Pergamon

0042-6989(95)00192-1

Vision Res., Vol. 36, No. 6, pp. 869–887, 1996

Printed in Great Britain.
0042-6989/96 $15.00 + .00

Contribution of Area MT to Perception of Three-dimensional Shape: a Computational Study

GIEDRIUS T. BURAČAS,*† THOMAS D. ALBRIGHT*

Received 9 March 1995; in revised form 11 July 1995

Successful recognition and manipulation of objects in one's visual environment is critically dependent upon the ability to recover three-dimensional (3D) surface geometry from two-dimensional (2D) retinal images. The relative motion of image features, caused by relative displacement of object and observer, has characteristic properties that betray components of the 3D source geometry (distance, tilt, slant and curvature) and is among the most valuable sources of information used for 3D surface recovery by the primate visual system. We have considered the behavior of motion-sensitive neurons in primate visual cortex and found that their properties closely resemble those of differential motion operators that can be used to formally characterize the 3D shape of a smooth moving surface. Our analysis has led us to identify a set of three orders of filters for differential motion detection. These filters behave in a manner that is strikingly similar to the spatial and velocity tuning profiles of a sub-population of neurons—those possessing antagonistic motion surrounds—in the middle temporal visual area (MT). On the basis of this analysis, we suggest that MT neurons subserve 3D surface recovery from relative motion cues.

Shape-from-motion Receptive field surrounds Area MT Neural computing Differential motion filters

INTRODUCTION

The primate visual system affords effective behavioral interaction with a three-dimensional (3D) world. This ability is contingent upon recovery of the 3D spatial relationships between objects and the shapes of their surfaces from the two-dimensional (2D) spatio-temporal properties of retinal images. Multiple 2D image properties, including shading, motion and binocular positional disparity contain implicit information about environmental structure and are available as cues for recovery of that structure. It has long been recognized (von Helmholtz, 1867; Gibson, 1950) that the relative motion of retinal image features is among the richest of such cues. Of particular value is the pattern of retinal motion induced by the relative movement of an observer and a rigid environmental object/surface [i.e. either the observer moves or there is a moving object(s) present in his/her visual field]—a pattern that is commonly termed 'optical flow.' Brain mechanisms responsible for 3D surface recovery from optical flow (i.e. 'shape-from-motion' mechanisms) are suggested by both neurophysiological data and computational considerations. Our goal has been to establish that motion-sensitive neurons in the middle temporal visual area (area MT) of primate cortex are endowed with computational properties that could allow them to estimate shape-related parameters from optical flow.

Many computational approaches to the shape-from-motion (SFM) problem have as a foundation the fact that the 3D surface of an environmental object can be characterized or represented by standard geometrical descriptors, such as tilt, slant and curvature. Although such descriptors are formally computable from the first- and second-order spatial derivatives of the surface function, the projective formation of the retinal image renders the explicit surface function unavailable. Because in most cases the velocity vector field is proportional (to a first approximation) to the surface function, it is nonetheless possible to compute geometrical shape descriptors by applying differential operators to the flow field itself. The first-order properties of the optic flow and the possibility that the primate visual system has adopted a strategy for estimating these properties have been subjects of study for many years. It has, for example, been shown that the linear portion (i.e. the planar portion) of the velocity vector field can be parsed into independent components of local rotation, expansion/contraction and shear (Koenderink & Van Doorn, 1975). These components of optical flow allow affine estimation of surface orientation. The second-order components of the optic flow can be used to compute affine estimates of shape invariants of smooth surfaces, such as principal normal

* Vision Center Laboratory, The Salk Institute, P.O. Box 85800, San Diego, CA 92138–9216, U.S.A.
† To whom all correspondence should be addressed [*Email* giedrius @ salk.edu].

curvatures (Koenderink & van Doorn, 1992; Dijkstra, 1994).

Neurophysiological data thought to bear on this issue are mainly those providing evidence for neurons that detect differential motion within a receptive field. Of particular interest are neurons that possess *center-surround motion antagonism*. Such neurons typically respond optimally when direction/speed of motion in the central or 'classical' receptive field (CRF) differs from that present in a larger 'surround' receptive field (SRF). Neurons of this sort have been found to exist in the optic tectum of pigeons (Frost, 1978), area 17 of cat visual cortex (Hammond, 1981), and in the area MT of primate visual cortex (Allman, Miezin & McGuinnes, 1985; Tanaka, Hikosaka, Saito, Yukie, Fukada & Iwai, 1986; Born & Tootell, 1992; Xiao, Marcar, Raiguel & Orban, 1994). We suggest herein that the behavior of such neurons corresponds to differential motion filters of the type required for extraction of 3D shape descriptors from the optic flow field.

Of greatest relevance to the present discussion are neurons expressing center–surround interactions in area MT. This visual area is part of the parietal cortical stream (Ungerleider & Mishkin, 1982; Livingstone & Hubel, 1988), and it receives weighty projections from primary visual cortex (area V1) (Ungerleider & Mishkin, 1979) and from the magnocellular compartments of area V2 (DeYoe & Van Essen, 1985). By comparison with other cortical visual areas, MT possesses a high proportion of neurons that are selective for direction and speed of motion (Zeki, 1974; Albright, 1984). Considerable evidence indicates that this visual area is a principal component of the neural substrate for motion processing resulting in a veridical description of optical flow [see Albright (1993) for review].

Area MT was implicated in SFM by the apparent loss of this facility following selective cortical ablation (Siegel & Andersen, 1987). We have explored the possibility that the motion antagonism of receptive fields in area MT may be involved in SFM perception. A consideration of analytical solutions to the SFM problem, viewed in the light of neurophysiological data, has led us to identify response properties of MT neurons that are ideally suited to 3D surface recovery.

Unlike other work on neurophysiologically plausible SFM mechanisms, we utilize data about the spatial profiles of receptive fields in area MT. The study closest to our vantage point (Droulez & Cornilleau-Perez, 1990) offered a limited treatment of surface recovery from optic flow. By contrast, our model provides a rich set of descriptors for surface form. These descriptors can subserve the estimation of relative distance, slant, and curvature. In addition, an early formulation of our model (Buračas & Albright, 1994) predicted the existence of receptive fields with oriented inhibitory lobes, which were subsequently found to be characteristic of many MT neurons (Xiao, Marcar, Raiguel & Orban, 1994).

The paper is organized in the following way: we begin by formalizing the physical relationship between 3D surface shape and the attendant velocity flow field. We then develop a simplified treatment of the SFM problem, in which emphasis is placed upon surface characterization using geometric shape descriptors (tilt, slant and curvature). From an abstract computational perspective, there is considerable flexibility in the means by which shape descriptors can be computed. Because our interest is limited to biologically plausible mechanisms, we have used known properties of cortical neurons to constrain the possibilities. To that end, we review the response properties of motion-sensitive neurons in primate cortical visual area MT, and we use those properties to define a set of idealized motion filters for surface characterization. We commence by considering neurons having *radially symmetric* surrounds and we develop differential motion filters that capture essential features of the neurons. The subsequent section is dedicated to a discussion of the computational advantages afforded by *oriented* receptive field surrounds, which have recently been discovered to exist in area MT (Xiao *et al.*, 1994).

We then proceed with a detailed examination of the properties of the selected motion filters and develop a model of surface characterization based upon these operators. Again, we begin with analysis of the differential-geometric properties of the simpler radially symmetric filter. We then switch to oriented filters and discuss computational schemes whereby outputs of these filters can be combined to render surface orientation and invariant shape descriptors. The model is compared with alternative neurophysiologically plausible approaches to the SFM problem. We close with a demonstration of the performance of our model when provided with realistic input.

RELATIVE MOTION AS A CUE FOR RELATIVE DEPTH

In this section we show that the retinal velocity field caused by translational displacement of a 3D surface constitutes a good approximation to the surface function itself. The zeroth-, first- and second-order structure of the ensuing velocity vector field for an arbitrary motion of a smooth surface was analyzed in detail by Koenderink and van Doorn (1992). In contrast to these authors, in order to make our arguments clear, we will constraint our analysis to the ecologically prevalent translational motion. If we assume an observer fixating a point on the passing surface, this type of motion can be decomposed into a translational component tangential to the direction of gaze, and rotation along the axis perpendicular to the direction of gaze (see Fig. 1). Later we will argue that the requirement of eye-tracking can be relaxed without consequences to our approach. Because, in the case of translational displacement, only the relative motion between an observer and the object matters, we will not distinguish between observer self-motion and object motion.

The velocity field caused by object–observer displace-

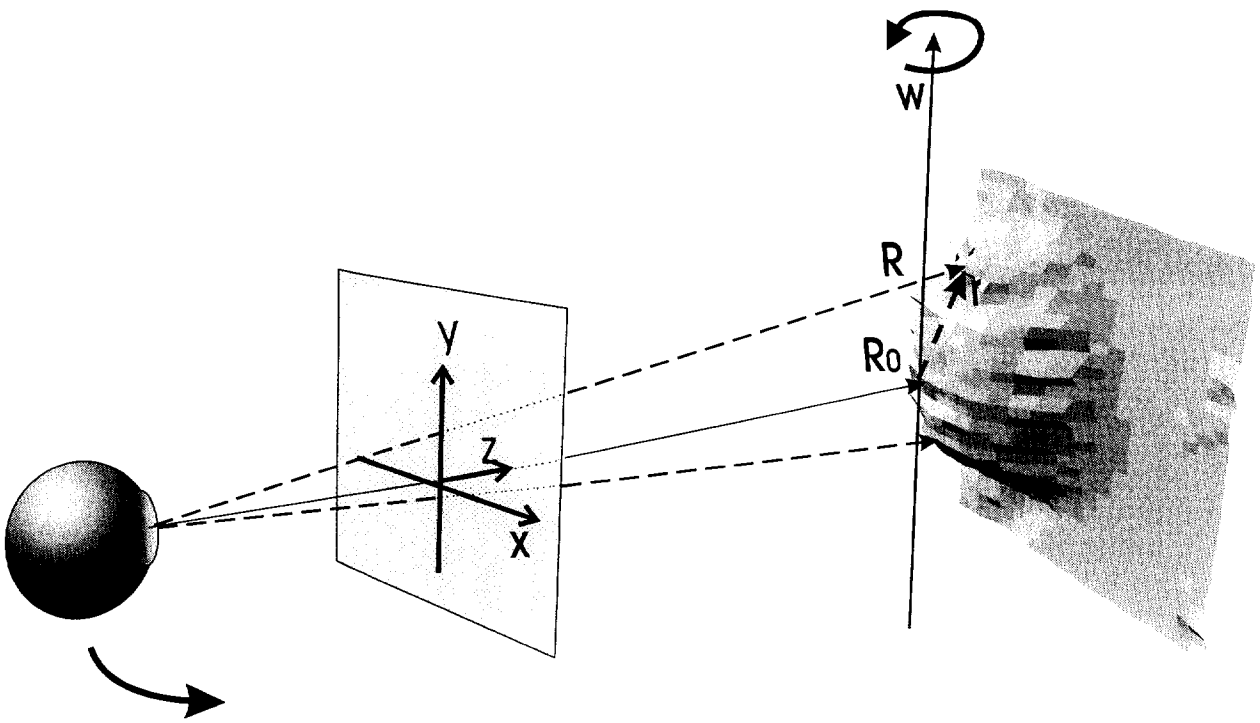

FIGURE 1. Coordinate system assumed in the model. Translation of an observer (represented by eye at left) with respect to a surface (**R**, right) containing the point of ocular fixation ($\mathbf{R}_0$) causes a pattern of retinal motion which can be decomposed into a rotational component (**w**) along the axis perpendicular the direction of sight and translation in depth [see equation (4)]. The retinal image projection is modeled by the fronto-parallel plane **X–Y**; the **Z**-axis is perpendicular to the fronto-parallel plane; **r** represents a surface positional vector in the coordinate system with the origin moved to the fixation point.

ment can then be expressed in terms of relative depth $z(x,y)$:

$$\mathbf{V} = -\mathbf{t} - \mathbf{w} \times (\mathbf{R} - \mathbf{R}_0) = -\mathbf{t} - \mathbf{w} \times \mathbf{r}, \quad (1)$$

where $\mathbf{w} = [w^x, w^y, 0]$ is the *effective rotation vector* (superscripts indicate vector components) of a surface $\mathbf{r} = [x,y,z(x,y)]$; **R** is a position vector of a surface point, $R_0 = [0,0,z_0]$ is a position vector of the fixation point; $\mathbf{t} = [0,0,t^z]$ is a translational component along the z-axis. (See Fig. 1 for an explanation of the coordinate system.)

