

Robust velocity computation from a biologically motivated model of motion perception

Alan Johnston^{1*}, Peter W. McOwan^{1,2} and Christopher P. Benton¹

¹*Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK*

²*Department of Mathematical and Computational Sciences, Goldsmiths College, New Cross, London SE14 6NW, UK*

Current computational models of motion processing in the primate motion pathway do not cope well with image sequences in which a moving pattern is superimposed upon a static texture. The use of non-linear operations and the need for contrast normalization in motion models mean that the separation of the influences of moving and static patterns on the motion computation is not trivial. Therefore, the response to the superposition of static and moving patterns provides an important means of testing various computational strategies. Here we describe a computational model of motion processing in the visual cortex, one of the advantages of which is that it is highly resistant to interference from static patterns.

Keywords: motion transparency; pedestal; motion blind; optic flow; MT; speed perception

1. INTRODUCTION

Separating a motion signal from a static pattern signal provides a challenge to current models of motion perception. Motion opponency has been thought of as a means of assuring that motion mechanisms are insensitive to static pattern. Two influential motion detection models, the Reichardt correlation model (Reichardt 1961; Foster 1971; Van Santen & Sperling 1984, 1985) and the motion energy model (Adelson & Bergen 1985), incorporate a stage in which the outputs of systems tuned to opposite directions of motion are subtracted (Barlow & Levick 1965). For the purposes of the discussion of opponency we can consider the Reichardt motion detector and the opponent stage of the motion energy model as functionally equivalent. The motion energy model specifies the construction of linear filters which are orientated in space–time and are therefore tuned to a particular speed and direction of motion. The outputs of even and odd symmetrical filters tuned to a particular direction of motion are squared and summed and responses of units tuned to opposite directions are subsequently subtracted. Motion opponency ensures a zero response to static pattern but opponency *per se* does not produce a general immunity to the addition of static pattern. If the component filters are sensitive to static pattern, the nonlinear squaring operation ensures that the motion opponent response to movement is influenced by the static pattern signal.

Opponent motion energy, as defined by Adelson & Bergen (1985), is proportional to the square of the stimulus contrast. To ensure that stimuli of different contrasts which are moving at a particular speed appear to move at the same speed there needs to be some kind of normalization (Adelson & Bergen 1985; Heeger *et al.* 1996; Simoncelli & Heeger 1998). Contrast normalization

schemes usually use or include a measure of static pattern contrast and this provides another significant route by which static pattern can influence motion computation. A generic model for computing velocity based on a contrast-normalized, opponent combination of space–time-orientated filters is described in Bruce *et al.* (1996). Figure 1 shows the response of the model to a moving sine wave in the presence of a static sine wave pedestal. It is clear that a small amount of static pattern can significantly affect the measurement of speed.

It is possible to remove the influence of static pattern by introducing a linear band-pass temporal filtering stage prior to motion analysis. The problem with this strategy, which essentially removes all low temporal frequencies, is that it would significantly reduce signal strength for slowly moving patterns. The lower threshold for detecting motion can be as small as 0.02°s^{-1} at the fovea (Johnston & Wright 1983), which corresponds to an orientation in space–time of only 1° (scaling space–time such that $1^\circ \text{s}^{-1} = 45^\circ$). It would be difficult to imagine a biologically plausible temporal filter that would remove the influence of static pattern while retaining sensitivity to very slow motion. This suggests that the influence of static pattern is removed actively by special mechanisms or operations within the cortical motion pathway rather than by early filtering operations. This view is consistent with neuropsychological studies (Baker *et al.* 1991; McLeod *et al.* 1996) and lesion studies in primates (Newsome & Pare 1988) showing that damage to the motion area, V5–MT and its analogue in the human brain increases sensitivity to both dynamic and static noise with little effect on low-level spatio-temporal pattern detection thresholds (Hess *et al.* 1989; Pasternak & Merigan 1994). It is difficult to account for how damage to an extrastriate area, containing neurons which do not themselves respond well to static pattern, could allow the effects of static pattern to be revealed if

*Author for correspondence (a.johnston@ucl.ac.uk).

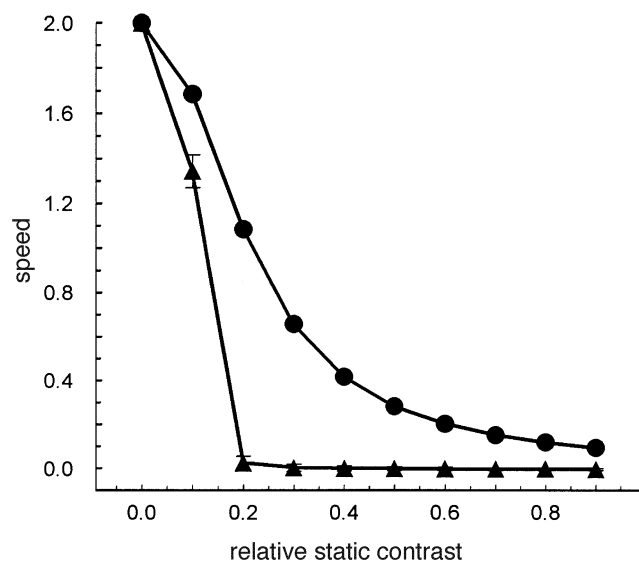


Figure 1. Response of a motion energy computation to a sine grating moving at 2° s^{-1} as a function of the contrast of a static sine wave which has been added to the motion signal. The abscissa shows relative contrast of a static 10 c deg^{-1} pattern added to a 3 c deg^{-1} sine grating drifting at 2° s^{-1} . The computed results are the output of the normalized opponent energy stage of the model described on p. 185 of Bruce *et al.* (1996) (\blacktriangle) and the output of the multichannel gradient model as described in Johnston & Clifford (1995) (\bullet). Error bars are \pm s.e. Both models use low-pass and band-pass temporal filters. If the motion sequence was subject to temporal differentiation prior to motion energy calculation or if only band-pass temporal filters are used, the influence of static pattern can be removed at an early stage. However, this reduces sensitivity to slow motion and is inconsistent with the neuropsychological and neurophysiological evidence for active strategies for noise reduction at a late stage in motion processing (see text).

static signals are filtered out at an early stage in motion processing.

Standard spatio-temporal gradient techniques (Fennema & Thompson 1979; Horn & Schunck 1981; Lucas & Kanade 1981; Sobey & Srinivasan 1991; Heeger & Simoncelli 1995) deliver a measure of image velocity but they also suffer from sensitivity to static pattern (Van Santen & Sperling 1984). Figure 1 shows the effects of adding static pattern for the one-dimensional (1D) multichannel gradient model (Johnston *et al.* 1992; Johnston & Clifford 1995). Typically, two-dimensional (2D) space + time gradient methods recover the component of motion in the direction of the gradient of the spatio-temporal image. This is the component which is orthogonal to the isobrightness contours in the image. The fact that this direction is often not the true direction of motion of the pattern is generally referred to as the 'aperture problem' (Hildreth 1984). The introduction of static pattern provides additional difficulties because the image gradient can be dominated by the static structure rather than by the moving pattern.

