PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Efficient Coding of Natural Time Varying Images in the Early Visual System

Michael P. Eckert and Gershon Buchsbaum

Phil. Trans. R. Soc. Lond. B 1993 **339**, 385-395
doi: 10.1098/rstb.1993.0038

Efficient coding of natural time varying images in the early visual system

MICHAEL P. ECKERT AND GERSHON BUCHSBAUM†

Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, 220 S. 33rd Street, Philadelphia, Pennsylvania 19104-6315, U.S.A.

SUMMARY

We investigate the hypothesis that the early visual system efficiently codes natural time varying images, first by tracking part of the image, then by matching the spatiotemporal properties of the neural pathway to those of the tracked image. A representation for the time varying image is formulated which consists of two spatiotemporal components, a velocity field component and a stationary component. We show, using digitized sequences of natural images, that the spatiotemporal spectrum and other attributes of the image markedly differ before and after tracking. The temporal frequency bandwidth and velocity distribution of the velocity field component are diminished in the region of tracking and broaden with increasing eccentricity from this region. On the other hand, the spectrum of the stationary component is unaffected by tracking. Comparison of the properties of the tracked image to those of the M and P pathways suggests that each pathway transmits different attributes of the tracked image. A retinal architecture which varies with eccentricity also matches the properties of the tracked image.

1. INTRODUCTION

Natural images contain spatiotemporal information comprised of motion and other time varying details such as flicker. Motion in the retinal image includes object motion in the visual scene, observer motion, and eye motion. Image motion presents a significant problem for efficient coding and representation of images in the visual system. The visual system must code and interpret the visual scene while accounting for objects moving at velocities which may exceed the temporal limitations of visual system processing.

The visual system confronts this complex time varying signal with two spatiotemporal mechanisms: eye movements and the spatiotemporal filter arrays known as the M and P pathways. In this paper, we examine how eye movements and the M and P pathways conjoin to make an efficient coder of the time varying image. Eye movements limit the temporal bandwidth of images by reducing the range of velocities reaching the fovea. M and P pathways efficiently carry image components, modified by eye movements, for analysis at cortical levels. The spatiotemporal properties of ganglion and lateral geniculate cells which form the M and P pathways have been extensively investigated in recent years (Kaplan & Shapley 1982; Hicks *et al.* 1982; Derrington & Lennie 1984; Blakemore & Vital–Durand 1986; Crook *et al.* 1988; Lee *et al.* 1989*a*; Purpura *et al.* 1990), especially the role of these pathways in coding spatiotemporal image components (Shapley & Perry 1986; Merigan 1986; Merigan 1989; Merigan & Maunsell 1990; Merigan *et al.* 1991; and Schiller *et al.* 1990). Generally, the M pathway is associated with fast temporal changes and the P pathway with high spatial acuity and colour, although there is considerable overlap across spatial and temporal frequencies.

The idea that the visual system efficiently codes the visual scene is not a new one (Barlow 1961, 1981; Snyder *et al.* 1977; Srinivisan *et al.* 1982; Buchsbaum & Gottschalk 1983; Laughlin 1983; Field 1987; Tsukamoto *et al.* 1990; Watson 1990; Derrico & Buchsbaum 1991). Under the efficient coding hypothesis, the purpose of retinal processing is to transmit visual information as effectively as possible to higher visual centers. This means that the visual system optimizes its coding strategy, given the physiological constraints of limited dynamic range of nerves, noise, and limited spatial and temporal bandwidths.

A general block diagram of the coding system under investigation is presented in figure 1. The coder is comprised of two components, the pre-retinal eye movements, modelled as a linear time variant filter, and the retinal spatiotemporal pathways, modelled as linear time invariant filters. Investigation of efficiency and other properties of the coder requires an understanding of the signal environment in which it operates. For the visual system, the environment is an observer freely viewing natural images. We begin by investigating the spatiotemporal spectrum of the time varying image (the input, $I(\boldsymbol{u},t)$, in figure 1) and the effects of tracking on the spatiotemporal spectrum and other properties of the image (the signal at the retina, $I_r(\boldsymbol{u},t)$, in figure 1). A representation of natural images

† To whom correspondence should be sent.

