the optical flow computed as the solution to equation (4) is close to the true motion field (as $|dE/dt|$ and the $|\partial u_i/\partial x_j|$ must be small). This suggests that the two different schemes provide similar results due to the relatively simple spatial structure of the motion field of rigid surfaces. Figure 3*b* shows the optical flow obtained through the technique described in (Horn & Schunck 1981), and associated with, the third frame of the sequence of figure 2. It is evident that the vector field of figure 3*b* is qualitatively correct and quantitatively similar to the optical flow of figure 3*a*.

(*c*) *A matching method*

A third approach to the computation of optical flow is based on the matching of features from two images (see, for example, Hildreth (1984) and Faugeras *et al.* (1987)). A possible method (Poggio *et al.* 1986) assumes that optical flow is locally constant, that is, for each point, the displacement of nearby points under the optical flow is the same. At each point in the image, under each integer displacement, the images in the two frames are compared and a measure of the matching between points is computed and summed over a small region. This can be interpreted as matching small patches from the first image with small patches in the second. The resulting flow field is spatially coherent as a result of the fact that the support regions, the patches, have large overlap. The displacement is chosen to maximize the matching measure over all displacements. Accuracy can be improved by interpolating the matching measure. Once again the conditions under which this method produces an optical flow similar to the motion field are that the image brightness is nearly stationary ($dE/dt \simeq 0$) and that spatial variation of the motion field is locally negligible ($\partial u_i/\partial x_j \simeq 0$). The optical flow of figure 3*c* has been obtained by means of this method. A computational advantage of matching approaches is that they perform equally well even if the dynamic range and the intensity distribution of the image brightness pattern are far from being optimal.

(*d*) *A shape constraint method*

In the attempt to combine the simplicity of both the constraint equation (6) and the spatial structure of the motion field of rigid surfaces, a method has been proposed in which optical flow is estimated as the solution to a highly overconstrained system of linear algebraic equations (Verri & Campani 1990). The image plane is divided into overlapping regions in which it is assumed that the image brightness is stationary and that the motion field is at most linear. Let $\boldsymbol{x}_0$ be the central point of a region r which contains N points. Then at each point $\boldsymbol{x}$ of r we have

$$\nabla E \cdot \boldsymbol{v}(\boldsymbol{x}_0) + \nabla E \cdot \mathbf{M}(\boldsymbol{x} - \boldsymbol{x}_0) = -E_t, \tag{7}$$

which is a linear equation for $\boldsymbol{v}(\boldsymbol{x}_0)$ and $\mathbf{M}$ (whose entries are meant to be computed at $\boldsymbol{x}_0$). It is possible to use standard least mean square techniques to solve the overconstrained system of N equations, such as equation (7), for six unknowns (the two components of $\boldsymbol{v}(\boldsymbol{x}_0)$ and the four entries of $\mathbf{M}$). As a result, both the motion field and its first-order spatial derivatives can be estimated at the same time. Again it should be noticed that the conditions for which the method is likely to provide good estimates of the motion field coincide with the previous cases. It can easily be seen that the optical flow of figure 3*d*, which has been computed throuth this method, is qualitatively correct and very similar to the optical flows obtained through the techniques previously described (see figure 3*a*,*b*,*c*, respectively).

4. FIRST-ORDER PROPERTIES OF THE MOTION FIELD

This section analyses the relation between the linear part, or the first-order spatial properties of the motion field, and some relevant features of the viewed motion. From these properties it is possible to obtain a qualitative understanding of the three-dimensional motion, to distinguish between translation, rotation, and general motion, and to recognize different moving objects. Moreover, relevant quantitative parameters, such as time-to-collision and angular velocity, can be robustly estimated.

It may be useful to give an intuitive notion of first-order properties before going into the mathematical details. A vector field, which is not constant, has a different orientation and amplitude at each position. An important property of a vector field is the way in which the orientation and amplitude change at each position. When changes of the amplitude and direction of vectors are the same for similar displacements, first-order properties of the field are constant all over the field. In this case the essence of the vector field is captured by the rate of change of the amplitude and direction, or, in other words, by first-order properties. When changes of the amplitude and direction vary significantly at each position, it is also necessary to consider higher-order properties. The main reason why first-order properties of the motion field are usually adequate to describe the motion field is that the motion field of solid opaque objects is usually

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

coherent and its rate of change is almost constant. In the presence of a relative motion between two different objects, the motion field can be discontinuous at some locations and therefore it is also necessary to detect discontinuities.

(*a*) *Singular Points*

The rate of change of a vector field $\boldsymbol{v}$ is usually described by the jacobian matrix $\mathbf{M}$, which has already been introduced in equation (5). The linear approximation of the motion field in the region of a point $\boldsymbol{x}$ can be described by the 2×2 matrix $\mathbf{M}$ computed at $\boldsymbol{x}$, which fully captures the first-order properties of the field. If $\boldsymbol{x}$ is a singular point, that is, a point in which the motion field vanishes, the matrix $\mathbf{M}$ could be useful to distinguish between different kinds of motion and to estimate motion parameters. If $\boldsymbol{x}$ is not a singular point, the matrix $\mathbf{M}$ could still be used to understand qualitative properties of the viewed motion and segment the image in the different moving surfaces. Let us consider the case in which $\boldsymbol{x}$ is a singular point first.

The qualitative nature of three-dimensional motion (like translation or rotation) can be described by looking at the temporal evolution of the spatial structure of the motion field in the region of the singular points. Intuitively, translation is often associated with expanding (or contracting) motion fields, and rotation with circulating motion fields around a singular point (which may or may not be visible). Let us study these assumptions in details.

(i) *Translation*

If a surface is translating in space with velocity $\boldsymbol{T}$, the resulting motion field $\boldsymbol{v}$ has, at most, one singular point $\boldsymbol{p} = (p_1, p_2, f)$ (which can be obtained by substituting $\boldsymbol{V} = \boldsymbol{T}$ in equation (2) and setting $\boldsymbol{v} = 0$) such that (Verri *et al.* 1989):

1. The spatial location of $\boldsymbol{p}$ on the image plane does not change over time;
2. The heading direction is related to $\boldsymbol{p}$ as for $i = 1,2$

$$p_i = f(T_i/T_3)$$

3. The matrix $\mathbf{M}$ at $\boldsymbol{p}$ is a multiple of the identity matrix, that is, $\mathbf{M} = \lambda \mathbf{I}$. In the language of the theory of dynamical systems, $\boldsymbol{p}$ is always a non-degenerate focus with the only eigenvalue λ of $\mathbf{M}$ given by

$$\lambda = T_3/X_3 = 1/\tau, \tag{8}$$

where τ is the time-to-collision.

From these properties it is evident that useful information on translational motion can be easily obtained from the singular point of the motion field.

(ii) *Rotation*

In the case of a rotating surface it is useful to distinguish between two different kinds of singular points. Let $\boldsymbol{p}$ be the singular point perspective projection of a point $\boldsymbol{P}$ in the three-dimensional space and $\boldsymbol{V}(\boldsymbol{P})$ the velocity of $\boldsymbol{P}$. A point $\boldsymbol{p}$ can be a singular point either because $\boldsymbol{P}$ lies on the rotation axis (that is, $\boldsymbol{v}(\boldsymbol{p}) = 0$ and $\boldsymbol{V}(\boldsymbol{P}) = 0$), or because $\boldsymbol{V}(\boldsymbol{P})$ lies on the straight line which goes through $\boldsymbol{P}$ and the centre of projection (that is, $\boldsymbol{v}(\boldsymbol{p}) = 0$ but $\boldsymbol{V}(\boldsymbol{P}) \neq 0$). A singular point of the first kind is named an immobile point.

Let us briefly discuss the case in which $\boldsymbol{p}$ is an immobile point (a more general analysis of the singular points of rotation can be found in Verri *et al.* (1989)). If the rotation axis is orthogonal to the plane tangential to the surface at $\boldsymbol{P}$, then the singular point $\boldsymbol{p}$ is a centre and the motion field in the region of $\boldsymbol{p}$ is tangential to closed orbits. In this case, which can be named orthogonal rotation, the angular velocity ω can be written in terms of the complex conjugate eigenvalues λ and $\bar{\lambda}$ of the matrix M computed at $\boldsymbol{p}$ as

$$\omega^2 = \lambda\bar{\lambda}.$$

In all the other cases, the spatial structure of the motion field in the region of $\boldsymbol{p}$ changes depending on the angle θ between the rotation axis and the unit normal to the surface at $\boldsymbol{p}$ and on the relative position of the rotating surface with respect to the viewing point. As an example, figure 4 shows four rather different optical flows which can be produced by a rotating planar surface.

