

The ‘ecological’ probability density function for linear optic flow: Implications for neurophysiology

Jim Ivins

Department of Computer Science, Curtin University of Technology, GPO Box U 1987, Perth 6845, Western Australia; e-mail: ivinsj@cs.curtin.edu.au

John Porrill, John Frisby

Department of Psychology, University of Sheffield, Sheffield S10 2TP, UK

Guy Orban

Laboratorium voor Neuro- en Psycho-fysiologie, Katholieke Universiteit te Leuven, Campus Gasthuisberg, B 3000 Leuven, Belgium

Received 10 June 1998, in revised form 2 December 1998

Abstract. A theoretical analysis of the recovery of shape from optic flow highlights the importance of the deformation components; however, pure deforming stimuli elicit few responses from flow-sensitive neurons in the medial superior temporal (MST) area of the cerebral cortex. This finding has prompted the conclusion that MST cells are not involved in shape recovery. However, this conclusion may be unjustified in view of the emerging consensus that MST cells perform nonlinear pattern matching, rather than linear projection as implicitly assumed in many neurophysiological studies. Artificial neural models suggest that the input probability density function (PDF) is crucial in determining the distribution of responses shown by pattern-matching cells. This paper therefore describes a Monte-Carlo study of the joint PDF for linear optic-flow components produced by ego-motion in a simulated planar environment. The recent search for deformation-selective cells in MST is then used to illustrate the importance of the input PDF in determining cell characteristics. The results are consistent with the finding that MST cells exhibit a continuum of responses to translation, rotation, and divergence. In addition, there are negative correlations between the deformation and conformal components of optic flow. Consequently, if cells responsible for shape analysis are present in the MST area, they should respond best to combinations of deformation with other first-order flow components, rather than to the pure stimuli used in previous neurophysiological studies.

1 Introduction

Koenderink and van Doorn (1975) showed that optic flow over a small area of the visual field can be decomposed into two translational components, and four first-order elementary flow components (EFCs) shown in figure 1. The EFCs are circular motion or rot (rotation), radial motion or div (divergence), and two components of shearing motion def+ and def× (deformation). Together, rot and div are called the conformal components of flow; def+ and def× are the deformation components.

A computational analysis of the ‘shape from flow’ problem reveals that the deformation components are the least dependent on ego-motion.⁽¹⁾ [Koenderink (1986) showed

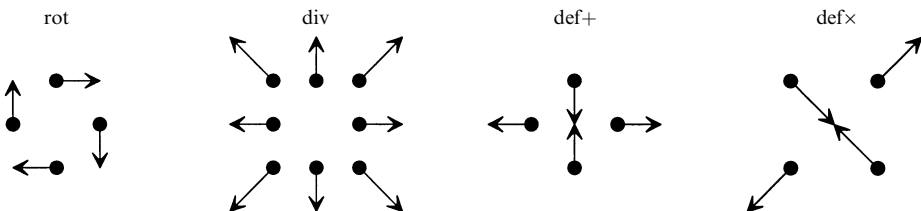


Figure 1. Elementary flow components. The four first-order components of optic flow: rotation, divergence, and two forms of shearing. The arrows indicate the displacement of a square relative to its centroid.

⁽¹⁾ In this context ‘shape’ refers to the local three-dimensional structure of surfaces in the environment.

that all four EFCs depend on the structure of the environment, and on the translational component of ego-motion in the frontoparallel plane; however, only the conformal components depend on other aspects of the relative movement between the observer and the environment.] Consequently, def^+ and def^\times are the most useful EFCs for recovering surface shape. Deformation-selective cells are therefore expected to be present in areas of the brain that compute shape from optic flow. This analysis has prompted the search for neurons that respond to deforming flow fields.

1.1 *The medial superior temporal area*

Neurons in the medial superior temporal (MST) area of the primate cerebral cortex have large receptive fields and are directionally selective for moving stimuli. Duffy and Wurtz (1991a) tested 220 MST neurons using a set of dynamic stimuli 100 deg in diameter. About a quarter of the cells responded primarily to one component of motion (planar, circular, or radial), one third responded to two components (plano-circular or plano-radial, but never circulo-radial), and one third responded to all three components. Although Duffy and Wurtz did not examine responses to deforming stimuli, their study is typical of many which suggest that MST neurons contribute to the analysis of large-field optic flow. However, it is not known for certain whether MST is involved in computing motion or shape (or both) from flow.

When a cell responds both to translation and to EFCs it is possible that the EFC response is caused either by positioning the stimuli off-centre or, when positioning is correct, by an asymmetry in the translation receptive field. An MST cell which is selective only for translation might therefore give a continuum of responses to translation, rotation, and divergence stimuli. To avoid this possibility, Lagae et al (1994) compared the receptive field maps for translation and for EFCs in 82 cells from the MST area. Direction selectivity for an EFC was position invariant in 40% of the cells; these were considered EFC-selective. Most of the EFC-selective cells responded to a single component, sometimes combined with translation. However, only 3 EFC-selective cells responded to deforming stimuli; these cells also responded to either translation or rotation. The fact that relatively few deformation-selective cells were found was interpreted as evidence that MST is not involved in recovering shape from flow. However, Lagae et al used *pure* deformation stimuli which, as we shall argue in subsections 1.2 and 1.3, may not be appropriate. The first argument is based on an analysis of input selectivity; the second is based on consideration of the optic-flow environment and hence on the probabilities of various combinations of EFCs occurring naturally.

1.2 *Neural architecture: pattern matching versus projection*

There is an important difference between the notion of a neuron selective for deformation as specified by Lagae et al (1994), and the notion of a deformation projector which is implicit in many other studies. The first is a nonlinear template for a limited range of optic flow which may include translation and conformal components in addition to deformation. The second is a linear projector which responds only to the deformation component of any input, ignoring translation and conformal flow. This is similar to the distinction made by Duffy and Wurtz (1991b) between the *direction mosaic* and *vector field* hypotheses for flow selectivity.⁽²⁾

Figure 2a shows the four-dimensional EFC space represented in two dimensions, with the horizontal axis devoted to the deformation (Def) components and the vertical

⁽²⁾The direction mosaic (pattern-matching) hypothesis states that the receptive fields of MST cells contain direction-selective subfields which match the local directions of motion within optic-flow fields. The vector field (projection) hypothesis states that the receptive fields are uniquely sensitive to distributed properties of planar, circular, or radial flow. Duffy and Wurtz (1991b) therefore examined the optic-flow selectivity of small subfields within the large receptive fields of 160 MST neurons; however, the results were not entirely consistent with either of the hypotheses.