The component velocities $v^x = u$ and $v^y = v$ of the velocity field under perspective projection can be calculated from (see Appendix for derivation):

$$u = -\frac{w^y z}{z_0} - \frac{-xt^z - w^x xy + w^y x^2}{z_0^2},$$
$$v = \frac{w^x z}{z_0} - \frac{-yt^z + w^y xy - w^x y^2}{z_0^2} \quad (2)$$

Here we assume $z_0 \gg z$, since, under natural viewing conditions, the distance to surface z_0 is typically much larger than distance variation on the surface of interest; this assumption permits replacement of $z_0 + z$ by z_0.

For an observer undergoing an arbitrary translation, the effective rotation vector is perpendicular to the direction of eye fixation. Without loss of generality, we can assume that $\mathbf{w} = [0,w^y,0] = [0,w,0]$, i.e. the observer is moving tangentially to ground. This assumption corresponds to rotation of the coordinate system around the z-axis (i.e. the axis of gaze) such that the y-axis becomes parallel to the vector **w**. The generality of our approach is supported by the fact that the representation of directions in area MT is approximately isotropic (Albright, Desimone & Gross, 1984), and we can always select a subset of orthogonally tuned neurons to represent velocity components corresponding to the rotated coordinate axes. Equation (2) then reduces to:

$$u = -\frac{w^y z}{z_0} - \frac{-xt^z + w^y x^2}{z_0^2},$$
$$v = -\frac{-yt^z + w^y xy}{z_0^2}. \quad (3)$$

When we take up the topic of specific motion operators (see below), it will be convenient to neglect the second-order term of equation (3), which is small compared to $1/z_0$ in the vicinity of the fixation point ($x,y \ll z_0$). Thus:

$$u = -wz/z_0 + O(1/z_0^2) \approx -z(x,y) \cdot w/z_0,\ v = O(1/z_0^2). \quad (4)$$

It is readily apparent from equation (4) that the horizontal velocity component u is proportional to relative distance $z(x,y)$. The latter can thus be recovered from the unidirectional velocity field up to the scaling factor z_0/w. This relation permits us to treat optic flow as a scalar direction—an approximation that represents the

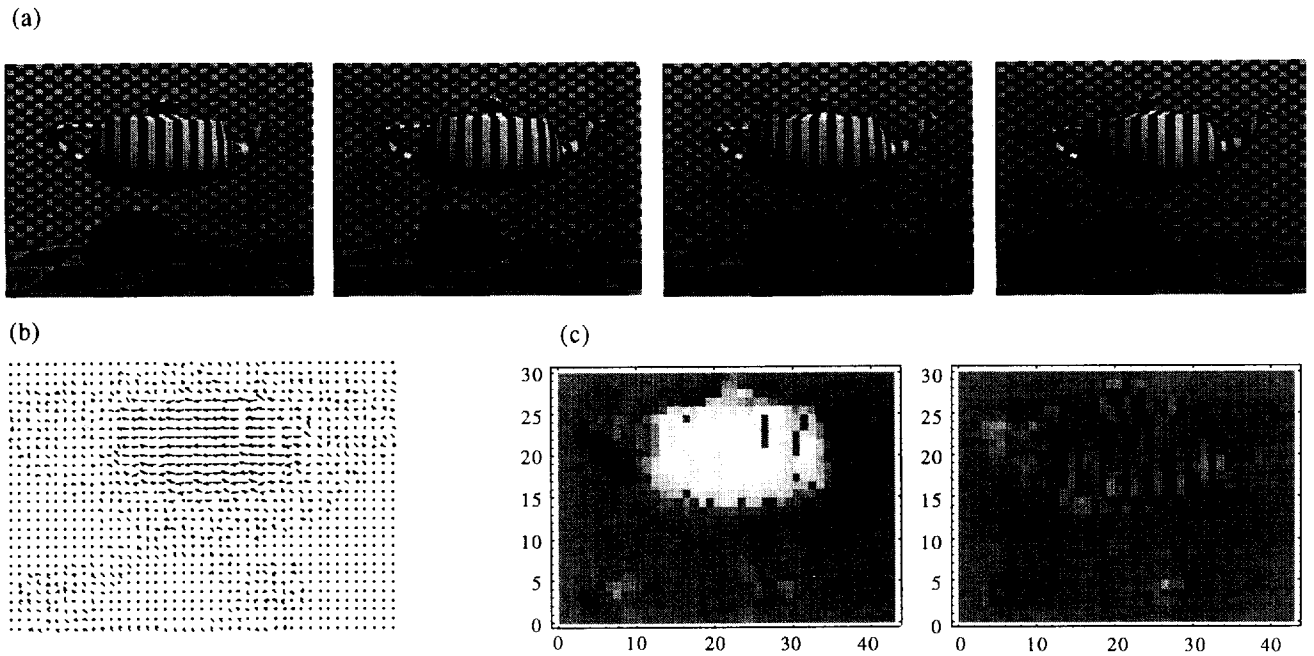

FIGURE 2. Optical flow produced by textured surfaces: (a) shows four frames of the movie of a rotating striped teapot used to create the flow field (b), which was calculated using an algorithm developed by Nagel (1987); (c) gray-scale images of horizontal (left) and vertical (right) components of the optic flow. Note that information about depth is found almost exclusively in the horizontal component.

retinal velocity field rather well within the central 20 deg. This point is emphasized in Fig. 2, which illustrates relative depth estimation for a realistic dynamic scene (a rotating teapot). Four frames of the rotating teapot are shown in Fig. 2(a). Teapot surface texture induces a flow field supporting dense sampling, which is essential for detecting small differences in velocities of neighboring surface points [Fig. 2(b)]. The flow field has been computed using a multi-scale velocity extraction algorithm developed by Nagel (1987). The $v^x = u$ and $v^y = v$ components of the velocity flow field are shown as gray-scale images in Fig. 2(c). Despite the deformation of the velocity field due to perspective projection, the component u, which is perpendicular to the vector of effective rotation w, carries nearly all shape-related information. By contrast, the v component is small and noisy, owing to imperfect estimation of the velocity field. Discarding the second-order terms in equation (4) effectively discounts relative depth information contained in the expansion/contraction component of the velocity field, caused by approaching/receding of a surface (velocity component t_z). This step can be justified by the fact that human observers do not appear to use this component in inferring shapes of moving objects (Hoffman, 1982; Ullman, 1979).

REDEFINING THE SHAPE-FROM-MOTION (SFM) PROBLEM

The shape-from-motion problem has been couched traditionally as a problem of recovering the 3D positions of features on the surface of an object, given geometric constraints provided by multiple, movement-related, 2D snapshots of the object (e.g. Ullman, 1979). While satisfactory in a formal computational sense, this approach fails to draw upon the power of spatio-temporally oriented filters in mammalian visual cortex, which compute motion energy in a continuous way (Adelson & Bergen, 1986). An alternative approach to the SFM problem (Longuet-Higgins & Prazdny, 1980; Koenderink & van Doorn, 1992), which co-opts the neurobiological legacy, relies instead upon the instantaneous values of the retinal velocity vector field. Retinal velocity field differentials, caused by differential motion of surface features, are used to compute standard descriptors for 3D shape. This approach appears neurobiologically sound, as the requisite local measurements of image velocity are known to be represented at an early cortical stage. It is, moreover, consistent with evidence indicating that human shape-from-motion perception involves a process of surface interpolation between moving surface features (e.g. Hussain, Treue & Anderson, 1989; Treue, Andersen, Ando & Hildredth, 1995). In light of these considerations, we find it useful to view the SFM problem as one of characterizing the interpolated surface. Our objective in this section is to formalize the derivation of geometric descriptors that can be used for surface characterization.

In order to characterize a visible surface one must accommodate both intrinsic surface properties and the spatial relationship between surface and observer. For a smoothly varying surface, these properties/relationships can be adequately represented by knowledge of:
(1) distance of the surface relative to the observer's plane of ocular fixation; (2) orientation of the surface relative to the observer (tilt and slant); and (3) surface curvature

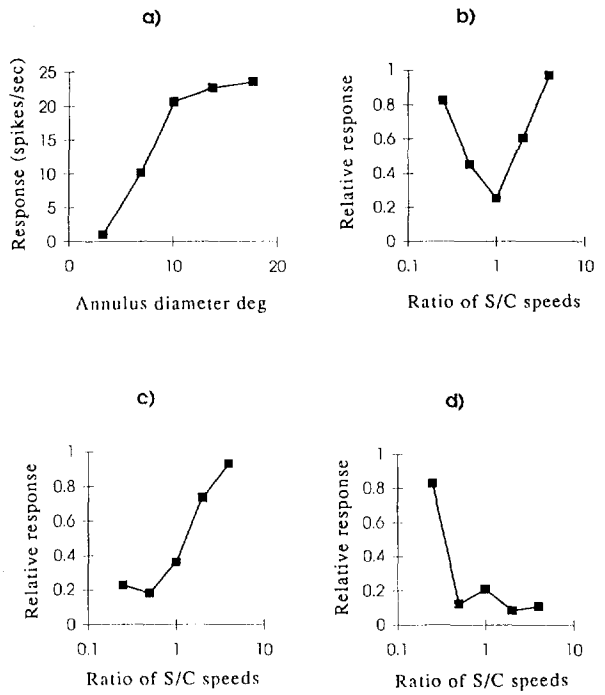

FIGURE 3. Four types of RF surrounds encountered in visual area MT: (a) synergistic surround: responses increase with increasing diameter of textured moving field [based on data published in Born and Tootell (1992)]; (b) V(U)-shaped speed tuning for background texture moving in the same direction. The tuning curve was mapped using a bar moving through the CRF at optimal velocity. Responses are plotted in units relative to the CRF response elicited without the textured background; (c) monotonically increasing; and (d) monotonically decreasing surround tuning curves; also compare the prediction in Fig. 5d with Fig. 4 [(b)–(d) are based on data published in Tanaka *et al.* (1986)]. S/C, surround/center ratio.

(convexity vs concavity). In concert, these surface descriptors provide the 3D information needed for navigation and object manipulation.

The 'relative depth' of a surface can be defined as a smooth function of a positional vector $z(x,y) = \mathbf{k} \cdot (\mathbf{R} - \mathbf{R}_0)$ expressed relative to the point of fixation $\mathbf{R}_0$ [see equation (1); Fig. 1]; $\mathbf{k}$ is a unit vector in the direction of the z-axis. It contains implicit information about all higher-order descriptors of the surface shape.

Both tilt (τ) and slant (ψ) can be recovered from the gradient vector of the surface function $\nabla z = [z_x, z_y]^T$:

$$\Psi = \arctan(|\nabla z|) = \arctan\left(\sqrt{z_x^2 + z_y^2}\right),$$

$$\tau = s \cdot \arctan(z_y/z_x). \qquad (5)$$

Subscripts indicate partial derivatives of the relative depth-function $z(x,y)$ and the value of $s = \pm 1$ depends on the quadrant to which the gradient vector points.

From differential geometry we know that the normal curvature $\kappa(\varphi)$ in the direction φ is the local measure of surface shape, which is invariant with respect to translation and rotation of the coordinate system. Rotating the coordinate system such that the x-axis becomes aligned with direction φ, allows surface curvature to be expressed in terms of the partial first- and second-order derivatives z_i, z_{ij} in a rather simple way (see Appendix for justification):

$$\kappa(\varphi = 0) = \frac{z_{xx}}{\sqrt{1 + z_x^2 + z_y^2(1 + z_x^2)}}. \qquad (6)$$

Knowledge of Gaussian curvature $\kappa_g = \kappa_{min}\kappa_{max}$ (κ_{min} and κ_{max} are the two principal normal curvatures) allows a means to discern hyperbolic surfaces ($\kappa_g < 0$) from parabolic ($\kappa_g > 0$) and cylindrical ($\kappa_g = 0$) surfaces. It is also useful to know the mean curvature $\bar{\kappa} = (\kappa_{min} + \kappa_{max})/2$, whose sign distinguishes parabolic convex surfaces ($\bar{\kappa} > 0$) from parabolic concave surfaces ($\bar{\kappa} < 0$). Because the projected curvature increases with distance of a surface from an observer, it is constructive to separate the scale-invariant shape descriptor $\Omega = (2/\pi) \arctan [2\bar{\kappa} / (\kappa_{max} - \kappa_{min})]$ from the scale-variant shape component $\Theta = \sqrt{(\kappa_{max}^2 + \kappa_{min}^2)}$ quantifying the magnitude of curvature, as introduced by Koenderink and van Doorn (1992) [see also Dijkstra (1994), p. 104].