(iii) *General motion*

The case of general motion is more complicated. Although it is well known that any rigid motion can be instantaneously decomposed into a translational and a rotational term, it is not evident whether this classical result of kinematics is relevant to motion

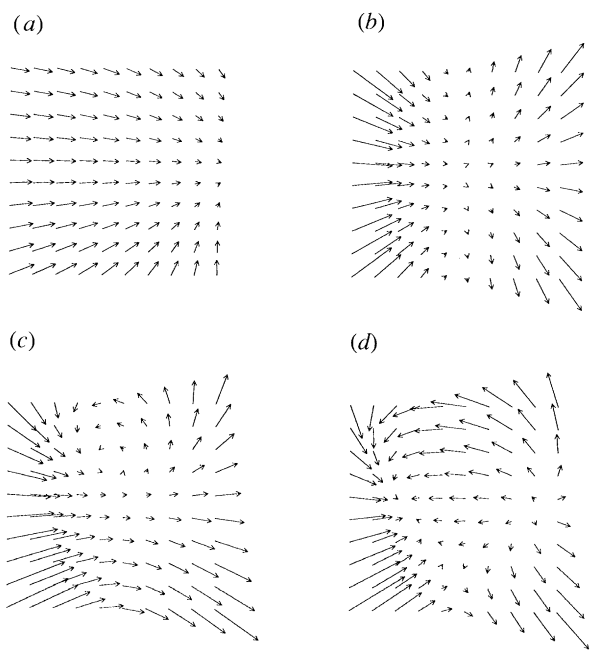


Figure 4. Optical flows produced by a rotating planar surface. (*a*) Optical flow of a planar surface orthogonal to the ground plane and rotating around a vertical axis which lies outside the field of view. (*b*), (*c*) and (*d*) Optical flows of a planar surface nearly orthogonal to the ground plane and rotating around a vertical axis which lies on the optical axis.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

perception or not. Let us restrict the present analysis of general motion to the simple but interesting case in which the viewed object is moving on a flat surface S. In this case, which can be termed passive navigation, the rotation axis is orthogonal to both the surface S and the translational component of motion and useful information can again be obtained from the singular points of the motion field.

Firstly, the three-dimensional motion is instantaneously indistinguishable from a rotation around a fixed axis. Therefore, the singular points of the motion field must lie on the straight line of the image plane which is parallel to S and contains the projection of the optical axis. At a singular point $\boldsymbol{p}$ we can write

$$Det\mathbf{M} = \frac{Tr^2\mathbf{M}}{4} - \frac{1}{4}\left(\frac{\boldsymbol{\alpha} \times \boldsymbol{V} \cdot \boldsymbol{\omega}}{\boldsymbol{\alpha} \cdot \boldsymbol{V}}\right)^2, \tag{9}$$

where $Det\mathbf{M}$ and $Tr\mathbf{M}$ are the determinant and the trace of the matrix $\mathbf{M}$ respectively, $\boldsymbol{\alpha}$ is the unit vector normal to the moving surface at $\boldsymbol{P}$, $\boldsymbol{\omega}$ is the angular velocity, and $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{P})$. From equation (9) it follows that the eigenvalues of the matrix $\mathbf{M}$ are always real. This constraint can be useful in the analysis of complex motion from singular points (Verri *et al.* 1989).

To conclude, the understanding of motions such as rotations and constrained general motion requires a fairly deep knowledge of mathematics of a type which is probably beyond the relatively simple stages of motion perception. It may thus be argued that, at these preliminary stages, biological visual systems are mainly tuned for the purpose of recognizing translational motion and orthogonal rotation. More general kinds of motion are likely to require more sophisticated analysis and processing.

(*b*) *Elementary components*

Let us now consider the case in which the matrix $\mathbf{M}$ of equation (5) is computed at $\boldsymbol{x}$, where $\boldsymbol{x}$ is not a singular point. According to a classical theorem of the theory of deformable bodies (Helmholtz 1858; Koenderink & Van Doorn 1975), the spatial structure of a vector field over a sufficiently small patch can be described as the sum of a rigid translation, a uniform expansion, a pure rotation, and two components of shear. Therefore, the motion field $\boldsymbol{v}'$, at a point $\boldsymbol{x}'$ in a sufficiently small region of $\boldsymbol{x}$, can be written as

$$\boldsymbol{v}' = \boldsymbol{v} + \mathbf{M}(\boldsymbol{x}' - \boldsymbol{x}), \tag{10}$$

where $\boldsymbol{v}$ is the motion field at $\boldsymbol{x}$ and the matrix $\mathbf{M}$ is meant to be computed at $\boldsymbol{x}$. It is clear that in equation (10) the rigid translation is given by the term $\boldsymbol{v}$. As $\boldsymbol{x}$ is not a singular point, then $\boldsymbol{v} \neq 0$. The matrix $\mathbf{M}$ can be written as

$$\mathbf{M} = \alpha\mathbf{I}_1 + \beta\mathbf{I}_2 + \gamma_1\mathbf{I}_3 + \gamma_2\mathbf{I}_4,$$

where

$$\mathbf{I}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \mathbf{I}_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \mathbf{I}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \mathbf{I}_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};$$

and $\alpha = (\mathbf{M}_{11} + \mathbf{M}_{22})/2$, $\beta = (\mathbf{M}_{12} - \mathbf{M}_{21})/2$, $\gamma_1 = (\mathbf{M}_{11} - \mathbf{M}_{22})/2$, and $\gamma_2 = (\mathbf{M}_{12} + \mathbf{M}_{21})/2$. The matrices $\mathbf{I}_i$, $i = 1, \ldots, 4$ correspond to the terms of uniform expansion, pure rotation, and components of shear, respectively. Because $\mathbf{I}_1$ and $\mathbf{I}_2$ are left unchanged for arbitrary orthogonal transformation, the quantities α and β are independent of the system of coordinates. Consequently, the amount of uniform expansion and pure rotation are intrinsic properties of the motion field (Koenderink & Van Doorn 1975). On the contrary, because $\mathbf{I}_3$ and $\mathbf{I}_4$ are not invariant, γ_1 and γ_2 do not describe intrinsic properties of the motion field. An invariant measure of the amount of shear can be given by $\gamma = \sqrt{\gamma_1^2 + \gamma_2^2}$ (because $\alpha^2 + \beta^2 + \gamma_1^2 + \gamma_2^2$ is invariant).

The relevance of elementary components to the analysis of visual motion stems from the fact that the motion field is usually well approximated by a linear vector field over rather large patches of the image plane (see § 2). In other words, the amount of expansion, rotation, and shear is a nearly piecewise constant function over the image plane. This often makes it possible to segment the motion field into regions where the relative amount of expansion, rotation or shear is larger than a fixed value. These regions may correspond to different moving objects, and can be used to describe the observed motion and identify motion discontinuities qualitatively (see § 6).

5. SEEING MOVING OBJECTS

Seeing moving objects is a very sophisticated property of vision systemss. In this section, the scope of the paper is properly defined and clarified.

Firstly, it has to be said that the present research is restricted to the analysis of rigid (or piecewise rigid) and opaque objects. This processing stage, which we may call motion perception, includes the computation and analysis of optical flow and the understanding of a number of relatively simple three-dimensional motions, like translational motions (which may also cause expanding or contracting flows) and orthogonal rotations. Therefore, deformable and transparent objects are beyond the scope of this research. In the brain, the visual pathway we are looking at goes from the visual areas V1 (and V2) to the MT and MST areas (Livingstone & Hubel 1988; Zeki & Shipp 1988; De Yoe & Van Essen 1988).

The mathematical results presented in § 2, § 3 and § 4 suggest that an adequate account of motion perception is likely to be composed of: (i) the location and identification of the singular points of optical flow; (ii) the computation of the linear terms of optical flow at every location of the image plane; in other words, the estimation of the average component of translation, rotation, expansion, and shear; and (iii) the detection of optical flow discontinuities.

In § 4 it has been shown that the information obtained in the first and second points makes it possible to distinguish between translation and rotation, and to evaluate three-dimensional motion parameters like the time-to-collision and angular velocity. The information obtained in the third point, instead, can be used to understand the complexity of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the viewed scene and identify the different moving objects (Francois & Bouthemy 1990). Let us now test the proposed scheme in computer vision. The biological implications are considered in § 7.

6. MOTION PERCEPTION IN MACHINES

In this section, experiments are reported in which the proposed scheme is tested on natural images. The experiments suggest that optical flow can be computed from real images and that the proposed representation of motion perception can be effectively obtained.

Figure 5*a* illustrates a frame of a sequence of images in which the viewing camera was moving towards the scene (composed of a plant in the foreground and two puppets in the background) while the puppets were moving away. Figure 5*b* shows the optical flow computed according to the technique described by Uras *et al.* (1988). The singular points are shown in figure 5*c*, and the optical flow segmentation is shown in figure 5*d*. The singular points determine the direction and sign of the apparent motion of the viewing camera (lower point of figure 5*c*, focus of expansion) and the motion of the puppets (upper point of figure 5*c*, focus of contraction). Correspondingly, the homogeneous region in light grey of figure 5*d* roughly identifies the part of the image which moves consistently with camera motion, whereas the darker area of figure 5*d* corresponds to the moving puppets. This segmentation is obtained through a stochastic relaxation technique described by Rognone *et al.* (1992).