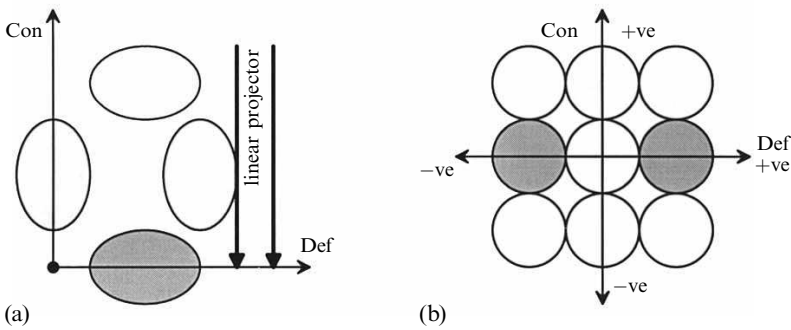


Figure 2. Pattern matchers and projectors. (a) The receptive fields of four pattern-matching cells (ellipses) in EFC space. One cell is selective for pure deformation (shaded); two of the others respond to mixtures of deformation with conformal components. The ‘receptive field’ of a deformation projector is also shown (arrows). The projector ignores conformal components, effectively projecting optic flow onto the horizontal (Def) axis so that its response is governed solely by the deformation components. (b) A coarse coding of EFC space with a 3×3 array of pattern-matching cells; those selective for pure deformation are shaded.

axis devoted to the conformal (Con) components. A pattern-matching neuron selective for pure deformation would have its receptive field oriented along the horizontal axis. The response of this cell would depend on both the deformation and the conformal components of its input. [Such tuned cells can be modelled by hyper-radial basis functions with ellipsoidal receptive fields—see Girosi et al (1995).] In contrast, a linear projector for deformation would give the same response regardless of the conformal components of its input. The receptive field of a deformation projector is well-localised in Def space, so that deformation must be present to elicit a response; however, it is not well-localised in Con space, so the response is not affected by rotation or divergence. It is difficult to see how a projector could be built other than by OR-ing together the outputs of many pattern-matching cells. The common assumption that pure deformation stimuli are optimal for identifying deformation-selective cells is only justified if the cells behave like linear projectors. However, several recent studies have suggested that MST cells behave more like nonlinear pattern matchers, responding to a limited range of input flow patterns—for example, see Orban et al (1992), and Perrone and Stone (1994, 1998).⁽³⁾ Experiments using pure deformation stimuli are therefore only suitable for identifying cells selective for deformation with little or no translation, rotation, or divergence; other deformation-selective cells will be much less responsive to pure stimuli. For example, a pattern-matching cell with mixed selectivity for def+ combined with rot will exhibit small responses to both components in isolation; however, it will respond most strongly to a mixture of the two.

A simple combinatorial analysis suggests that pattern-matching cells selective for pure deformation will be relatively rare. For example, a coarse coding of EFC space by cells which each register conformal and deformation components as positive, zero, or negative would give the partition shown in figure 2b. In this two-dimensional illustration only $3 - 1 = 2$ out of $3^2 = 9$ cells are selective for pure deformation. In four dimensions only $3^2 - 1 = 8$ out of $3^4 = 81$ cells would be selective for pure deformation—those in the (def+, def×) plane, excluding the central zero-flow cell. This proportion obviously decreases with finer (uniform) coding of the input space, which is considered in the next subsection.

⁽³⁾The concept of optic-flow templates (pattern matchers) has been around for some time—for example, see Saito et al (1986) and Tanaka et al (1986, 1989). However, it was not widely accepted until recently because the details of how such templates could be constructed were not formalised. Perrone and Stone (1994, 1998) eliminate this deficit by describing a convincing simulation which demonstrates the possible role of MST in template-based heading estimation.

1.3 *The 'ecological' PDF for linear optic flow*

Consider a neural module that performs some information-processing task. In a natural environment some inputs will be more likely than others. The probability that the module will receive an input in a small volume of the input space is proportional to that volume, with the coefficient being the local probability density function (PDF) of the inputs. If the neural module acts as an intermediate representation of the input, with its output being used by a range of other modules, then it should represent the input so as to maximise the amount of information carried by each output. Linsker (1989) has shown that when the output representation is a winner-take-all place-coding scheme (using neurons with Gaussian receptive fields) the configuration which maximises information is one in which activation of each node is equally likely. This configuration can be achieved by unsupervised learning algorithms—for example, see Kohonen (1988)—which encode input–output relationships by placing more nodes where the relationships are more complex. The density of neurons in a volume of input space should therefore be proportional to the input PDF there.⁽⁴⁾

As an organism develops, it will establish a PDF for the six components of linear flow generated by moving through its environment. Individual neurons should arrange themselves to sample this 'ecological' PDF optimally. This assumption leads to the prediction that 'nonecological' flow space which is devoid of inputs will not be populated at all, while space that is dense with flow inputs will be well populated by neurons. Pure deformation fields tend to arise only from very unlikely combinations of motion and shape parameters, and also tend to have very short durations. [For example, when approaching a planar surface, the apparent expansion can only be removed if the surface rotates to foreshorten one dimension at exactly the rate necessary to leave pure deformation. (Lines perpendicular to the rotation axis must contract at twice the rate that lines parallel to the axis expand.) Such motion will soon result in the plane being parallel to the line-of-sight, at which time the flow vanishes.] Cells selective for pure deformation will therefore have very little work to do, and so are likely to be rare. Thus, rather than being selective for pure deformation, MST cells might instead be tuned for mixtures of deformation with other flow components.

1.4 *Aims*

The prevalent theory in the neurophysiological literature is that MST plays an important role in motion analysis—for example, see Tanaka et al (1989) or Perrone and Stone (1994, 1998). However, a complete description of the characteristics shown by a neural population depends crucially on at least three pieces of information: the nature of the task(s) for which the cells are used, the architecture of the computational units, and the environment in which the computation is performed. In particular, the distribution of responses shown by flow-selective cells is likely to be governed by their pattern-matching architecture and by the distribution of inputs in EFC space. No argument based on the observed abundance, or otherwise, of deformation-selective cells is safe until both these characteristics have been considered. Hence the conclusion that MST does not contribute to shape analysis may be unjustified.

From the arguments developed in this section it seems that a computational study of the 'ecological' PDF for optic flow might provide interesting quantitative data suitable for comparing with, and making predictions about, biological findings. The remaining

⁽⁴⁾ This conclusion is invalid if the output is used for a very restricted range of tasks. For example, consider an error-correcting cell such as a 'falling-over' detector. The situations in which this cell will respond have low probability, but are vitally important. In general, supervised training regimes which minimise the error in performing a task will not lead to equiprobable neuron configurations. However, other things being equal, neurons should still be more dense where the PDF is high than where it is low. The broader the range of tasks to which a module contributes, the safer this conclusion is.

sections of this paper therefore describe a Monte-Carlo simulation designed to generate the PDF for linear optic-flow components in PDF in ecologically plausible settings; a summary of this work is given by Ivins et al (1998).

The study by Lagae et al (1994) is used to investigate the possibility that the ‘ecological’ PDF of a perceptual stimulus (in this case optic flow) may be crucial in determining the characteristics of cells which respond to that stimulus. Predictions about deformation selectivity based on the PDF of optic flow may or may not provide some insight into the function of MST cells. More generally, however, understanding the ecology of any perceptual stimulus may prove invaluable for interpreting experimental results in psychophysics, neurophysiology, and related disciplines. In this wider context, the relationship between deformation, MST, and shape recovery is a side issue, albeit a very interesting and important one.

