Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Physiology - Paris 97 (2003) 325–334
Journal of Physiology Paris
www.elsevier.com/locate/jphysparis

Biological computation of image motion from flows over boundaries

A. Johnston [a,*], P.W. McOwan [b], C.P. Benton [c]

[a] Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK
[b] Department of Computer Science, Queen Mary University of London, Mile End Road, London, UK
[c] Department of Experimental Psychology, University of Bristol, 8 Woodlands Road, Bristol, UK

Abstract

A theory of early motion processing in the human and primate visual system is presented which is based on the idea that spatio-temporal retinal image data is represented in primary visual cortex by a truncated 3D Taylor expansion that we refer to as a jet vector. This representation allows all the concepts of differential geometry to be applied to the analysis of visual information processing. We show in particular how the generalised Stokes theorem can be used to move from the calculation of derivatives of image brightness at a point to the calculation of image brightness differences on the boundary of a volume in space–time and how this can be generalised to apply to integrals of products of derivatives. We also provide novel interpretations of the roles of direction selective, bi-directional and pan-directional cells and of type I and type II cells in V5/MT.

Keywords: Optic flow; Cortex; Differential forms; Vision; Motion

1. Introduction

Three basic strategies have guided the development of models of motion computation in biological vision systems—the spatio-temporal correlation strategy [1], the motion energy strategy [2] and the spatio-temporal gradient strategy [3–5]. The first two are essentially template schemes. The aim is to construct a biological mechanism that will respond best to its target feature. In this case the target feature is image motion in a particular direction at a particular velocity. Ideally the response should reflect the quality of the match between the stimulus and the motion template, but typically the response of correlation and energy models also depends on image contrast unless there is an additional stage of contrast normalisation [2,6]. Because the output of any detector does not indicate speed of motion but rather the degree of correlation with the template, speed computation proceeds by interpreting, via some further computational strategy, the population response of a range of detectors varying in their tuning characteristics [7,8]. The third strategy is different in that speed is calculated directly from the ratio of the temporal and spatial derivatives at each point in space and time. Since speed is computed as a ratio of the outputs of spatio-temporal filters, both of which vary with contrast, the calculation delivers contrast invariance at no additional cost [5,9]. Although perceived speed is not linearly related to contrast, contrast can influence perceived speed, particularly at low temporal and low spatial frequencies [10]. Johnston et al. [11] discuss how these effects of perceived speed may be explained with reference to the threshold behaviour of a population of differentiating filters. Johnston et al. [12] show that all three of these strategies have difficulty in computing image speed in the presence of static noise.

The aim of this paper is to introduce a fourth strategy for motion computation in biological systems, based on the computation of flow of some quantity over a closed boundary in the image in some time interval. First we need to draw a distinction between the computation of the motion of objects, a property attached to an object, and the motion sensed or measured at a particular locus in visual space. We are taught that the motion of an object is computed as its change in position for some time interval. However it is unlikely that this is done explicitly, i.e. we do not have clear biological models of how we measure distance between two points, measure a

* Corresponding author. Tel.: +20-7679-5310; fax: +20-7436-4276.
E-mail address: a.johnston@ucl.ac.uk (A. Johnston).

0928-4257/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jphysparis.2003.09.016

temporal interval associated with the translation, or attach the ratio of these measures to each object in the field of view. An alternative is to define a region in the image and measure the flow of material in and out of the region. The region can be identified with the receptive field of a neurone at some point in the visual pathway.

We have described the operation of this approach in a procedural way in previous reports [5,12–14]. The aim of this paper is to express the mathematical operations in a geometrical form, which will allow a more pictorial view of the operations in the model, and which will make explicit the idea of motion computation from flow. The geometrical approach will provide a clearer path towards the generalisation of the model to higher dimensional inputs [15]. Since we will be estimating information on a space–time boundary by integrating image derivatives, the most appropriate geometrical tools to use are differential forms and exterior differentiation, as these naturally accommodate the generalised Stokes theorem, which expresses the relationship between integrals of derivatives over a volume and integrals over the boundary of the volume [16]. Since in a biological model, we cannot have access to multiple temporal samples simultaneously, will use Taylor's theorem to estimate values within the space–time volume from derivatives taken at a point. Because the desire to express the model in a more general form inevitable leads to greater abstraction, we will also indicate, where appropriate, how the model might be implemented neurophysiologically.

First we consider how to compute motion from the flow of image brightness as a means of introducing differential forms and the notation we will use which follows Weintraub [16]. We will then introduce the truncated Taylor expansion as a means of representing image brightness. Once we have described the extraction of the basic measures of image speed for a particular reference frame we will rotate the reference frame to derive vectors of speed measures for each image point in 2D space–time. These measures will then be combined to give separate estimates of image speed and direction [12].

2. Flow of image brightness

Consider the bounding contour of a rectangular region in the image. As the image flows over a segment of the contour, it can be considered to develop a spatio-temporal trace in a plane orthogonal to the image plane (Fig. 1A). The total flow of light over this segment of the contour can be calculated by integrating the function over the spatio-temporal trace shown as S–T in Fig. 1. The difference between the flow over opposite sides of the rectangular region delivers the flow out of the box minus the flow into the box for the selected direction—the net flow in a particular direction. We can attribute a different sign to flow from out-to-in and flow from in-to-out so that the sum of the flow over the rectangular contour will be zero in the case of uniform laminar flow.