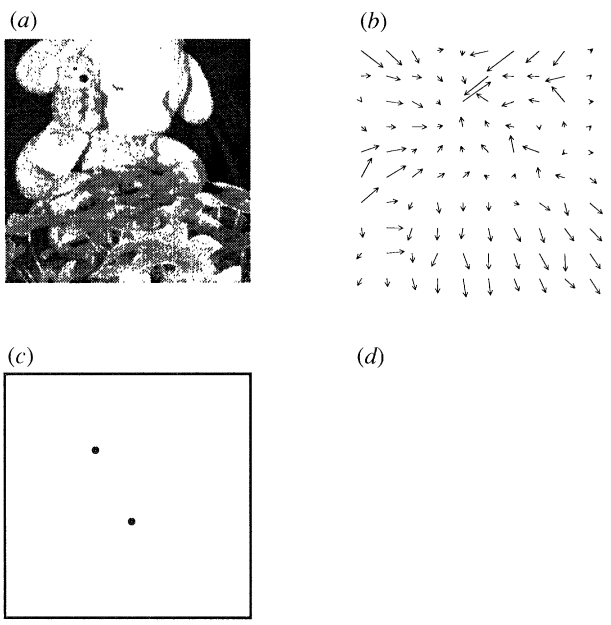


Figure 5. Detection of relative motion. (*a*) A frame of a sequence in which the puppets are moving away from the camera, while the camera was translating towards the plant. (*b*) The optical flow associated with the frame of (*a*) and computed through the spatial gradient constancy method (Uras *et al.* 1988). (*c*) The two singular points which have been detected in the optical flow of (*b*). (*d*) The motion segmentation which has been obtained through a technique described Rognone *et al.* (1992). The darker area indicates a region of contraction and corresponds to the puppets motion, the lighter area a region of expansion and corresponds to the rest of the viewed scene.

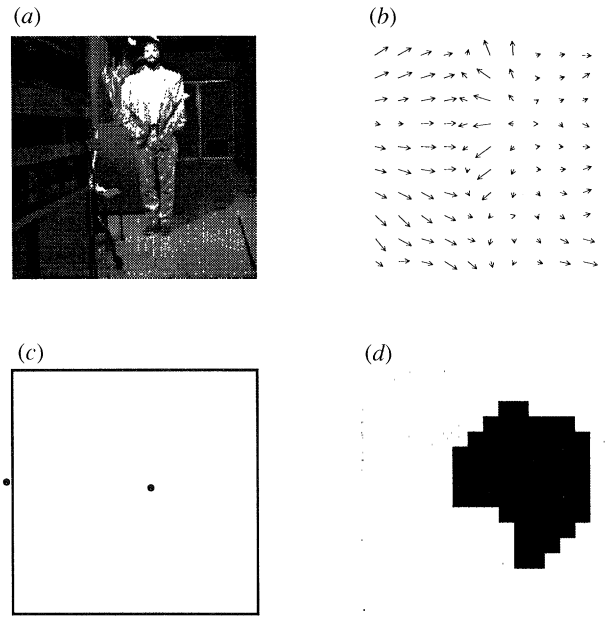


Figure 6. A second example of detection of relative motion. (*a*) A frame of a sequence in which a person was moving towards the camera, while the camera was translating toward the scene. (*b*) The optical flow computed by using a shape constraint method (Verri & Campani 1990). (*c*) The singular points (both of them are foci of expansion) of the optical flow of (*b*). The focus of expansion which corresponds to the camera motion lies just outside the image boundaries. (*d*) The segmentation of the optical flow of (*b*). The darker area corresponds to the moving person, whereas the lighter area identifies the part of the scene which moves rigidly with respect to the viewing camera.

Figure 6*a* shows a frame of a second sequence of real images in which a person was moving towards the camera, while the camera was translating towards the scene (a corridor of our laboratory). The optical flow computed through the same method as above is shown in figure 6*b*. Figure 6*c*,*d* identifies the singular points and the regions which correspond to the different motions. In this case, both of the singular points of figure 6*c* are foci of expansion. The darker area of figure 6*d* roughly corresponds to the motion of the person (the associated singular point lies near the centre of figure 6*c*), whereas the area in light grey corresponds to the camera motion (the associated focus of expansion lies just outside the image frame of figure 6*c*).

Similar results obtained on other image sequences suggest that the relative motion between the viewer and the scene can be recognized from the segmentation of first-order properties of optical flow (Rognone *et al.* 1992). Quantitative motion estimates, like the time-to-collision between the viewing system and the point in the scene imaged in the focus of expansion, can be computed through equations like equation (8). In the example of figure 6, the eigenvalue λ can be estimated by integrating over quite a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

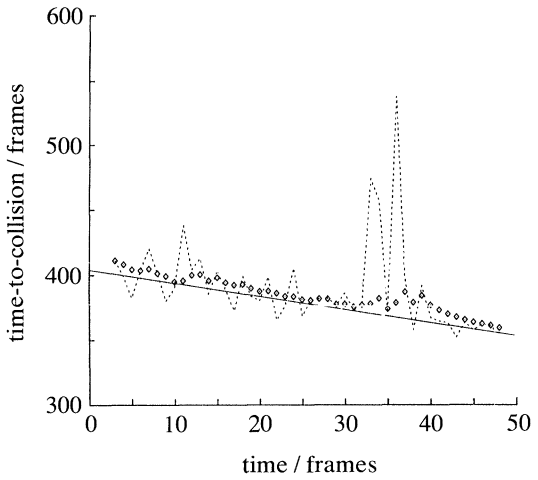


Figure 7. Accuracy of the computation of the time-to-collision from optical flow data. The time-to-collision is computed from a sequence of 50 frames (four of which are shown in figure 2). The solid line is the true time-to-collision. The broken line indicates the time-to-collision estimated through equation (8). Also shown are results obtained through a Kalman-filtering procedure described in De Micheli *et al.* (1992) (diamonds).

large number of independent motion measurements, and the obtained time-to-collision is thus reasonably accurate.

In some cases, such as passive navigation, it may be possible to improve the accuracy of the estimation through the use of appropriate Kalman filtering. Figure 7 shows the estimated time-to-collision which has been computed over a long sequence of images (the solid line reproduces the true time-to-collision). It is evident that the estimated time-to-collision (broken line) may be affected by large errors and become point-wise unreliable. More stable and accurate estimates, instead, can be obtained by means of a simple Kalman filtering procedure (diamonds) (De Micheli *et al.* 1992).

7. BIOLOGICAL IMPLICATIONS

Let us now discuss some biological implications of the computational properties of motion perception presented in the previous sections. Firstly, the nature of the computation of optical flow in biological vision systems is analysed. Then, a basic strategy which seems to be used by the brain in motion perception is described. Finally, the issue of whether or not the brain actually needs, and therefore really computes, optical flow is raised.

(*a*) *The nature of the computation of optical flow*

The analysis of the mathematical properties of the motion field of rigid objects (see § 2, § 3 and § 4) has shown that the first-order spatial properties of the motion field almost completely characterize the entire vector field. Therefore, the problem of the computation of image motion does not need to go beyond a first-order analysis. In addition, as a result of the property of structural stability, the proposed scheme tolerates some difference between the true motion field and the estimated optical flow. These facts have important consequences.

Firstly, the computation of optical flow does not appear to be very difficult. The 'aperture problem' (Marr & Ullman 1981) can easily be solved in several different ways (see the various methods described in § 3). Secondly, because of the strong constraints on the spatial changes of the optical flow of rigid bodies, the computation of optical flow does not require the use of rather sophisticated techniques such as those provided by regularization theory (Tikhonov & Arsenin 1977; Bertero *et al.* 1988). Therefore, it is difficult to find a particular technique which the brain must adopt. In addition, a very accurate estimation of the true motion field is hardly necessary. Finally, the three-dimensional motion parameters can be recovered robustly from the singular points of an optical flow, which is not exactly equal to the true motion field.

(*b*) *Spatial integration*

The presented analysis and electrophysiological evidence (Livingstone & Hubel 1988; Zeki & Shipp 1988; De Yoe & Van Essen 1988) suggest a possibly basic strategy adopted by the brain in motion perception. It appears that the processing of visual motion is subject to massive integration over a larger portion of the visual field. The receptive field of neurons in V1 is usually not larger than 1° or 2°, whereas the average receptive field size is about 15° in MT (Maunsell & Van Essen 1983; Ungerleider & De Simone 1986) and much larger (about 50°) in MST (Saito *et al.* 1986; Tanaka *et al.* 1986; Boussaud *et al.* 1990). In § 6 it has been shown that the role of spatial integration in machine vision is to improve accuracy in the estimation (and understanding) of the viewed motion. Therefore, spatial integration can be a good strategy for motion perception common to both artificial and natural visual systems. However, because of the different circuitry, the actual extent of spatial integration in artificial and natural visual systems could be different.

(*c*) *Does the brain compute optical flow?*

According to the computational model previously described, motion perception consists of two steps. In the first step, optical flow is computed, whereas in the second step its first-order spatial properties, like singular points and elementary flow components, are extracted and used to accomplish several relatively simple visual tasks. Although there is little doubt that the brain uses some kind of optical flow information when looking at moving images (Movshon *et al.* 1985), it is not obvious whether the brain computes and represents optical flow in a specific cortical area or not.

Neurons in the visual area V1, for example are orientation and directional selective, and have been described as having the 'component directional selectivity' (Movshon *et al.* 1985). Many neurons in the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

visual area MT, instead, have been reported as 'pattern direction selective'. Therefore, MT seems to be the area in which the aperture problem is solved. However, the receptive fields of MT and MST neurons are very large, and the representation of optical flow information in MT and MST must be fundamentally different from the mathematical notion of planar vector field. The analysis of § 2 and § 3 suggests that the first-order spatial properties of optical flow, instead of the optical flow itself, are likely candidates for the representation of motion information at the level of the areas MT and MST. This is because the first-order spatial properties of the motion field of rigid objects are almost piecewise constant over rather large regions of the field of view and can thus be conveniently described by cells with large receptive fields (see § 2 and figure 1).

These arguments are consistent with the hypothesis that the visual areas MT and MST build up a representation of first-order properties of optical flow but do not compute optical flow in a mathematical sense.