2 Method: Monte-Carlo simulation

The PDF for linear optic-flow components was approximated by Monte-Carlo methods—see Press et al (1992). This section describes the main features of the simulation, which is (very crudely) intended to resemble a primate moving through a ‘forest’ and ‘savannah’ environment; further mathematical details are given in the appendix.

2.1 *Simulating the environment*

For comparison with existing neurophysiological data, only the zero-order and first-order components of optic flow were simulated. These are often the highest-order components considered robust enough to be useful—for example, see Verri et al (1992). Furthermore, restricting the study to linear flow offers the advantage that an environment can be simulated with planes as shown in figure 3.

The simulated environment is typically based on a ground plane $100\text{ m} \times 100\text{ m}$ square, surrounded by four planar walls, with a ‘sky’ plane at a height of 100 m. Exploration is restricted to the ground plane, so the sky is always effectively at infinity.

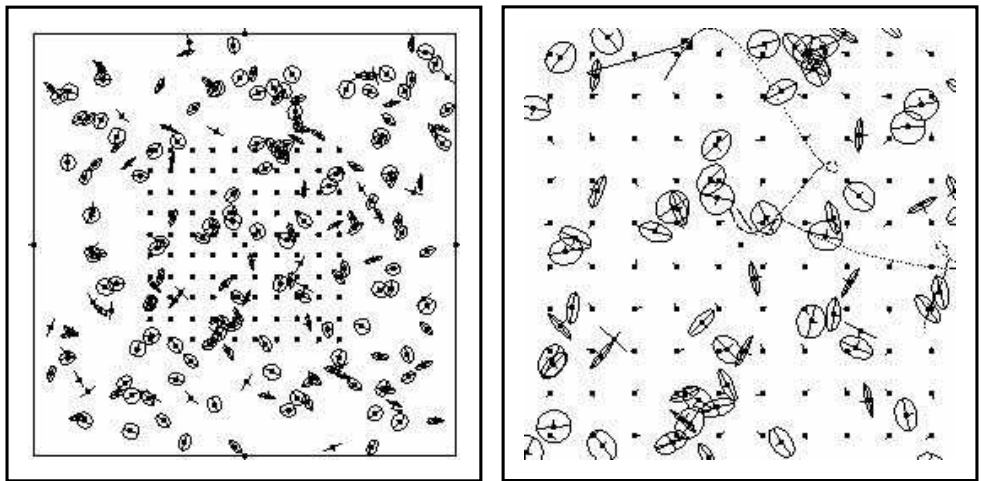


Figure 3. Ego-motion in a simulated environment. (a) An aerial view of a typical simulated planar environment. The ground plane is littered with 200 disks, each 2 m in diameter. The central area is covered by a canopy of planes each 5 m across arranged in a 10×10 grid (represented by black dots). (b) An enlarged view of the central part of the ground plane. The current position of the eye is shown (top centre) as a black square from which two vectors project; previous positions are shown as a trail of dots. The longer vector shows the line-of-sight from the eye to the current fixation point; the shorter vector indicates the current velocity.

The environment is littered with between 50 and 200 planar disks arranged with random positions and orientations;⁽⁵⁾ each disk is 2 m in diameter with its centre 1 m above the ground plane. There is also a 10×10 grid of planes which form a canopy over the central $50 \text{ m} \times 50 \text{ m}$ area to obscure the sky when viewed from below. Each canopy plane is 5 m across and has a random orientation within 30° of horizontal, with its centre 10 m above the ground plane. [The canopy planes might represent trees 10 m high; they are not exactly horizontal because, when viewed from below, a tree consists of numerous ‘planes’ (leaves) with random orientations.]

The optic-flow mechanisms described in subsection 2.3 respond to all of the bounding planes (the four walls, ground plane, and sky), and all of the planar disks near the ground plane and in the ‘forest’ canopy. However, the simulation can be varied in a number of ways—for example, by altering the number of disks near the ground plane or in the canopy (possibly removing these features altogether) or by altering the characteristics of each disk, such as its size, position, and orientation.

2.2 Simulating motion of the eye

The kinematic chain used to model ego-motion consists of a single eye that can pitch and yaw, connected to a head that has position \mathbf{t} (coincident with the eye) and three translational degrees of freedom:

$$\mathbf{t} = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}, \quad \text{where } y(t) = h_0 + h_1 \sin(\omega_1 t). \quad (1)$$

The height $y(t)$ of the eye above the ground plane is h_0 (typically 1 m) plus a sine-wave oscillation with amplitude h_1 (typically 0.5 m) intended to simulate the head-bobbing and posture changes that occur during locomotion over uneven terrain.

Deterministic ego-motion over the (x, z) ground plane is specified directly by two Fourier series of the form:

$$\frac{1}{2}a_0 + \sum_{k=1}^n a_k \cos(k\omega_2 t) + b_k \sin(k\omega_2 t). \quad (2)$$

The coefficients ω_1 , ω_2 , a_0 , a_k , b_k , and n are chosen randomly but constrained so that the eye cannot leave the environment, and the translational and angular velocities stay within realistic limits. The maximum translational velocity is less than 4 m s^{-1} , and the maximum angular velocity is under $\pi/2 \text{ rad s}^{-1}$. Motion is always forward in the simulation; backward motion (like falling over) is regarded as an exceptional case. The overall behaviour can be regarded as pseudo-random exploration of the environment.

The *pose* of the eye is specified by a rotation matrix \mathbf{R} composed of orthonormal vectors \mathbf{e}_x , \mathbf{e}_y , and the line-of-sight \mathbf{e}_z :

$$\mathbf{R} = \begin{pmatrix} \mathbf{e}_x \\ \mathbf{e}_y \\ \mathbf{e}_z \end{pmatrix}. \quad (3)$$

As the eye moves, it fixates a point in the world using Fick (‘gun turret’) movements with two degrees of freedom—horizontal and vertical orientation; there is no cyclo-torsion. The pose of the eye is therefore completely specified by its position \mathbf{t} and the fixation point \mathbf{f} :

$$\mathbf{e}_z = \mathbf{f} - \mathbf{t}, \quad \mathbf{e}_x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \times \mathbf{e}_z, \quad \mathbf{e}_y = \mathbf{e}_z \times \mathbf{e}_x. \quad (4)$$

⁽⁵⁾In this context ‘orientation’ refers to the ‘pose’ of a plane—the direction of its normal vector (and hence its roll, pitch, and yaw), which is selected at random.

The fixation point is tracked until either it is too close to the eye, or the line-of-sight becomes too eccentric (more than 45°) relative to the direction of motion. (There is no obstacle-avoidance mechanism, so planes less than 0.5 m from the eye become invisible, prompting re-fixation; this strategy eliminates the unnatural flows that would arise from moving through objects.) Re-fixation on the nearest plane in the direction of motion occurs when either of these constraints is violated. Vertical eye orientation is chosen randomly each time the eye re-fixates, with a pitch range of $\pm 45^\circ$ above and below horizontal.