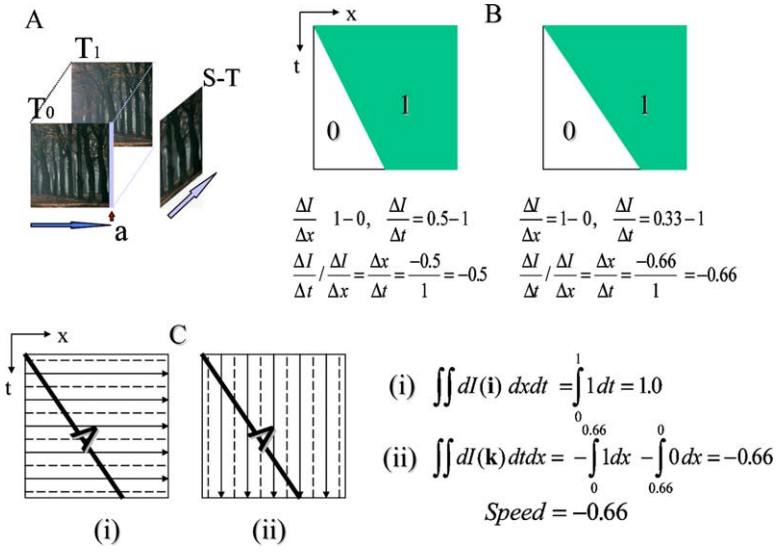

Fig. 1. (A) The input to the visual system as a space–time solid. T_0 indicates the first frame of the sequence and T_1 is the last frame. Flow of image data over (a) generates the plane S–T. (B) A space–time plot showing a bright region moving to the right with a speed of 0.5 pixels/frame. The difference in image brightness of the normalised integral on the vertical boundaries is 1. The difference in image brightness of the normalised integral on the horizontal boundary is −0.5. Thus the speed is −0.5, which is computed from the ratio of the net change in the amount of light in the region over the time interval to the amount of light flowing out of the region minus the amount flowing in over the spatial boundary of the region. (C) Integrating over the one-form in (i) gives the number of contours cut in the spatial direction. This divides the number of contours cut in the time direction to give the speed. The one-form is indicated by a line with an arrow showing the direction of the gradient.

In the case of images this corresponds to the movement of a uniform grey field.

From the fundamental theorem of calculus we have

$$f(b,y,t) - f(a,y,t) = \int_a^b f_{,x}(x,y,t)\,\mathrm{d}V_x = \int_a^b \omega(\mathbf{i})\,\mathrm{d}V_x \quad (1)$$

where $\omega = \mathrm{d}f = f_{,x}\,\mathrm{d}x + f_{,y}\,\mathrm{d}y + f_{,t}\,\mathrm{d}t$ is the gradient of the image brightness, a one-form. The expression $\omega(\mathbf{i})$ indicates the contraction (inner product) of the one-form, ω, on the unit vector $\mathbf{i} = [1,0,0]^{\mathrm{T}}$. The comma subscript in (1) denotes partial differentiation of the image brightness $f(x,y,t)$. The basis one-forms and basis vectors combine to give $\mathrm{d}x(\mathbf{i}) = 1$, $\mathrm{d}y(\mathbf{j}) = 1$, and $\mathrm{d}t(\mathbf{k}) = 1$. All other combinations are equal to zero. The unit one-forms $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}t$ are basis one-forms and are to be distinguished from $\mathrm{d}V_x$ which indicates the ordinary integral taken in the direction x. In other treatments [17] the vector gradient is set to zero. Here the brightness gradient is more properly identified as a one-form and one-forms and vectors are combined via contraction to give scalar values e.g. $\omega(\mathbf{i}) = f_{,x}(x,y,t)$. The picture one has in mind when thinking about the contraction of a vector on a one-form is the action of walking up a hill in the direction of the vector [18]. The height of the hill can be represented by a field of one-forms, which form a contour map of the hill. The contraction of the vector on the one-form gives the number of contour lines cut by the vector, which is the height gained (Fig. 1C). If the image brightness is being represented then the contraction gives the brightness difference between the tip and the tail of the vector. The fundamental theorem is the simplest version of the Generalised Stokes Theorem, which relates differences on boundaries of a region to integrals over the derivative of a function defined on the region.

The calculation in (1) gives the brightness difference at a single point in y and time taken over an interval of x stretching from a to b. Integrating over one face of the box $[c\ldots d,\ e\ldots f]$ delivers the integral of the two-form [16,18] describing the flow of image brightness per unit spatio-temporal area over the segment $[c\ldots d]$ in the time interval $[e\ldots f]$

$$A\,\mathrm{d}y\,\mathrm{d}t = \int_e^f \int_c^d \left[\int_a^b \omega(\mathbf{i})\,\mathrm{d}V_x\right]\mathrm{d}V_{y,t} \quad (2)$$

The other components of the two-form, φ, are recovered from $\omega(\mathbf{j})$ and $\omega(\mathbf{k})$, giving

$$\varphi = A\,\mathrm{d}y\,\mathrm{d}t + B\,\mathrm{d}t\,\mathrm{d}x + C\,\mathrm{d}x\,\mathrm{d}y \quad (3)$$

with C representing the difference in the total image intensity over the region at time f minus that at time e. Speed can be computed from the change in the amount of light over the temporal interval divided by the change in the amount of light over some spatial interval (Fig. 1B and C). Hence speed can be calculated as $C/\sqrt{A^2 + B^2}$, although this ratio is ill-conditioned when the denominator is zero. More intuitively, if light is conserved, as it is in rigid motion, one relates the change in brightness in a region over time to the change in brightness over a spatial interval. This is conceptually similar to the gradient approach but the computation is done on differences over extended intervals rather than on gradients at a point (Fig. 1B and C). Procedurally, if the integration is eliminated the technique defaults to the standard gradient method [12].

Gupta and Kanal [19] have recently developed an effective computer vision algorithm for velocity field computation based on the constraint outlined above. They explicitly compute integrals of image brightness over the boundaries of local space–time volumes. They then construct a series of equations which are solved to recover the parameters of an affine model assumed to describe the local flow. However, in a biological system one does not have access to a space–time block of raw image values since the visual system does not have a temporal buffer within which to keep a record of brightness values in over an extended period of time. Information about brightness values on the boundary of a space–time volume would need to be constructed from the activity of neural assemblies in visual cortex. This can be accomplished by representing the local image structure by a local Taylor expansion, differentiating this representation and then integrating over a local space–time volume [5,12]. We also need to generalise the approach to computation of flow outlined above to accommodate the Taylor representation and to deliver robust well-conditioned computation.

3. The jet representation in the visual cortex

Following Koenderink and Van Doorn [20–22], Young [23] and Johnston [5], we consider simple cells in primary visual cortex to approximate Gaussian derivatives of various orders, allowing a truncated Taylor approximation of the image about a point in space–time—a k-order jet. This allows all the machinery of differential geometry to be applied to visual computation. Bruce et al. [24] discusses the neurophysiological and psychophysical evidence in favour of this view of neural spatial processing and Johnston and Clifford [25] show that the temporal filters of the human visual system can be characterised as differentials of Gaussians in log time. Both Werkhoven and Koenderink [26] and Otte and Nagel [27] have used Taylor expansions of the image to compute motion but these techniques involve inversion of large matrices, which we would like to avoid in a biological model.

We choose a primary direction and then construct a "jet vector" containing the values of the nth-order approximation, which has been truncated by eliminating terms above first-order in time and above first-order in the direction orthogonal to the primary direction (Fig. 2A and B). First-order derivatives provide the minimum differential structure that we need to allow prediction of the image structure at adjacent locations in space or at different points in time.