8. THE ANALYSIS OF OPTICAL FLOW IN THE BRAIN

Recent electrophysiological experiments (Tanaka & Saito 1989; Andersen *et al.* 1990, 1991; Lagae *et al.* 1991; Duffy & Wurtz 1991*a*,*b*; Lagae 1991) have shown that units in area MST are tuned to expanding and rotating stimuli. In this section, the main properties of these units are summarized. Then, a simple model of MST cells, which is consistent with anatomy and electrophysiology, is discussed in detail.

(*a*) *Properties of MST neurones*

Many neurons of area MST, which seems to be devoted to motion analysis, are very good candidates for the processing of optical flow information in the brain. In what follows, the main properties of these neurons are listed.

1. Receptive field size: the size of the receptive field of MST units is rather large, ranging from 30° to as much as 100° (Tanaka & Saito 1989), and does not increase with eccentricity (Tanaka & Saito 1989).
2. Stimulus selectivity: MST units appear to be selectively tuned to rotation, expansion, shear, and combinations of these stimuli (Andersen *et al.* 1990). Furthermore, many of these neurons are also sensitive to translation in a given direction, suggesting a continuum of response selectivity (Duffy & Wurtz 1991*a*,*b*).
3. Position invariance: the response and selectivity of MST units to a given stimulus does not change appreciably when the stimulus is displaced with respect to the centre of the receptive field (Lagae *et al.* 1991; Duffy & Wurtz 1991*b*; Lagae 1991). The response amplitude of a unit tuned to expansion, for example, decreases by about 50% for a relative displacement of 30°.
4. Nonlinearity: MST cells do not extract an optical flow component from a complex stimulus (Lagae 1991; Orban *et al.* 1992). A unit tuned to rotation, for example, decreases its response when an expansion of increasing strength is added to a rotating stimulus.

Properties 1, 2, and 3 suggest that these units detect first-order properties of the optical flow, almost in the mathematical sense of § 3. A neuron with a response highly dependent on the position of the stimulus across the receptive field can be thought of as detecting a singular point (e.g. a focus of expansion). Instead, a neuron whose response is highly position invariant appears to be tuned to a elementary flow component (e.g. expansion). Interestingly, it seems that the brain analyses optical flow in a way which does not use the decomposition of a complex flow into the linear superposition of simple components. This is also confirmed by the existence of MST units which are selectively tuned to spirals (Andersen *et al.* 1990). Let us now discuss a simple model of the organization of the receptive field of cells in MST which takes into account the properties above.

(*b*) *Modelling MST neurons*

This model has been inspired by a previous model suggested by Tanaka *et al.* (1989). It is assumed that

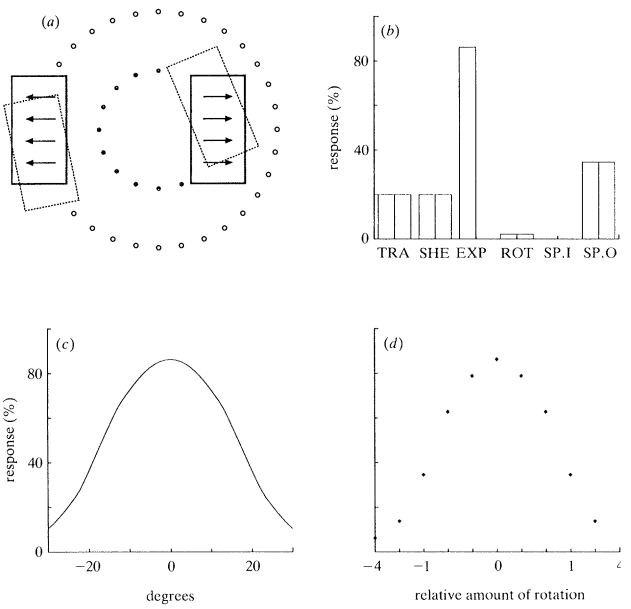


Figure 8. Model of the organization of the receptive field of an MST neuron tuned to expansion. (*a*) The neuron is receiving inputs from 16 MT-like units located on an inner ring (10° of radius) and 32 MT-like untis on an outer ring (20° of radius). The rectangular frames denote the spatial extension of some of these units, and the arrows indicate the unit (and subunit) preferred direction. (*b*) Stimulus selectivity. Normalized responses of the model cell to pure translation (TRA), shear (SHE), expansion (EXP), rotation (ROT), contraction and rotation (SP.I), and expansion and rotation (SP.O), in the forward and reversed direction, respectively. (*c*) Position invariance. Normalized response to an expanding stimulus as a function of the displacement between the receptive field centre and the focus of expansion. (*d*) Non-linear behaviour. Normalized responses to a stimulus composed of expansion and clockwise rotation (−) or anticlockwise rotation (+).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the input to these cells is composed of units with properties similar to those of cells in area MT. These units, or MT-like units, have a rectangular receptive field of about $10° \times 20°$, and are directionally selective (Maunsell & Van Essen 1983). Each of these units is composed of several identical and equally spaced subunits (probably V1 cells) with the same directional selectivity. The spacing between subunits is assumed to be 1°.

If ϕ is the angle between the optical flow direction and a fixed axis at the location $\boldsymbol{x}$ on the receptive field of the unit u, the response R_u is assumed to be

$$R_u = \sum_{\text{subunits}} f(\phi - \phi_p),$$

where the summation extends to all the subunits of u, ϕ_p (the angle between the preferred direction p and the same fixed axis), and $f(\cdot)$ is a tuning function of the type

$$f(y) = \exp(-y^2/\sigma^2) \qquad (11)$$

where σ is about 30°. Finally, the cell response, R_{cell}, is simply

$$R_{\text{cell}} = \Theta\left(\sum_{\text{units}} R_u\right),$$

with $\Theta(\cdot)$ being a suitable threshold function. In the computer experiments, the function Θ was a logistic function.

The behaviour of the model cell depends primarily on two factors: (i) the spatial organization of the units within the receptive field of the cell; and (ii) the 'shape' of the tuning function.

The type of selectivity of the cell is essentially a result of the first factor, that is, the arrangement of the MT-like units within the receptive field, whereas the extent to which the properties of selectivity, position invariance, and nonlinearity hold depends primarily on the second factor. A tuning function of the type of equation (11), in agreement with experimental evidence, produces a response which is moderately position invariant and shows a clear nonlinear behaviour (Orban *et al.* 1992). A tuning function with negative side lobes (ideally a 'cosine' function), however, produces an almost perfectly position-invariant response and a linear behaviour (Poggio *et al.* 1991).