2.3 Simulating the retina

The optic flow on the retina is coarsely sampled at 9 or 25 positions with a regular 3×3 or 5×5 grid of receptors. The grid spacing is 20 deg of visual angle, so a 5×5 grid covers ± 40 deg from the fovea. The whole grid of receptors therefore spans 80 deg vertically and horizontally, which is roughly the size of an MST receptive field. (MST neurons can have receptive field diameters of up to 100 deg.) However, the receptors are not intended to simulate MST neurons; they merely sample the PDF for optic flow in the simulated environment. The grid was introduced to investigate variations in flow distribution with retinal location; for simplicity, however, this aspect of the analysis will not be emphasised.

Gaze stabilisation is calculated for the central receptor which is regarded as a fovea. The flow calculation is performed by ray-tracing from each receptor to find the closest plane along the line-of-sight; optic-flow components are then calculated directly for that plane by the formula of Longuet-Higgins and Prazdny (1980) as shown in the appendix. Each of the cells in figure 4 shows the optic flow at one of the point receptors in a 3×3 grid. The central vector in each cell gives the translational flow at

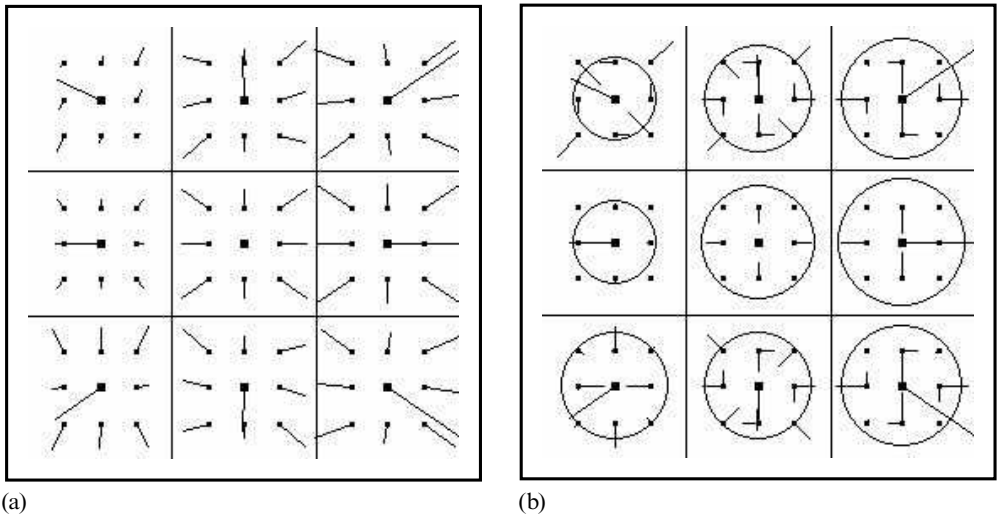


Figure 4. Linear flow components. Two graphical representations of a typical optic-flow field from the Monte-Carlo simulation. Receptors are arranged in a 3×3 grid of retinal positions at 20 deg offsets around the central fovea. For each receptor, a central vector indicates the translation (this is zero at the fovea); the surrounding vectors show the first-order flow. (a) The starting points of the eight radial vectors join to form a square; the ends of these vectors form a parallelogram (an affine distortion of the square), the shape of which indicates the first-order flow registered by the receptor. (b) An invariant decomposition of this flow. The rot component is shown by tangent vectors arranged at N, E, S, and W compass points. A circle passing through the starting points of the rotation tangents indicates zero divergence; a larger circle indicates positive div and a smaller circle indicates negative div. Deformation vectors are arranged at N, E, S, and W for def+; and at NE, SE, SW, and NW for def \times .

that receptor; the surrounding vectors in (a) show (magnified) first-order flow, while the vectors and circles in (b) show the equivalent EFCs. These displays change dynamically as the environment is explored, revealing the wealth of information available from the optic-flow simulation. In reality an environment consists of many microfacets, and optic flow at a retinal receptor is obtained by integrating over all such facets in its receptive field. However, receptor size is not an important issue when using a small number of planar facets to simulate an environment. A receptive field in the simulation will rarely overlap multiple planes, so it can be treated as a single point. This simplification sidesteps the issue of integrating flow to simulate a large receptive field, and with good reason: an obvious method, such as averaging over many point receptors, is probably an oversimplification of a complicated nonlinear process. Note that the simulation is conservative as far as the recovery of correlations between flow components is concerned, in that results are more random than would be obtained in reality. (Integration over microfacets would tend to increase correlations rather than decrease them, being biased towards the production of planes frontoparallel to the line-of-sight.)

3 Results

Flow was generated over a simulated time of 5 min in each of ten different environments. To avoid aliasing artifacts, flow samples were calculated randomly, once every second on average. (The time step for numerical differentiation was 0.04 s, giving a maximum possible sampling rate of 25 Hz.) Results from many different environments have been compared, and those shown in this section are typical. A single flow sample from one receptor produces a single dot in each graph.

3.1 Translation

Figure 5, which has the same layout as the grids in figure 4, shows the joint PDF for x and y translation, broken down by receptor. The receptor at the fovea receives negligible translation because the optic behaviour is restricted to fixation with occasional (blind) saccades. In contrast, receptors at the periphery receive large translational flows.

For simplicity, the analysis of first-order components in the remainder of this section excludes the translation components; note, however, that pure deformations can only occur when the translation components are both zero. Translation is considered further in subsection 4.3.

3.2 Elementary flow components

Since the recovered EFCs form a four-dimensional data set, graphical presentation is problematic. Pairs of variables are therefore shown with the aid of two-dimensional density plots, which are sufficient to illustrate the correlations in the data. Each plot is an approximation to the marginal distribution of the variables chosen—that is, the first-order PDF with the other two variables integrated out.

Figure 6a shows the joint PDF of the conformal components. There is no obvious correlation between rotation and divergence, though divergence is predominantly positive because the eye is usually moving towards visible planes. Occasionally, negative divergence is produced when the angle between the direction of motion and the line-of-sight is large (near 45°); under these circumstances it is possible for a peripheral receptor to move away from the plane it sees. The PDF is dense around the origin as is necessary for the presence of pure deformations (which can only occur when both rotation and divergence are zero). However, these samples might simply represent flows for which all EFCs are small.