First- and second-order partial derivatives of a surface function are thus sufficient to compute surface curvature and orientation. Since the retinal velocity field caused by an object undergoing displacement relative to an observer is, to a first approximation, proportional to the surface function [equation (4)], differentiation of the surface can be replaced by differentiation of the velocity field components, with the constraint that the scaling factor and the sign of relative depth must be recovered using other visual cues or modalities.

NEURONAL ELEMENTS FOR DIFFERENTIAL MOTION DETECTION

Two decades ago, Nakayama and Loomis (1974) proposed that direction-tuned neurons with antagonistic motion surrounds could provide a mechanism for the extraction of relative depth from the velocity flow field. This prediction was supported by the subsequent discovery of neurons with antagonistic motion surrounds in the visual systems of many species. Of particular interest for our theoretical formulation of the SFM problem are directionally tuned neurons in primate visual area MT. In designing optimal differential motion filters for the computation of 3D surface shape, we have adhered to biological constraints given by the known response properties of these neurons. Before describing the differential motion filters, we will briefly review the relevant neurophysiological data.

Three types of RF center–surround interaction in area MT

Selectivity for direction and speed of motion are salient features of the vast majority of neurons in visual area MT of the primate cerebral cortex (Zeki, 1974; Maunsell & Van Essen, 1983; Albright, 1984; Rodman & Albright, 1987). In addition to these basic selectivities, which are characteristic of responses elicited by visual stimulation of the CRF, the responses of many MT neurons can be

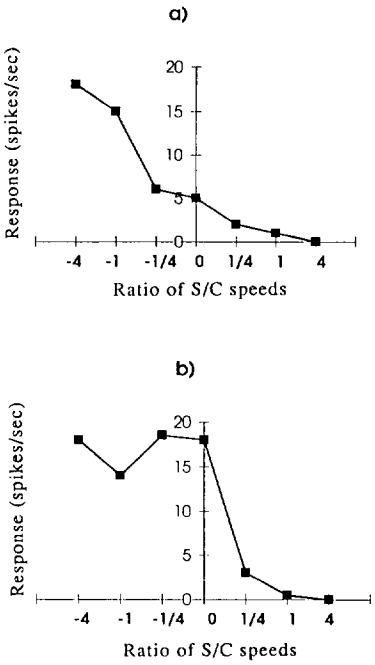

FIGURE 4. Variations in surround velocity tuning for monotonically decreasing tuning type, as observed by Lagae *et al.* (1989): (a) data from a single MT neuron tested with surround velocities including direction reversal. Response increases when the surround texture moves in the direction opposite to the central stimulus. Spontaneous activity level ≈ 2 spikes/sec; and (b) data for another cell, tested under the same conditions, and showing no modulation by negative surround motion. Spontaneous activity level ≈ 8 spikes/sec. Variation across these tuning curves can be formalized in terms of bias [bias < 0 for cell (a), and >0 for cell (b)] and 'temperature' T [which controls the slope of the curve, see equation (9)]. Based on data published in Lagae *et al.* (1989).

modulated by appropriate simultaneous stimulation of a larger surrounding region of the visual field, known as the 'non-classical' (Allman *et al.*, 1985; Tanaka *et al.*, 1986) or SRF. By definition, stimulation of the surround alone is incapable of eliciting a neuronal response; the function of the surround appears to be exclusively modulatory in nature. Such modulatory surrounds typically abut the CRF and can extend over a radius up to 7–10 times larger than the radius of CRF (Allman *et al.*, 1985). An MT neuron may possess one of several different types of modulatory surrounds. We will argue that each plays a critical role in the extraction of 3D surface shape from the retinal velocity field.

Receptive field surrounds can be divided into *antagonistic* and *synergistic* subtypes (Fig. 3) (Born & Tootell, 1992). Roughly speaking, antagonistic SRF activation will *suppress* the normal CRF response when the SRF and CRF are stimulated with a moving pattern of the same direction and speed. Likewise, the normal CRF response will be *enhanced* when the SRF and CRF are stimulated with moving patterns of different directions and speeds. For example, the response of a typical cell possessing an antagonistic SRF and preferring rightward CRF motion may be enhanced when the SRF is stimulated with motion in the opposing (leftward) direction, and suppressed when the SRF is stimulated with motion in the same (rightward) direction. Furthermore, antagonistic SRFs have been shown to be smoothly tuned for direction, with the preferred direction typically opposing that of the CRF (Allman *et al.*, 1985). The modulatory behavior of the synergistic SRF is precisely the converse, i.e. the response to motion in the CRF is facilitated by the same motion in the SRF.

In addition, it was found that antagonistic SRFs can be subdivided into several subtypes according to their speed tuning curves. For example, 44% of neurons recorded in the owl monkey (*Aotus trivirgatus*) (Allman *et al.*, 1985) possessed antagonistic SRFs. One third of those exhibited *V(or U)-shaped* tuning curves [Fig. 3(b)]. These neurons were suppressed maximally by SRF motion of the same velocity as that in the central RF; decreasing or increasing SRF speed had disinhibitory effects. They thus appeared to favor speed differences between center and surround, regardless of the sign of these differences. For the remaining sample of MT neurons, inhibition decreased with background speed (Allman *et al.*, 1985). A complementary study by Tanaka *et al.* (1986) performed on the old-world Japanese macaque (*Macaca fuscata*), found MT neurons with both: (1) antagonistic SRFs tuned for speed in a V(U)-shaped fashion (50% of sample) [Fig. 3(b)]; and (2) monotonically increasing velocity tuning curves [Fig. 3(c)].

In addition, Tanaka *et al.* (1986) reported neurons with monotonically *decreasing* speed tuning curves [Fig. 3(d)]. Although these authors used only preferred motion in the CRF when characterizing such neurons, it seems likely that they correspond to those neurons that are facilitated by opposing motion in surrounds (Allman *et al.*, 1985). This idea can be visualized as a smooth extension of the surround–modulation speed tuning curve into the domain of negative speeds. Our argument is supported by recent experiments that employed SRF stimuli moving both in the direction of CRF stimuli as well as in the opposite direction (Lagae, Gulyas, Raiguel, & Orban, 1989). It was found that a subset of MT neurons ['antiphase conditionally direction selective' neurons, Lagae *et al.*, (1989)] had monotonically decreasing speed tuning when surround direction was the same as in the CRF, but a surround stimulus moving in the null direction caused a substantial facilitation of response compared to the condition of central stimulation only [Fig. 4(a,b)]. This type of surround tuning is consistent with the opponent surround cells hypothesized by Nakayama and Loomis (1974) and reported by Allman *et al.* (1985), which are facilitated by opposing surround motion. Neurons with monotonically tuned SRFs are sensitive to a signed speed difference and the magnitude of modulation of response by SRF is approximately proportional to the speed difference. This observation underlies the model proposed herein.

In summary, the center–surround interactions of RFs in area MT can be broadly classified as synergistic or antagonistic. Antagonistic SRFs can be further subdi-

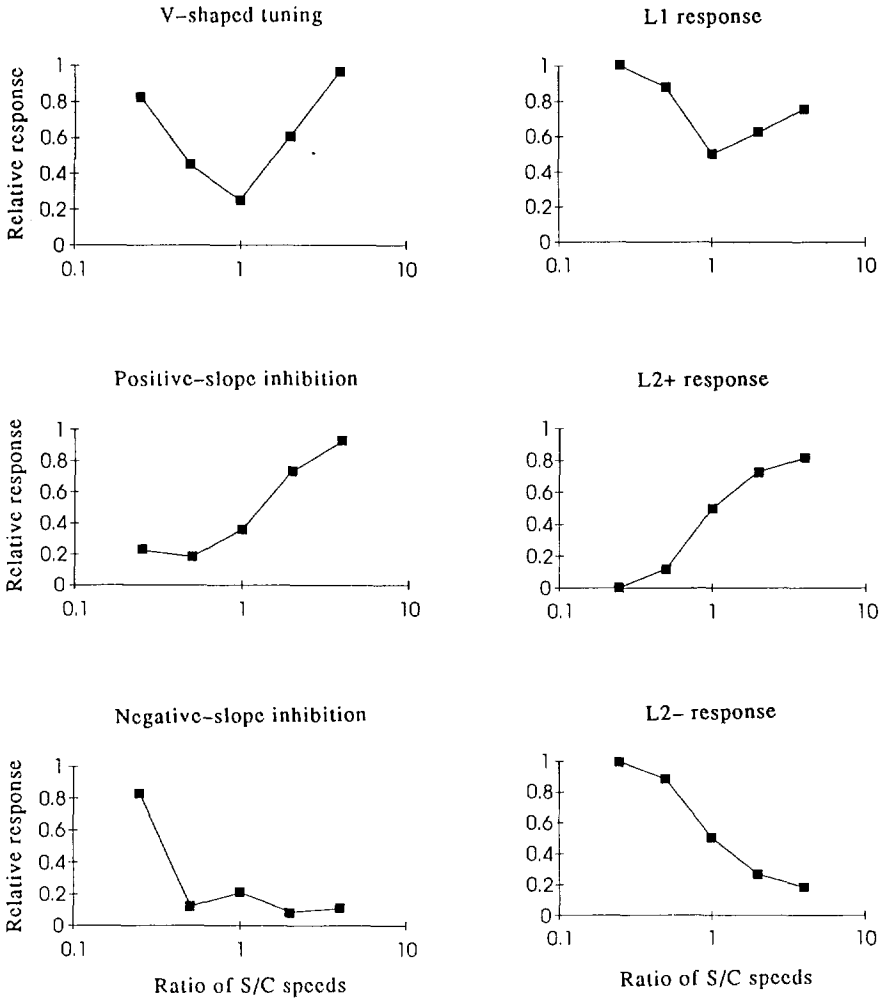

FIGURE 5. Comparison of data and model velocity tuning curves for center–surround interaction. Left column: data illustrating types of center–surround interactions seen in MT by Tanaka *et al.* (1986) (replotted from Fig. 3b–d). Right column: model tuning curves illustrating types of center–surround interaction produced by equations (8*b*) and (8*c*).

vided into those possessing either V-shaped or monotonic (increasing and decreasing) velocity tuning.

CONSTRUCTING RADIALLY SYMMETRICAL DIFFERENTIAL MOTION FILTERS

Our goal in this section is to identify a basis set of differential motion filters that can be used to compute descriptors of 3D surface geometry. To gain biological plausibility, the specific choice of filter properties has been inspired by the differential-motion sensitivities of MT neurons (described above). Only the receptive field surrounds that are capable of eliciting *suppressive* modulation are consistent with the type of differential motion detection needed for computation of 3D surface shape. For completeness, however, we will also treat synergistic surrounds. In addition, currently available data (Lagae *et al.*, 1989; Born & Tootell, 1993) suggest that the modulatory influence of SRF on CRF activation is additive in nature (i.e. speed tuning in the CRF is not affected by stimulation in SRF), and we will employ this constraint in the design of idealized differential motion detectors.