Phil. Trans. R. Soc. Lond. B (1993) **339,** 385–395
Printed in Great Britain

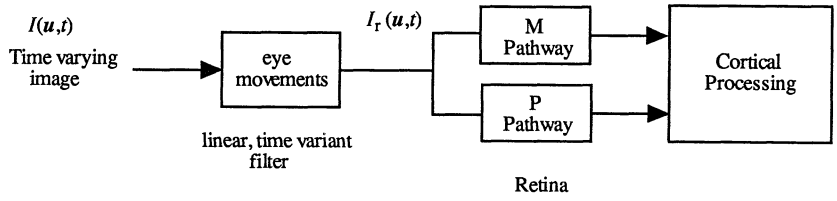


Figure 1. Block diagram of the flow of visual information through the visual system. The original image, $I(\boldsymbol{u},t)$, is filtered by eye movements. Eye movements are modeled as a linear time variant filter. The image reaching the retina, $I_r(\boldsymbol{u},t)$, is a filtered version of the original image. The M and P pathways, which are time invariant filters, further process the image before it reaches the cortex for analysis.

is formulated as a combination of a velocity field component and a stationary component which have markedly different spatiotemporal spectra. We then apply a tracking algorithm to natural image sequences to investigate how these two image components are affected. As expected, tracking reduces the temporal frequency bandwidth of the velocity field component at the point of tracking (Jain & Jain 1981; Girod 1987), but the temporal frequency bandwidth and velocity distribution broaden with increasing eccentricity from the point of tracking. The spectrum of the stationary component is not affected by tracking.

We discuss the properties of the M and P pathways and how they match the spatiotemporal components of images after tracking. We give special attention to calculating the effect of tracking in the region of tracking and at increasing eccentricities from it. This is needed to evaluate the advantage of a retinal architecture with a velocity tuning which changes with eccentricity.

2. SPATIOTEMPORAL SPECTRUM OF TIME VARYING IMAGES AND THE EFFECT OF TRACKING

(a) *Spectrum and velocity distribution of images*

Spatiotemporal variations in a visual scene generally arise from motion. On the image plane of the retina, motion can be approximated with a two-dimensional velocity field. The velocity field assigns a translational velocity vector to each point in space, and characterizes time variations resulting from geometric motion in the scene, including rotation, dilation, and affine deformations commonly found with perspective projections of three-dimensional motion. The velocity field serves as a basis for many computational models of human motion processing (Adelson & Bergen 1985; van Santen & Sperling 1985; Watson & Ahumada 1985; Heeger 1987). However, the velocity field cannot account for spatiotemporal changes of the image such as flicker and the photometric effects of motion (Pentland, 1991). To include all spatiotemporal changes, we model images as a combination of two uncorrelated components, a velocity field component and a stationary component.

$$I(\boldsymbol{u},t) = I(\boldsymbol{u} - \int_{t_0}^{t} \boldsymbol{v}(\boldsymbol{u},t')\mathrm{d}t', t_0) + s_s(\boldsymbol{u},t), \qquad (1)$$

where $I(\boldsymbol{u},t)$ is the intensity at spatial point $\boldsymbol{u}$ and time t, $I(\boldsymbol{u},t_0)$ specifies the initial image intensity, $\boldsymbol{v}(\boldsymbol{u},t)$ is the velocity field assigning a velocity vector, $\boldsymbol{v} = (v_x,v_y)$, to each point of space and time, and $s_s(\boldsymbol{u},t)$ is the stationary component.

The stationary component can be formed by removing local translational motion. Conceptually, this operation is analogous to filtering the image with a space–time variant filter which removes the velocity field component. The residual spatiotemporal intensity variations, the stationary component, will consist of flicker of the illuminant, the photometric effects of motion, and the occlusion and disocclusion at the edges of moving objects. These spatiotemporal effects are biologically relevant. For example, photometric motion provides depth and three-dimensional structure information about the image (Pentland 1991), and occlusion effects provide information about the location of object edges and relative depth. While these effects are caused by object motion, they cannot be removed by translational shifts of image intensity, and thus cannot be incorporated into the velocity field component.

The space and time variant spatiotemporal spectrum derived from the model of equation (1) is

$$S(\boldsymbol{u},t,\boldsymbol{k},f) = S(\boldsymbol{k})\delta(f - \boldsymbol{v}(\boldsymbol{u},t)\cdot\boldsymbol{k}) + S_s(\boldsymbol{k},f), \qquad (2)$$

where $\delta(\ )$ is the Dirac delta function, $\boldsymbol{k} = (k_x,k_y)$ is a two-dimensional spatial frequency vector, f is temporal frequency, $S(\boldsymbol{k})$ is the spatial power spectrum of $I(\boldsymbol{u},t_0)$, and $S_s(\boldsymbol{k},f)$ is the spatiotemporal spectrum of the stationary component. By definition, the velocity field and stationary component are uncorrelated, so the spectrum is the sum of the two components. In local spatiotemporal regions, the energy of the velocity field component exists on a plane in the three dimensions of frequency space (2 dimensions spatial, 1 dimension temporal), where the local velocity determines the orientation of the plane (Watson 1983; Watson & Ahamada 1985). A highly ordered structure does not exist for the stationary component which is distributed throughout spatiotemporal frequency space. Figure 2 illustrates the differences between the spectra of these components in a local spatiotemporal neighborhood.