Thus, none of the standard biologically motivated models cope well with the superposition of moving and static patterns. The aim of this paper is to consider how we might extend our earlier model (Johnston *et al.* 1992; Johnston & Clifford 1995) to provide predictions

for 2D image sequences in a way that allows for the active removal of the influence of static noise at a late stage in the algorithm. A truncated 2D space + time Taylor series expansion is introduced as a model of the representation of image structure in the visual cortex. Measures of image speed and inverse speed are computed, using a generalization of a technique described in previous work, for a range of directions rotating around the point of interest. These functions of direction are subsequently combined to give an estimate of the image velocity. Having outlined the basic structure of the model, the response of the model to the addition of static pattern is described and the method by which invariance with respect to static pattern is achieved is discussed.

2. THE TAYLOR EXPANSION REPRESENTATION IN THE VISUAL CORTEX

It is generally accepted that simple cells in V1 act like spatio-temporal linear filters (De Valois & De Valois 1988) but there is no general agreement about the exact form and computational role of these linear filters. Following Koenderink (1988), Koenderink & van Doorn (1987, 1992) and Young & Lesperance (1993) we consider simple cells in the primary visual cortex to approximate Gaussian derivatives of various orders. Blurring and differentiation of images can be accomplished by these differential operators, as described by Koenderink & van Doorn (1987). Thus, the outputs of appropriate simple cells can provide various orders of partial derivatives of the blurred image, allowing a truncated Taylor approximation of the image in the neighbourhood of a point in space-time. A Taylor expansion provides a very rich description of image structure at each point in the visual field allowing, amongst other things, the approximation of image brightness values at adjacent points in space and time. This characterization of the action of simple cells as computing partial derivatives goes beyond the usual conception of the role of linear filtering as selecting out spatio-temporal Fourier components of image sequences.

The Taylor representation requires a bank of linear filters, taking derivatives in two spatial directions x and y and in time t . The spatial filters are tuned to different spatial frequencies. Filters extracting higher derivatives of the blurred image have more lobes in their receptive fields and are therefore tuned to higher spatial frequencies. Those incorporating temporal differentiation have transient temporal characteristics (Johnston *et al.* 1992). The neurophysiological and psychophysical evidence in favour of the Gaussian derivative model of neural spatial processing is discussed in Bruce *et al.* (1996) and Johnston & Clifford (1995) showed that the temporal filters of the human visual system can be characterized as differentials of Gaussians in log time. Both Werkhoven & Koenderink (1990) and Otte & Nagel (1995) used Taylor expansions of the image to compute motion, but their techniques involved inversion of large matrices, which we would like to avoid in a biological model. All the operations in the current model can be achieved by combining the outputs of linear, orientated spatio-temporal filters through addition, multiplication and division. Thus, the mathematical

algorithm introduced here can in principle be implemented by neural systems in the visual cortex.

We chose a primary direction, x , corresponding to a particular orientation column in V1 (Hubel & Wiesel 1974) or direction column (Albright *et al.* 1984) in V5-MT and then constructed a vector, the n th component of which contains the value of the n th-order term of the Taylor approximation. Each component of the vector has been truncated by eliminating terms above first order in time and above first order in the direction y , orthogonal to the primary direction. Differentiation in the y spatial direction generates linear filters with end-zone inhibition (Hubel & Wiesel 1965). The expansion is truncated to ensure that, in the model as a whole, there are no more than three linear temporal filters and no greater spatial complexity in the filters than end stopping at both ends. We can approximate the brightness about the point $\mathbf{P}=(x,y,t)$ from the Taylor expansion at that point in space-time (Lang 1987). Writing out the first three terms of the truncated Taylor expansion explicitly, we have

$$\begin{aligned} f(x+p, y+q, t+r) &= [f(x,y,t)] + [f_x(x,y,t)p + f_y(x,y,t)q + f_t(x,y,t)r] \\ &+ \frac{1}{2!} [f_{2x}(x,y,t)p^2 + f_{xy}(x,y,t)2pq + f_{xt}(x,y,t)2pr \\ &+ f_{yt}(x,y,t)2qr] + \dots \text{higher order terms,} \end{aligned} \quad (1)$$

where $\mathbf{H}=(p,q,r)$ is a vector from \mathbf{P} to the point in the visual field at which we require our approximation. Each of the partial derivatives, denoted by subscripts in equation (1), are represented by the output of a linear spatio-temporal cortical filter with a receptive field centred at the location \mathbf{P} in the image. The variables p , q and r are weights on the filter outputs which allow the approximation of image brightness at any point $\mathbf{P}+\mathbf{H}$ in the space-time neighbourhood of \mathbf{P} .

We group the terms by order of approximation, as indicated by the square brackets in equation (1), to form a vector which is our basic representation of image structure. Let

$$\mathbf{k}(x,y,t) = (k_0(x,y,t), k_1(x,y,t), k_2(x,y,t), \dots, k_n(x,y,t))^T \quad (2)$$

be the vector-valued function associated with the Taylor expansion at \mathbf{P} , where each term corresponds to one of the bracketed terms within equation (1). The superscript T denotes the vector transpose. Note that \mathbf{k} is really a function of p , q , r as well as x , y , t but we can assume \mathbf{H} is fixed for the present without loss of generality to simplify the notation. The derivative of the vector function $\mathbf{k}=(k_0, k_1, k_2, \dots, k_n)^T$ is given by the matrix

$$\begin{aligned} \mathcal{J} &= \mathbf{D}\mathbf{k}(x,y,t) = (\mathbf{k}_x(x,y,t), \mathbf{k}_y(x,y,t), \mathbf{k}_t(x,y,t)) \\ &= \begin{bmatrix} k_{0,x} & k_{0,y} & k_{0,t} \\ k_{1,x} & k_{1,y} & k_{1,t} \\ \vdots & \vdots & \vdots \\ k_{n,x} & k_{n,y} & k_{n,t} \end{bmatrix} \end{aligned} \quad (3)$$

where \mathbf{D} is the derivative operator as defined in equation (3). Subscripts in the matrix index the components of the vectors as well as indicating partial differentiation. For the motion computation only the values of the derivatives of the terms in $\mathbf{k}(x,y,t)$ need to be represented in the visual cortex.

From these basic measures we can compute the matrix product

$$\mathcal{J}^T \mathcal{J} = \begin{bmatrix} \mathbf{k}_x \cdot \mathbf{k}_x & \mathbf{k}_x \cdot \mathbf{k}_y & \mathbf{k}_x \cdot \mathbf{k}_t \\ \mathbf{k}_y \cdot \mathbf{k}_x & \mathbf{k}_y \cdot \mathbf{k}_y & \mathbf{k}_y \cdot \mathbf{k}_t \\ \mathbf{k}_t \cdot \mathbf{k}_x & \mathbf{k}_t \cdot \mathbf{k}_y & \mathbf{k}_t \cdot \mathbf{k}_t \end{bmatrix}. \quad (4)$$

We have shown previously (Johnston *et al.* 1992) that the sign of scalar product terms of the form $\mathbf{k}_x \cdot \mathbf{k}_t$ depends upon the direction of motion irrespective of image polarity, mirroring the behaviour of many directionally selective neurons which respond in the same way to moving light and dark bars. This matrix is integrated over a spatio-temporal volume $R = a \geq p \geq b, c \geq q \geq d$ and $e \geq r \geq f$ to give the matrix

$$M = \int_e^f \int_c^d \int_a^b \mathcal{J}^T \mathcal{J} \, dpdqdr = \begin{bmatrix} \mathbf{x} \cdot \mathbf{x} & \mathbf{x} \cdot \mathbf{y} & \mathbf{x} \cdot \mathbf{t} \\ \mathbf{y} \cdot \mathbf{x} & \mathbf{y} \cdot \mathbf{y} & \mathbf{y} \cdot \mathbf{t} \\ \mathbf{t} \cdot \mathbf{x} & \mathbf{t} \cdot \mathbf{y} & \mathbf{t} \cdot \mathbf{t} \end{bmatrix}. \quad (5)$$

As in earlier models (Johnston *et al.* 1992; Johnston & Clifford 1995) integration is implemented by indexing the parameters, here p , q and r and summing over the resulting inner products. The inner product $\mathbf{x} \cdot \mathbf{y}$ indicates that the terms in the first vector are those generated by differentiating the basic representation of image structure with respect to x , and the terms in the second vector are generated by differentiating the basic representation with respect to y .