For clarity of exposition we will exploit the fact that the broad, cosine-like direction tuning curve of the typical MT neuron betrays a simple relationship between stimulus motion and neuronal response. Specifically, the response of a neuron i to a stimulus in the CRF can be expressed as the inner product between the actual velocity vector $\mathbf{V}$ of a stimulus feature projecting to the unit vector representing a neuron's preferred direction of motion $\mathbf{n}_{pref}$ (Zhang, Sereno & Sereno, 1993):

$$\mathbf{u}_i = \mathbf{n}_{pref} \cdot \mathbf{V}_i = |\mathbf{V}_i| \cos(\varphi_{pref} - \varphi_v), \qquad (7)$$

where φ_{pref} is the preferred direction of a neuron and φ_v is the direction of the stimulus. This simplified representation can be justified within any linearly approximated part of the speed tuning function. We also assume that a veridical representation of the retinally projected velocity vector field $\mathbf{V}(x,y)$ is available in area MT. This assumption rests on the evidence that many MT neurons appear to represent true velocity, rather than components of local motion perpendicular to a moving edge

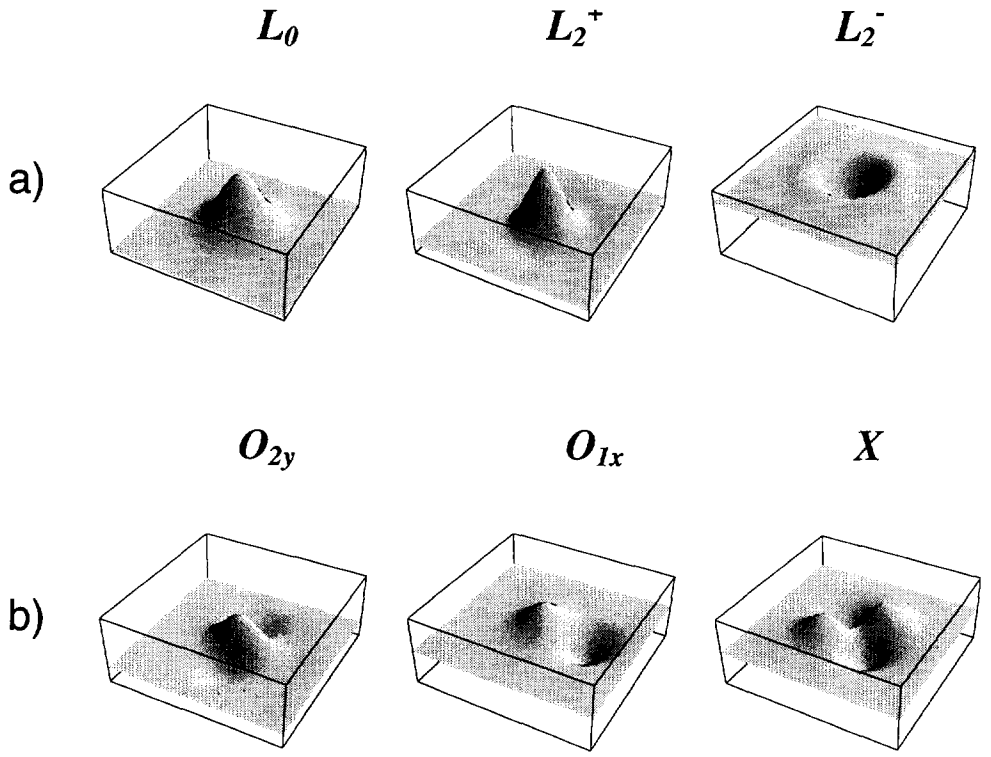

FIGURE 6. (a) Spatial profiles of kernels of radially symmetrical 'fuzzy' differential operators. The zeroth-order operator L_0 only blurs, while the second-order operators can enhance motion contrast isotropically. (b) Kernels of oriented differential operators: O_{2y} [or $O_2(\pi/2)$] calculates the blurred second-order derivative along the y-axis, O_{1x} [or $O_1(0)$] calculates the blurred first-order derivative along the x-axis. The righthand kernel (X) can yield the estimate of a mixed second-order derivative in x and y directions.

(Movshon, Adelson, Gizzi & Newsome, 1985; Rodman & Albright, 1987).

Discrete differential-motion filters

Idealization of the properties of MT neurons relevant to 3D surface characterization yields motion filters of three critical orders. We propose that these filters function in the computation of 3D surface descriptors for: (1) distance relative to point of ocular fixation; (2) surface orientation; and (3) surface curvature. The properties of these filters can be captured by equations involving summation of inputs over discretized center and surround receptive field subregions.

The zeroth-order filter, characterized by *synergistic* center–surround interactions, serves to compute the distance of the surface relative to the distance of the point of ocular fixation. This filter class can be modeled by:

$$l_0 = S\left\{\sum_{i\in N}[u_c + u_s(i)]\right\}, \qquad (8a)$$

where u_k ($k = c, s$) are projections of the velocity field to the direction perpendicular to the vector of effective rotation **w** [see equation (6)]; here, and in subsequent filter equations (8b,c) and (10a–c), $u_s(i) = 0$ if $u_c = 0$ and $u_s = \mathbf{n}_{pref} \cdot \mathbf{V}_s$ if $u_c \neq 0$, as required by the constraint that receptive field surrounds must be 'silent'; the summation is over surround neighborhood N of u_s. The sigmoidal activation function:

$$S(x) = [\tanh(T \cdot x + b) + 1]/2, \qquad (9)$$

imposes the bounds for neuronal response $0 < S(x) < 1$. The coefficient T controls the slope of the sigmoidal function and b is a bias term determining the activation threshold value.

The first-order filter, characterized by V(U)-shaped *antagonistic* surround tuning [Fig. 3(b)], serves to compute the orientation of the surface relative to the observer (i.e. tilt and slant). This filter class can be modeled by applying a squaring nonlinearity:

$$l_1 = S\left\{\sum_{i\in N}[u_c - u_s(i)]^2\right\}. \qquad (8b)$$

The discrete equation of this form satisfies the constraint that surround inhibition is maximal when the speeds in the center and surround are equal: $u_c = u_s$.

The second-order filter, characterized by both monotonically ascending and descending tuning curves for surround antagonism [Fig. 3(c,d)], serves to compute surface curvature. This filter class can be expressed in

terms of straightforward summation of the speed difference between center and surround:

$$l_2^{\pm} = S\left\{\sum_{i \in N} \pm [u_c - u_s(i)]\right\}. \qquad (8c)$$

The l_2^+ filter responds to convex surfaces, and is silent for concave surfaces, while l_2^-, conversely, detects concavities of surfaces. The l_2^- filter property of facilitation by a surround stimulus moving in the direction opposite to the center stimulus can be achieved by appropriate choice of the bias term b in the sigmoidal function. In equations (8b) and (8c) we choose the slopes of speed tuning curves in the CRF to be equal to those in the surround, which is essential for obtaining balanced filters. The tuning curves for monotonically decreasing (l_2^-) SRF speed tuning [Fig. 4(a,b); Lagae *et al.*, 1989] suggest that neither bias b (which determines the position of sigmoidal bending of the activation function S), nor tuning-slope T are fixed across MT neurons allowing some flexibility in selecting these parameters for our differential motion filters.

Figure 5 compares the responses of our filters to their biological correlates. Note that, although no curve-fitting has been performed, there is strong qualitative agreement between the neurophysiological data [Fig. 5(a–c)] and the behavior of equations (8b,c) [Fig. 5(d–f)]. We have not presented simulated velocity tuning for equation (8a) because appropriate neurophysiological data are unavailable for comparison. Qualitatively, however, the tuning curve for equation (8a) is similar to that for L_2^+, but shifted upwards (+1) to reflect facilitation.

Continuous differential-motion filters

Although they capture the essential properties of center–surround interactions that we wish to promote, the discrete nature of the filters described above does not reflect realistic receptive field spatial profiles. We will now substitute continuous analogs, which are more consistent with known biological constraints.

We assume, first of all, that synaptic weights for both center and surround receptive fields decline as a Gaussian function $G(\sigma)$ of distance from the CRF center and that σ is different for center and surround: $\sigma_c < \sigma_s$. Thus, by applying convolution with Gaussians, equations (8a) and (8c) can be rewritten:

$$L_0(i,j) = S\{G(\sigma_c)*u(i,j) + G(\sigma_s)*u(i,j)\}, \qquad (10a)$$

$$L_2^{\pm}(i,j) = S\{\pm [G(\sigma_c)*u(i,j) - G(\sigma_s)*u(i,j)]\}, \qquad (10c)$$

(see Appendix for explication of notation). The spatial profiles of the operator kernels are shown in Fig. 6(a). The continuous nonlinear L_1 filter can be defined if equivalence to l_1 [equation (8b)] is observed only up to the second-order term of the power series for $u(i,j)$:

$$L_1(i,j) = S\{G(\sigma_c)*u^2(i,j) + G(\sigma_s)*u^2(i,j) - c_1[G(\sigma_c)*u(i,j)][G(\sigma_s)*u(i,j)]\}; \qquad (10b)$$

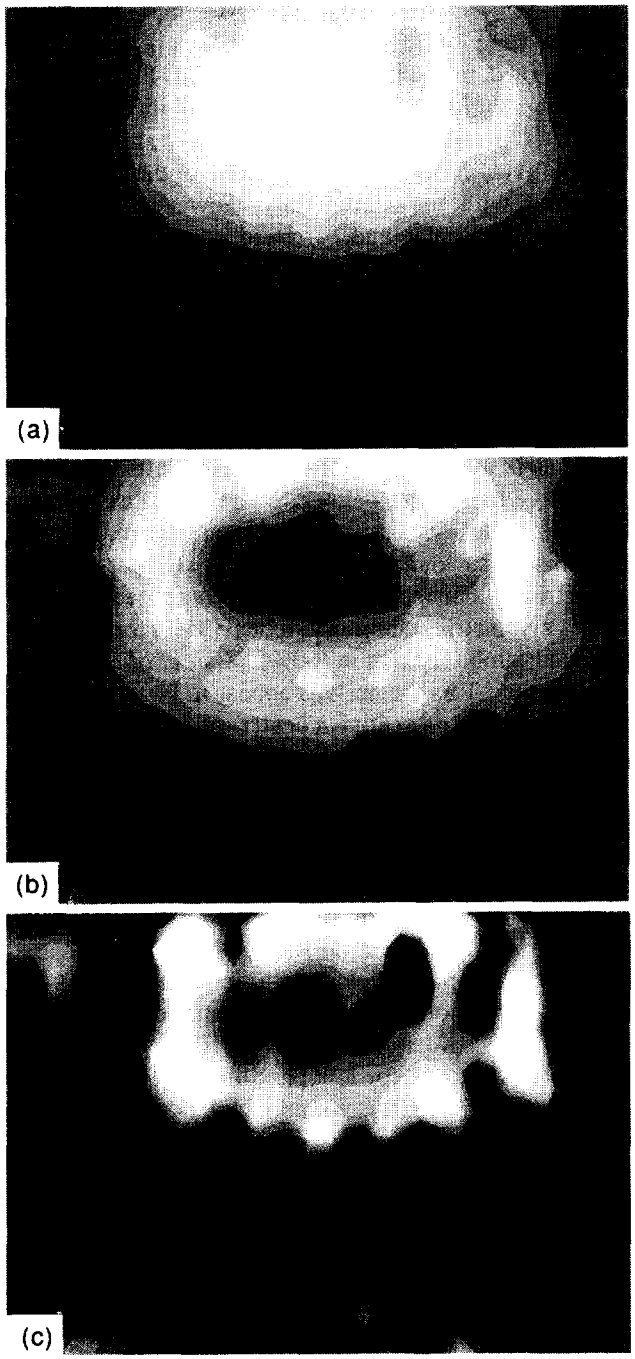

FIGURE 7. Application of the model MT filters to a realistic dynamic scene: (a) the response of l_0 filter to the horizontal component of the velocity field, is followed by responses of l_1 and l_2 filters [(b) and (c), respectively]. The iso-depth and iso-slant gray-level values in (a) and (b) are consistent with the shape of the teapot. The image in panel (b) contains peak values along the upper edge, which corresponds to maximal slant of the teapot in this area. The response in (c) is calculated from both convex and concave operators ($l_2^+ - l_2^-$) to convey both the positive and negative components of the second-order response. The noise in the image is attributable to the coarse texture of the velocity field inducing surface.

$u^2(i,j)$ corresponds to full-wave rectification [formerly used to model response properties of complex neurons in area V1 (Spitzer & Hochstein, 1985)]. Note that the form of the equation is that of an expanded quadratic binomial;

$c_1 = 2/\mathrm{Erf}^2(\frac{n}{\sqrt{2}})$ is the balancing constant, and Erf( ) is the error function. The balancing constant secures the zero-response to a uniform velocity field. It depends on integration limits $\{-n\sigma,n\sigma\}$ for convolution integrals. The spatial profiles of the L_0 and L_2 filters are illustrated in Fig. 6(a) (L_1 is not displayed because of the complexity associated with the nonlinearity; however the spatial profile of the center and surround components of this filter is identical to that for L_2).

Motion filter refinements: oriented receptive field surrounds

Following their discovery, it was assumed that RF surrounds in area MT were radially symmetric about the CRF. More recently, a deliberate test of this assumption in rhesus monkeys has revealed that the majority of antagonistic surrounds in area MT are not radially symmetric; rather they appear to be composed of either axially symmetric inhibitory lobes (35%) or a single elongated inhibitory region (60%) (Xiao *et al.*, 1994). In addition, the CRFs themselves appear to be elongated in many cases, thus presenting a center–surround sensitivity and spatial interaction profile that is striking similar to receptive field profiles in area V1. (This structural similarity suggests certain functional parallels between the operations carried out on velocity flow fields in area MT and the operations carried out on distributions of luminance in V1—a possibility that will be explored more fully in the Discussion.)

Oriented receptive fields with inhibitory lobes are known to correspond to fuzzy differential operators (e.g. Werkhoven & Koenderink, 1990). The term 'fuzzy' refers to the blurring (or low-pass filtering) of an image by differential filters. It is common practice to characterize these RFs in terms of Hermite polynomials, i.e., derivatives of the Gaussian function (Beitas & Kirvelis, 1987; Koenderink & van Doorn, 1987). We have chosen a less restrictive approach, however, that does not automatically place constraints on the widths of excitatory and inhibitory regions. This approach renders greater conformity with neurophysiological evidence.