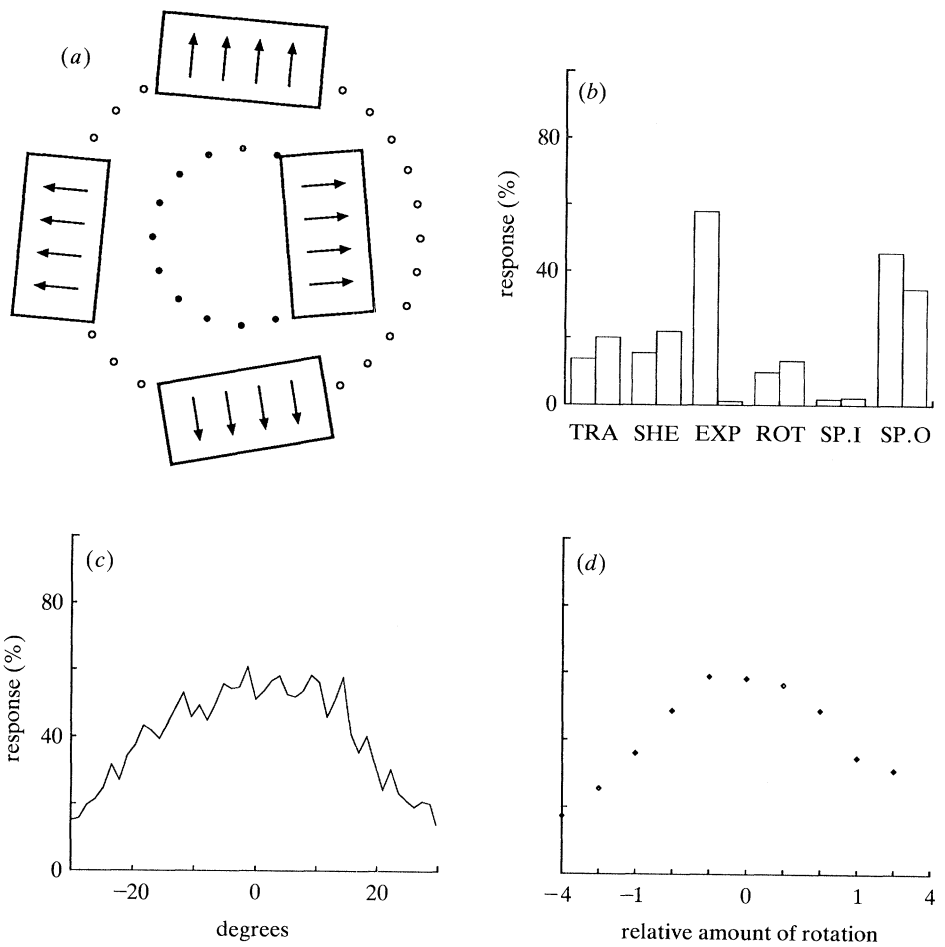


Figure 9. The proposed model does not depend critically on the exact distribution of the unit and subunit preferred direction. (*a*) The preferred direction of each unit of the model of figure 8*a* is perturbed by adding zero-mean white Gaussian noise with standard deviation equal to 15° (a randomly generated angle is added to each preferred direction). As a result, the preferred direction of each unit is not necessarily along the line which joins the unit centre with the cell centre. (*b*), (*c*), and (*d*) Stimulus selectivity, position invariance and nonlinear behaviour, respectively, for the model cell of (*a*).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Figure 8*a* illustrates the organization of the MT-like units in the receptive field of a model cell tuned to expansion. The MT-like units are located in two concentric rings (16 units in the inner and 32 in the outer ring) and the radius of each ring is 10° and 20°, respectively. The preferred direction of each unit is shown by the solid arrows. Figure 8*b* shows the response selectivity for large stimuli, which cover the entire receptive field, and obtained with the tuning function of equation (11). The responses are given in percentage of maximum response, normalized to unity, and both the preferred and reversed direction of motion are shown (i.e. preferred–null direction of translation, expansion–contraction; clockwise–anti-clockwise rotation, etc.). Figure 8*c* illustrates the property of position invariance when the focus of expansion is displaced over the receptive field. Finally, figure 8*d* reproduces the change of the response when a rotation of increasing intensity is added to a fixed amount of expansion. The parameters of the threshold function were chosen to closely match the behaviour of MST cells, as described by Orban *et al.* (1992), particularly with respect to position invariance and nonlinearity. Unlike a previous model (Tanaka *et al.* 1989), the property of position invariance is obtained without the repetition of similar arrangements of neurons.

The regular arrangement of units within the receptive field of figure 8*a* is not critical for the model. The addition of random noise in the unit preferred direction (see figure 9*a*) does not affect either of the properties of selectivity, position invariance, or non-linearity, as can be clearly inferred from figure 9*b*,*c*,*d*, respectively.

A cell tuned to expansion and translation (Duffy & Wurtz 1991*a*) can be modelled by adding a 'bias' along the preferred direction of translation to each subunit of a cell tuned to expansion. The preferred direction of each unit is thus obtained as a weighted average between the preferred direction of translation and the radial direction. Figure 10*a* shows an example in which the model cell is obtained by adding a 'bias' in the horizontal direction from left to right to the model cell of figure 8. The model cell of figure 10*a* is tuned to expansion and horizontal translation (see figure 10*b*) and still has a remarkable degree of position invariance (see figure 10*c*) and nonlinear behaviour (see figure 10*d*).

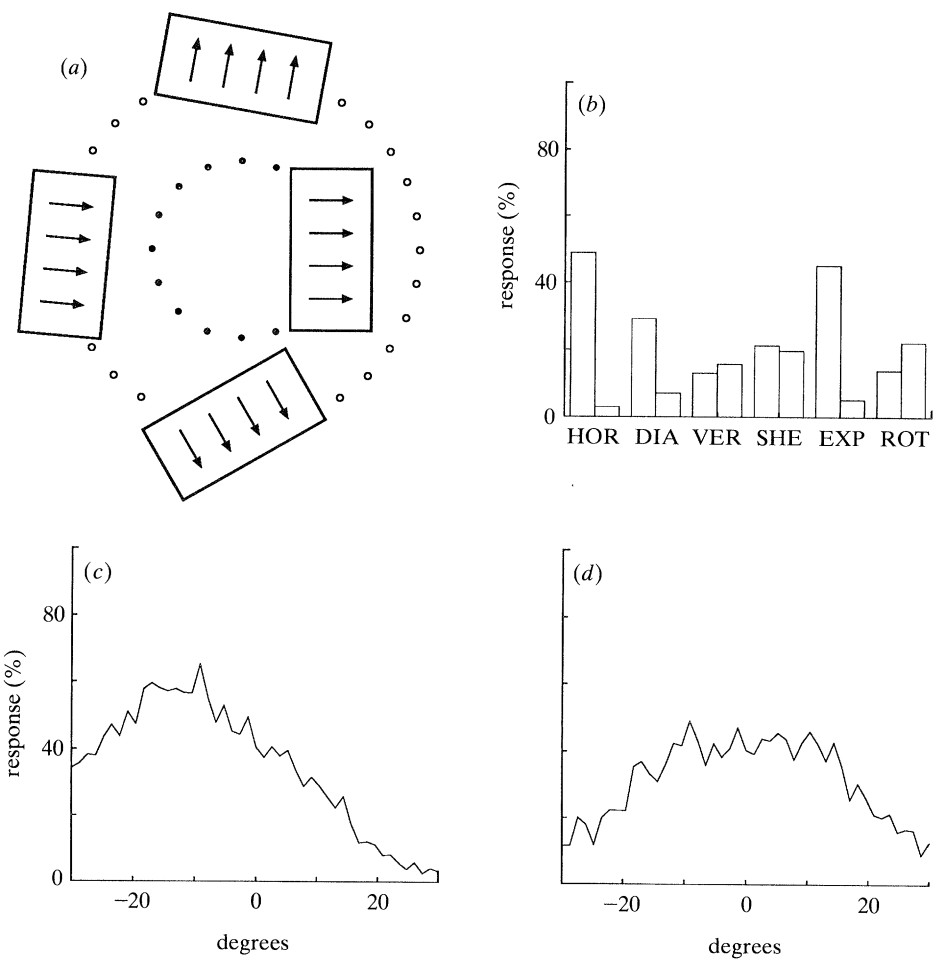


Figure 10. Model of the organization of the receptive field of an MST neuron tuned to expansion and horizontal translation. (*a*) At each unit of the model cell, the preferred direction is obtained as a weighted average between the horizontal (from left to right) and the radial direction. In the experiments, equal weight was used and random noise was added as in the model cell of figure 9. (*b*), (*c*), and (*d*) Stimulus selectivity, position invariance and non-linear behaviour, respectively, for the model cell of (*a*).

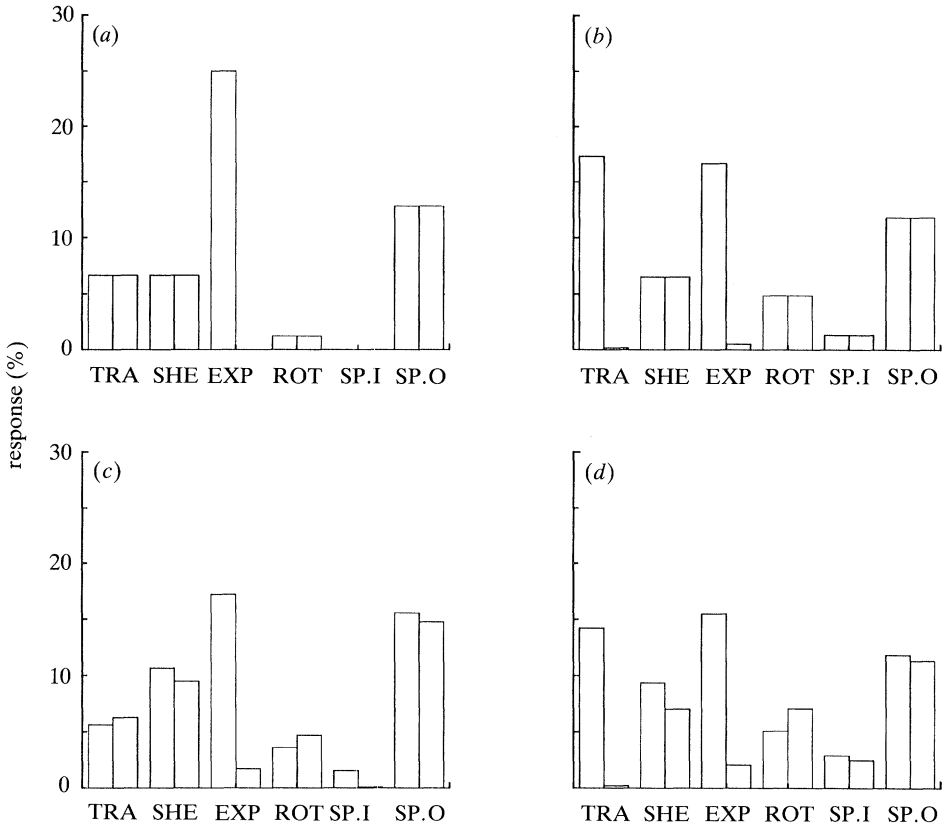


Figure 11. Selectivity and position invariance in the presence of smaller stimuli. (*a*) The response of the cell of figure 8*a* to stimuli of 30° of radius centred on the receptive field. (*b*) The response of the cell of figure 10*a* to the same stimuli. (*c*) The response of the cell of figure 8*a* to stimuli displaced by 30° with respect to the receptive field centre. (*d*) The response of the cell of figure 10*a* to the same stimuli.

To conclude, the selectivity and position invariance illustrated in figures 8 and 10 were obtained with large stimuli completely covering the receptive field of the cell. However, even if the cells of figures 8 and 10 are tested with relatively small stimuli, the cell selectivity and the property of position invariance do not significantly change when the stimuli move across the receptive field. This is illustrated in figure 11. Figure 11*a*,*b* shows the stimulus selectivity of the model cells of figures 8*a* and 10*a*, respectively, when the stimuli of 30° of diameter are on the centre of the receptive field. Figure 11*c*,*d* shows the responses which are obtained when the same stimuli are displaced by 10° with respect to the receptive field centre.