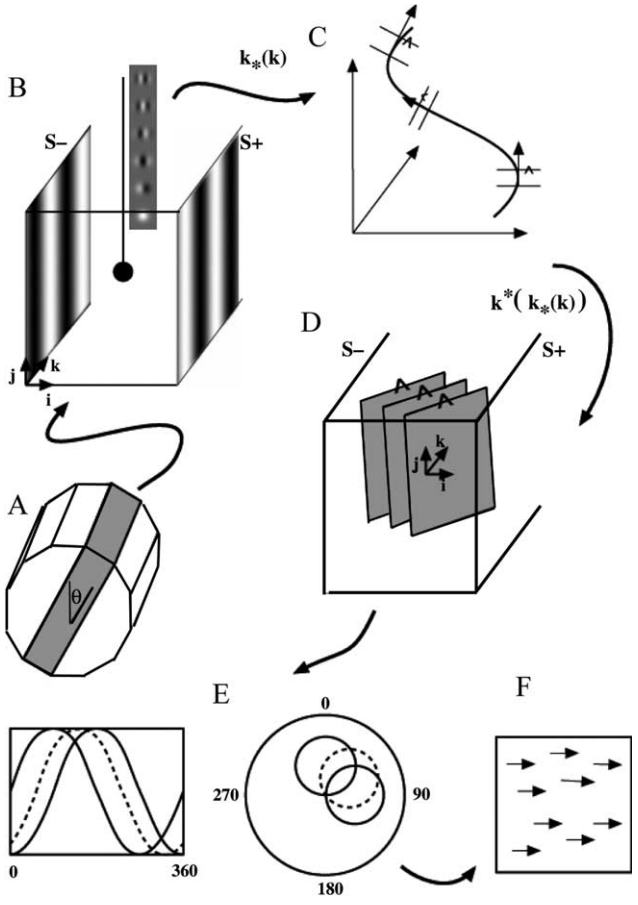

Fig. 2. (A) A space–time faceted cylinder representing the spatio-temporal scope of the calculations in the model. (B) The input sequence is represented as a space–time volume. One slab of the space–time cylinder represents the scope of a single direction/orientation column in A. The space–time volume is sampled by a bank of linear spatio-temporal filters which form a representation of the image. Only a few of the 24 filters used in the simulations are shown. The sine gratings on the side of the box represent a history of the flow over the horizontal sides generated by a moving grating. (C) The filter outputs are grouped to form a vector representation of a truncated Taylor series expansion of the image. The basic vectors shown are orthogonal and form a three-dimensional space but this is simply for purposes of illustration. By changing the parameters of this jet vector we can draw curves in the jet space and then differentiate with respect to parameters to give tangent vectors to the curves. For 1D stimuli the curves are coincident. The one-forms corresponding to tangent vectors are indicated by parallel lines with an arrow to indicate direction. (D) To integrate the one-form field along the curves the one-forms are "pulled back" to the input space to provide a variable one-form field. The number of planes pierced by vectors parallel to the coordinate vectors delivers the integral. (E) Speed and inverse speed measures are extracted for each of the slabs in A. For 1D patterns the radial functions are sinusoidal and are shown here as linear and polar plots. The polar plots show positive values only. (F) Combination of the radial functions of speed and inverse speed, measured in primary and orthogonal directions at each point in the image, delivers the velocity field map.

We can approximate the brightness about the point, $\mathbf{P} = (x, y, t)$, from the Taylor expansion at that point in space–time.

$$f(\mathbf{P}+\mathbf{H}) = f(\mathbf{P}) + g_1(\mathbf{H}) + g_2(\mathbf{H}) + g_3(\mathbf{H}) + \cdots + R_r \tag{4}$$

where

$$g_n = \frac{1}{n!}(\mathbf{H}\cdot\nabla)^n f(\mathbf{P})$$

is the polynomial approximation of degree n [28], and $\mathbf{H}$ is the position vector $\mathbf{H} = (p, q, r)$. Introducing coordinates, and writing out the first two terms explicitly, we have

$$\begin{aligned} f(x+p, y+q, t+r) = f(x,y,t) + [&f_{,x}(x,y,t)p \\ &+ f_{,y}(x,y,t)q + f_{,t}(x,y,t)r] \\ &+ g_2(p,q,r) + \cdots \end{aligned} \tag{5}$$

Notice one gets the same result if we differentiate the equation in (5) with respect to x, y, t or p, q, r. We can think of the vector associated with the Taylor expansion as a function which maps the point (p, q, r) in the parameter space (Fig. 2B), which has the standard Cartesian basis vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$, onto a point in an n-dimensional space in which each coordinate axis delimiting the space serves to represent the value of the nth-order derivative of the local structure of the image intensity surface (Fig. 2C). Since the space contains a representation of a truncated Taylor expansion of the image we will refer to it as the jet space.

Let

$$\begin{aligned} &(a_1, a_2, a_3, \ldots, a_n) \\ &\quad = k(p,q,r) \\ &\quad = (c, g_1(p,q,r), g_2(p,q,r), \ldots, g_n(p,q,r)) \end{aligned} \tag{6}$$

be a vector associated with the Taylor expansion, where k is a vector valued function of p, q and r and $c = f(x, y, t)$. By varying the parameters p, q and r we can draw curves in the jet space. The tangent vectors to the curves in the jet space, induced by the push forward of the basis vectors in the parameter space, are given by

$$\begin{aligned} k_*(\mathbf{i}) &= k_{,p}(p,q,r) \\ &= (0, g_{1,p}(p,q,r), g_{2,p}(p,q,r), \ldots, g_{n,p}(p,q,r)) \\ k_*(\mathbf{j}) &= k_{,q}(p,q,r) \\ &= (0, g_{1,q}(p,q,r), g_{2,q}(p,q,r), \ldots, g_{n,q}(p,q,r)) \\ k_*(\mathbf{k}) &= k_{,r}(p,q,r) \\ &= (0, g_{1,r}(p,q,r), g_{2,r}(p,q,r), \ldots, g_{n,r}(p,q,r)) \end{aligned} \tag{7}$$

respectively. The subscripted star denotes the push forward operation. Push forward refers to a mapping from the parameter space onto the jet space. The process is complemented by the pull back operation, which enables differential forms in the jet space to be mapped back to the parameter space.