Figure 6b shows the joint PDF of the deformation components. Again, there is no obvious correlation between the two components. Because the ground plane forms the lower bound of the environment, it is more often than not visible between 0.5 and 1.5 m below the simulated eye (though it is sometimes obscured by the obstacles which

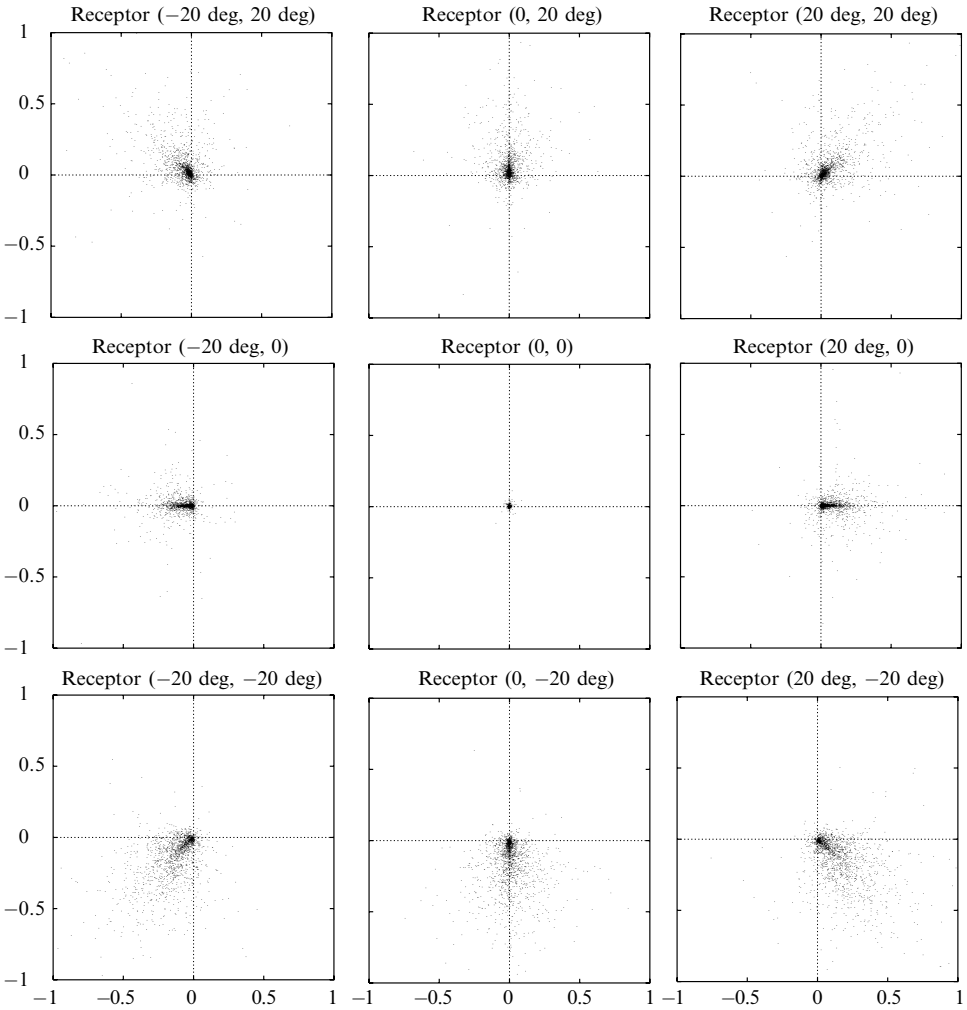


Figure 5. Translation. The translation at each of eight peripheral receptors arranged at 20 deg offsets around the central fovea. The horizontal axes show x -translation, and the vertical axes show y -translation. There is almost no translation at the fovea (the eye is usually fixating); however, peripheral receptors receive large translational flows.

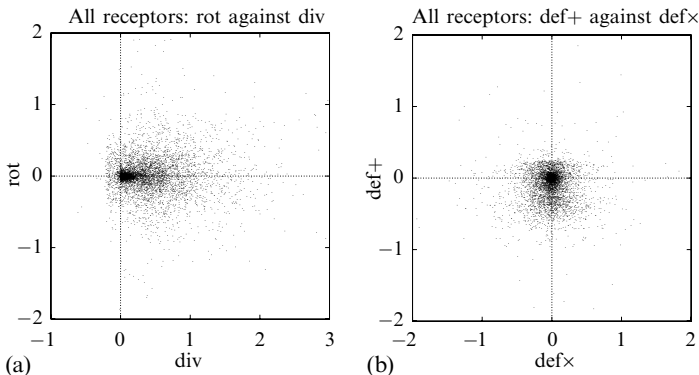


Figure 6. Conformal and deformation components. (a) Marginal distribution of conformal components (all receptors). There is no correlation between these components; however, div tends to be positive. (b) Marginal distribution of deformation components (all receptors). There is no correlation between these components; however, $def+$ tends to be negative.

are randomly scattered 1 m above its surface); def+ therefore tends to be negative [see equation (A9) in the appendix].

Figure 7 supports the hypothesis that pure deformations occur very rarely; it shows the magnitudes of the conformal and deformation components, computed as follows:

$$|\text{Con}| = (\text{rot}^2 + \text{div}^2)^{1/2}, \quad |\text{Def}| = (\text{def}+^2 + \text{def}\times^2)^{1/2}.$$

Almost all flow samples lie below the diagonal line $\text{rot}^2 + \text{div}^2 = \text{def}+^2 + \text{def}\times^2$ on which these magnitudes are equal. The upper half of the plot is almost unoccupied, confirming that deformation components rarely occur unless accompanied by larger conformal components. In contrast, relatively pure rotation or divergence (or at least a deformation-free mixture of the two) are quite likely to occur. Pure deformation-selective cells, with receptive fields positioned along the vertical axis of this plot, would receive very few appropriate inputs; they are therefore likely to be rare. Deformation-selective MST cells are, instead, likely to respond best to mixtures of def+ or def \times with other first-order flow components.

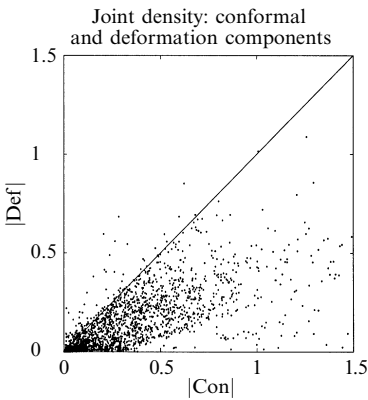


Figure 7. Conformal and deformation magnitudes. Marginal distribution of conformal (Con) and deformation (Def) magnitudes (for all receptors). The horizontal axis shows the combined rot and div components of each flow measurement; the vertical axis shows the combined def+ and def \times components. The Con component is usually larger than the corresponding Def component.