9. DISCUSSION

The aim of this paper was to analyse some computational aspects of motion perception in biological and computer vision systems. Let us summarize the results obtained and discuss their relevance to the understanding of visual motion.

A mathematical analysis has shown that the spatial structure of optical flow is usually rather simple because, in most cases, a piecewise linear vector field is sufficient to approximate optical flow over large areas of the visual field. This observation has a few computational and biological implications. Firstly, the estimation of optical flow from the time-varying image brightness is not a difficult problem, and many different (and simple) methods for the computation of optical flow produce similar results. Secondly, because the first-order spatial properties of optical flow are almost piecewise constant, a coarse representation of these properties is usually adequate. Therefore, the increasingly large size of the receptive fields of the cells of the viual areas MT and MST, which seem dedicated to motion analysis, suggests the hypothesis that first-order properties of optical flow (like elementary flow components), and not the optical flow itself, are computed and represented in the processing of visual motion.

The segment of motion perception which is discussed in this paper can be described as the perception of coherent motion, which can best be analysed with units with a large receptive field. The perception of the motion of a small spot of light, which avoids the aperture problem, can be obtained with units with a small receptive field, such as those in the visual area V1. As a consequence, it is likely that motion perception may require the combined use of units with a small receptive field (as those in V1) and units with large receptive fields (as those in V5), as suggested by the mutual connections between these two areas (Shipp & Zeki 1989*a*,*b*).

The characteristic properties of stimulus selectivity and position invariance of MST cells has been explained by means of a very simple arrangement of MT-like

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

units in the receptive field of these cells. According to the model, the nonlinear behaviour of MST cells, which are tuned to specific elementary flow components but do not extract these components from complex flows, primarily originates from the shape of the tuning function (which controls the response in the preferred and null direction of the MT-like units). Consequently, it is possible to explain the behaviour of MST cells in terms of the direction mosaic hypothesis (Tanaka *et al.* 1989), which is in agreement with the properties of MT cells, the main input to MST cells.

To conclude, it is evident that, through the analysis of first-order properties of optical flow, natural and artificial vision systems seem to rely on very similar tools for the processing of visual motion. This is hardly surprising because the physical properties of image formation and the data available to both of the systems are almost identical. An interesting difference, however, can be found when looking more closely at the analysis of the elementary components of optical flow. Artificial vision systems can be designed to exploit linear algebra and the Helmholtz theorem (an arbitrary motion can be obtained by means of a suitable linear combination of a translation with four elementary components). In the brain, instead, MST cells do not behave as linear detectors, but as nonlinear detectors of the AND type. This computationally less-efficient strategy adopted by the brain can probably be explained in terms of the evolutionary need to differentiate between different kinds of stimuli quickly to detect and avoid danger.

We thank H. Barlow and H. Hildreth, for very useful comments on the manuscript, and T. Poggio and G. A. Orban for very fruitful discussions. This work was partly supported by grants from the EEC (ESPRIT II VOILA), ESPRIT B.R.A. Insight Project 3001, EEC BRAIN Project No. 88300446/JUI, Progetto Finalizzato Trasporti PROMETHEUS and Progetto Finalizzato Robotica. Clive Prestt checked the English.

REFERENCES

Adelson, E.H. & Bergen, J.R. 1985 Spatiotemoral energy models for the perception of motion. *J. opt. Soc. Am. A* **2**, 284–299.

Andersen, R.A., Snowden, R.J., Treue, S. & Graziano, M. 1990 Hierarchical processing of motion in the visual cortex of monkey. In *The brain.* (*Cold Spring Harb. Symp. quant. Biol.* **55**), pp. 741–748. Cold Spring Harbor Laboratory Press.

Andersen, R.A., Graziano, M. & Snowden, R.J. 1991 Selectivity of area MST neurons for expansion/contraction and rotation motions. *Invest. Ophthal. vis. Sci.* **32**, 823.

Barlow, H.B. & Levick, R.W. 1965 The mechanism of directional selectivity in the rabbit's retina. *J. Physiol., Lond.* **173**, 477–504.

Bertero, M., Poggio, T. & Torre, V. 1988 Ill-posed problems in early vision. *Proc. Inst. elect. Electron Engrs* **76,** 869–889.

Boussaud, D., Ungerleider, L.G. & Desimone, R. 1990 Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in he macaque monkey. *J. comp. Neurology* **296**, 462–495.

Campani, M. & Verri, A. 1990 Computing optical flow from an overconstrained system of linear algebraic equations. In *Proc. Third Int. Conf. Comput. Vis.*, Osaka (ed. Makoto Nagao), pp. 22–26. Los Alamitos: IEEE Computer Society Press.

Campani, M. & Verri, A. 1992 Motion analysis from first order properties of optical flow. In *Comput. Vis. Graph. Image Proc.: Image Understanding*. (In the press.)

De Micheli, E., Uras, S. & Torre, V. 1992 The accuracy of the computation of optical flow and of the recovery of motion parameters. *IEEE Trans. Patt. Anal. Machine Intell.* (In the press.)

De Yoe, E.A. & Van Essen, D.C. 1988 Concurrent processing streams in monkey visual cortex. *Trends Neurosci.* **11**, 219–226.

Duffy, C.J. & Wurtz, R.H. 1991 Sensitivity of MST neurons to optic flow stimuli I: a continuum of response selectivity to large field stimuli. *J. Neurophysiol.* **65**, 1329–1345.

Duffy, C.J. & Wurtz, R.H. 1991 Sensitivity of MST neurons to optic flow stimuli II: mechanisms of response selectivity revealed by small field stimuli. *J. Neurophysiol.* **65**, 1346–1359.

Faugeras, O.D., Lustman, F. & Toscani, G. 1987 Motion and structure from motion from point and line matches. *Proc. First Int. Conf. Comput. Vis.*, London (ed. J. M. Brady and A. Rosenfeld), pp. 25–34. Washington: Computer Society Press of the IEEE.

Fenmena, C.L. & Thompson, W.B. 1979 Velocity determination in scenes containing several moving objects. *Comput. Vis. Graph. Image Proc.* **9**, 301–315.

Francois, E. & Bouthemy, P. 1990 The derivation of qualitative information in motion analysis. *Image Vis. Comput. J.* **8**, 279–287.

Gibson, J.J. 1950 *The perception of the visual world.* Boston: Houghton Mifflin.

Haralick, R.M. & Lee, J.S. 1983 The facet approach to optic flow. In *Proceedings Image Understanding Workshop*, Arlington (Virginia) (ed. L. S. Baumann), pp. 84–93. Arlington. Science Applications International Corporation Report Number SAIC-83/940.

Hassenstein, B. & Reichardt, W. 1956 Systemtheoretische Analyse der Zeit-, Reihenfolgen and Vorzeichenauswertung bei der Bewgungsperzeption der Russelkafers. *Chlorophanus. Z. Naturf.* **IIb**, 513–524.

Helmholtz, H. 1858 Uber Integrale der hydrodynamischen Gleichungen welche den Wirbelwegungen entsprechen. *Crelles J.* **55**, 25.

Hildreth, E.C. 1984 The computation of the velocity field. *Proc. R. Soc. Lond.* B **221**, 189–220.

Hirsch, M.W. & Smale, S. 1974 *Differential equations, dynamical systems, and linear algebra.* New York: Academic Press.

Horn, B.K.P. & Schunck, B.G. 1981 Determining optical flow. *Artif. Intell.* **17**, 185–203.

Koenderink, J.J. & van Doorn, A.J. 1975 Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta* **22**, 773–791.

Lagae, L. 1991 A neurophysiological study of optic flow analysis in the monkey brain. Ph.D. thesis, Leuven, Belgium.

Lagae, L., Xiao, D., Raiguel, S., Maes, H. & Orban, G.A. 1991 *Invest. Ophthal. vis. Sci.* **32**, 823.

Livingstone, M. & Hubel, D. 1988 Segregation of form, color, movement, and depth. Anatomy, Physiology and perception. *Science, Wash.* **240**, 740–749.

Longuet-Higgins, H.C. 1984 The visual ambiguity of a moving plane. *Proc. R. Soc. Lond.* B **223**, 165–175.

Longuet-Higgins, H.C. & Prazdny, K. 1981 The interpretation of moving retinal images. *Proc. R. Soc. Lond.* B **223**, 165–175.

Marr, D. & Ullman, S. 1981 Directional selectivity and its use in early vision processing. *Proc. R. Soc. Lond.* B **211**, 151–180.

Maunsell, J.H.R. & Van Essen, D.C. 1983 Functional properties of neurons in middle temporal visual area of the macaque monkey I: selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* **49**, 1127–1147.

Movshon, J.A., Adelson, E.H., Gizzi, M.S. & Newsome, W.T. 1985 The analysis of moving visual patterns. *Pont. Acad. Scient. Scipta Varia* **54**, 117–151.

Nagel, H.H. 1983 Displacement vectors derived from 2nd order intensity variations in image sequences. *Comput. Vis. Graph. Image Proc.* **21**, 85–117.

Negahdaripour, S. & Horn, B.K.P. 1987 Direct passive navigation. *IEEE Trans. Patt. Anal. Machine Intell.* **9**, 168–176.

Orban, G.A., Lagae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H. & Torre, V. 1992 First order analysis of optical flow in monkey brain. *Proc. natn. Acad. Sci. U.S.A.* (In the press.)

Poggio, T., Little, J. & Gamble, E. 1986 Parallel optical flow. *Nature, Lond.* **301**, 375–378.