The tangent vectors in the jet space (Fig. 2C) indicate the change in the representation induced by a change in spatial position in the image or a change in time. The dimensionality of the tangent space depends upon the image structure. For the rigid motion of one-dimensional patterns, such as sine-wave gratings, the differential structure changes in the same way, but at different rates, for increments in the p, q and r parameter values. Thus the tangent vectors induced by increments in the parameter space lie along the same curve in the jet space but typically have different lengths. The rigid motion of two-dimensional images gives rise to a two-dimensional surface in the jet space. In the case of non-rigid motion, the space is three-dimensional. The tangent vectors are linearly independent and they therefore form a local frame parallel to the parameter lines induced by increments of p, q and r, respectively. Given this variability, dependent upon the image structure, we need to develop a method for recovering motion that does not rely on assumptions about the dimensionality of the space.

The advantages of moving to this higher dimensional space (Fig. 2C) compared to simply using the image brightness gradient is that at points at which the brightness partial derivatives are zero, which occurs at peaks, troughs and points of inflection in the image brightness surface one of the components of the tangent vectors is set to zero. However the relative lengths of the tangent vectors in the now $n-1$-dimensional space remain the same. Consider two parallel vectors in three space, the relative lengths of the shadows they cast on any two-dimensional plane remain the same. Computation can only break down when all the components of the jet vector are zero, in which case the image structure must be uniformly grey and motion cannot be computed. However we need to return to ordinary 2D space–time to implement the integration over the 2D space–time region that forms part of our flow calculation. This can be accomplished by "pulling back" the information from the jet space by projection on tangent vectors.

4. Pullbacks of one-forms in the jet space

A general one-form in the tangent space can be written as

$$\omega = A^1 \mathrm{d}a^1 + A^2 \mathrm{d}a^2 + A^3 \mathrm{d}a^3 + \cdots + A^n \mathrm{d}a^n \tag{8}$$

where A^n are coefficients. We can convert a tangent vector in the jet space to a one-form (Fig. 2C) using the sharp operator, #.

$$\phi = \#(k_*(\mathbf{k})) \tag{9}$$

where k is the function defined in (5) and $\mathbf{k}$ is a unit vector. In this case the use of the sharp operator is largely a semantic exercise since the coefficients of ϕ are the same as those of $k_*(\mathbf{k})$.

The pullback of this one-form under the mapping $k(p,q,r)$ is indicated by the one-form $k^*(\phi)$, which has components

$$A = k^*(\phi)(\mathbf{i})\,\mathrm{d}p, \quad B = k^*(\phi)(\mathbf{j})\,\mathrm{d}q, \quad C = k^*(\phi)(\mathbf{k})\,\mathrm{d}r.$$

A one-form in 2D space–time can be visualised as parallel planar sheets with an associated direction (Fig. 2D). Contraction of a vector on a one-form delivers a value depending upon the number of sheets pierced by the vector, the sign of which depends upon whether the directions of the one form and vector agree (+) or are opposed (−). We will call the one-forms corresponding to the tangent vectors generated by the pushforward of the $\mathbf{i}$ and $\mathbf{j}$ unit vectors ψ and χ respectively.

Since these one-forms are in ordinary 2D space–time we can integrate over a volume as before. Consider the integral of M, a vector of one-forms, which we use in place of the gradient in the flow calculations in Section 2,

$$M = \begin{bmatrix} k^*(\psi)(\mathbf{i}) & k^*(\psi)(\mathbf{j}) & k^*(\psi)(\mathbf{k}) \\ k^*(\chi)(\mathbf{i}) & k^*(\chi)(\mathbf{j}) & k^*(\chi)(\mathbf{k}) \\ k^*(\psi)(\mathbf{i}) & k^*(\psi)(\mathbf{j}) & k^*(\psi)(\mathbf{k}) \end{bmatrix} \begin{bmatrix} \mathrm{d}p \\ \mathrm{d}q \\ \mathrm{d}r \end{bmatrix} \tag{10}$$

over the volume R. This integral delivers a vector of two forms

$$\int_R M \otimes (\mathbf{i},\mathbf{j},\mathbf{k})\,\mathrm{d}V_{x,y,t} = \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{bmatrix} \begin{bmatrix} \mathrm{d}q\,\mathrm{d}r \\ \mathrm{d}r\,\mathrm{d}p \\ \mathrm{d}p\,\mathrm{d}q \end{bmatrix}, \tag{11}$$

The outer product, $\otimes$, with $(\mathbf{i},\mathbf{j},\mathbf{k})$ indicates integration in the p, q, and r direction respectively. Thus A_1 is the integral of the first one-form taken in the direction indicated by $\mathbf{i}$. The resulting vector of two-forms can be used to compute the components of velocity and inverse velocity with reference to the frame $(\mathbf{i},\mathbf{j},\mathbf{k})$.

5. Flow of an implicit potential field

The variable one-form field generated by the pullback of the one-form corresponding to the tangent in the p direction can be envisaged as composed of local parallel planes in the parameter space (Fig. 2D). The sign of these one-forms depends upon the direction of motion rather than image structure, unlike the brightness gradient, the sign of which depends upon whether image intensity is increasing or decreasing. Speed can be computed by relating the number of planes cut by vectors in the spatial and temporal coordinate directions,

e.g. integrating over r gives the number of planes pierced by a vector whose length is that of the temporal dimension of the integration zone. The p component of the flow over an image segment parallel to the one-form field can be calculated from these three measures as detailed below. The q component of the flow can be calculated in an analogous way from the second two-form field.

If there is a potential field corresponding to one of the one-forms in (10), integrating terms in the first row over p, q and r respectively should give the same function which we will call $h(\mathbf{H})$. To find this function requires some straightforward but tedious analysis that was accomplished with the aid of the Maple symbolic mathematics package. For $f(\mathbf{H}) = f(\mathbf{v}\cdot\mathbf{H})$; $\mathbf{v} = (a,b,c) = f(ap+bq+cr)$ where $\mathbf{v}$ is a vector specifying the speed and direction of rigid motion, we compute the first three terms of the series

$$h(\mathbf{H}) = a\mathbf{D}f(\mathbf{v}\cdot\mathbf{P})^2(\mathbf{v}\cdot(\mathbf{H}-\mathbf{P})) + \frac{1}{3}a\mathbf{D}_2 f(\mathbf{v}\cdot\mathbf{P})^2 \times (\mathbf{v}\cdot(\mathbf{H}-\mathbf{P}))^3 + \frac{1}{20}a\mathbf{D}_3 f(\mathbf{v}\cdot\mathbf{P})^2(\mathbf{v}\cdot(\mathbf{H}-\mathbf{P}))^5 \tag{12}$$

where $\mathbf{D}_n$ is the nth-order differential operator. In the specific example of a moving sine grating, $\mathbf{P} = \mathbf{0}$; $f = \sin$ gives us $h(\mathbf{H}) = a(ap+bq+cr) + \frac{1}{20}a(ap+bq+cr)^5$.