The spatiotemporal spectrum (equation 2) can also be defined in terms of a velocity distribution, which is a probability distribution (histogram) of velocities. The velocity distribution is formed by sampling the velocity field through time at every spatial point in

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

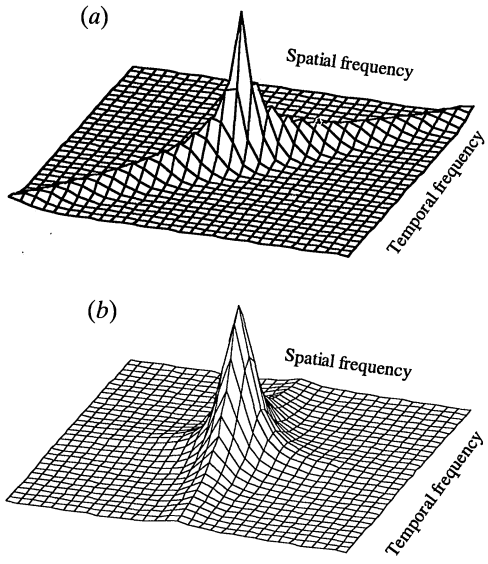


Figure 2. The velocity field and stationary components of time varying images occupy different regions of spatiotemporal frequency space. (*a*) The energy of translational motion lies along a line with a slope equal to velocity in spatiotemporal frequency space (a plane when spatial frequency is two dimensional). This figure illustrates an idealized case where the velocity field component has a constant translational velocity of 1 in the appropriate spatial and temporal frequency units. (*b*) The stationary component is modeled as the product of separable functions of spatial and temporal frequency which fall off inversely with spatial and temporal frequency.

the image. The result is a set of distributions, each representing the velocity variability in a region of space. Formation of the velocity distribution removes the functional dependence of time from the spatiotemporal spectrum, so it becomes a time averaged, but spatially localized, spatiotemporal spectrum. The formulation of the spectrum based on the velocity distribution is:

$$S(\boldsymbol{u},\boldsymbol{k},f) = h_{\mathrm{v}}(\boldsymbol{u},\boldsymbol{v})\mathrm{S}(\boldsymbol{k}) + S_{\mathrm{s}}(\boldsymbol{k},f), \quad (3)$$

where $h_{\mathrm{v}}(\boldsymbol{u},\boldsymbol{v})$ is the distribution of velocity, $\boldsymbol{v}$, at spatial point, $\boldsymbol{u}$, and $\mathrm{S}(\boldsymbol{k})$ is the spatial power spectrum of the image.

(*b*) *Effect of tracking on the velocity field and spatiotemporal spectrum*

Eye movements can be modeled as a linear, time variant filter introduced between the image scene and the retina (figure 1). Inclusion of eye movements introduces a single time varying vector component to the velocity field. As eye movements are fully described by introducing a term to the velocity field, they do not modify the stationary component. The spatiotemporal spectrum after eye movements is:

$$S_{\mathrm{r}}(\boldsymbol{u},t,\boldsymbol{k},f) = S(\boldsymbol{k})\delta(f - [\boldsymbol{v}(\boldsymbol{u},t) - \boldsymbol{v}_{\mathrm{c}}(t)] \cdot \boldsymbol{k}) + S_{\mathrm{s}}(\boldsymbol{k},f), \quad (4)$$

where $S_{\mathrm{r}}(\boldsymbol{u},t,\boldsymbol{k},f)$ is the spatiotemporal spectrum which has been filtered by eye movements, and $\boldsymbol{v}_{\mathrm{c}}(t)$ is the eye velocity at time, t.