Poggio, T., Verri, A. & Torre, V. 1991 Green theorems and qualitative properties of the optical flow. *MIT AI Lab. Memo 1289.*

Rognone, A., Campani, M. & Verri, A. 1992 Identifying multiple motions from optical flow. In *Proc. 2nd European Conference on Computer Vision*, S. Margherita (ed. G. Sandini) (*Lecture Notes in Computer Science* **588**), pp. 258–266. Berlin: Springer Verlag.

Saito, H., Yukio, M., Tanaka, K., Hikosaka, K., Fukada, Y. & Iwai, E. 1986 Integration of direction signals of image motion in superior temporal sulcus of the macaque monkey. *J. Neurosci.* **6**, 145–157.

Shipp, S. & Zeki, S. 1989*a* The organization of connections between areas V5 and V1 in macaque monkey visual cortex. *Eur. J. Neurosci.* **1** (4), 309–332.

Shipp, S. & Zeki, S. 1989*b* The organization of connections between areas V5 and V2 in macaque monkey visual cortex. *Eur. J. Neurosci.* **1** (4), 333–354.

Tanaka, K., Fukada, Y. & Saito, H. 1989 Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62**, 642–656.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y. & Iwai, E. 1986 Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J. Neurosci.* **6**, 134–144.

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

Tichonov, A.N. & Arsenin, V. 1977 *Solution of ill-posed problems*. New York: John Wiley.

Torre, V. & Poggio, T. 1978 A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond.* B **202**, 409–416.

Tretiak, O. & Pastor, L. 1984 Velocity estimation from image sequences with second order operators. In *Proc. 7th Int. Conf. on Patt. Recognition*, Montreal, pp. 16–19. Los Angeles: IEEE Computer Society Press.

Ungerleider, L.G. & Desimone, R. 1986 Cortical connections of visual area MT in the macaque. *J. comp. Neurol.* **248**, 190–222.

Uras, S., Girosi, F., Verri, A. & Torre, V. 1988 A computational approach to motion perception. *Biol. Cybern.* **60**, 79–87.

Verri, A. & Aicardi, F. 1990 Limit cycles of the two-dimensional motion field. *Biol. Cyber.* **64**, 141–144.

Verri, A. Girosi, F. & Torre, V. 1989 Mathematical properties of the two-dimensional motion field: from singular points to motion parameters. *J. opt. Soc. Am.* A **6**, 698–712.

Verri, A., Girosi, F. & Torre, V. 1990 Differential techniques for optical flow. *J. opt. Soc. Am.* A **7**, 912–922.

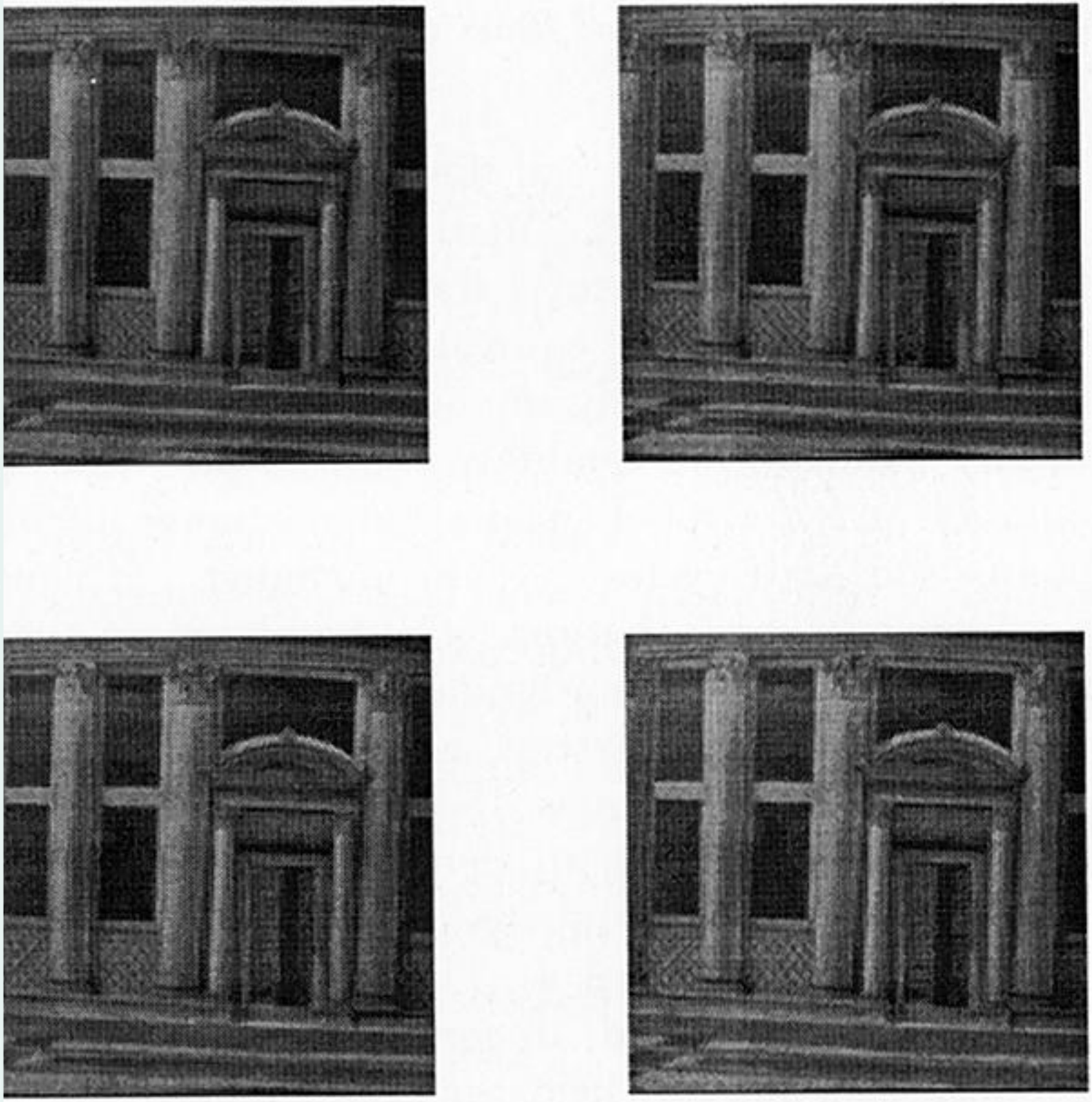
Verri, A. & Poggio, T. 1989 Motion field and optical flow: qualitative properties. *IEEE Trans. Patt. Anal. Machine Intell.* **11**, 490–498.

Zeki, S.M. 1974 Functional organisation of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *J. Physiol., Lond.* **236**, 549–573.

Zeki, S. & Shipp, S. 1988 The functional logic of cortical connections. *Nature, Lond.* **335**, 311–317.

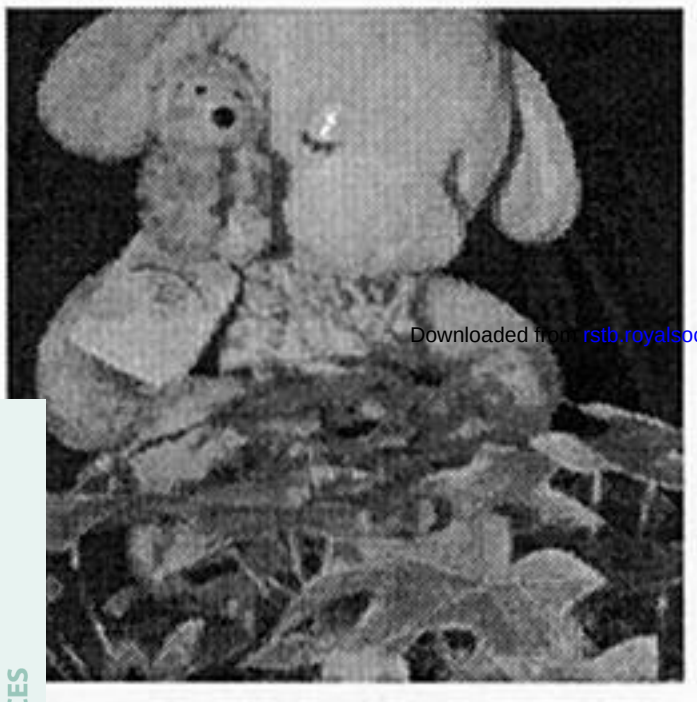
Received 14 February 1992; accepted 20 March 1992

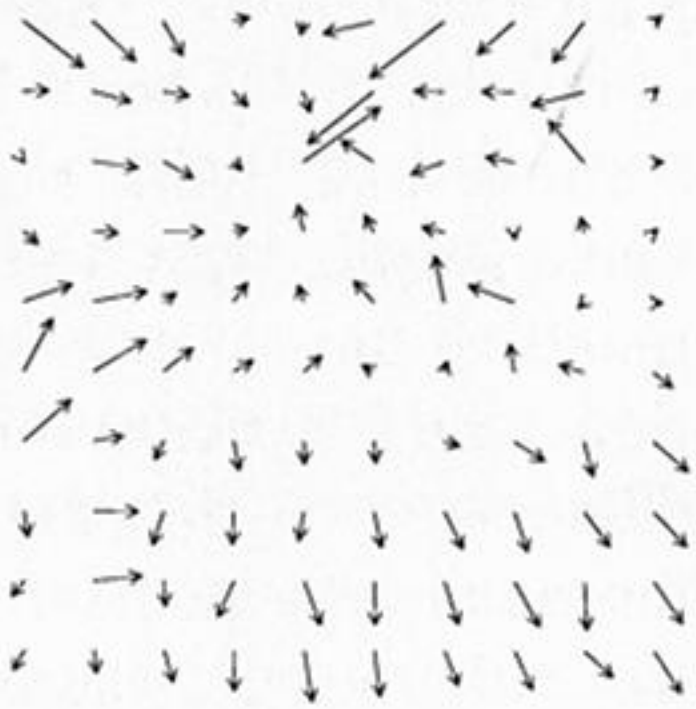
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