From this matrix we can recover measures of image speed estimated in two orthogonal spatial directions by computing the ratios $\mathbf{x} \cdot \mathbf{t} / \mathbf{x} \cdot \mathbf{x}$ and $\mathbf{y} \cdot \mathbf{t} / \mathbf{y} \cdot \mathbf{y}$. This essentially generalizes our previous model (Johnston & Clifford 1995) to include an additional spatial dimension. Note that these ratios are well conditioned, since the denominator is equal to the squared magnitude of a vector, e.g. $\mathbf{x} \cdot \mathbf{x} = |\mathbf{x}|^2$. This scalar product is only zero when all the terms of the vector are zero, i.e. when the image is uniform. In this situation we compute zero divided by zero, which we define to be zero.

3. EXTRACTING SPEED AND INVERSE SPEED

We now consider computing these speed measures concurrently for a range of primary directions, corresponding to a range of orientation-direction columns (Hubel & Wiesel 1974; Albright *et al.* 1984) in the primate visual system. It is convenient to introduce a notation for speed $\hat{\mathbf{s}}$ and inverse speed $\check{\mathbf{s}}$ vectors. We may construct a vector, $\hat{\mathbf{s}} = (\hat{s}_{\parallel}, \hat{s}_{\perp})$, whose components are speed and orthogonal speed. This vector is computed at m different orientations θ around a point in the image. Raw speed measures, e.g. $\mathbf{x} \cdot \mathbf{t} / \mathbf{x} \cdot \mathbf{x}$, are infinite for directions parallel to isobrightness contours (figure 2a-c). However, we can define well-conditioned directional speed vectors,

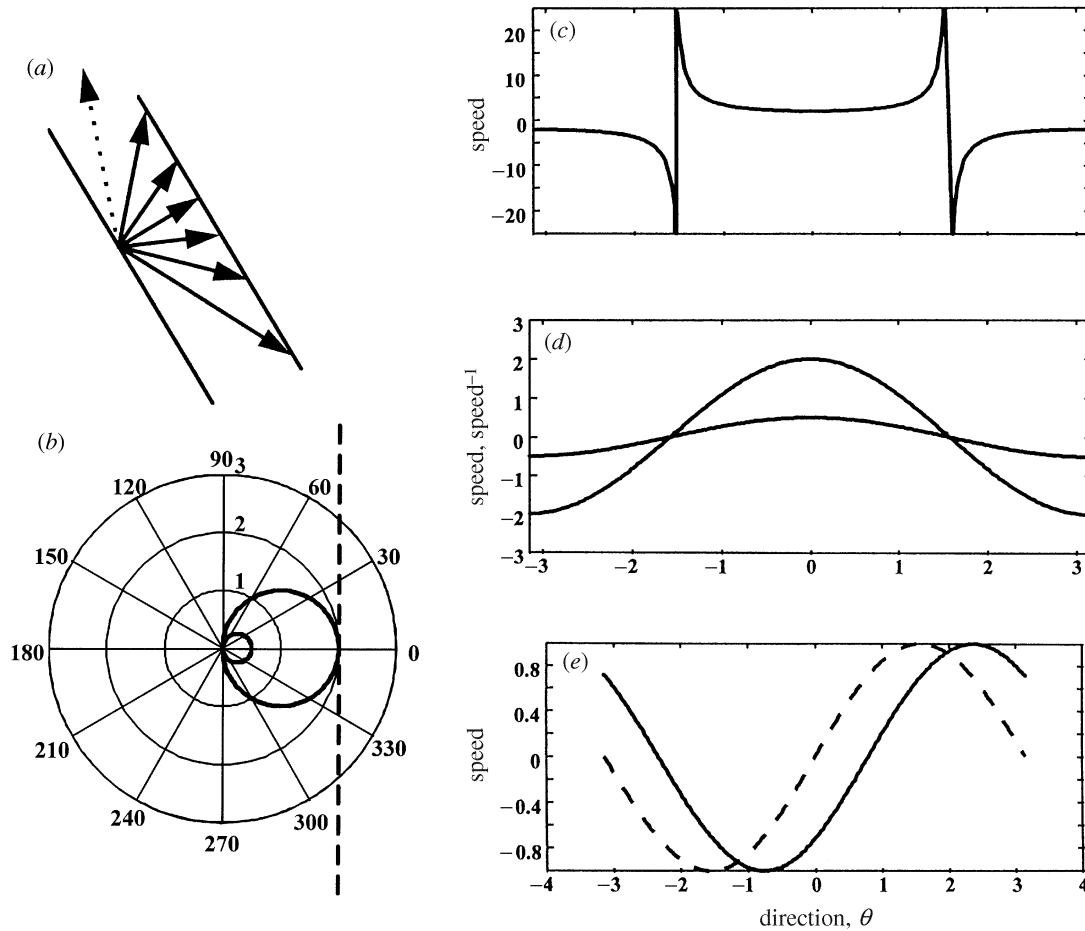


Figure 2. Speed and inverse speed computed as a function of direction for a moving 1D pattern in a 2D image. (a) Two frames of a sequence showing a moving line are superimposed. Because of the aperture problem all motion vectors shown indicate possible translations of the line. However, here we want to consider computed speed as a function of direction. (b) The dotted line shows raw speed as a function of direction in a polar plot. We can think of this as illustrating the speed (represented as distance from the origin) as we change the direction in which it is computed or, alternatively, speed measured in a single direction as we change the orientation of the line. The same data is shown as a linear plot in (c) where it is clearer that computed velocity along the line is infinite. The ordinate is in radians and the data are plotted over 2π radians. Inverse speed is plotted as a small circle in (b) and as a low amplitude cosine wave in (d). The amplitude reflects the inverse speed. The larger circle is the result of plotting the speed measures from equation (6) which also appears as the higher amplitude linear plot in (d). (e) The result of plotting speed as in equation (6) or inverse speed for a line moving at 1° s^{-1} rotated to a direction slightly clockwise with respect to the fiducial reference frame (shown as a dotted line). It is clear from this analysis that speed can be computed from the amplitude of these direction functions and direction of motion can be computed as a phase angle.

$$\hat{\mathbf{s}}(\theta) = \sqrt{\frac{2}{m}} \left[\frac{\mathbf{x}_\theta \cdot \mathbf{t}_\theta}{\mathbf{x}_\theta \cdot \mathbf{x}_\theta} \left(1 + \left(\frac{\mathbf{x}_\theta \cdot \mathbf{y}_\theta}{\mathbf{x}_\theta \cdot \mathbf{x}_\theta} \right)^2 \right)^{-1}, \right. \\ \left. \frac{\mathbf{y}_\theta \cdot \mathbf{t}_\theta}{\mathbf{y}_\theta \cdot \mathbf{y}_\theta} \left(1 + \left(\frac{\mathbf{x}_\theta \cdot \mathbf{y}_\theta}{\mathbf{y}_\theta \cdot \mathbf{y}_\theta} \right)^2 \right)^{-1} \right]. \quad (6)$$