Hence, in this case, we have an implicit function, the flow of which is computed, which is an increasing function in the direction of motion modified by squared derivatives of the image brightness at the point of interest.

If we were to compute image motion directly by integrating the image brightness on the boundary of a 2D space–time volume difficulties can arise. To take our example of a moving sine-wave grating, if the period of the sine function is the width of the region and it moves one cycle in the temporal window then the net change on the boundaries is zero. However, notice that we are not integrating over the one-forms representing the brightness gradient. Since the calculation we are using is equivalent to integrating over a function, which is positive and always increasing, the method does not suffer from this form of ill-conditioning.

6. Extracting speed and inverse speed

Now we consider computing these measures concurrently for a range of primary directions corresponding to a range of orientation/direction columns [29,30] in the primate visual system. It is convenient to introduce a notation for speed and inverse speed vectors. The speed, $\widehat{s}$ for a coordinate system with orientation θ is given by

$$\widehat{s}(\theta) = \left[\frac{C_1(\theta)}{\sqrt{(A_1^2(\theta)+B_1^2(\theta))}}\cos(\phi), \frac{C_2(\theta)}{\sqrt{(A_2^2(\theta)+B_2^2(\theta))}}\sin(\phi)\right] \tag{13}$$

where ϕ is orientation of local image contours with respect to the coordinate frame θ. This can be re-written in a way that allows well-conditioned calculation.

$$\widehat{s}(\theta) = \left[\frac{C_1(\theta)}{A_1(\theta)}\left(\frac{1}{1+\left(\frac{B_1(\theta)}{A_1(\theta)}\right)^2}\right), \frac{C_2(\theta)}{B_2(\theta)}\left(\frac{1}{1+\left(\frac{A_2(\theta)}{B_2(\theta)}\right)^2}\right)\right] \tag{14}$$

We will denote the first column vector as $\widehat{s}_{\parallel}$ and the second as $\widehat{s}_{\perp}$ to indicate the speed measures taken in the principle and orthogonal directions respectively. We can also calculate inverse speed, $\breve{s}$ from the third two-form, which after some manipulation gives us.

$$\breve{s}(\theta) = \left[\frac{A_3(\theta)}{C_3(\theta)}, \frac{B_3(\theta)}{C_3(\theta)}\right] \tag{15}$$

Radial functions are calculated over 2π radians. Fig. 2E illustrates that, for a translating grating, the functions are sinusoidal. In general, the translation component is antisymmetric with a period of 2π radians. Local divergence components simply add to the mean level of the radial function. To recover the translation component and remove any divergence we force the integrals of the radial functions to be zero. This can be accomplished by extracting the fundamental Fourier coefficients by projecting onto fiducial sine and cosine functions. We construct normalised cosine and sine vectors

$$\mathbf{F}(\theta) = (F_{\parallel}(\theta), F_{\perp}(\theta)) = \sqrt{\frac{2}{n}}[\cos(\theta), \sin(\theta)]. \tag{16}$$

This matrix both forms a reference frame for the computation of direction of motion and extracts the fundamental Fourier coefficients of the radial functions. Speed is computed as

$$S^2 = \frac{\left|\begin{pmatrix} \widehat{s}_{\parallel}\cdot F_{\parallel} & \widehat{s}_{\parallel}\cdot F_{\perp} \\ \widehat{s}_{\perp}\cdot F_{\parallel} & \widehat{s}_{\perp}\cdot F_{\perp}\end{pmatrix}\right|}{\left|\begin{pmatrix} \widehat{s}_{\parallel}\cdot \breve{s}_{\parallel} & \widehat{s}_{\parallel}\cdot \breve{s}_{\perp} \\ \widehat{s}_{\perp}\cdot \breve{s}_{\parallel} & \widehat{s}_{\perp}\cdot \breve{s}_{\perp}\end{pmatrix}\right|} \tag{17}$$

where, for example $\widehat{s}_{\parallel}\cdot F_{\parallel}$ is the scalar product first column of $\widehat{s}$ and the first column of $\mathbf{F}$. Direction is given by

$$\text{direction} = \tan^{-1}\left(\frac{(\widehat{s}_{\parallel}+\breve{s}_{\parallel})\cdot F_{\parallel} - (\widehat{s}_{\perp}+\breve{s}_{\perp})\cdot F_{\perp}}{(\widehat{s}_{\parallel}+\breve{s}_{\parallel})\cdot F_{\perp} + (\widehat{s}_{\perp}+\breve{s}_{\perp})\cdot F_{\parallel}}\right) \tag{18}$$

Speed and direction are represented separately but can combined to deliver the local vector field (Fig. 2F).

Speed is computed as a ratio of determinants or a ratio of areas. Ratios of scalars are straightforward. For vectors relative projective lengths serve in place of standard ratios as in $\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}$, or equivalently $\omega(\mathbf{u})$ when expressed as the contraction of a vector on a one-form. Eq. (17) can be interpreted as a relative projected area. It can also be written as the contraction of a bivector on a two-form $(\omega \wedge \beta)(\mathbf{u}, \mathbf{v})$ where

$$\omega = [\breve{s}_{\parallel} \cdot \hat{s}_{\parallel}, \breve{s}_{\parallel} \cdot \hat{s}_{\perp}], \quad \beta = [\hat{s}_{\perp} \cdot \breve{s}_{\parallel}, \hat{s}_{\perp} \cdot \breve{s}_{\perp}],$$
$$\mathbf{u} = \begin{bmatrix} \hat{s}_{\parallel} \cdot F_{\parallel} \\ \hat{s}_{\parallel} \cdot F_{\perp} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \hat{s}_{\perp} \cdot F_{\parallel} \\ \hat{s}_{\perp} \cdot F_{\perp} \end{bmatrix}. \tag{19}$$

7. Simulations

In contrast with approaches which introduce additional constants or rely on noise to condition equations [6,31,32] the model provides accurate speed measures for both 1D and 2D spatial patterns. For a grating moving at 2°/s the direction indicated is orthogonal to motion contours, the average speed is correct to two significant figures with zero variance. Turning to the aperture problem. Fig. 3A and B shows the results of the model for a translating Gaussian patch. One can see that motion parallel to image contours can be recovered for 2D spatial patterns reasonably well, without an additional regularisation stage [4]. The model also provides an implicit solution to the recovery of plaid motion and speed [33]. We measured computed speed for symmetric plaids in which component speed was held constant. Pattern speed was manipulated by changing the angle between the components. The model gives the correct direction of motion and the correct IOC speed (Fig. 3C).