Eye movements have the effect of shifting all velocities in the image by the eye velocity. In the case of tracking, eye velocity is set equal to the velocity at spatial point, $\boldsymbol{u}_0$.

$$\boldsymbol{v}_{\mathrm{c}}(t) = \boldsymbol{v}(\boldsymbol{u}_0,t). \quad (5)$$

Tracking has the effect of minimizing the spread of the velocity distribution, $h_{\mathrm{v}}(\boldsymbol{u},\boldsymbol{v})$, which decreases the temporal bandwidth of the signal in the neighborhood of the tracked point, $\boldsymbol{u}_0$. For perfect tracking, the velocity distribution becomes a delta function at point, $\boldsymbol{u}_0$, and we have the spatiotemporal spectrum,

$$S(\boldsymbol{u}_0,t,\boldsymbol{k},f) = S(\boldsymbol{k})\delta(f) + S_{\mathrm{s}}(\boldsymbol{k},f). \quad (6)$$

Because tracking compensates for the velocity field component in the region of $\boldsymbol{u}_0$, the temporal variations are contributed by the stationary component, $S_{\mathrm{s}}(\boldsymbol{k},f)$.

For points away from $\boldsymbol{u}_0$, the spectrum is weighted by the velocity distribution,

$$S(|\boldsymbol{u}-\boldsymbol{u}_0|,\boldsymbol{k},f) = h_{\mathrm{v}}(|\boldsymbol{u}-\boldsymbol{u}_0|,\boldsymbol{v})S(\boldsymbol{k}) + S_{\mathrm{s}}(\boldsymbol{k},f), \quad (7)$$

where $h_{\mathrm{v}}(|\boldsymbol{u}-\boldsymbol{u}_0|,\boldsymbol{v})$ is a velocity distribution which broadens with increasing eccentricity, $|\boldsymbol{u}-\boldsymbol{u}_0|$. In the region around $\boldsymbol{u}_0$, tracking narrows the velocity distribution, thereby reducing the variability of the spatiotemporal spectrum and the temporal frequency bandwidth. With increasing eccentricity from the point of tracking, the spatiotemporal spectrum will have a broader velocity distribution and larger temporal frequency bandwidth. The degree to which tracking narrows the velocity distribution away from the point $\boldsymbol{u}_0$ depends on the spatial correlation of the velocity field, which is a measure of the change of the velocity field across space. For a highly correlated velocity field which changes slowly through space, tracking can reduce the velocity distribution at relatively large distances from the point of tracking.

Implementation of tracking invariably requires a feedback loop which estimates position and velocity from a time delayed input and past expectations (Stark *et al.* 1962; Lisberger *et al.* 1987; Steinman *et al.* 1990). This means that eye velocity can be set only to an estimated value of image velocity, rather than the true image velocity. As a result, the velocity distribution in the tracked region, $h_{\mathrm{v}}(0,\boldsymbol{v})$, will have a spread related to the effectiveness of tracking. Tracking effectiveness is a signal related phenomenon, so that highly predictable motion will be more effectively tracked than unpredictable motion (Stark *et al.* 1962; Barnes & Lawson 1989). However, tracking of 'real world' motion can be quite accurate (Steinman *et al.* 1990), usually maintaining a foveal velocity of less than 1–2 deg s^{-1}.

3. THE EFFECT OF TRACKING ON DIGITIZED IMAGE SEQUENCES

We calculated the velocity field, velocity distribution, and spatiotemporal spectrum of four real world image sequences before and after tracking objects in the sequences. The sequences (256 pixels × 256 pixels × 64 frames at 8 bits per pixel, 30 frames per second with no scene cuts) were taken from a video disk which

Table 1. *Description of sequences*

sequence number	sequence description
IJ10833	Jungle scene with some three-dimensional object motion and a small amount of camera motion
IJ12426	Man walking. Some camera motion to keep man centred in visual field. The result is significant amounts of background motion
IJ01300	Man talking while moving head occasionally. No camera motion. Some slight motion in the background
IJ04454	Storm scene. No camera motion, but large amounts of non-rigid three-dimensional motion from waves. Intensity changes from lightning

contained scenes from movies. The frame rate of 30 frames per second limits the maximum estimated temporal frequency bandwidth of the images to 15 Hz. However, image energy drops off quickly with temporal frequency, and we found that signal energy is concentrated below 10 Hz. This suggests that any aliasing introduced by the sampling rate has little effect on the estimated spectrum. Each sequence was selected to contain varying levels of motion activity to form the broadest possible ensemble of images with the small sample size (see table 1). The velocity field for each frame was estimated by minimizing the squared difference between 24 pixel × 24 pixel blocks in two sequential frames of the sequence (Jain & Jain 1981). The same method was used to track selected regions in the sequence (see figure 3). While this (minimization of sum of squared diferences) algorithm is unlikely to be the method used by the visual system,