The $m \times 2$ matrix in equation (6) is normalized by a factor depending on the number of directions, m . By including multiplications involving terms computing the orientation of image structure as a function of direction, e.g. $\mathbf{x} \cdot \mathbf{y} / \mathbf{x} \cdot \mathbf{x}$, we can ensure that for 1D spatial stimuli, equation (6) delivers a sinusoidal function of direction, the amplitude of which is directly related to speed (figure 2*b,d*). We can also calculate inverse speed, $\check{\mathbf{s}}$ from terms in the matrix in equation (5):

$$\check{\mathbf{s}}(\theta) = \sqrt{\frac{2}{m}} \left(\frac{\mathbf{x}_\theta \cdot \mathbf{t}_\theta}{\mathbf{t}_\theta \cdot \mathbf{t}_\theta}, \frac{\mathbf{y}_\theta \cdot \mathbf{t}_\theta}{\mathbf{t}_\theta \cdot \mathbf{t}_\theta} \right). \quad (7)$$

For 1D spatial stimuli this delivers a sine function with an amplitude is directly related to inverse speed (figure 2*b,d*).

Koenderink & van Doorn (1976) showed that a velocity field can be decomposed into a translation component and differential components: divergence and curl, plus two components of affine shear. For a pure translation the sum of both the components of $\hat{\mathbf{s}}(\theta)$ over π radians will be zero. A local divergence will result in a non-zero sum of $\hat{s}_\parallel(\theta)$ and a local curl will result in a non-zero sum of $\hat{s}_\perp(\theta)$ over π radians. To recover the translation component and remove the differential components we can force the integral of the directional

speed functions to be zero by extracting the fundamental Fourier coefficients. This is achieved by projecting onto fiducial sine and cosine functions. We construct normalized cosine and sine vectors

$$\mathbf{F}(\theta) = ((F_{\parallel}(\theta), F_{\perp}(\theta)) = \sqrt{2/m} [\cos(\theta), \sin(\theta)]. \quad (8)$$

This matrix forms both a fiducial reference frame, in terms of angle θ , for the computation of direction of motion and allows for the extraction of the fundamental Fourier coefficients of the directional speed functions. Speed squared is computed as a ratio of determinants:

$$S^2 = \frac{\begin{vmatrix} \hat{s}_{\parallel} \cdot F_{\parallel} & \hat{s}_{\parallel} \cdot F_{\perp} \\ \hat{s}_{\perp} \cdot F_{\parallel} & \hat{s}_{\perp} \cdot F_{\perp} \end{vmatrix}}{\begin{vmatrix} \check{s}_{\parallel} & \check{s}_{\perp} \\ \check{s}_{\perp} & \check{s}_{\parallel} \end{vmatrix}}, \quad (9)$$

where, for example, $\hat{s}_{\parallel} \cdot F_{\perp}$ is the scalar product of the first column of $\hat{\mathbf{s}}$ and the second column of \mathbf{F} . The denominator takes the value of one for rigid motion of simple patterns and can vary from point to point in the image to compensate for conditions in which the computed speed and inverse speed are not exact inverses. The denominator can be zero when $\hat{s}_{\parallel} = c \hat{s}_{\perp}$, which would be the case for a pure divergence; however, if this relation holds the numerator is also zero and, as in the similar situation noted above, the indeterminacy is resolved by rule. It is our assumption that direction is coded in the visual system as a phase angle (figure 2*e*) by pairs of cells encoding the projection on fiducial sine and cosine functions, respectively. However, for the purpose of the simulations direction is computed explicitly as

$$\text{direction} = \tan^{-1} \left(\frac{(\check{s}_{\parallel} + \hat{s}_{\parallel}) \cdot F_{\perp} + (\check{s}_{\perp} + \hat{s}_{\perp}) \cdot F_{\parallel}}{(\check{s}_{\parallel} + \hat{s}_{\parallel}) \cdot F_{\parallel} - (\check{s}_{\perp} + \hat{s}_{\perp}) \cdot F_{\perp}} \right). \quad (10)$$

There are only six parameters in the model. Three are required to define the spatio-temporal parameters of the blur kernel, which is differentiated to generate all of the linear filters used in the computation. Thus, the relationship between the filters is highly constrained, but this degree of constraint is reflected in the shape of the temporal filters (Johnston & Clifford 1995) measured psychophysically by Hess & Snowden (1992). These three blur kernel parameters map the spatio-temporal scale of the model onto physical space–time and were fixed in previous work (Johnston & Clifford 1995). They provide a means of calibrating the model to deliver measures of speed in deg s^{-1} (ca. $1^{\circ} = 128$ pixels and $1\text{s} = 128$ pixels). An additional parameter sets the number of spatial derivatives in the primary direction (five), the fifth parameter defines the number of orientations–directions sampled (24), which also determines the aspect ratio of the spatio-temporal integration zones and the sixth defines the extent of these integration zones (11 pixels).

4. SIMULATIONS

First we establish that the model provides accurate results for simple translating patterns, sine gratings and checkerboards. For a grating moving at 2°s^{-1} the direc-

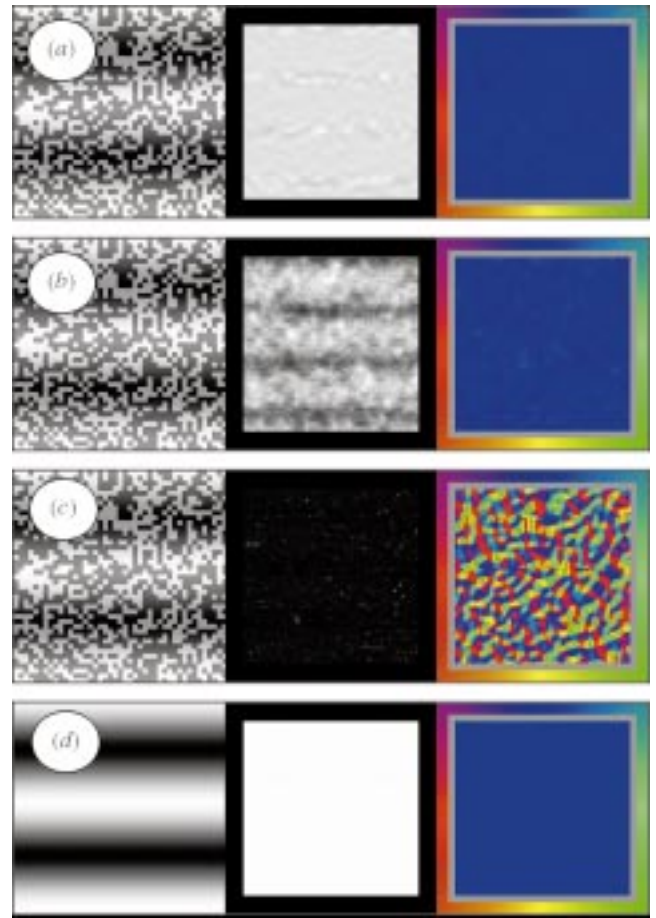


Figure 3. The figure shows dense direction and speed images computed for an arbitrarily chosen frame in a motion sequence. The stimulus frame is shown on the left. In the centre a speed map is displayed which is scaled to the full brightness range. The narrow border (shown in black) is set to a speed of zero. In the rightmost column the corresponding direction map is plotted, with direction coded by colour which should be read with reference to the colour wheel. (a) Results of the full model for a grating moving upwards (2 c deg^{-1} and 2° s^{-1} ; calibrated at $1 \text{ image} = 1^{\circ}$, $1 \text{ image} = 1 \text{ s}$ and $\text{image size} = 128 \text{ pixels} \times 128 \text{ pixels}$; Johnston & Clifford 1995) superimposed on static binary noise. Computed mean speed = 2.00 and $\text{s.d.} = 0.05$. (b) Model output using just the information on the numerator of equation (9) for the stimulus in (a) (computed mean speed = 1.32 and $\text{s.d.} = 0.39$). (c) As in (a) but without the spatio-temporal integration stage of the model (computed mean speed = 0.97 and $\text{s.d.} = 1.87$). (d) Results for the model used in (c) without the added static pattern (computed mean speed = 2.00 and $\text{s.d.} = 0.0$).