Since the model does not resort to regularisation in order to solve the aperture problem, we should expect to

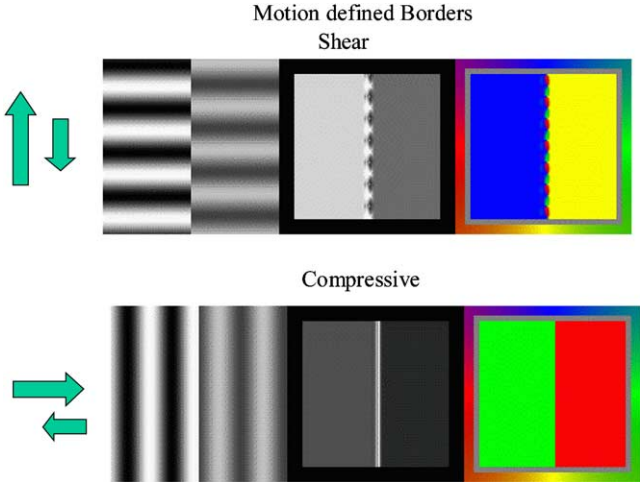

Fig. 4. Compression and sheer boundaries for sine gratings illustrating the presence of sharp boundaries even though the blur kernel is defined over a 23×23 pixel region (128×128 pixel images). The leftmost pictures show a frame from the sequence. The central pictures shows speed coded in terms of brightness and the third pair of images shows direction. Differences in the contrast of the stimulus pattern have no effect on the computed velocity.

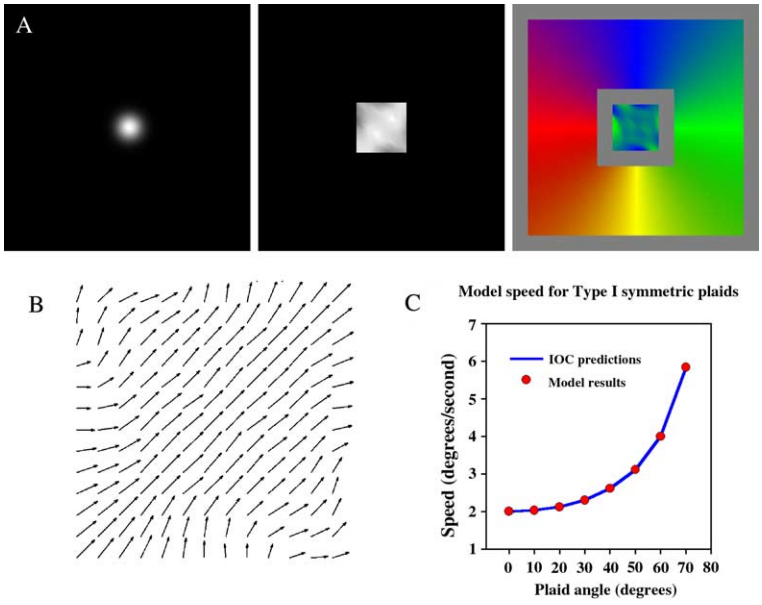

Fig. 3. The figure shows dense direction and speed images computed for an arbitrarily chosen frame in a motion sequence. The stimulus frame is shown on the left. In the centre a speed map is displayed which is scaled to the full brightness range. In the rightmost column the corresponding direction map is plotted, with direction coded by colour which should be read with reference to the colour wheel. (A) A Gaussian patch moving up and to the right. (B) A vector plot of the data in A. (C) Computed speed (maximum speed) for a symmetric plaid as a function of the IOC speed. Component speed is constant and plaid speed is manipulated by varying the angular difference between the components.

be able to generate well defined motion boundaries. Fig. 4 illustrates the results for shearing and compression boundaries. Boundaries are well defined, which would allow subsequent mechanisms to segment the scene. The precision of the boundary is almost the resolution of the input array, calibrated at 30″/pixel, even though the linear filters at the first stage of analysis are defined in a 23×23 pixel kernel.

8. Mapping the model onto the physiology

The identification of early filtering operations, and what we have characterised here as the pullback of tangent vectors in the jet space, with directionally selective simple and complex cells has been made in earlier work [5]. Individual cells in visual cortex may have many roles depending upon the goal of the system to which they are recruited. In addition to the variable number of differentiating filters in the primary direction, three temporal filters and no, mono-lateral and bi-lateral end inhibition provide the minimum neural substrate which will allow the calculation of image gradients in the spatio-temporal neighbourhood about the centre of the aggregate receptive field.

The theory predicts the existence of cells encoding inverse speed. Evidence for inverse speed coding exists in a number of reports of directionally selective neurones that are "low pass" in relation to speed i.e. they reduce their firing rate with increased image speed [34–37]. To our knowledge the functional role of these cells has not received much attention prior to this. The model also envisages that speed and inverse speed is rate coded. The observation that most cells in V5 and V4 are "tuned" to image speeds of 64°/s and that very few are tuned to velocities under 4°/s [38] suggests a rate coding scheme for speed, particularly if one takes into account the upper temporal frequency cut-off of the visual system, at around 60 Hz, [39] will limit the response range of speed coded cells.

It is proposed that the radial functions in Fig. 2E are represented by the responses of V5/MT cells tuned to different directions and that these outputs are combined to give a rate coded velocity signal. Albright [30] classified V5/MT neurones as type I if the optimal direction of motion was orthogonal to their preferred orientation, measured with flashing bars, and as type II if the preferred direction of motion was parallel to their preferred orientation. Here, motion in the primary direction, $\widehat{s}_{\parallel}$ & $\breve{s}_{\parallel}$ in (14) and (15), is computed by spatial differentiation in a direction orthogonal to orientation tuning and motion in the orthogonal direction is computed via "end stopped" units, $\widehat{s}_{\perp}$ & $\breve{s}_{\perp}$ in (14) and (15), filters in which additional spatial derivatives are taken parallel to the preferred orientation of the filter. In practice, the preferred direction of motion will be slightly offset from the preferred orientation. Thus filters computing motion in the primary direction appear to correspond to Albright's type I cells and those in the orthogonal direction to type II. In Fig. 5 we plot the response of speed coded units, which measure speed in the principle and orthogonal directions, as a function of the direction of motion of gratings and plaids. Both classical pattern cell behaviour and component cell behaviour could be elicited from the same directionally selective primary or orthogonal units (Fig. 5C–F). There is considerable variability in the radial functions for plaid stimuli dependent upon the chosen position on the plaid, which is consistent with the variability found in physiological data [40]. The observed behaviour does not require the combination of signals from different "direction columns". Albright, Desimone and Gross [30] report a small percentage of V5/MT cells with bi-directional and pan-directional tuning functions. No clear functional role has been attributed to these neurons. Simulations of direction tuning functions (Fig. 5B) for the pairs of products in the two determinants in (17) delivered two bi-directional units (upper determinant) and two pan-directional cells (lower determinant). Pan and Bi are in equal proportions as predicted here ($\approx$45 of each type in their sample). If we assume a direction column separation of 15 degrees, the population data of Albright et al. indicate a ratio of 2:2:1 relating the total number of pan-directional and bi-directional cells to the number of directionally selective cells (around 20 for each orientation) in a single direction column. Since we have four types of units in each column (primary/speed, primary/inverse speed, orthogonal/speed, orthogonal/inverse speed), we would have predicted a ratio of 2:2:4. It is possible that 30 degree separations for direction columns might be more appropriate or that some classes of cells (inverse speed or orthogonal direction) may have been missed due to the choice of stimulus, or the excess numbers of Bi and Pan cells in the population data indicate additional pooling of directional cells not present in the current model. Further work will be needed to determine the relationship between Bi and Pan cells and directional cells.