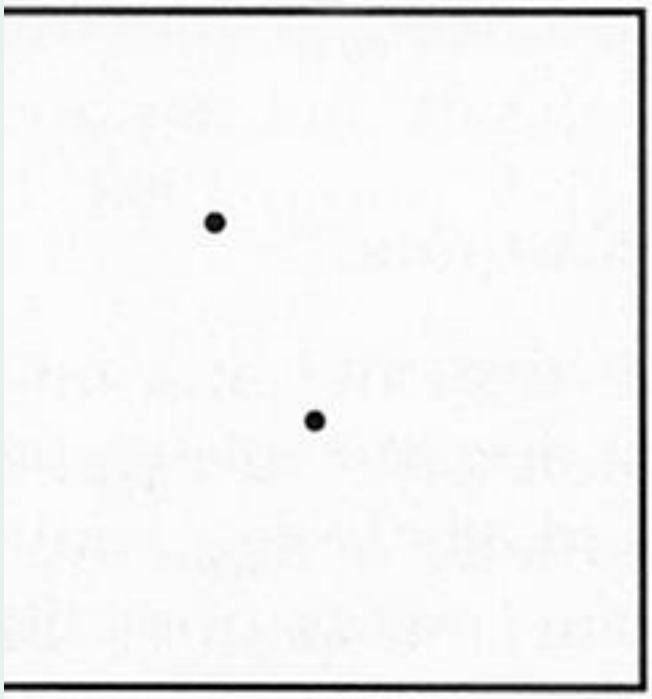


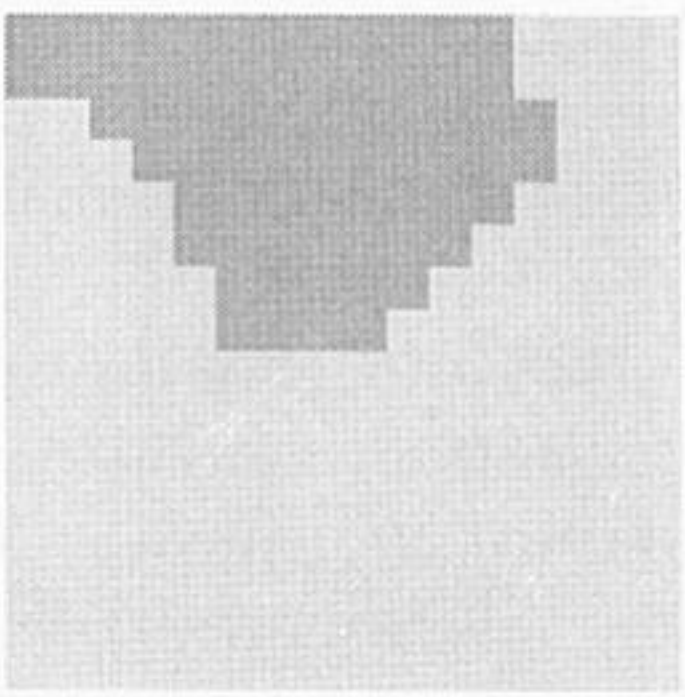
gure 2. Four frames of a sequence (from upper left to wer right) taken while the viewing camera is translating ong the optical axis toward a picture posted on the wall. nages were grabbed by means of a PULNIX TM46 mera and an Imaging Technology board FG100. Each nage consists of 256 × 256 pixels.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

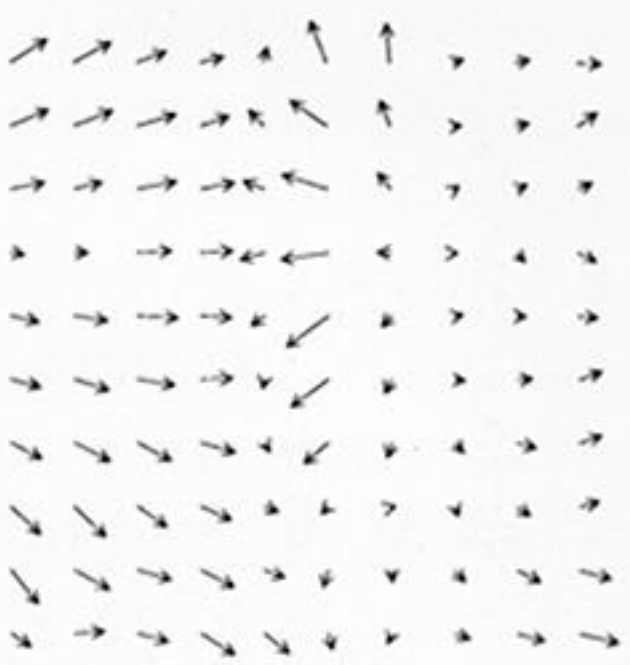


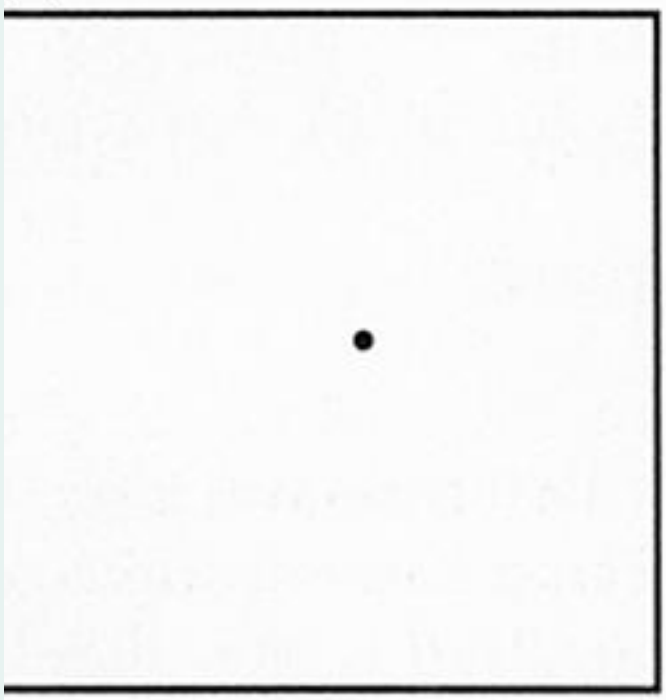


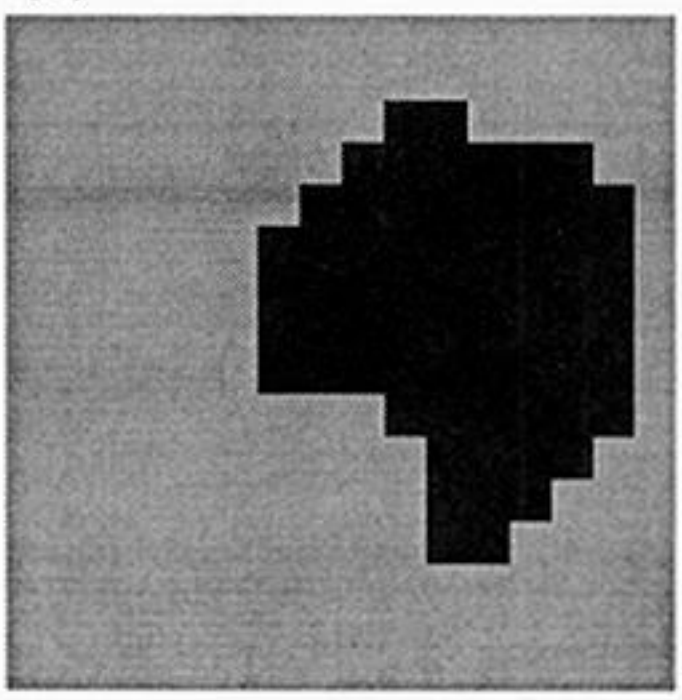




gure 5. Detection of relative motion. (*a*) A frame of a quence in which the puppets are moving away from the mera, while the camera was translating towards the plant.) The optical flow associated with the frame of (*a*) and mputed through the spatial gradient constancy method Jras *et al.* 1988). (*c*) The two singular points which have en detected in the optical flow of (*b*). (*d*) The motion gmentation which has been obtained through a technique scribed Rognone *et al.* (1992). The darker area indicates a gion of contraction and corresponds to the puppets otion, the lighter area a region of expansion and corre- onds to the rest of the viewed scene.







igure 6. A second example of detection of relative motion.) A frame of a sequence in which a person was moving wards the camera, while the camera was translating ward the scene. (*b*) The optical flow computed by using a iape constraint method (Verri & Campani 1990). (*c*) The ngular points (both of them are foci of expansion) of the otical flow of (*b*). The focus of expansion which corre- onds to the camera motion lies just outside the image oundaries. (*d*) The segmentation of the optical flow of (*b*). he darker area corresponds to the moving person, whereas e lighter area identifies the part of the scene which moves gidly with respect to the viewing camera.