tion indicated is orthogonal to the brightness contours and the average speed is correct to two significant figures with zero standard deviation. For a moving checkerboard (four pixels per square, 2° s^{-1}) the average speed and standard deviation are 1.98 and 0.03, respectively. In figure 3*a* we show the effects of adding binary static noise. The response of the model to a grating moving at 2° s^{-1} superimposed on a static random binary noise pattern is largely unaffected by the presence of the static pattern. This is examined in more detail in figure 4. The averaged speed (figure 4*a*) and standard deviation (figure 4*b*) are shown for a grating moving at 4° s^{-1} superimposed on binary noise, as a function of the Michelson contrast ratio

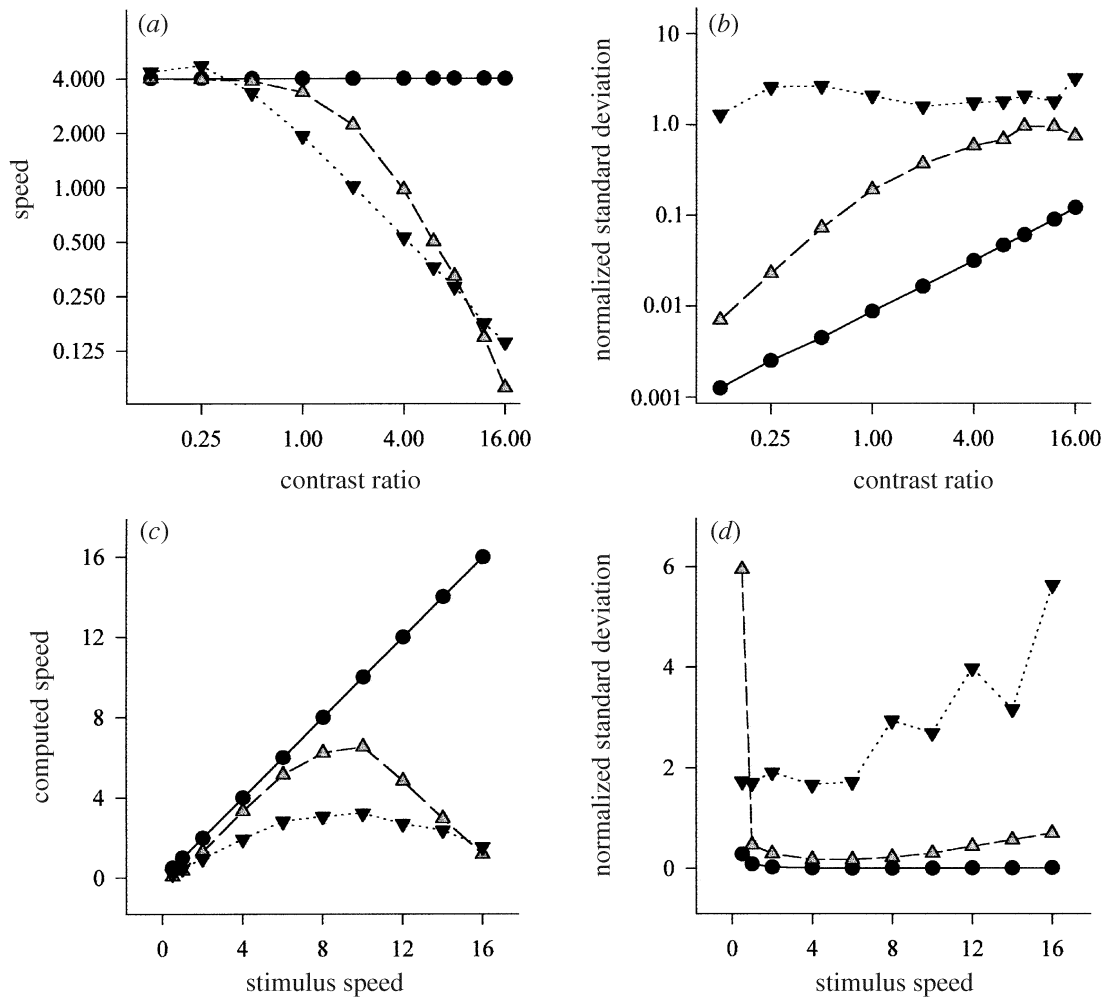


Figure 4. (a) Response of the full model (\bullet), the numerator of equation (9) (Δ) and the full model minus the integration stage (\blacktriangledown) to a moving grating (2 c deg^{-1} and 4° s^{-1}) in the presence of a static binary noise as a function of the relative contrast of the static pattern. (b) Normalized standard deviation for the data in (a). The standard deviation of the responses within a single frame is divided by the mean computed speed to show the variability of the computed result over space. (c) Response of the full model (\bullet), the numerator of equation (9) (Δ) and the full model minus the integration stage (\blacktriangledown) to a moving grating in the presence of a static binary noise (contrast ratio 1:1) as a function of the speed of the grating. (d) Normalized standard deviation for the data in (c).

(relative amplitude) of the moving sine wave and static binary noise. It is clear that the calculated mean speed is essentially invariant with respect to the contrast of the static pattern. The normalized standard deviation is small but increases with the contrast of the static pattern.

Figure 4c shows that the model recovers speed accurately in the presence of noise (Michelson contrast ratio, 1:1) over a range of speeds. Since the true speed of the sine grating is constant throughout the frame, low standard deviations indicate a consistent measure of speed irrespective of the structure of the underlying static pattern (figure 4b,d). Figure 4 shows that the model is able to recover the speed and direction of motion of a sine wave grating in the presence of 2D binary noise, even in situations in which the contrast of the noise is greater than the signal by a factor of 16.