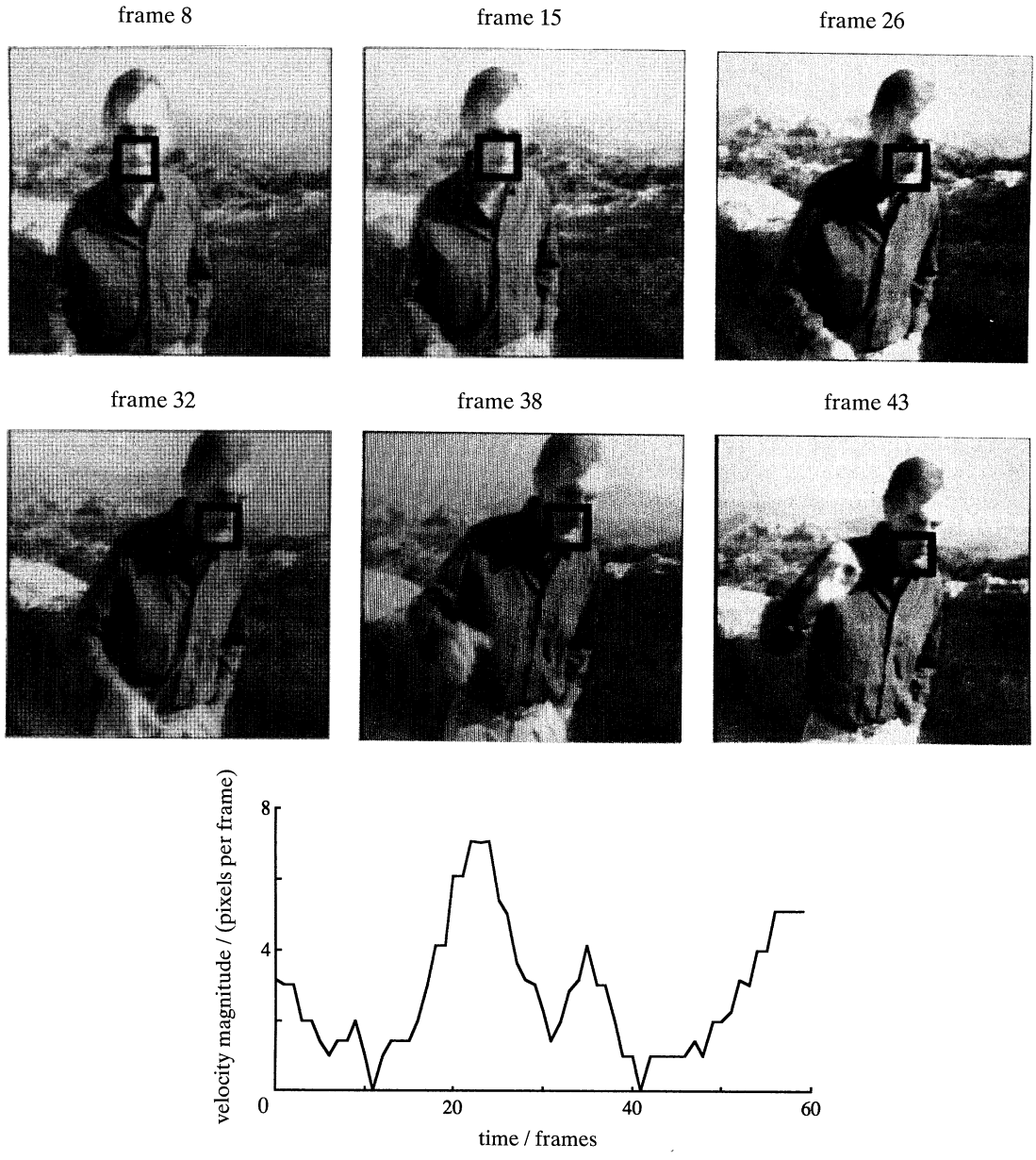


Figure 3. A typical image sequence over which the spatiotemporal statistics were analysed (sequence IJ12426). The black box indicates a region which was selected for tracking. The graph illustrates the magnitude of the velocity of the tracked region as a function of time.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

it tracks a region of the image and keeps it centred, as does smooth pursuit.