The nature of the coupling between conformal and deformation components is clearly seen in figure 8 which reveals some interesting correlations. Figure 8a shows the joint PDF of div and def+. The distribution of def+ is asymmetrical (tending to be negative), and is inversely correlated with div (which is usually positive). This may at least partly reflect the fact that the term $N_y v_y$ appears (with different signs) in the formulae for both components given in the appendix; the term $N_x v_x$ which also

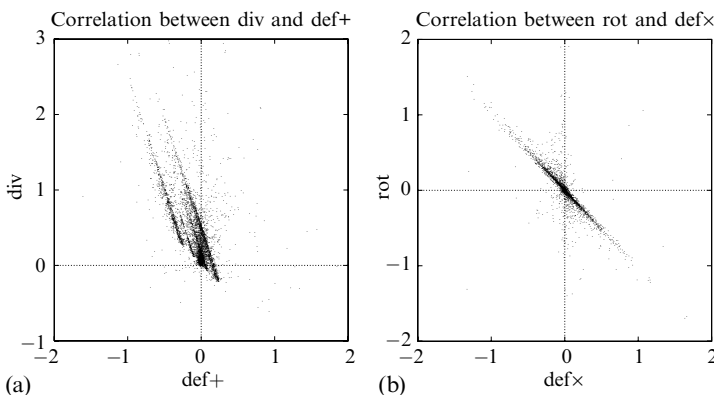


Figure 8. Correlations between EFCs. (a) Marginal distribution of div and def+, which are inversely correlated; the divergence usually exceeds the associated deformation in magnitude. (b) Marginal distribution of rot and def \times , which are inversely correlated; the rotation usually exceeds the associated deformation in magnitude.

appears in these formulae is often small in comparison. Figure 8b shows the joint PDF of rot and def \times . The distribution of def \times is symmetrical, and is inversely correlated with rot. This may at least partly reflect the presence of the term $N_y v_x$ in the formulae for both components.

Figure 8 not only specifies the types of flow mixtures that should occur, it also suggests the ratios for these mixtures: div will generally be faster than def+; rot will generally be faster than def \times . (Note that there is no correlation between rot and def+, or between div and def \times .)

3.3 Principal component analysis

The most representative combinations of EFCs were recovered quantitatively by principal component analysis. The required covariance matrix was calculated with the use of a cropped, symmetric data set. Cropping was necessary because high-speed approaches close to obstacles generated a small number of exceptionally large flow measurements; the 2% of measurements lying more than 3 standard deviations from the mean were therefore deleted. The cropped data set was made symmetric about the origin by adjoining the negative of each flow sample to the set. Before being made symmetric, the flow distribution was skewed in some dimensions, and so was not well described by its covariance.

The covariance matrix of the cropped, symmetric data reveals correlations between EFCs similar to those illustrated in the graphical analysis in the previous subsection:

$$\text{Correlation} \begin{pmatrix} \text{rot} \\ \text{div} \\ \text{def+} \\ \text{def}\times \end{pmatrix} = \begin{bmatrix} 1 & -0.11 & 0.09 & -0.99 \\ & 1 & -0.96 & 0.11 \\ & & 1 & -0.10 \\ & & & 1 \end{bmatrix}.$$

The principal components can be recovered as the eigenvectors of the covariance matrix:

$$\begin{array}{c} \text{rot} \\ \text{div} \\ \text{def+} \\ \text{def}\times \end{array} \begin{matrix} 1 & 2 & 3 & 4 \\ \begin{pmatrix} -0.03 \\ 0.96 \\ -0.27 \\ 0.04 \end{pmatrix} & \begin{pmatrix} 0.77 \\ 0.05 \\ -0.01 \\ -0.64 \end{pmatrix} & \begin{pmatrix} 0.04 \\ 0.27 \\ 0.96 \\ 0.05 \end{pmatrix} & \begin{pmatrix} 0.64 \\ -0.03 \\ -0.06 \\ 0.76 \end{pmatrix} \end{matrix}$$

The eigenvectors are arranged in order of decreasing contribution to variance, given by the eigenvalues: 0.042, 0.019, 0.004, 0.001. The first two components therefore describe over 90% of the variance.

The first principal component is mainly a combination of positive div (expansion) and negative def+ with a third of the magnitude. The second principal component is mainly a mixture of rot and def \times with equal magnitude and opposite sign. These components can be visualised graphically by combining elementary flow fields with coefficients taken from the appropriate eigenvector. The resulting fields are approximately an asymmetric expansion and a horizontal shear, as shown in figure 9.

The second principal component is particularly interesting, given the apparent predominance of spiral-tuned units found in the dorsal division of the MST area (MSTd) by Graziano et al (1994). [The study tested whether MSTd neurons decompose optic flow into discrete channels for translation, rotation, and divergence, by searching for MSTd cells preferentially tuned to spiral stimuli combining both rotation and expansion/contraction. Many of the MSTd cells responded to spiral stimuli, suggesting that decomposition of flow does not occur. Instead, the authors suggested that there is a continuum of patterns to which MSTd cells are selective, which agrees with the suggestion that MST cells behave more like pattern matchers than projectors.] It is possible that some of the spiral-tuned MSTd cells might actually have been tuned to flow patterns similar to the horizontal-shear principal component described above.

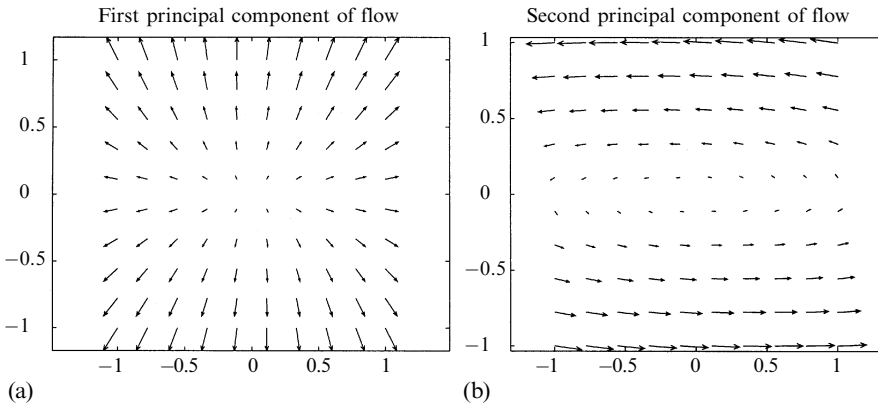


Figure 9. Principal components of first-order optic flow. (a) The first principal component is approximately an asymmetric expansion. (b) The second principal component is approximately a horizontal shear.

4 Discussion

We argue that an accurate description of the characteristics shown by a population of neurons depends on at least three pieces of information: the nature of the computational task performed by the cells (for example, motion or shape analysis), the neural architecture (for example, pattern matcher or projector), and the environment in which the cells operate. The key assumption when considering the impact of the environment, which is based on the properties of artificial pattern-matching neural networks, is that the input PDF plays a significant role in determining the distribution of selectivity in a neural population.

The argument is illustrated by considering the recent unsuccessful search for deformation-selective cells in the MST area of the primate cerebral cortex. Arguments for the existence of such cells were based on computational models of the recovery of shape from optic flow, which highlighted the importance of the deformation components in the recovery process. However, Lagae et al (1994) found that primate MST neurons are not sensitive to pure deformation stimuli, which seems to indicate that MST is not involved in recovering shape from flow. We challenge this notion by arguing that pure deforming stimuli are rare during normal ego-motion, and that neurons selective for pure deformation should be correspondingly rare.