For the discrete case of oriented filters, the general form of equations (8a–c) is preserved; only the neighborhood of summation is changed from radially to axially symmetric (second-order differential) or antisymmetric (first-order differential) (Buračas & Albright, 1994). The continuous second-order operators can be defined as a difference of radially symmetrical and elongated Gaussians. For example, for differentiation in the direction of the y-axis ($\varphi = \pi/2$) we have:

$$O_2(i,j,\pi/2) = S\{[G(\sigma_s,\sigma_s) - G(\sigma_s,\sigma_c)]*u^2(i,j)\},$$

(see Fig. 6(b) for examples of spatial profiles). Inspired by neurophysiological evidence (Xiao *et al.*, 1994), we assume that these oriented filters, as well as the ones described below, are defined for every direction of differentiation $\varphi\in[-\pi,\pi]$. Likewise, by analogy with equation (10c), the axially symmetric (even) first-order differential operator (for differentiation in the direction of the x-axis, $\varphi = 0$) is:

$$O_1^e(i,j) = S\{G(\sigma_c,\sigma_s)*u^2(i,j) + G(\sigma_s,\sigma_s)*u^2(i,j) - c_1^e[G(\sigma_c,\sigma_s)*u(i,j)][G(\sigma_s,\sigma_s)*u(i,j)]\}.$$

Here c_1^e is the balancing constant. The anti-symmetric first-order operators can be modeled by a difference of 2D Gaussians with shifted maxima. For example, differentiation in the direction of the x-axis is:

$$O_1^{o\pm}(i,j) = S\{\pm[G(\sigma_c,\sigma_c)*u(i+\sigma_c,j) - G(\sigma_c,\sigma_c)*u(i-\sigma_c,j)]\}.$$

The sign $\pm$ indicates that, to support rectification by the sigmoidal function, we need to postulate filters that independently transfer the positive and negative values. The distance between the peaks of the two Gaussian components of $2\sigma_c$ is set within a range of physiologically plausible values. Although standard deviations in the two orthogonal directions are equal in the equation, they may assume different values. Although not yet proven experimentally, the distinction between the even and odd first-order differential operators is important for computing the curvature of moving surfaces, as will be shown below. Furthermore, although we have described differential motion operators oriented only in cardinal directions, they can be defined for any direction of differentiation. Below we will argue that these operators function as fuzzy directional derivatives of velocity field components.

FILTER PROPERTIES AND PERFORMANCE

To summarize, three orders of differential motion filters have been defined for the purpose of computing descriptors of 3D surface geometry. On the one hand, their design incorporates constraints dictated by the response properties of cortical neurons that encode differential motion. On the other hand, the lack of data on certain parameters has left us some freedom to choose values that optimize performance of our model filters, but simultaneously conform to a principle of maximal simplicity. For example, when analyzing response properties of the filters, we choose T and b such, that the range of inputs to the filters falls approximately within the linear portion of the sigmoidal function [equation (9)].

When presented with a velocity flow field, the zeroth-order filter is expected to 'respond' to velocity components that reflect the relative distance between the source surface and the plane of ocular fixation. The first-order filter is expected to respond to velocity components that reflect the orientation (tilt and slant) of the surface relative to the observer. The second-order filter is expected to respond to velocity components that reflect the degree and sign of surface curvature (concave vs convex). In concert, these filters can make a significant contribution to the characterization of surface geometry.

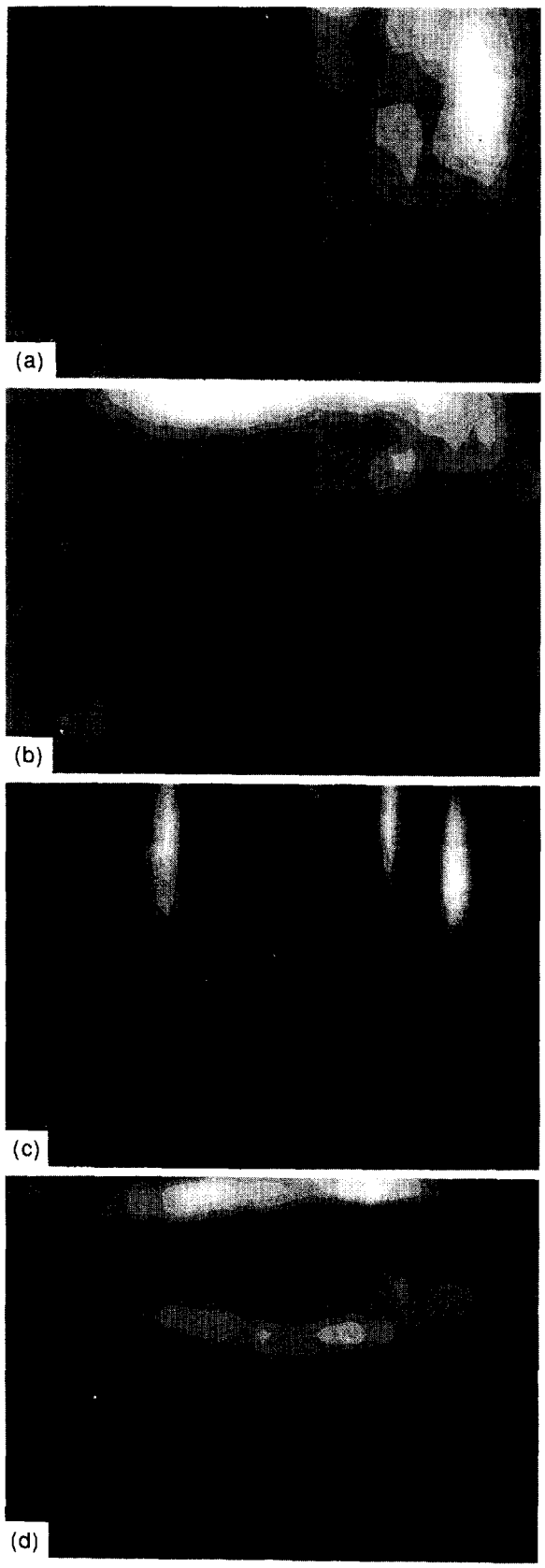

FIGURE 8. Responses of oriented filters to the flow field induced by the teapot of Fig 2: (a) and (b): responses of filters differentiating in the x- and y-directions. The responses from both positive and negative filters are pooled [$O_1^{0+}(0)$–O_1^{0-} (0) in (a) and $O_1^{0+}(\pi/2)$–O_1^{0-} ($\pi/2$) in (b)]. These responses peak where slant is maximal for the direction preferred by the filters; (c) and (d): responses of second-order oriented filters. Note that the response of O_{2x} [or $O_2(0)$] is noisier than that of O_{2y} [or $O_2(\pi/2)$] due to vertical orientation of texture.

(In order to obtain a reliable response from any of these filters, it is, of course, essential that the source surface possesses sufficiently dense surface texture—a constraint that is generally fulfilled by the properties of natural surfaces).

We will now proceed with a formal and detailed description of the properties of each of the motion filters, followed by an illustration of their application to (sur-) realistic visual input.

Computing 3D surface parameters

In further development of the radially symmetric motion operators, we will exploit only the simplified relation for the velocity flow field given in equation (4). Expanding $z(x,y)$, from equation (4), into a power series around an arbitrary point and truncating above the second-order term yields:

$$u(x,y) = -\frac{w}{z_0} z(x,y) = -\frac{w}{z_0}(ax^2 + by^2 + cxy + dx + ey) \text{ and } v(x,y) = 0, \tag{11}$$

where $a = z_{xx}/2$, $b = z_{yy}/2$, $c = z_{xy}$, $d = z_x$, $e = z_y$ are expansion coefficients (see Appendix for elaboration of relations between the surface representation and velocity field).

Operations of radially symmetric filters

Zeroth-order filter. Application of L_0 to $u(x,y)$ removes high spatial frequencies, but otherwise the result does not differ from $u(x,y)$—i.e. $L_0 * u$ covaries with the lower frequencies of $u(x,y)$. Thus this zeroth-order filter blurs the velocity field, effectively attenuating noise caused by sparsely distributed features. This point can be seen by comparing the visual input [Fig. 2(c), left panel]—the 2D retinal velocity field $u(x,y)$ elicited by surface displacement—with the output of L_0 [Fig. 7(a)—a representation of the output of this and the following filters, shown in Figs 7 and 8, can be interpreted as firing-rate maps for a sheet of neurons containing a single filter type]. The resulting low-pass filtered 'neural image' conveys the relative-depth map of the 3D surface more reliably than does the unfiltered pattern of the horizontal velocity component.

First-order filter. When applied to planar stimuli $u_p(x,y) = (dx + ey)\frac{w}{z_0}$, it can be shown by direct evaluation of the convolution integral that L_1 exhibits properties of a squared first-order differential operator:

$$L_1 * u_p \propto -(d^2 + e^2) \cdot C_1(\sigma_c^2 - \sigma_s^2) \cdot \frac{w}{z_0} = -D^2 \tilde{z}_p \cdot C_1(\sigma_c^2 - \sigma_s^2) \cdot \frac{w}{z_0}, \tag{12}$$

where $D^2 = (\frac{\partial}{\partial x})^2 - (\frac{\partial}{\partial y})^2$ is the first-order symmetric differential operator, commonly used in robot vision (Horn, 1986); $C_1(\sigma_c^2 - \sigma_s^2)$ is a function of σ_c, σ_s and does not depend on properties of stimulus u_p. The tilde sign in $\tilde{z}_p$ indicates the fact that the estimate of the differential

refers to the low-passed version of the surface function. Owing to its isotropic nature, L_1 extracts only the scalar value $\frac{w}{z_0} C_1 |\nabla \tilde{z}|^2$ (the weighted squared norm of the gradient vector), which bears a direct relationship to the slant Ψ of the moving surface. The (blurred version of the) latter can be directly computed by substituting the output of L_1 in equation (5): $\tilde{\Psi} = \arctan \frac{z_0}{w\sqrt{C_1}} \sqrt{L_1 * u_p}$. Note that the proportionality ($\propto$) in equations (12) and (13) holds within the linear portion of the sigmoidal activation function S.

An interesting property of this filter is that it is invariant with respect to the direction of tilt of the inducing surface. Indeed, when applied to the velocity field arising from our rotating teapot, L_1 peaks at the locations of maximal slant of the teapot surface [Fig. 7(b)]—i.e. close to the boundary contour of the teapot—regardless of surface orientation. Closer examination of Fig. 7(b) reveals that the filter response is greatest near the top of the teapot image, due to the maximal slant in that region. Absence of response to the velocity field in the central region of the teapot correlates correctly with the near-zero slant of the teapot surface.

Second-order filter. Application of L_2 to $u(x,y)$ from equation (11) yields:

$$L_2 * u \propto -(2a + 2b) \cdot C_2(\sigma_c^2 - \sigma_s^2) \cdot \frac{w}{z_0} = -\nabla^2 \tilde{z} \cdot C_2(\sigma_c^2 - \sigma_s^2) \cdot \frac{w}{z_0}. \quad (13)$$

As expected, L_2 exhibits properties of the second-order space-differential operator—Laplacian. $C_2(\sigma_c^2 - \sigma_s^2)$ is a constant that depends only upon the widths of center and surround Gaussians.

The output of L_2 is related to surface curvature in a non-trivial way. It can be shown that mean curvature can be computed from:

$$\bar{\kappa} = \frac{\nabla^2 z + z_{xx} z_y^2 + z_{yy} z_x^2}{2(1 + D^2 z + z_x^2 z_y^2)\sqrt{1 + D^2 z}}, \quad (14)$$

(see Appendix for derivation). At points of singularity $z_x = z_y = 0$—i.e. where the tangent plane is perpendicular to the direction of gaze, equation (14) reduces to:

$$\bar{\kappa} = \frac{1}{2} \nabla^2 z = \frac{z_0}{2w} \nabla^2 u. \quad (15)$$

Thus, at these points of singularity, the output of L_2 is proportional to the mean curvature of the velocity-field-inducing surface: $\tilde{\bar{\kappa}} \propto \frac{z_0}{2w} L_2 * u$ (recall that the tilde denotes the 'blurred' estimate of the mean curvature). The second-order analysis of the flow field induced by the rotating teapot is illustrated in Fig. 7(c). Large response troughs are traceable to the noisy estimate of the velocity field [Fig. 2(c)], which is, in turn, tied to surface texture density.