We examined the effect of tracking on the velocity distribution in the region of tracking and at increasing eccentricities from that region. The velocity distribution of each 64 frame sequence was calculated from the velocity field as the frequency of occurrence of the velocity magnitude. To calculate the change of the velocity distribution with eccentricity, a region of interest was selected and tracked by shifting the entire image for each frame to maintain the tracked region in the same spatial location (figure 3). The velocity distribution and average velocity for the tracked image were then computed as a function of eccentricity from the point of tracking. The results are presented in figures 4 and 5. Before tracking, the velocity distribution and average velocity vary across the spatial extent of the image, but we found no trend from sequence to sequence. This is expected, as there is no reason why one part of the scene should experience more motion than any other part. When tracking is performed, the velocity distribution at the point of tracking is a delta function (for perfect tracking), and broadens with eccentricity. Higher velocities occur more frequently with increasing eccentricity from the region of tracking. The increase in average velocity with eccentricity was accompanied by a corresponding increase in the standard deviation, reflecting the fact that the variability of the velocity distribution increases with eccentricity. This result was consistent across all four sequences examined, although the degree to which tracking reduced the velocity distribution away from the point of tracking varied depending on the spatial correlation of the velocity field in that particular sequence. At the largest eccentricities, the velocity distribution after tracking can exceed the distribution before tracking. This can occur because the velocity distribution at large eccentricities is the vector sum of two uncorrelated velocity components, the image velocity and the velocity of eye movements.

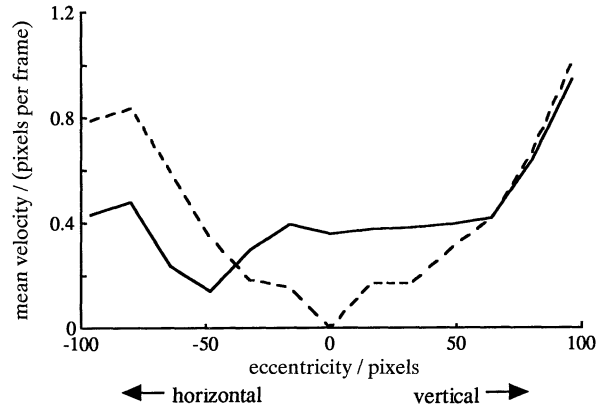


Figure 5. The average velocity in sequence IJ01300 before and after tracking (solid and dashed line respectively) as a function of eccentricity from the tracked point. The average was computed over the full 64 frames of the sequence. After tracking, the average velocity in the region of tracking drops to zero, and there is a regular increase in average velocity with eccentricity from the point of tracking. At large eccentricities, the average velocity after tracking can be larger than the average velocity before tracking.

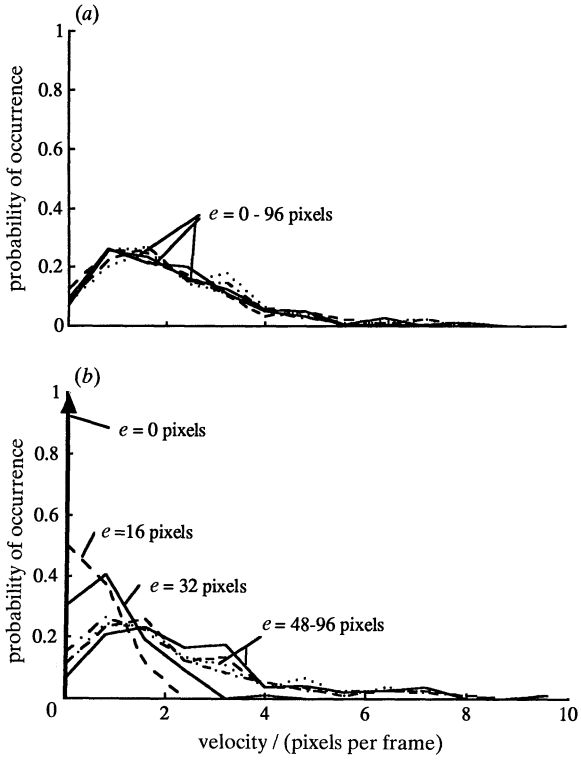


Figure 4. The velocity distribution of sequence IJ12426 as a function of eccentricity from the point of tracking. The distribution was computed from a 64 frame sequence. The curves represent the distribution at eccentricities of 0, 16, 32, 48, 64, 80, 96 pixels from the point of tracking. (*a*) Before tracking, the velocity distribution does not depend on eccentricity. The standard deviation for the curves is about 1.5 pixels per frame at all eccentricities. (*b*) After tracking, the distribution varies with eccentricity, from a delta function at the point of tracking, to a broad distribution at the largest eccentricity. The standard deviation increases with eccentricity and is 0, 0.51, 0.75, 1.6, 1.7, 1.71, 1.8 pixels per frame, respectively, for the eccentricities shown.