We have not, as yet, incorporated the changes in receptive field size seen between V1 and V5/MT. However, we envisage the later stage of the theory to provide a model of information processing in area V5/MT. It is this level that provides for the encoding of plaid speed and direction, well-specified motion defined boundaries and resistance to static noise [12]. Support for a late stage of noise reduction comes from neuropsychological evidence that neurological damage to V5 results in severe degradation of the capacity to report both the direction of movement of random dots [41] and the ability to detect the presence of biological motion from point light displays [42] in the presence of *static* noise dots. Lesioning of macaque V5 leads to raised motion co-

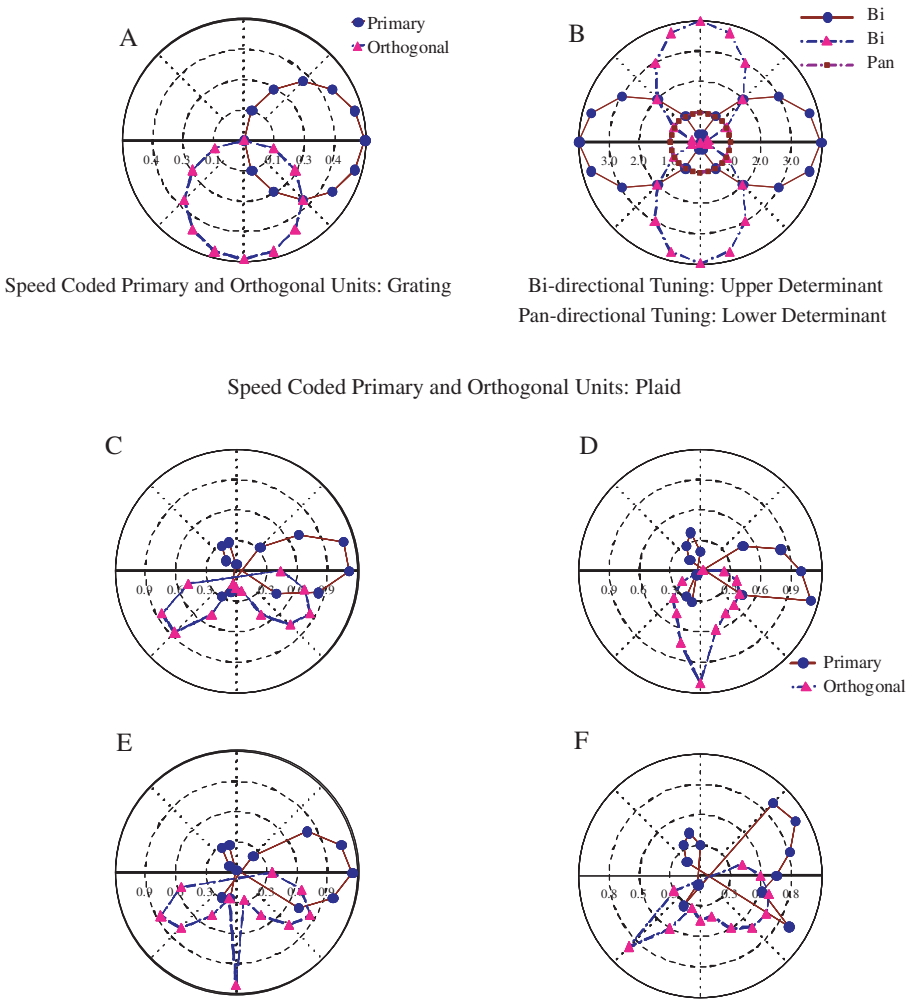

Fig. 5. (A) Polar plots of direction tuning curves for units computing speed in primary and orthogonal directions for gratings. Both units are tuned to rightward motion. Curves for the orthogonal unit are rotated by 90°. (B) Directional tuning curves for products in the upper determinant of the speed quotient (bi-directional) and lower determinant (pan-directional). For gratings one of the pan-directional cells would give a zero-response. (C–F) Varieties of directional tuning curves for plaids showing component type and pattern type behaviour. In each case the units compute the same function—the variation is due to change in position on the plaid.

herence thresholds, defined as the percentage of dots in a random dynamic display moving in the same direction necessary to support direction discrimination [43], and impairment in the ability to use motion defined boundaries as a cue to form discrimination [44].

9. Conclusion

A model based on the computation of image flow over a bounded region of the image identified with an aggregate receptive field can account for many aspects of 2D motion perception in a computationally robust way. The model recovers accurate optic flow information without recourse to regularisation and is noise resistant. The theory requires a hierarchy of computational components that map well onto the known properties of motion sensitive neurones and the computational architecture of the primate visual system.

References

[1] W. Reichardt, in: W.A. Rosenblith (Ed.), Sensory Communication, Wiley, New York, 1961, pp. 303–317.
[2] E.H. Adelson, J.R. Bergen, Spatiotemporal energy models for the perception of motion, J. Opt. Soc. Am. A 2 (1985) 284–299.
[3] C.L. Fennema, W.B. Thompson, Velocity determination in scenes containing several moving objects, Comput. Graph. Image Process. 9 (1979) 301–315.
[4] B.K.P. Horn, B.G. Schunck, Determining optical flow, Artif. Intel. 17 (1981) 185–203.
[5] A. Johnston, P.W. McOwan, H.A. Buxton, Computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells, Proc. R. Soc. London B 250 (1992) 297–306.
[6] D.J. Heeger, E.P. Simoncelli, in: L. Harris (Ed.), Spatial Vision in Humans and Robots, Cambridge University Press, Cambridge, 1995.
[7] D.J. Heeger, Model for the extraction of image flow, J. Opt. Soc. Am. A 4 (1987) 1455–1471.
[8] E.P. Simoncelli, D.J. Heeger, A model of neuronal responses in visual area MT, Vis. Res. 38 (1998) 743–762.