We previously imposed the constraint that the observer must track a single point on the surface of a moving object. However, eye rotation caused by tracking of a moving surface serves to compensate the transversal component of translational motion. It should be apparent that both L_1 and L_2 filters discard the constant transversal component of optical flow. This fact was previously noted by Carman (1991), who suggested that the primary role of radially symmetric antagonistic surrounds in area MT may be to discard the full-field motion component. It follows that we can lift the eye-tracking constraint without altering the conclusions of the previous section.

Extracting vector quantities

Equations (12) and (13) document the fact that only *scalar* quantities can be extracted by the radially symmetric operators described here. A proper description of surfaces requires *vector* quantities, however, which involve directional derivatives, like the gradient $[z_x, z_y]^T$. We previously noted that this requirement could be fulfilled by MT neurons with oriented receptive fields and non-uniform surrounds, and we predicted the existence of such neurons (Buračas & Albright, 1994). The recent discovery of MT neurons with these critical properties (Xiao *et al.*, 1994) permits us to complete our treatment of the shape-from-motion computation, by justifying the inclusion of directional differential operators.

Oriented filters, analogous to the ones described above, yield responses proportional to 'fuzzy' directional first- and second-order derivatives. For example, as when such filters are applied to a function $f(x,y)$ in the direction of the x ($\varphi = 0$) axis: $O_1^o(0) * f(x,y) \propto f_x$, $O_2(0) * f(x,y) \propto f_{xx}$ and $O_1^e(0) * f(x,y) \propto f_x^2$. Again, the estimate of the derivative is blurred. The cause of this blurring becomes evident by recalling that the differentiation of a blurred image corresponds to convolving the original image with a blurred kernel of a differential operator.

Since the retinal velocity field relates to the surface function in a simple way, as stated by equation (4) for the first-order approximation, we can readily obtain the relation between the outputs of the oriented differential-motion filters and the surfaces that induce optical-flow. From equations (A3) and (A4) (Appendix) we deduce:

$$\tilde{z}_x = -\frac{z_0}{w} \tilde{u}_x \propto -\frac{z_0}{w} O_1^0(0) * u,$$
$$\tilde{z}_y = -\frac{z_0}{w} \tilde{u}_y \propto -\frac{z_0}{w} O_1^0(\pi/2) * u, \quad (16a)$$
$$\tilde{z}_x^2 = -\frac{z_0^2}{w^2} \tilde{u}_x^2 \propto -\frac{z_0^2}{w^2} O_1^e(0) * u.$$

Here $O_1^0(O)$ indicates the response pooled over both positive and negative asymmetrical first-order oriented differential operators $O_1^{0+}(O)$ and $O_1^{0-}(O)$ in direction $\varphi = 0$ (tilde again denotes that the operators yield a low-pass filtered estimate of derivatives). Simulated responses of these operators to the velocity field induced by our rotating teapot are shown in Fig. 8(a,b). One can see that the responses of these operators peak for surfaces with high slant values tilted in the direction of differentiation.

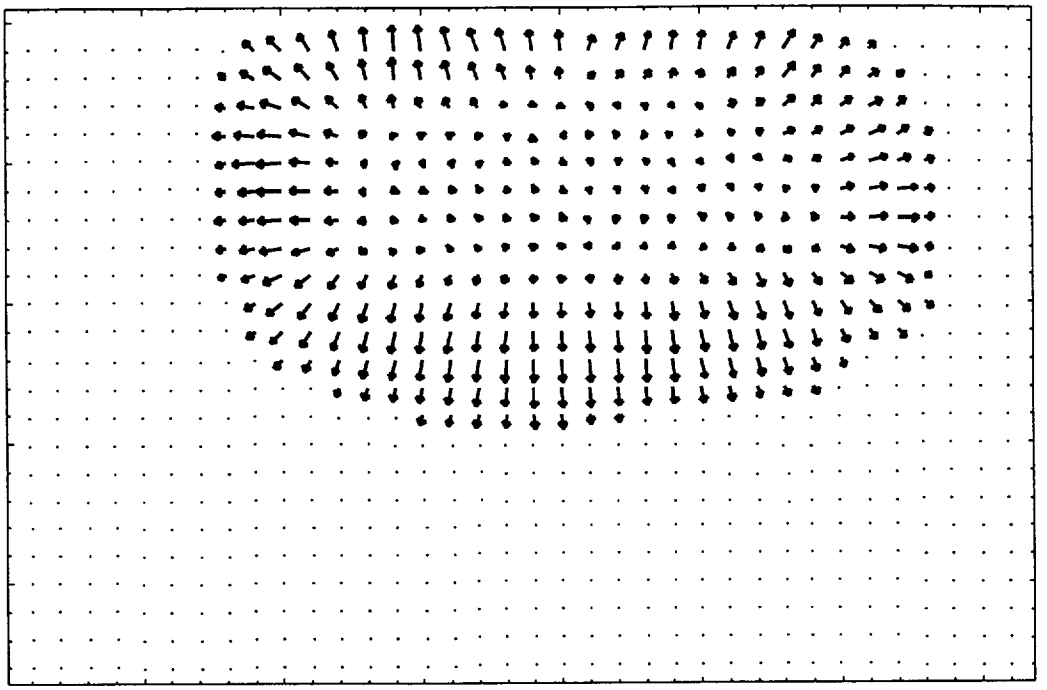

FIGURE 9. Gradient vector field calculated for the teapot surface from a set of orthogonal oriented differential-motion filters of the type developed in our model. Vectors correctly approximate the overall surface orientation. Note that the estimate is spatially blurred by low-pass filtering of convolution kernels.

Accordingly, the second-order derivatives are extracted using the second-order O-filters:

$$\begin{aligned}\tilde{z}_{xx} &= -\frac{z_0}{w}\tilde{u}_{xx} \propto -\frac{z_0}{w}O_2(0) * u,\\ \tilde{z}_{yy} &= -\frac{z_0}{w}\tilde{u}_{yy} \propto -\frac{z_0}{w}O_1^0(\pi/2) * u,\\ \tilde{z}_{xy} &= -\frac{z_0}{w}\tilde{u}_{xy} \propto -\frac{z_0}{w}X(0) * u\end{aligned} \qquad (16b)$$

(here the values in parentheses near the O symbol indicate the direction of differentiation, i.e., O stands for differentiation in the direction of the x-axis, and $\pi/2$ the direction of the y-axis). Figure 8(c,d) depicts an example of the outputs of two oriented second-order filters. The 'response' of the $O_2(0)$ filter is corrupted by noise arising from discontinuities of teapot texture in the horizontal direction [Fig. 8(c)]. Since the velocity field in the vertical direction is sampled more densely, the response of the $O_2(\pi/2)$ filter represents the second-order differential more faithfully. The bottom expression in equation (16b) was motivated by the recent finding (Xiao *et al.*, 1994) that some cells in area MT appear to possess receptive fields with spatial profiles that are consistent with the mixed differential operator X [see Fig. 6(b), righthand panel, for a spatial profile of the X kernel]. As indicated in the Appendix [equation (A4), bottom row, center column], computation of this derivative allows direct estimation of the rate of effective rotation of the object, provided that the distance to the surface is known.

Computations of curvature are greatly simplified if curvature is defined by directions of maximal and minimal curvatures [which are always orthogonal; see Appendix, equations (A1), (A2) for details], rather than as a combination of derivatives in cardinal directions. For this reason, it is advantageous to employ oriented differential operators that are defined over the full range of directions of differentiation $\varphi \in \{0,2\pi\}$. This distribution of differential motion filters seems to hold for the primate visual system: neurophysiological evidence demonstrates a lack of correlation in area MT between preferred direction of motion and orientation of receptive field profile (Xiao *et al.*, 1994), indicating the potential for differentiation in any spatial direction for every preferred direction of motion.

The MT-like filters L_1 and L_2, although reconstructable from oriented filters, may be an important shortcut in shape-from-motion computation, since the magnitudes proportional to their outputs show up in formulas for curvature. The normal surface curvature can be readily extracted using combinations of oriented and symmetrical filters. For example, the normal curvature in a direction of angle φ can be estimated from:

$$\begin{aligned}\tilde{\kappa}(\varphi) &= \frac{\tilde{z}_{\varphi\varphi}}{(1 + D^2\tilde{z})^{1/2}(1 + \tilde{z}_{\varphi}^2)}\\ &= \frac{\alpha O_2(\varphi) * u}{(1 + \beta L_1 * u)^{1/2}(1 + \gamma O_1^e(\varphi) * u)};\end{aligned} \qquad (17)$$

here α,β and γ are constants independent of stimulus properties (they absorb w and z_0), $*$ denotes convolution.

The scale-invariant shape index, introduced earlier,

can also be described using outputs of our filters (Koenderink & van Doorn, 1992):

$$\Omega = \frac{2}{\pi}\arctan\left[\frac{(z_{\varphi\varphi} + z_{\phi\phi}) + (z_{\varphi\varphi}z_{\phi}^2 + z_{\phi\phi}z_{\varphi}^2)}{(z_{\varphi\varphi} - z_{\phi\phi}) + (z_{\varphi\varphi}z_{\phi}^2 - z_{\phi\phi}z_{\varphi}^2)}\right]. \quad (18)$$

Here the directional derivatives in direction of the two normal principal curvatures are employed. Inserting the equations for partial derivatives from (15a) and (15b) will yield the desired expression.

PUTTING IT ALL TOGETHER

Thus far we have demonstrated that the filter characteristics of MT neurons have computational properties that are potentially useful for extracting 3D surface shape from optical flow. Critical as this potential contribution of MT may be, it should be clear that there is no evidence to support a proposal that MT neurons yield a complete motion-derived description of shape invariants. Our novel contribution is simply a demonstration that MT provides differential motion operators *capable* of extracting shape-from-motion; its operations do not complete the object-centered shape description. We suggest that the solution to the SFM problem may be completed in another, MT recipient, cortical area (e.g. areas V4, MST, FST, VIP), although there exists no unequivocal evidence on this point.

In the remainder of this section we briefly sketch a scheme for implementing the computations of differential-geometric descriptors by interactions in a set of neural networks. The co-existence of differential motion operators of several orders in a single visual area suggests that both first-order surface descriptors (tilt and slant) are computed in parallel with the second-order descriptors (e.g. curvature or scale-invariant analogue of curvature). Surface slant can be readily estimated from the filter L_1, as suggested in equation (5). Interestingly, since the arctangent function is identical to the hyperbolic tangent below the fifth-order term of power series expansion, the former function can be approximated by our sigmoidal transfer function S of equation (9) (see Appendix for details). The tilt can be computed by performing a cross-orientation search for the maximally responding first-order oriented filter O_1 at every spatial location. This is readily implementable by a winner-take-all neural network allowing competition between all differently oriented first-order filters. The orientation of the 'winning' filter betrays the direction of the tilt. The utility of this approach is illustrated by the example in Fig. 9, which shows an estimate of the gradient vector (with negative sign) for a surface function. The estimate was calculated from the horizontal component of the velocity field induced by our rotating teapot. Calculations were based on the output of two sets of orthogonal filters: $O_1^{0+}(0)$–$O_1^{0-}(0)$ and $O_1^{0+}(\pi/2)$–$O_1^{0-}(\pi/2)$. Note, that the resulting estimate of surface gradient is equivalent to direct estimation of the maximal directional derivative by means of winner-take-all network.

Similarly, the direction of maximal normal curvature could be determined by a winner-take-all network that stages a competition within a population of neurons estimating the value of $z_{\varphi\varphi}/(1 + z_{\varphi}^2)$. The response of the differential operator that is aligned perpendicularly to the maximally responding filter is also needed for a complete description of surface curvature as it yields an estimate of the minimal normal curvature. Since the implementation of a network estimating of the minimal normal curvature by means of a 'loser-take-all' network may be difficult; we suggest that the latter can be computed as a difference between the outputs of the maximally responding oriented and radially symmetric filters, e.g. if the maximally responding filter is oriented along the x-axis, we have: $\tilde{u}_{yy} \propto \tilde{u}_{xx} - L_2$ and $\tilde{u}_y^2 \propto \tilde{u}_x^2 - L_1$.

From the above passage it should be obvious that we can arrive at ecologically relevant representations of moving surfaces with little effort, given the responses of differential motion filters. We thus advocate the view that MT neurons are more than just faithful speedometers measuring velocities of retinally projected visual features. They also have properties suitable for computing the structure of surfaces carrying these features.

DISCUSSION

We have suggested, in computational terms, a functional role for the receptive field surrounds characteristic of neurons in visual area MT. Our approach has allowed us to explain and integrate diverse and seemingly contradictory data into a coherent theoretical framework, in which MT neurons are predicted to participate in the computation of object shape. Our analysis demonstrates that the differential-motion filters developed herein closely approximate the experimentally observed properties of MT neurons. These filters convey estimates of space differentials, essential in computing both surface orientation and curvature, as well as other ecologically important parameters of optical flow.