A Monte-Carlo simulation was used to generate the joint PDF for linear optic-flow components, as seen by an ambulating primate in a pseudo-realistic environment. This ‘ecological’ PDF revealed significant correlations which suggest that when deformations occur in natural flow fields they are usually combined with conformal components. In particular, $\text{def}+$ is most likely to occur when divergence is already present in the flow, while $\text{def}\times$ tends to occur when rotation is already present. Artificial neural models suggest that flow-selective cells should partition the optic-flow space according to the input PDF. Thus if deformation-selective neurons are present in MST (or in any other area of the primate cortex), they should respond best to hybrid stimuli containing mixtures of deformation and conformal components.⁽⁶⁾

⁽⁶⁾ Previous work has shown that separating the components of flow might be important in carrying out biologically useful tasks. This strategy is used in computational systems that attempt to extract heading, where rotational information is factored out to leave translation—for example, Longuet-Higgins and Prazdny (1980), and Heeger and Jepson (1990). Hence the computational task may exert pressure to separate flow components, even if they are correlated in the environment. However, even if pure $\text{def}+$ or $\text{def}\times$ are computed at some point in the neural processing of optic flow, a pure deformation stimulus will only be useful if the pattern-matching cells responsible for the flow input can detect it.

Even if the MST area is involved in shape analysis, it is unlikely that cells selective for pure deforming fields will be present there. Hence, the failure to find large numbers of cells selective for pure deformation cannot be used as evidence that MST is not involved in shape analysis. This finding clearly demonstrates the importance of the 'ecological' PDF in determining the characteristics of a neuron population.

4.1 *Recovering shape and motion*

Several recent studies have suggested that MST is involved in analysing motion from optic flow, a process for which deformation is not important.

Perrone and Stone (1994) proposed a computational model in which optic flow is processed by specialised detectors acting as templates for specific instances of self-motion. The detectors respond to global optic flow by sampling image motion over a large portion of the visual field through networks of local motion sensors with properties similar to those of neurons found in the middle temporal (MT) area of primate extrastriate visual cortex. These detectors were designed to extract self-translation (heading) and self-rotation, as well as the scene layout (relative distances) ahead of a moving observer. Perrone and Stone (1998) subsequently compared MST responses with those of detectors from two different configurations of the model under matched stimulus conditions. The results indicated that characteristic physiological properties of MST neurons can be explained by the template model. These findings suggest that MST neurons are well suited to support self-motion estimation from optic flow via a direct encoding, with individual neurons in the MST area acting as heading detectors.

Nevertheless, evidence that MST is involved in motion analysis does not exclude the possibility that it is also involved in shape recovery. Furthermore, the argument that the 'ecological' PDF may be crucial in determining the characteristics of cell populations is valid regardless of the function performed by the cells.

4.2 *Varying the simulation*

The basic form of the EFC correlations can be recovered analytically; however, this involves several rigid assumptions about the distribution and independence of motion and shape parameters; the Monte-Carlo method is much more flexible. Nevertheless, a variety of simplifications were used in the simulation—for example, only two kinds of obstacle (large and small planes) were present in the environment, and motion was deterministic rather than random.

In an environment composed of planar disks with random positions and orientations in a three-dimensional volume most of the patterns described in section 3 are no longer present. However, patterns arising from the motion of the eye are preserved to some extent, most notably the positive divergence due to forward motion. These findings highlight the importance of basic features of the natural environment, such as the fact that the ground plane is always relatively close to the observer whereas the sky is always relatively distant. Aside from these basic features, the exact form of the environment has little impact on the results of the simulation. For example, removing the central canopy or altering the motion parameters in equations (1) and (2) had little effect on the overall results.

The environment could be made more realistic simply by adding more obstacles; likewise, the simulated motion could be made more realistic by including a more complicated kinematic chain and by adding ocular roll (cyclotorsion) to the pitch and yaw. (Note that using Listing's law for cyclotorsion would simply introduce very small additional rotation components.) However, it seems unlikely that additional complexity would alter the finding that pure deformations are unlikely to occur.

4.3 Predictions

Adding def+ and def \times to a translation stimulus produced by a random-dot display generates a compelling impression of surface slant in human observers—see Meese et al (1995), and Freeman et al (1996). This observation raises the question whether or not a mixture of deformation and translation would produce a response in MST neurons. For simplicity, the simulation did not examine translation in detail. However, peripheral receptors are nearly always exposed to some translational flow (see figure 5) so it is likely that a suitable deformation stimulus could include translation, though this is not the case at the fovea. [Note that many of the EFC-selective cells examined by Lagae et al (1994) also responded to translation.] However, the suitability of mixing translation and deformation without conformal components is unclear given that deformation is unlikely to arise in the simulation unless conformal flow is also present.

The principal component analysis suggests that the most appropriate EFC stimuli for examining deformation selectivity would be either a horizontally dominated asymmetric expansion or a horizontal shear. These components might form the basis of appropriate stimuli for experimental work (the first is def+ mixed with div; the second is def \times mixed with rot). Physiological mechanisms in MST might be tuned to these combinations if MST is involved in shape analysis. Further neurophysiological investigations will be necessary to determine whether this is the case in the primate visual cortex.

References

- Duffy C J, Wurtz R H, 1991a “Sensitivity of MST neurons to optic flow stimuli. 1: a continuum of response selectivity to large-field stimuli” *Journal of Neurophysiology* **65** 1329–1345
- Duffy C J, Wurtz R H, 1991b “Sensitivity of MST neurons to optic flow stimuli. 2: Mechanisms of response selectivity revealed by small-field stimuli” *Journal of Neurophysiology* **65** 1346–1359
- Freeman T C A, Harris M G, Meese T S, 1996 “On the relationship between deformation and perceived surface slant” *Vision Research* **36** 317–322
- Girosi F, Jones M, Poggio T, 1995 “Regularisation theory and neural network architectures” *Neural Computation* **7** 219–269
- Graziano M S A, Andersen R A, Snowden R J, 1994 “Tuning of MST neurons to spiral motions” *Journal of Neuroscience* **14** 54–67
- Heeger D J, Jepson A D, 1990 “Visual perception of three-dimensional motion” *Neural Computation* **2** 129–137
- Ivins J, Porrill J, 1995 “A PILUT study of the optic flow seen by a monocular primate moving through a simulated planar environment”, AIVRU Memo No. 97, Department of Psychology, University of Sheffield, Sheffield, UK
- Ivins J, Porrill J, Frisby J, Orban G, 1998 “The probability density function for linear optic flow components” *Proceedings of the Fourteenth International Conference on Pattern Recognition, August 16–20, 1998, Brisbane, Australia* Eds A K Jain, S Venkatesh, B C Lovell (Los Alamitos, CA: IEEE Computer Society) volume 1, 795–798
- Koenderink J J, 1986 “Optic flow” *Vision Research* **26** 161–180
- Koenderink J J, Doorn A J van, 1975 “Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer” *Optica Acta* **22** 773–791
- Kohonen T, 1988 *Self-organisation and Associative Memory* 2nd edition (New York: Springer)
- 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** 1597–1626
- Linsker R, 1989 “How to generate ordered maps by maximising the mutual information between input and output signals” *Neural Computation* **1** 402–411
- Longuet-Higgins H C, Prazdny K, 1980 “The interpretation of a moving retinal image” *Proceedings of the Royal Society of London, Series A* **208** 385–397
- Meese T S, Harris M G, Freeman T C A, 1995 “Speed gradients and the perception of surface slant: analysis is two-dimensional not one-dimensional” *Vision Research* **35** 2879–2888
- 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” *Proceedings of the National Academy of Science of the USA* **89** 2595–2599
- Perrone J A, Stone L S, 1994 “A model of self-motion estimation within primate extrastriate visual cortex” *Vision Research* **34** 2917–2938