As discussed in §1 standard motion energy and spatio-temporal gradient models are influenced by the presence of static spatial pattern. To demonstrate which features of the model presented here produce invariance with respect to static pattern, we first disabled the integration stage,

equation (5), while keeping the rest of the structure the same. In this degraded form the model is essentially a spatio-temporal gradient model in which differentials of various orders are combined with equal weight. One can see in figure 3c that this degraded version is sensitive to static pattern even though the same spatio-temporal filters are used as in the full model. As expected, the computed direction appears to be dominated by the static pattern. The computed speed field contains many large isolated spikes. In figure 4a we see that the average computed speed is reduced as the relative contrast of the noise is increased and that speed is generally underestimated (figure 4c). The normalized standard deviation, which is a measure of the relative variation in the output, is highest in the case in which the integration stage is disabled. This manipulation also demonstrates that the motion energy configuration is sensitive to static pattern since one can replace any of the inner products in equation (4) with a motion energy-like computation (Van Santen & Sperling 1985; Adelson & Bergen 1986) using the relation

$$\mathbf{k}_x \cdot \mathbf{k}_t = \frac{1}{4} (|\mathbf{k}_x + \mathbf{k}_t|^2 - |\mathbf{k}_x - \mathbf{k}_t|^2), \quad (11)$$

where, for example, \mathbf{k}_x is a vector of filters with an extra spatial derivative in the x spatial direction and \mathbf{k}_t is a vector of spatio-temporal filters with an extra derivative in the time direction. Then $\mathbf{k}_x + \mathbf{k}_t$ becomes a vector of filters which are space-time orientated and $\mathbf{k}_x - \mathbf{k}_t$ becomes their mirror symmetrical partners. See Bruce *et al.* (1996) for an illustrated discussion of this relationship. Further degrading the model by reducing the number of spatio-temporal filters to bring its structure closer to the standard energy model has little effect. Figure 3*d* demonstrates that the spatio-temporal gradient configuration gives good results for sine wave gratings when static noise is removed.

If the integration stage of the model is restored and the results for the numerator in equation (9) alone are plotted we see that the response of the model to the rigid motion of gratings is also degraded (figures 3*b* and 4) but less so than when the integration stage is disabled. However, there is still a significant reduction in computed speed particularly at high speeds and a dependence on static pattern contrast. A decrease in the number of directions sampled further degraded performance but a decrease in the number of higher order spatial filters included in the model had little effect. Results of the simulations showed that the integration stage and the quotient calculated in equation (9) were critical to the success of the model in delivering invariance to static pattern.

5. ANALYSIS

Let $k(x, y, t) = f(x, y, t) + g(x, y, t)$ be the linear addition of a moving, $f(x, y, t)$, and static, $g(x, y, t)$ pattern. Because of the linearity of differentiation and the inner product, the matrix in equation (4) can be rewritten as

$$\begin{aligned} \mathcal{J}^T \mathcal{J} &= \begin{bmatrix} \mathbf{k}_x \cdot \mathbf{k}_x & \mathbf{k}_x \cdot \mathbf{k}_y & \mathbf{k}_x \cdot \mathbf{k}_t \\ \mathbf{k}_y \cdot \mathbf{k}_x & \mathbf{k}_y \cdot \mathbf{k}_y & \mathbf{k}_y \cdot \mathbf{k}_t \\ \mathbf{k}_t \cdot \mathbf{k}_x & \mathbf{k}_t \cdot \mathbf{k}_y & \mathbf{k}_t \cdot \mathbf{k}_t \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{f}_x \cdot \mathbf{f}_x & \mathbf{f}_x \cdot \mathbf{f}_y & \mathbf{f}_x \cdot \mathbf{f}_t \\ \mathbf{f}_y \cdot \mathbf{f}_x & \mathbf{f}_y \cdot \mathbf{f}_y & \mathbf{f}_y \cdot \mathbf{f}_t \\ \mathbf{f}_t \cdot \mathbf{f}_x & \mathbf{f}_t \cdot \mathbf{f}_y & \mathbf{f}_t \cdot \mathbf{f}_t \end{bmatrix} \\ &\quad + \begin{bmatrix} \mathbf{g}_x \cdot \mathbf{f}_x & \mathbf{g}_x \cdot \mathbf{f}_y & \mathbf{g}_x \cdot \mathbf{f}_t \\ \mathbf{g}_y \cdot \mathbf{f}_x & \mathbf{g}_y \cdot \mathbf{f}_y & \mathbf{g}_y \cdot \mathbf{f}_t \\ 0 & 0 & 0 \end{bmatrix} \\ &\quad + \begin{bmatrix} \mathbf{f}_x \cdot \mathbf{g}_x & \mathbf{f}_x \cdot \mathbf{g}_y & 0 \\ \mathbf{f}_y \cdot \mathbf{g}_x & \mathbf{f}_y \cdot \mathbf{g}_y & 0 \\ \mathbf{f}_t \cdot \mathbf{g}_x & \mathbf{f}_t \cdot \mathbf{g}_y & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{g}_x \cdot \mathbf{g}_x & \mathbf{g}_x \cdot \mathbf{g}_y & 0 \\ \mathbf{g}_y \cdot \mathbf{g}_x & \mathbf{g}_y \cdot \mathbf{g}_y & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned} \quad (12)$$

The term \mathbf{f}_x , for example, denotes the partial derivative of the vector derived from the Taylor expansion of the function $f(x, y, t)$, as defined in equation (1) with respect to the parameter x . Note $\mathbf{g}_t = 0$ since $g(x, y, t)$ is the static pattern and the partial derivatives of \mathbf{g} with respect to time are zero. Since integration is a linear operation the integrals of terms on the left-hand side, computed by

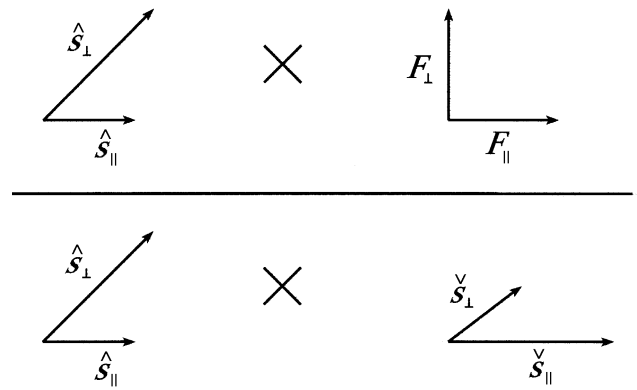


Figure 5. Since determinants of matrices can represent areas, the calculation of the speed in equation (9) can be interpreted as the ratio of the products of areas spanned by the vectors shown above. The vectors are m -dimensional, having the same number of elements as the number of direction columns. Since the speed directional vectors $\hat{\mathbf{s}} = (\hat{s}_{\parallel}, \hat{s}_{\perp})$ contribute to both the numerator and denominator the measure is determined primarily by the inverse speed vectors which are relatively invariant with respect to the addition of static pattern.

indexing the parameters p, q and r as described above, are equal to the sum of the integrals of the corresponding terms on the right-hand side. The decomposition shows that the inverse speed measures (equation (7)), which result from ratios of integrals of terms in the third row, depend almost entirely on the moving pattern, $f(x, y, t)$ and not on the static pattern, $g(x, y, t)$. The two terms on the bottom row of the third matrix, $\mathbf{f}_t \cdot \mathbf{g}_x, \mathbf{f}_t \cdot \mathbf{g}_y$, add to the numerators of the inverse speed ratios. For these terms to be close to zero the vectors should be either be uncorrelated or equally likely to be positively or negatively correlated over the spatio-temporal extent of the integration zone. Suppose the brightness of the static pattern is increasing in a particular spatial direction (positively signed) then the brightness change over time induced by the moving pattern is equally likely to be positively or negatively signed. Since oppositely signed products will cancel in the integration, the sum of these terms over the spatio-temporal volume is likely to be small. The denominator, $\mathbf{k}_t \cdot \mathbf{k}_t$, of the ratios in equation (7) is entirely unaffected by static pattern. Thus, inverse speed as calculated here provides a robust measure even in the presence of static pattern. The directional speed computations also include inner products between the static and moving functions which can be expected to be small, but are corrupted by static pattern terms from the fourth matrix. Substituting into equation (6) we have

$$\hat{\mathbf{s}} = \frac{\int_R (\mathbf{f}_x \cdot \mathbf{f}_t + \mathbf{g}_x \cdot \mathbf{f}_t)}{\int_R (\mathbf{f}_x \cdot \mathbf{f}_x + 2\mathbf{g}_x \cdot \mathbf{f}_x + \mathbf{g}_x \cdot \mathbf{g}_x)} \left(1 + \left(\frac{\int_R \mathbf{k}_x \cdot \mathbf{k}_y}{\int_R \mathbf{k}_x \cdot \mathbf{k}_y} \right)^2 \right)^{-1}. \quad (13)$$

The term $\mathbf{g}_x \cdot \mathbf{g}_x$ reduces the computed directional speed to an extent which is roughly proportional to the contrast of the static pattern.