The changes in the velocity distribution as a result of tracking have corresponding effects in the spatiotemporal frequency spectrum (equation 7). In figure 6 we compare the spatiotemporal spectrum of images before and after tracking. The spectrum was computed in spatially and temporally localized blocks of size 32 pixels × 32 pixels × 16 frames at the point of tracking for a viewing distance of four screen heights (1 screen height = 256 pixels). For purposes of comparison, 1 pixel ≈ 1/15 deg and 1 pixel per frame ≈ 2 deg s^{-1} on a 256 pixel × 256 pixel image at a standard viewing distance of 4 screen heights with a frame rate of 30 frame per second. Most of the energy is concentrated below 10 Hz and 4 cycles per degree and diminishes quickly above these frequencies. After tracking, the temporal bandwidth of the spatiotemporal spectrum is greatly reduced. This is consistent with the changes in the velocity distribution after tracking (figure 4*b*). A reduction in the velocity distribution decreases the temporal bandwidth of the spatiotemporal spectrum.

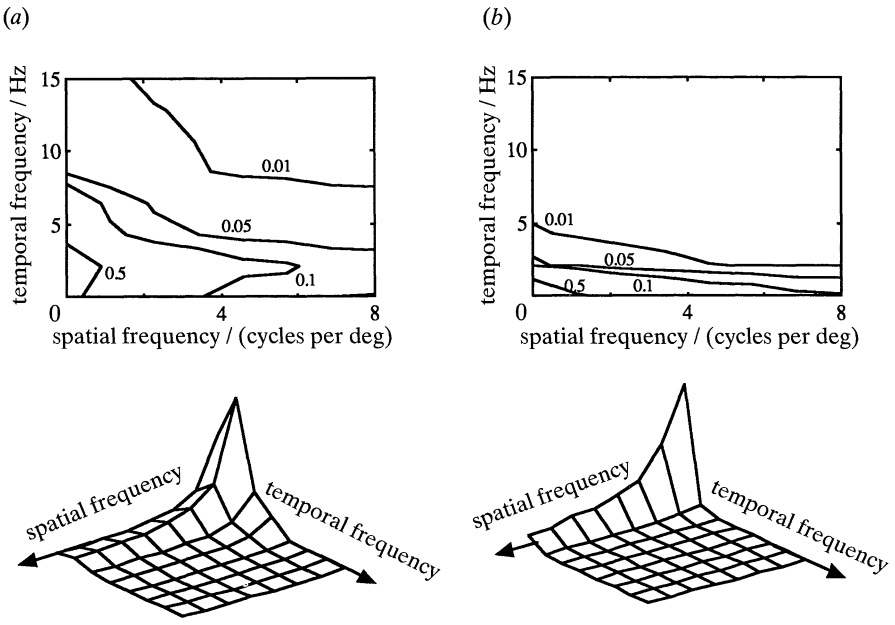


Figure 6. The spatiotemporal spectrum of the image before and after tracking computed for a $32 \times 32 \times 16$ spatiotemporal block from sequence IJ12426. The contour plots have lines at 0.01, 0.05, 0.1, 0.5 of the maximum values. The bottom figures are surface plots of the corresponding spectra. (*a*) Before tracking, the image has a large temporal bandwidth due to occasional large velocities. (*b*) After tracking, spatiotemporal energy is concentrated in lower temporal frequencies. The units of cycles per degree for spatial frequency axis were determined by using a viewing distance of 4 scene heights from the image. The range of spatial frequencies will vary for different viewing distances, but the range of temporal frequencies will not vary with viewing distance.

Figure 7 shows how tracking modifies the instantaneous temporal frequency bandwidth in the region of tracking. The instantaneous temporal frequency bandwidth was computed from the frame to frame correlation, ρ_τ, using a correlation model of the form $\rho_\tau = e^{-\alpha|\tau|}$ (Jayant & Noll 1984; Eckert *et al.* 1992). For a frame rate of 30 frames per second, the temporal bandwidth can be computed as $\alpha = -30 \log(\rho_\tau)$. Before tracking, the signal has a highly variable temporal frequency bandwidth, due to variable object velocity. After tracking, the temporal bandwidth is small and less variable. The temporal bandwidth is not zero after tracking, however, since tracking does not remove all time variations, only those which result from translational motion at the tracked point. Figure 7 highlights the two ways tracking affects temporal frequency bandwidth: (i) tracking greatly reduces the temporal frequency bandwidth during high velocity motion, and (ii) before tracking, the temporal frequency bandwidth fluctuates widely, depending on the velocity, but after tracking, the temporal frequency bandwidth in the region of tracking has only small fluctuations.