We have confined our attention to the components of optic flow induced by translational motion. However, the proposed differential motion filters would serve as well for optic flow arising from arbitrary piecewise-rigid motion, because generic estimation of shape-from-motion involves first- and second-order differentiation of velocity vector-field components (i.e. computation of second-order deformation tensor) (Koenderink & van Doorn, 1992).

Comparison with neural solutions for computing shape-from-shading

Our ideas about the neural computation of shape-from-motion in area MT have a direct parallel with the proposition by Lehky and Sejnowski (1989) for a solution to the shape-from-shading (SFS) problem. These investigators trained a neural network to compute curvature of Lambertian surfaces using only shading information. One consequence of training was the emergence of units

possessing receptive fields with oriented spatial sensitivity profiles. The receptive fields appeared remarkably similar to those known to exist in area V1, which have traditionally been thought to function as line or edge detectors, but can also perform the first-order (odd bi-lobed RFs) or second-order (even tri-lobed RFs) differentiation. For a Lambertian surface, irradiance is proportional to the gradient of the surface function and can be used for recovering curvature directly from first-order derivatives of the shading profile. Likewise, solving the curvature-from-motion problem also involves first- and second-order differentiation, albeit of velocity vector field components, not luminance. As we suggest herein, this differentiation is performed by RFs with spatial profiles very similar to those in V1.

The striking similarity of RF profiles in areas V1 and MT supports the possibility that computing differentials—contrast—may be a universal organizing principle of topographically arranged sensory areas. This idea resonates with the suggestion of Barlow and Foldiak (1989) that the primary function of the cortex is redundancy reduction via decorrelation of sensory input. For smooth surfaces the low-order (zeroth, first, second and few subsequent ones) terms of power series expansion coincide with principal components. Thus, representing surfaces in a form of low-order differentials leads to redundancy reduction.

Motion contrast enhancement with RF surrounds

Besides responding to smooth variations in speed, RFs with antagonistic surrounds (in general, and in the case of differential motion filters) are also responsive to edges. Our model thus implies that MT neurons must be responsive to motion-contrast defined edges. For example, L_1 responds vigorously at the surface boundary of the rotating teapot [Fig. 7(c)] due to both motion contrast and slant, which achieves a maximum at the boundary. Note that L_2 starts responding abruptly at the contour, thus, giving a highly localized signal about its position. Thus, within the distance of 2–3 σ_s from the edge of a surface, the filter response confounds curvature-or slant-related variations of velocity field with depth discontinuity. This confound may be neutralized by noting that L_2 responds with a zero-crossing at the motion-contrast boundary and by confining the shape analysis to regions beyond 2–3 σ_s from the edge of a surface. Notably, neurons tuned to kinetic (motion-contrast defined, either of shearing, or compression/expansion type) edges have not been found in area MT (e.g. Marcar, Raiguel, Xiao, Maes & Orban, 1991). We believe, that this failure can be explained by the fact that the relevant experimental manipulations have been largely confined to the classical RFs, which alone are devoid of motion-contrast detecting properties.

Viability of the model: experimental quirks and data shortcomings

Our model implies that the first-order differential-motion filters should be monotonically tuned to motion gradients, while the second-order filters should be tuned to second-order differentials. Treue and Andersen (1993) measured the responses of MT neurons to motion gradients and, indeed, found that approximately one-third of the sampled neurons were selective for the direction of the gradient, although this selectivity was rather weak. Our model leads us to attribute this weak selectivity to the nature of stimuli used in the study, which were confined to the classical receptive field. Had these investigators used motion gradients that spanned center and surround RF subdivisions, we expect that gradient selectivity would have been more pronounced.

Our analysis of SFM in area MT was based on compilations of data across several primate species, including the old-world rhesus monkeys (*Macaca mulatta*) and Japanese macaques (*Macaca fuscata*), as well as the new-world owl monkey (*Aotus trivirgatus*). It should be noted, however, that existing evidence for *oriented* receptive field surrounds—a critical feature of the model—comes exclusively from *M. mulatta* (Xiao *et al.*, 1994). Without this property, the shape-from-motion computation is limited to an estimation of slant and mean curvature at singular surface points (see above). The species generality of non-radially symmetrical surrounds is suggested by evidence for the existence of asymmetrical SRF profiles in the Claire-Bishop cortical area of cat (the feline analog of primate MT) (Rukšenas, Šatinskas & Stabinyte, 1994).

The spatial resolution for SFM estimation of the computational scheme proposed herein is limited by the size of RF surrounds in area MT. While early data from anesthetized animals indicated discouragingly large SRFs (the radius of SRF being 7–10 times larger than CRF; Allman *et al.*, 1985), more recent studies of surround/center ratios in alert macaques have yielded much smaller values (SRF 2–4 times larger than CRF; Born & Tootell, 1993; Xiao, personal communication). The latter value constrains the resolution of SFM estimation in fovea to 2–4 deg, since the diameter of foveal RFs of MT neurons is about 1 deg (Albright & Desimone, 1987).

Effects of nonlinearities in speed tuning curves

We have assumed that local inputs to the MT-like motion filters are (to a first approximation) semi-linear functions of stimulus speed. It is known, however, that many MT neurons are broadly tuned for speed with Gaussian-shaped tuning profiles (Maunsell & Van Essen, 1983; Felleman & Kaas, 1984; Rodman & Albright, 1987). The equations of our model [(8a–c), (10a–c) etc.] can be made compatible with this neuronal behavior by approximating the monotonically raising and descending parts of tuning functions by straight lines and applying the model to each linear piece. We believe that the general result will hold true for non-linear basis velocity tuning functions as well. However, representation of geometry-related parameters must then be distributed among sub-populations of neurons, each of which represents an interval of speed values.

Alternative approaches in neural-like computation of shape-from-motion

An alternative neural-like curvature estimation from optic flow scheme was proposed by Droulez and Cornileau-Perez (1990). Their method estimates the spin variation of optic flow [e.g. in the direction of the x-axis $SV(0) = v^y_{xx}$, the second partial x derivative of the vertical, i.e., y, velocity vector component], which nicely accounts for the asymmetry in 3D shape perception of a cylindrical surface. This result can also be explained as follows: because differential motion detecting neurons are not known to respond in the absence of classical RF motion, there will be zero response along the whole axis of a *vertical* cylinder rotating along the axis positioned on its surface. For a horizontal cylinder, by contrast, the zero response will only be obtained at the point of fixation. This difference may result in underestimation of curvature for the vertical cylinder. Furthermore, the *SV* relates to curvature through the slope-dependent parameter undermining the invariance of this descriptor. Also, equation (17) suggests that in order to obtain the scale-invariant shape index one also needs to compute the second-order derivatives in the direction parallel to the velocity vector (e.g. v^x_{xx}) along with the first-order partial derivatives. Our inclusion of oriented spatial differential operators thus extends the concepts of Droulez and Cornileau-Perez (1990) by suggesting that the shape from motion mechanism computes first- and second-order *partial* derivatives in all directions. This predicted isotropy of differentiation direction is consistent with the neurophysiological data of Xiao *et al.* (1994), who found no correlation between preferred direction and a putative direction of integration of MT neurons.

In their paper on second-order optic flow, Koenderink and van Doorn (1992) have presented an excellent and exhaustive analysis of second-order structure of an arbitrary optical flow induced by a smooth surface under orthographic projection. They correctly derived the second-order dependency between the curvature of a surface and concomitant optic flow, and sketched out a hypothetical neural network possessing center–surround structure capable of estimating the second-order deformation tensor. However, lack of specificity about neuronal properties of their network limits the predictive power of their insightful analysis. Furthermore, confining this abstract network to computation of a set of second-order derivatives, rather than shape invariants, leaves the analysis of potential neuronal mechanisms for SFM incomplete. Our analysis suggests that the network proposed by Koenderink and van Doorn can be constructed from a set of oriented differential motion filters by pooling across filters differentiating in all directions and across all preferred directions. The epipolar direction is then represented by the preferred direction of the maximally excited neuron. In addition, we propose that the actual shape invariants (i.e. principal normal curvatures) may be estimated by setting competition between normalized outputs of oriented second-order filters.

Another alternative shape estimation scheme was proposed by Dijkstra (1994), who demonstrated that the Koenderink–van Doorn shape descriptors [equation (18)] can be obtained from gradients of velocity field divergence (*div*), curl (*rot*), and a double deformation def[*def*()]:

$$|S_e| = \frac{2}{\pi}\arctan\left\{\frac{|grad[div(v)] + J \cdot grad[rot(v)]|}{|def[def(v)]|}\right\},$$

where J is a rotation matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. It is known that area MST neurons are tuned for combinations of divergence and rotation (i.e., to spiral velocity flows) (Duffy & Wurtz, 1991; Graziano, Andersen & Snowden, 1994; Lagae *et al.*, 1994). Differentiating the spatial output map of such cells should yield values that would appear in the denominator of the above equation. However: (1) MST neurons are not as selective for deformation as they are for rotary and radial motion (Lagae *et al.*, 1994), which indicates that they are less responsive to surface shape than to surface orientation; and (2) the retinotopic representation in area MST is coarse, suggesting that neighborhood relations necessary for spatial differentiation may not be preserved, thus undermining the neurobiological plausibility of this computational scheme.

In summary, we propose that the RF surrounds in MT cause these neurons to function as differential operators. The described operators can be thought of as providing a continuous interpolation of cortically represented surfaces. The recent discovery of elongated RFs with flanking surrounds in area MT (Xiao *et al.*, 1994), which were predicted by our initial analysis (Buračas & Albright, 1994) strongly supports our theory of shape-from-motion computation. We have shown that RFs of this variety allow direct estimation of the orientation of velocity field-inducing surfaces. Computation of shape invariants (principal curvature, scale invariant shape descriptors) is notably advanced in this area as well. We do not suggest that the actual estimation of invariants is achieved in area MT; that operation may be carried out in higher tiers of the cortical hierarchy. We conclude, nonetheless, that MT neurons extract information from velocity fields in a manner that is relevant to forming representations of both 'where' and 'what'.

REFERENCES

Adelson, E. H. & Bergen, J. R. (1986). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A, 2*, 285–299.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology, 52*, 1106–1130.

Albright, T. D. (1993). Cortical processing of visual motion. In Walman, J. & Miles, R. A. (Eds), *Visual motion and its use in the stabilization of gaze* (pp. 177–201). Amsterdam: Elsevier.

Albright, T. D. & Desimone, R. (1987) Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Experimental Brain Research, 65*, 582–592.

Albright, T. D., Desimone, R. & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology, 51*, 16–31.

Allman, J., Miezin, F. & McGuinnes (1985). Stimulus specific responses from beyond the classical receptive field. *Annual Review of Neuroscience, 8*, 407–430.

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 Harbor Symposia on Quantitative Biology*, Vol. *LV* (pp. 741–748).

Barlow, H. B. & Foldiak, P. F. (1989). Adaptation and decorrelation in the cortex. In Durbin, R., Miall, C. & Mitchison, G. M. (Eds), *The computing neuron* (pp. 54–72). New York: Addison-Wesley.

Beitas, K. & Kirvelis, D. (1987) Orientational Hermite models of visual receptive fields. In *Visual systems* (pp. 74–87). Vilnius University Press.

Born, R. T. & Tootell, R. B. H. (1992). Segregation of global and local motion processing in primate middle temporal visual area. *Nature, 357*, 497–499.

Born, R. T. & Tootell, R. B. H. (1993). Center–surround interactions in direction-selective neurons of primate visual area MT. *Society of Neuroscience Abstracts, 19*, 315.5.

Buračas, G. T. & Albright, T. D. (1994). The role of MT neuron receptive field surrounds in computing object shape from velocity fields. *Advances in Neural Information Processing Systems, 6*, 969–976.

Carman, G. J. (1991). Relative depth from motion and stereo cues. *Society of Neuroscience Abstracts, 17*, 847.

Cornilleau-Perez, V. & Droulez, J. (1989). Visual perception of surface curvature: Psychophysics of curvature detection induced by motion parallax. *Perception & Psychophysics, 46*, 351–364.

DeYoe, E. A. & Van Essen, D. E. (1985). Segregation of efference connections and receptive field properties in visual area V2 of the macaque. *Nature, 317*, 58–61.

Dijkstra, T. (1994). Visual control of posture and visual perception of shape. Ph.D. thesis, The Netherlands.

Droulez, J. & Cornilleau-Perez, V. (1990). Visual perception of surface curvature. The spin variation and its physiological consequences. *Biology & Cybernetics, 62*, 211–224.