This analysis helps us to understand why the algorithm is insensitive to static pattern. In the current model speed is calculated as a ratio of determinants

(equation (9)). Both determinants can be interpreted as the products of the areas spanned by the two pairs of vectors (figure 5). The speed measures contribute to both numerator and denominator and cancel in taking the ratio. Thus, the final speed value is primarily computed on the basis of the inverse speed information which is relatively unaffected by static pattern. However, this relative invariance to static pattern depends upon the integration over the space–time volume. In the case of the rigid motion of smoothly varying pattern the denominator in equation (9) equals one and speed is given by the numerator.

6. DISCUSSION

We have described a method of computing velocity which is computationally robust in the presence of static pattern. The model computes directional speed and inverse speed from the derivatives of image structure, which is represented by a Taylor expansion and combines these measures via a ratio of determinants. We showed that the full model is virtually unaffected by the presence of static pattern. The model can be degraded to make the algorithm equivalent to a standard spatio-temporal gradient model which itself can be recast in the form of an energy model or Reichardt equivalent (Adelson & Bergen 1986; Bruce *et al.* 1996). Degrading the model in this way has a significant detrimental effect on computation of motion in the presence of static noise.

Informal observations do not indicate any substantive effect of static pattern on speed perception. Psychophysical investigations have concentrated more on direction discrimination for stimuli close to threshold. Van Santen & Sperling (1984) showed that direction discrimination judgements were unaffected by the addition of static pattern of equal or double the contrast of the moving pattern. Lu & Sperling (1995, 1996), using a similar technique, showed thresholds increased linearly (on log–log axes) with pedestal contrast for pedestal contrast ratios greater than 2. This increase in threshold is consistent with the model data in figure 4*b* which shows increased variability in the computed velocity field as pedestal contrast is increased. Zemany *et al.* (1998) have also shown phase-dependent pedestal effects on motion detection.

The importance of responses to motion in the presence of static pedestals as a means of selecting between models was highlighted by Van Santen & Sperling (1984) who argued that invariance to static pattern was predicted by the elaborated Reichardt model, but not by other methods including spatial correlation analysis between adjacent frames of a motion sequence or spatio-temporal gradient techniques. The prediction that the Reichardt model (or a motion energy equivalent) should be invariant to static pattern relies on the idea that the detector integrates its response over all preceding time (Reichardt 1961) or a temporal period exactly divisible by the wavelength of some periodic input pattern (Van Santen & Sperling 1985), although the infinite integration could be replaced by a leaky integrator (Foster 1971). This is a highly restrictive condition which will virtually never be met in natural image sequences. In addition, the output of the Reichardt detector is proportional to the square of

the contrast, so further processing is necessary to recover speed, including contrast normalization which may reintroduce dependence upon the static pattern.

There are some similarities in the way in which the model achieves invariance to contrast and invariance to static pattern. All of the divisions used to extract directional measures involve the projection of one partial derivative of the Taylor expansion vector onto another. These are self-normalizing operations with respect to contrast and therefore an active process of contrast gain control is not required to stabilize the system. Changing contrast will affect the numerator and denominator similarly (Johnston *et al.* 1992) and so contrast should only be expected to affect the computed velocity at low contrasts where the response of some of the neural components may fall below threshold. The final speed computation also involves a ratio. Factors affecting the speed functions will influence numerator and denominator similarly and so not influence the value of the quotient. One might reasonably ask, why include the speed measures if they have no influence? There may be a number of reasons for this but this architecture has the advantage of robustness in that when inverse speed is small, speed is large (and vice versa). Thus, taking the product on the denominator guards against a divide by zero problem.

The model requires the existence of neurons which encode inverse speed. There is a considerable amount of evidence for ‘low-pass’ speed-sensitive neurons in the visual system of the cat and monkey, which reduce their firing rate as speed is increased (Orban *et al.* 1981; Mikami *et al.* 1986; Rodman & Albright 1987; Lagae *et al.* 1993). The model envisages that speed is encoded in terms of firing rate. Although it is generally assumed population coding provides a better model of biological speed coding than rate coding, in fact the distribution of speed-tuned neurons in V5–MT is far from flat, which might be the expectation from a population code view point. Cheng *et al.* (1994) plotted the distribution of velocity-tuned cells in V5–MT and showed that the most prevalent class are neurons tuned to 32°s^{-1} . Relatively few neurons are tuned to less than 4°s^{-1} . Some tuned cells may be the precursors of the final velocity computation (figure 4). In addition, the reduction in firing rate at high velocities may be due to stimuli passing beyond the temporal frequency cut-off of temporal filters early in the motion pathway, which would render the moving stimulus invisible. Motion transparency, which has also been thought of as evidence for a population code (Simoncelli & Heeger 1998), can be thought of as resulting from grouping processes acting on local velocity signals (McOwan & Johnston 1996).

The model provides an effective explanation for the sensitivity to static noise shown by the motion-blind patient L.M. (Baker *et al.* 1991; McLeod *et al.* 1996). If static pattern was removed at the outset by temporal filtering there is no reason to expect that an extrastriate lesion should introduce enhanced sensitivity to this type of noise. The observation, that lesioning the model by removing spatio-temporal integration and the denominator of equation (6) increases sensitivity to static noise, leads us to speculate that the neural substrate of these processes may be located in extrastriate motion areas.

In summary, analysing the response to motion in the presence of static pattern provides an effective way of choosing between motion models. Biological models of motion perception typically involve some kind of motion energy or spatio-temporal gradient calculation. The squaring and product operations involved in these models makes it difficult to separate the influences of moving and static patterns. It is also the case, but not emphasized here, that models based on feature tracking (Del Viva & Morrone 1998) are subject to the problem of extracting the features of the moving pattern from the features of the static pattern. Early removal of static noise through band-pass temporal filtering was discounted because, if true, it would be difficult to explain increased sensitivity to static noise after lesions late in the motion pathway. We have shown that it is possible to remove the influence of static pattern on motion analysis at a late stage in motion computation. The two stages beyond the usual combination of linear filters described here, integration over a spatio-temporal volume and the combination of speed and inverse speed functions, provide a means of effectively reducing the influence of static pattern on motion analysis. Damage to these processes results in degraded performance in the presence of static noise.

The research was supported by a project grant from the EPSRC-BBSRC Mathematical Modelling Initiative. Dr McOwan was supported by a Wellcome Fellowship in Mathematical Biology.