- Perrone J A, Stone L S, 1998 “Emulating the visual receptive-field properties of MST neurons with a template model of heading estimation” *Journal of Neuroscience* **18** 5958–5975
- Press W H, Teukolsky S A, Vetterling W T, Flannery B P, 1992 *Numerical Recipes in C: the Art of Scientific Computing* 2nd edition (Cambridge: Cambridge University Press)
- Saito H, Yukie M, Tanaka K, Hikosaka K, Fukada Y, Iwai E, 1986 “Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey” *Journal of Neuroscience* **6** 145–157
- 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” *Journal of Neuroscience* **6** 134–144
- 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” *Journal of Neurophysiology* **62** 642–656
- Verri A, Straforini M, Torre V, 1992 “Computational aspects of motion perception in natural and artificial vision systems” *Biological Sciences* **337** 429–443

APPENDIX

Computation of linear optic flow

As outlined in subsection 2.2, exploratory motion is specified in terms of (x, z) coordinates in the ground plane by finite Fourier series. Given the simplicity of the motion and the fixation strategy, ego-motion (the translational and angular velocities of the eye) could be calculated analytically. However, motion parameters are calculated by numerical differentiation since this is much simpler and more general.

As the eye moves, both the position $\mathbf{t}(t)$ of its centre and the rotation matrix $\mathbf{R}(t)$ specifying its orientation are continually updated. If the eye pose at time t is $\mathbf{R}_1, \mathbf{t}_1$ and at time $t + dt$ is $\mathbf{R}_2, \mathbf{t}_2$, then for small dt the translational velocity \mathbf{v} and the components of angular velocity $\boldsymbol{\omega}$ with respect to the eye frame are:

$$\mathbf{v} = \frac{1}{t_2 - t_1} \mathbf{R}_1^{-1} (\mathbf{t}_2 - \mathbf{t}_1), \tag{A1}$$

$$\boldsymbol{\omega} = \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} = -\frac{1}{t_2 - t_1} \begin{pmatrix} Q_{yz} \\ Q_{zx} \\ Q_{xy} \end{pmatrix}, \quad \text{where } \mathbf{Q} = \mathbf{R}_1^{-1} \mathbf{R}_2.$$

Optic flow is calculated for each receptor separately, in Cartesian coordinates in which the z -axis is directed along the receptor line-of-sight and the x -axis is horizontal in the world (the y -axis is simply orthogonal to the other two).

Consider a point $\mathbf{q} = (x, y, z)^T$ lying on a planar disk with centre \mathbf{p} and normal \mathbf{n} ; the following relationships hold:

$$(\mathbf{q} - \mathbf{p}) \cdot \mathbf{n} = 0 \Rightarrow \mathbf{q} \cdot \mathbf{n} = \mathbf{p} \cdot \mathbf{n}. \tag{A2}$$

The closest plane along the receptor line-of-sight is transformed into receptor coordinates, and its equation is put into the form $n_x x + n_y y + n_z z = P$ which can be rewritten in terms of depth and retinal image coordinates. Setting $P = \mathbf{p} \cdot \mathbf{n}$ and dividing through by Pz the equation of the plane becomes:

$$\frac{1}{z} = N_x X + N_y Y + N_z, \quad \text{where } N_x = \frac{n_x}{P}, N_y = \frac{n_y}{P}, N_z = \frac{n_z}{P}. \tag{A3}$$

The apparent motion of the point \mathbf{q} produced by a translational velocity \mathbf{v} and an angular velocity $\boldsymbol{\omega}$ relative to the origin is given by:

$$\dot{\mathbf{q}} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{q}. \tag{A4}$$

Longuet-Higgins and Prazdny (1980) showed that if a point in the world such as \mathbf{q} has velocity \mathbf{v} and angular velocity $\boldsymbol{\omega}$ then the image velocity (optic flow) of the corresponding retinal image point $(X, Y) = (x/z, y/z)$ is given by:

$$\begin{aligned}\dot{X} &= \frac{v_x - Xv_z}{z} - XY\omega_x + (1 + X^2)\omega_y - Y\omega_z, \\ \dot{Y} &= \frac{v_y - Yv_z}{z} - (1 + Y^2)\omega_x + XY\omega_y + X\omega_z.\end{aligned}\tag{A5}$$

Substituting the expression for $1/z$ from equation (A3) into the equations of Longuet-Higgins and Prazdny (1980) gives:

$$\begin{aligned}\dot{X} &= (v_x - Xv_z)(N_z + N_x X + N_y Y) - XY\omega_x + (1 + X^2)\omega_y - Y\omega_z, \\ \dot{Y} &= (v_y - Yv_z)(N_z + N_x X + N_y Y) - (1 + Y^2)\omega_x + XY\omega_y + X\omega_z.\end{aligned}\tag{A6}$$

Keeping only terms in X, Y up to first order gives:

$$\begin{aligned}f_X &= N_z v_x + \omega_y + (N_x v_x - N_z v_z)X + (N_y v_x + \omega_z)Y, \\ f_Y &= N_z v_y - \omega_x + (N_x v_y + \omega_z)X + (N_y v_y - N_z v_z)Y.\end{aligned}\tag{A7}$$

The corresponding invariant first-order optic flow components (the EFCs) are defined as:

$$\begin{aligned}\text{rot} &= \frac{1}{2} \left(\frac{\partial f_Y}{\partial X} - \frac{\partial f_X}{\partial Y} \right), & \text{div} &= \frac{1}{2} \left(\frac{\partial f_X}{\partial X} + \frac{\partial f_Y}{\partial Y} \right), \\ \text{def}+ &= \frac{1}{2} \left(\frac{\partial f_X}{\partial X} - \frac{\partial f_Y}{\partial Y} \right), & \text{def}\times &= \frac{1}{2} \left(\frac{\partial f_X}{\partial Y} + \frac{\partial f_Y}{\partial X} \right).\end{aligned}\tag{A8}$$

From equation (A7) these components can be calculated as follows:

$$\begin{aligned}\text{rot} &= \omega_z + \frac{1}{2}(N_x v_y - N_y v_x), & \text{div} &= -N_z v_z + \frac{1}{2}(N_x v_x + N_y v_y), \\ \text{def}+ &= \frac{1}{2}(N_x v_x - N_y v_y), & \text{def}\times &= \frac{1}{2}(N_x v_y + N_y v_x).\end{aligned}\tag{A9}$$

These equations only apply at the fovea; however, re-projection can be used to obtain the decomposition elsewhere on the retina: flow is calculated in an imaginary eye (with the same focal point) facing in the required direction. See Ivins and Porrill (1995) for further details.