[9] A. Johnston, M.J. Wright, Matching velocity in central and peripheral vision, Vis. Res. 26 (1986) 1099–1109.
[10] P. Thompson, Perceived rate of movement depends upon contrast, Vis. Res. 22 (1982) 377–380.
[11] A. Johnston, C.P. Benton, M.J. Morgan, Concurrent measurement of perceived speed and speed discrimination threshold using the method of single stimuli, Vis. Res. 39 (1999) 3849–3854.
[12] A. Johnston, P.W. McOwan, C.P. Benton, Robust velocity computation from a biologically motivated model of motion perception, Proc. R. Soc. London B 266 (1999) 509–518.
[13] A. Johnston, C.P. Benton, P.W. McOwan, Induced motion at texture-defined motion boundaries, Proc. R. Soc. London B 266 (1999) 2441–2450.
[14] C.P. Benton, A. Johnston, A new approach to analysing texture-defined motion, Proc. R. Soc. London B 266 (2002) 2441–2450.
[15] E.H. Adelson, J.R. Bergen, in: M. Landy, J.A. Movshon (Eds.), Computational Models of Visual Processing, MIT Press, Cambridge, MA, 1991, pp. 1–20.
[16] S. Weintraub, Differential Forms: A Complement to Vector Calculus, Academic Press, San Diego, 1997.
[17] A. Mitiche, P. Bouthemy, Computation and analysis of image motion: a synopsis of current problems and methods, In. J. Comput. Vis. 19 (1996) 29–55.
[18] J.J. Koenderink, Solid Shape, MIT Press, Cambridge, MA, 1990.
[19] N. Gupta, L. Kanal, Gradient based image motion estimation without computing gradients, Int. J. Comput. Vis. 22 (1997) 81–101.
[20] J.J. Koenderink, A.J. van Doorn, Representation of local geometry in the visual system, Biol. Cybernet. 55 (1987) 367–375.
[21] J.J. Koenderink, Operational significance of receptive field assemblies, Biol. Cybernet. 58 (1988) 163–171.
[22] J.J. Koenderink, A.J. van Doorn, Receptive field assembly pattern specificity, J. Vis. Commun. Image Rep. 3 (1992) 1–12.
[23] R.A.Young, R.M. Lesperance, A physiological model of motion analysis for machine vision, Technical Report, General Motors Research Laboratories GMR-7878, 1993, pp. 1–76.
[24] V. Bruce, P.R. Green, M.A. Georgeson, Visual Perception: Physiology, Psychology & Ecology, third ed., Laurence Erlbaum Associates, Hove, 1996.
[25] A. Johnston, C.W.G. Clifford, A unified account of three apparent motion illusions, Vis. Res. 35 (1995) 1109–1123.
[26] P. Werkhoven, J. Koenderink, Extraction of motion parallax structure in the visual system 1, Biol. Cybernet. 63 (1990) 185–191.
[27] M. Otte, H. Nagel, Estimation of optical flow based on higher-order spatio-temporal derivatives in interlaced and non-interlaced image sequences, Artif. Intel. 78 (1995) 5–43.
[28] S. Lang, Calculus of Several Variables, Springer-Verlag, New York, 1987.
[29] D.H. Hubel, T.N. Wiesel, Sequence regularity and geometry of orientation columns in the monkey striate cortex, J. Comp. Neurol. 158 (1974) 267–294.
[30] T.D. Albright, R. Desimone, C.G. Gross, Columnar organisation of directionally selective cells in visual area MT of the macaque, J. Neurophysiol. 51 (1984) 16–31.
[31] A. Verri, M. Straforini, V. Torre, Computational aspects of motion perception in natural and artificial vision systems, Philos. Trans. R. Soc. London B 337 (1992) 429–443.
[32] M.V. Srinivasan, Generalised gradient scheme for the measurement of two-dimensional image motion, Biol. Cybernet. 63 (1990) 421–431.
[33] E.H. Adelson, J.T. Movshon, Phenomenal coherence of moving visual patterns, Nature 300 (1982) 523–525.
[34] G.A. Orban, H. Kennedy, H. Maes, Response to movement of neurones in areas 17 and 18 of the cat, J. Neurophsiol. 45 (1981) 1043–1058.
[35] A. Mikami, W.T. Newsome, R.H. Wurtz, Motion selectivity in macaque visual cortex I. Mechanisms of direction and speed selectivity in extrastriate area MT, J. Neurophysiol. 55 (1986) 1308–1327.
[36] H.R. Rodman, T.D. Albright, Coding of stimulus velocity in area MT of the macaque, Vis. Res. 27 (1987) 2035–2048.
[37] L. Lagae, S. Raiguel, G.A. Orban, Speed and direction selectivity of macaque middle temporal neurones, J. Neurophysiol. 69 (1993) 19–39.
[38] K. Cheng, T. Hasegawa, K.S. Saleem, K. Tanaka, 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 (1994) 2269–2280.
[39] D.H. Kelly, Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements, J. Opt. Soc. Am. 51 (1961) 422.
[40] J.A. Movshon, E.H. Adelson, E.H. Gizzi, W.T. Newsome, in: C. Chagass, R. Gattas, C. Gross (Eds.), Pattern Recognition Mechanisms, Springer, New York, 1985.
[41] C. Baker, R. Hess, J. Zihl, Residual motion perception in a motion blind patient, assessed with limited-lifetime random dot stimuli, J. Neurosci. 11 (1991) 454–461.
[42] P. McLeod, W. Dittrich, J. Driver, D. Perrett, J. Zihl, Preserved and impaired detection of structure from motion by a motion-blind patient, Vis. Cogn. 3 (1996) 363–391.
[43] W.T. Newsome, E.B. Pare, A selective impairment of motion perception following lesions of the middle temporal visual area (MT), J. Neurosci. 8 (1988) 2201–2211.
[44] V.L. Marcar, D.K. Xiao, S.E. Raiguel, H. Maes, G.A. Orban, Processing of kinetically defined boundaries in the cortical motion area MT of macaque monkey, J. Neurophysiol. 74 (1995) 1258–1271.