REFERENCES

- Adelson, E. H. & Bergen, J. R. 1985 Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.* **A2**, 284–299.
- Adelson, E. H. & Bergen, J. R. 1986 The extraction of spatio-temporal energy in human and machine vision. In *Proceedings of the IEEE workshop on motion: representation and analysis*, pp. 151–156. Charleston, SC, IEEE.
- Albright, T. D., Desimone, R. & Gross, C. G. 1984 Columnar organisation of directionally selective cells in visual area MT of the macaque. *J. Neurophysiol.* **51**, 16–31.
- Baker, C., Hess, R. & Zihl, J. 1991 Residual motion perception in a 'motion blind' patient, assessed with limited-lifetime random dot stimuli. *J. Neurosci.* **11**, 454–461.
- Barlow, H. B. & Levick, W. R. 1965 The mechanism of directionally selective units in rabbit's retina. *J. Physiol. Lond.* **178**, 477–504.
- Bruce, V., Green, P. R. & Georgeson, M. A. 1996 *Visual perception: physiology, psychology & ecology*, 3rd edn. Hove: Laurence Erlbaum Associates.
- Cheng, K., Hasegawa, T., Saleem, K. S. & Tanaka, K. 1994 Comparison of neuronal selectivity for stimulus speed, length and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *J. Neurophysiol.* **71**, 2269–2280.
- Del Viva, M. M. & Morrone, C. M. 1998 Motion analysis by feature tracking. *Vision Res.* **38**, 3633–3653.
- De Valois, R. L. & De Valois, K. 1988 *Spatial vision*. Oxford University Press.
- Fennema, C. L. & Thompson, W. B. 1979 Velocity determination in scenes containing several moving objects. *Comput. Graph. Image Process.* **9**, 301–315.
- Foster, D. H. 1971 A model of the human visual system in its response to certain classes of moving stimuli. *Kybernetik* **8**, 69–84.
- Heeger, D. J. & Simoncelli, E. P. 1995 Model of visual motion sensing. In *Spatial vision in humans and robots* (ed. L. Harris), pp. 367–392. Cambridge University Press.
- Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. 1996 Computational models of cortical visual processing. *Proc. Natl Acad. Sci. USA* **93**, 623–627.
- Hess, R. F. & Snowden, R. J. 1992 Temporal properties of human visual filters: number, shapes and spatial covariation. *Vision Res.* **32**, 47–60.
- Hess, R. H., Baker, C. L. & Zihl, J. 1989 The 'motion-blind' patient: low-level spatial and temporal filters. *J. Neurosci.* **9**, 1628–1640.
- Hildreth, E. C. 1984 The computation of the velocity field. *Proc. R. Soc. Lond.* **B221**, 189–220.
- Horn, B. K. P. & Schunck, B. G. 1981 Determining optical flow. *Artif. Intell.* **17**, 185–203.
- Hubel, D. H. & Wiesel, T. N. 1965 Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *J. Neurophysiol.* **28**, 229–289.
- Hubel, D. H. & Wiesel, T. N. 1974 Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J. Comp. Neurol.* **158**, 267–294.
- Johnston, A. & Clifford, C. W. G. 1995 A unified account of three apparent motion illusions. *Vision Res.* **35**, 1109–1123.
- Johnston, A. & Wright, M. J. 1983 Visual motion and cortical velocity. *Nature* **304**, 436–438.
- Johnston, A., McOwan, P. W. & Buxton, H. 1992 A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc. R. Soc. Lond.* **B250**, 297–306.
- Koenderink, J. J. 1988 Operational significance of receptive field assemblies. *Biol. Cybernet.* **58**, 163–171.
- Koenderink, J. J. & van Doorn, A. J. 1976 Local structure of movement parallax of the plane. *J. Opt. Soc. Am.* **66**, 717–723.
- Koenderink, J. J. & van Doorn, A. J. 1987 Representation of local geometry in the visual system. *Biol. Cybernet.* **55**, 367–375.
- Koenderink, J. J. & van Doorn, A. J. 1992 Receptive field assembly pattern specificity. *J. Vis. Comm. Image Rep.* **3**, 1–12.
- Lagae, L., Raiguel, S. & Orban, G. A. 1993 Speed and direction selectivity of macaque middle temporal neurones. *J. Neurophysiol.* **69**, 19–39.
- Lang, S. 1987 *Calculus of several variables*. New York: Springer.
- Lu, Z.-L. & Sperling, G. 1995 The functional architecture of human visual motion perception. *Vision Res.* **35**, 2697–2722.
- Lu, Z.-L. & Sperling, G. 1996 Contrast gain control in first- and second-order motion perception. *J. Opt. Soc. Am.* **A13**, 2305–2318.
- Lucas, B. D. & Kanade, T. 1981 An iterative image registration technique with an application to stereo vision. In *Proceedings of the seventh international joint conference on artificial intelligence*, pp. 674–679. Vancouver, BC.
- McLeod, P., Ditttrich, W., Driver, J., Perrett, D. & Zihl, J. 1996 Preserved and impaired detection of structure from motion by a 'motion-blind' patient. *Visual Cogn.* **3**, 363–391.
- McOwan, P. W. & Johnston, A. 1996 Motion transparency arises from perceptual grouping: evidence from luminance and contrast modulation motion displays. *Curr. Biol.* **10**, 1343–1346.
- Mikami, A., Newsome, W. T. & Wurtz, R. H. 1986 Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J. Neurophysiol.* **55**, 1308–1327.
- Newsome, W. T. & Pare, E. B. 1988 A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 2201–2211.
- Orban, G. A., Kennedy, H. & Maes, H. 1981 Response to movement of neurones in areas 17 and 18 of the cat. *J. Neurophysiol.* **45**, 1043–1058.

- Otte, M. & Nagel, H. 1995 Estimation of optical flow based on higher-order spatio-temporal derivatives in interlaced and non-interlaced image sequences. *Artif. Intell.* **78**, 5–43.
- Pasternak, T. & Merigan, W. H. 1994 Motion perception following lesions of the superior temporal sulcus in the monkey. *Cerebr. Cortex* **4**, 247–259.
- Reichardt, W. 1961 Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory communication* (ed. W. A. Rosenblith), pp. 303–317. New York: Wiley.
- Rodman, H. R. & Albright, T. D. 1987 Coding of stimulus velocity in area MT of the macaque. *Vision Res.* **27**, 2035–2048.
- Simoncelli, E. P. & Heeger, D. J. 1998 A model of neuronal responses in visual area MT. *Vision Res.* **38**, 743–762.
- Sobey, P. & Srinivasan, M. V. 1991 Measurement of optical flow by a generalized gradient scheme. *J. Opt. Soc. Am. A* **8**, 1488–1498.
- Van Santen, J. P. H. & Sperling, G. 1984 Temporal covariance model of human motion perception. *J. Opt. Soc. Am. A* **1**, 451–473.
- Van Santen, J. P. H. & Sperling, G. 1985 Elaborated Reichardt detectors. *J. Opt. Soc. Am. A* **2**, 300–321.
- Werkhoven, P. & Koenderink, J. 1990 Extraction of motion parallax structure in the visual system. 1. *Biol. Cybernet.* **63**, 185–191.
- Young, R. A. & Lesperance, R. M. 1993 *A physiological model of motion analysis for machine vision*. Technical Report GMR 7878. Warren, MI: General Motors Research Laboratories.
- Zemany, L., Stromeyer, C. F. III, Chaparro, A. & Kronauer, R. E. 1998 Motion detection on flashed, stationary pedestal gratings: evidence for an opponent-motion mechanism. *Vision Res.* **38**, 795–812.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.