Duffy, C. J. & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli, I and II. *Journal of Neuroscience, 56*, 1329–1345; 1346–1359.

Felleman, D. J. & Kaas, J. H. (1984). Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkey. *Journal of Neurophysiology, 52*, 488–513.

Frost, B. (1978). Moving background patterns after directionally specific responses of pigeon tectal neurons. *Brain Research, 151*, 599–603.

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

Graziano, M. S., Andersen, R. A. & Snowden, R. J. (1994). Tuning of MST neurons to spiral motions. *Journal of Neuroscience, 14*, 54–67.

Hammond, P. (1981). Simultaneous determination of directional tuning of complex cells in cat striate cortex for bar and texture motion. *Experimental Brain Research, 41*, 364–369.

Helmholtz, H. von (1867). *Handbuch der physiologischen optik*, 1st ed. Leipzig: Voss.

Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Societies of America, 72*, 888–892.

Horn, B. K. P. (1986) *Robot vision*. Reading, MA: MIT Press.

Hussain, M., Treue, S. & Andersen, R. A. (1989). Surface interpolation in three-dimensional structure-from-motion perception. *Neural Computation, 1*, 324–333.

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.

Koenderink J. J. & van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biological cybernetics, 55*, 367–375.

Koenderink, J. J. & van Doorn, A. J. (1992). Second-order optic flow. *Journal of Optical Society of America A, 9*, 530–538.

Lagae, L., Gulyas, B., Raiguel, S. & Orban, G. A. (1989). Laminar analysis of motion information processing in macaque V5. *Brain Research, 496*, 361–367.

Lagae, L., Maes, H., Raiguel, S., Xiao, D.-K. & Orban, G. A. (1994). Responses of macaque sts neurons to optic flow components: A comparison of areas MT and MST. *Journal of Neurophysiology, 71*, 1–30.

Lehky, S. & Sejnowski, T. (1988). Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature, 333*, 452–454.

Livingstone, M. S. & Hubel, D. H. (1988) Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240*, 740–749.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society London, B, 208*, 385–387.

Marcar, V. L., Raiguel, S. E., Xiao, D., Maes, H. & Orban, G. A. (1991). Do cells in area MT code the orientation of a kinetic boundary? *Society of Neuroscience Abstracts, 17*, 525.

Marr, D. & Hildredth, E. (1980). Theory or edge detection. *Proceedings of the Royal Society London, B207*, 187–217.

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. *Journal of Neurophysiology, 49*, 1127–1147.

Mortenson, M. E. (1985). *Geometric modeling*. New York: Wiley J.

Movshon, J. A., Adelson, E. A., Gizzi, M. & Newsome, W. T. (1985). The analysis of moving visual patterns. In Chagas, C., Gattass, R. & Gross, C. G. (Eds), *Study group on pattern recognition mechanisms* (pp. 117–151). Vatican City: Pontifica Academia Scientarum.

Nagel, H.-H. (1987). On estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence, 33*, 299–324.

Nakayama, K. & Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception, 3*, 63–80.

Rodman, H. R. & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the Macaque. *Vision Research, 27*, 2035–2048.

Rogers, S. & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*, 125–134.

Rukšėnas, O., Šatinskas, R. & Stabinytė, D. (1994). Inhibitory zones in the receptive fields of visual neurons determined by moving stimuli. *Perception (suppl.), 23*, 66.

Siegel, R. M. & Andersen, R. A. (1987) Motion perceptual deficits following ibotenic acid lesions of the middle temporal area in the behaving rhesus monkey. *Society of Neuroscience Abstracts, 12*, 1183.

Spitzer, H. & Hochstein, S. (1985). A complex-cell receptive field model. *Journal of Neurophysiology, 53*, 1266–1286.

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

Treue, S. & Andersen, R. A. (1993). Tuning of MT cells to velocity gradients. *Investigative Ophthalmology and Vision Science (Suppl.), 34*, 813.

Treue, S., Andersen, R. A., Ando, H. & Hildredth, E. C. (1995). Structure from motion: Perceptual evidence for surface interpolation. *Vision Research, 35*, 139–148.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ungerleider, L. G. & Mishkin, M. (1979). The striate projection zone in the superior temporal sulcus of *Macaca mulatta*: location and topographic organization. *Journal of Computational Neurology, 188*, 347–366.

Ungerleider, L. G. & Mishkin, M. (1982). Two cortical systems. In *Analysis of visual behavior*, Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W. (Eds), (pp. 549–586). Cambridge, MA: MIT Press.

Werkhoven, P. & Koenderink, J. J. (1990). Extraction of motion

parallax structure in the visual system I. *Biological Cybernetics, 63*, 185–191.

Wurtz, R. H. Yamasaki, D. S. Duffy, C. J. & Roy, J.-P. (1990). Functional specialization for visual motion processing in primate cerebral cortex. In *The brain. Cold Spring Harbor Symposia on Quantitative Biology*, Vol. *LV*, (pp. 717–728).

Xiao, D.-K., Marcar, V. L., Raiguel, S. E. & Orban, G. A. (1994). Does the surround really surround the classical receptive field of macaque MT cells? *Society of Neuroscience Abstracts, 20*, 773.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology, 236*, 827–832.

Zemel, R. S. & Sejnowski, T. J. (1995). Grouping components of three-dimensional moving objects in area MST of visual cortex. In *Advances in neural information processing systems*, Vol. 7, in press.

Zhang, K. C., Sereno, M. I. & Sereno, M. E. (1993). Emergence of position-independent detectors of sense of rotation and dilation with Hebbian learning—an analysis. *Neural Computation, 5*, 597–612.

Acknowledgements—We would like to thank Rich Zemel for leading us through the maze of realistic dynamical scene creation and optic flow estimation. We are grateful to George Carman, Lisa Croner, Steve Nowlan, Gene Stoner, and Rich Zemel for stimulating discussions and helpful comments on the manuscript. This project was sponsored by a grant from the National Eye Institute (TDA) and by scholarships from the Lithuanian Foundation and the Chapman Charitable Trust (GTB). The movie of a rotating teapot was created using POVRAY raytracing program.

APPENDIX

Vector field components [equation (1)]

The two components of the vector field under perspective projection are equal to (Horn, 1986):

$$u = \frac{\dot{x}}{(z+z_0)} - \frac{x\dot{z}}{(z+z_0)^2},\quad v = \frac{\dot{y}}{(z+z_0)} - \frac{y\dot{z}}{(z+z_0)^2}.$$

Equation (2) can be readily achieved by plugging the three vector components of equation (1) into the above equations and assuming that $\mathbf{w} = [w^x, w^y, 0]$ and $\mathbf{t} = [0,0,t^z]$.

Convolution

The standard convolution integral between an arbitrary function $u(i, j)$ and Gaussian:

$$G(\sigma) = G(x_0, y_0, \sigma) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right]$$

is defined as:

$$u(i,j) * G(\sigma_k) = \frac{1}{2\pi\sigma_k^2}\int\int_N u(i-x, j-y)G(x, y, \sigma_k)\mathrm{d}x\mathrm{d}y.$$

Normal curvature [equation 6]

The standard formula for normal curvature in direction of angle φ:

$$\kappa(\varphi) = \frac{z_{xx}\cos^2\varphi + 2z_{xy}\cos\varphi\sin\varphi + z_{yy}\sin^2\varphi}{\sqrt{1+z_x^2+z_y^2}(1+z_x^2\cos^2\varphi + z_y^2\sin^2\varphi + 2z_xz_y\cos\varphi\sin\varphi}, \quad \text{(A1)}$$

Mortenson (1985) greatly simplifies by noticing that the term in parentheses of the denominator is simply $1 + z_\varphi^2$ (i.e. squared derivative in direction of angle φ), and the expression in the numerator is simply $z_{\varphi\varphi}$. Equation (6) can be readily achieved from the standard formula (A1) for normal curvature by inserting $\varphi = 0$. The mean curvature for the particular case of the maximal normal curvature aligned with the x-axis is:

$$\overline{\kappa} = \frac{z_{xx}}{2\sqrt{1+z_x^2+z_y^2(1+z_x^2)}} + \frac{z_{yy}}{2\sqrt{1+z_x^2+z_y^2(1+z_y^2)}} = \frac{\nabla^2 z + z_{xx}z_y^2 + z_{yy}z_x^2}{2(1+D^2z+z_x^2z_y^2)\sqrt{1+D^2z}}, \quad \text{(A2)}$$

where $\nabla^2 z = z_{xx} + z_{yy}$ and $D^2 z = z_x^2 + z_y^2$. Now, for any direction φ and direction ϕ perpendicular to φ (i.e. $\phi = \varphi + \pi/2$) we can calculate the mean curvature by replacing z_{ij} and z_i derivatives ($i,j = x,y$) by z_{ij} and z_i with $i,j = \varphi,\phi$. Similarly, the scale-invariant shape index for $\kappa_{\min}$ and $\kappa_{\max}$ aligned with coordinate axes can be expressed as:

$$\Omega = \frac{2}{\pi}\arctan\left(\frac{\kappa_{\max}+\kappa_{\min}}{\kappa_{\max}-\kappa_{\min}}\right) = \frac{2}{\pi}\arctan\left[\frac{(z_{xx}+z_{yy})+(z_{xx}z_y^2+z_{yy}z_x^2)}{(z_{xx}-z_{yy})+(z_{xx}z_y^2-z_{yy}z_x^2)}\right],$$

and partial derivatives here can also be replaced by directional derivatives in both direction of the maximal curvature and the direction perpendicular to it [as in equation (16)].

Derivatives of the optical flow

Differentiating the optical flow components u and v in the x and y directions once yields:

$$u_x = -\frac{wz_x}{z_0} - \frac{-t^z + 2wx}{z_0^2},\quad u_y = -\frac{wz_y}{z_0},\quad v_x = -\frac{wy}{z_0^2},\quad v_y = -\frac{-t^z + wx}{z_0^2}. \quad \text{(A3)}$$

Combinations of these derivatives define the affine structure of the vector field (Koenderink and van Doorn, 1975). The derivatives of the vertical component v_x and v_y and the second term of u_x scale with distance from the fixation point and depend on z_0 as $O(z_0^{-1})$. Confining analysis to a limited visual angle around the fixation spot allows the assumption that $v_x = 0$ and $v_y = 0$; u_x and u_y then become proportional to z_x and z_y. Second-order derivatives can be estimated in a similar way:

$$u_{xx} = -\frac{w}{z_0}z_{xx} - \frac{2w}{z_0^2},\quad u_{xy} = -\frac{w}{z_0}z_{xy},\quad u_{yy} = -\frac{w}{z_0}z_{yy},\quad v_{xx} = 0,\quad \boxed{v_{xy} = -\frac{w}{z_0^2}},\quad v_{yy} = 0. \quad \text{(A4)}$$

We see that derivatives of u, after discarding $O(1/z_0^2)$ terms are proportional to corresponding derivatives of $z(x,y)$ with the coefficient $-w/z_0$. Thus, we can express the MacLauren expansion of $u(x,y)$ as in equation (11).

Interestingly, the mixed derivative of the vertical component can help estimate the value of angular velocity of effective rotation w: $w = -v_{xy}z_0^2$ [bottom center equation of (A4)].

After discarding the second-order term, equations (A3) and (A4) immediately suggest simplified solutions for partial derivatives of the surface $z(x,y)$, required in formulas for slant, tilt and curvature. If needed, z_x and z_y can be estimated more precisely, with the second-order term included:

$$z_x = -\frac{z_0}{w}(u_x - v_y) - \frac{x}{z_0} \text{ and } z_y = \frac{z_0}{w}(-u_y + v_x) + \frac{y}{z_0}, \quad \text{(A5)}$$

where u and v are as in equation (6). In this case, estimates of derivatives of the flow field can be received from the outputs of oriented filters, while the second term may be discarded for $x,y \ll z_0$.

Biologically plausible estimate of arctan

Both arctangent and hyperbolic tangent functions can be expressed as:

$$\arctan(x) \approx x - x^3 + O^5(x) \approx \tanh(x). \quad \text{(A6)}$$

The two functions can further be brought closer to each other by adjusting the T parameter of equation (9). Furthermore, since $\tanh(x) = [\tanh(x) - \tanh(-x)]/2$, we can approximate the arctangents, which show up in expressions for geometrical descriptors [equations (5), (17)] by a sum of two sigmoidal functions, one representing positive, and the other representing negative parts of arctan: $\arctan(x) \approx [S(x) - S(-x)]/2$. Here b is set to $b = 0$, and T minimizes $\arctan(x) - [S(x) - S(-x)]/2$. Thus, sigmoidal activation functions can serve as a biological substitute for arctan.