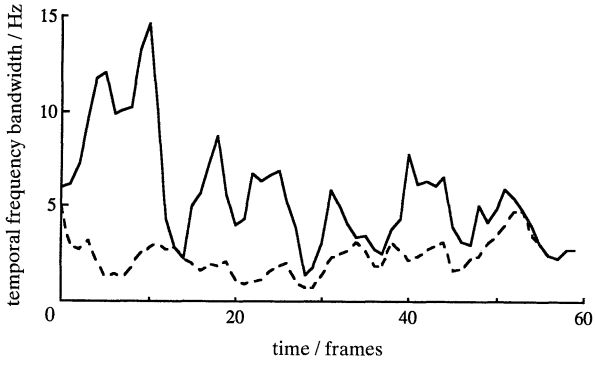


Figure 7. The effect of tracking on temporal frequency bandwidth for a 24 pixel $\times$ 24 pixel block from sequence IJ10833. The instantaneous temporal frequency bandwidth before tracking (solid line) and after tracking (dashed line) is computed from the frame to frame correlation. Before tracking, the signal has a highly variable temporal frequency bandwidth, due to variable object velocity. After tracking, the temporal bandwidth is small and does not vary significantly. Temporal frequency variations remain after tracking due to motion within the tracking region and the stationary component.

Tracking does not remove the velocity field component, but only shifts it to lower temporal frequencies. This shift will increase the energy share of the stationary component at high temporal frequencies. The relative share of the two components will vary from sequence to sequence and for different regions of tracking, depending on the amount of motion, and the degree to which time variations are represented by the velocity field component or by the stationary component. For the first three sequences in table 1, tracking of an object in the foreground reduces the average temporal bandwidth in the tracked region by 59%, 92%, and 56%, respectively. For these sequences, the large decrease of the temporal frequency bandwidth after tracking signifies that the velocity field component accounts for much of the signal energy at high temporal frequencies of the untracked image. The average temporal bandwidth of the last sequence (a

storm scene IJ04454) was reduced by only 22%. As tracking had little effect on the temporal bandwidth of this sequence, most time variations can be attributed to the stationary component. Large temporal intensity changes in this scene were due primarily to changing illuminant (lightning) and changing reflectance of light off ocean waves.

4. CODING BY THE VISUAL SYSTEM IN THE CONTEXT OF TRACKING

(*a*) *Advantages of tracking*

The velocity field of natural time varying images is signal dependent and variable. A scene may contain objects moving at high velocities, low velocities, or both. The corresponding spatiotemporal spectrum is also signal dependent and variable, with a large temporal bandwidth in the spatial regions which move at high velocities, and a small temporal bandwidth in slowly moving regions. A basic premise of coding theory is that a signal with a small bandwidth can be more efficiently coded than a signal with a large bandwidth (Jayant & Noll 1984). The coding efficiency is also affected by signal variability, because time-invariant coders (such as the retinal pathways) can be optimized only for a particular spectrum (Kassam & Poor 1983, 1985). The most efficiently coded signal is one with a small bandwidth and little or no variability. This corresponds to an image with little or no velocity, and thus a small temporal frequency bandwidth. Tracking with eye movements compensates for motion by matching eye velocity to the expected value of the image velocity in a region around the fovea. After tracking, the signal which actually reaches the retina (at the fovea) has a narrow velocity distribution and, therefore, a reduced temporal frequency bandwidth. A direct corollary of minimizing the temporal frequency bandwidth is a reduction in blur due to motion when the image is coded by fixed bandwidth time invariant channels. The role of eye movements in reducing blur was suggested before (Miller & Ludvigh 1962; Murphy 1978; Flipse *et al.* 1988) and follows from their role in the context of efficient coding.

Field (1987) showed that the spatial spectrum of natural images is scale invariant. This enables the visual system to use fixed, scene invariant, spatial filters to efficiently code a scene regardless of scale. The temporal spectrum and velocity distribution of natural time varying images are not scale invariant, and depend on the distance of moving objects from the observer. However, tracking maps the tracked region into the same temporal frequency and velocity distribution range regardless of the velocity (or scale) of the tracked region. Therefore, tracking provides a region of the retina with a virtually scale invariant signal in time and the coding advantages that accrue from the invariancy.

In addition to increased coding efficiency, tracking accentuates the importance of the stationary component in the temporal frequency domain in the region of tracking. Before tracking, this component is difficult to detect and isolate because it cannot easily be separated from the velocity field component. After tracking, the velocity field component is situated along the spatial frequency axis, and does not contribute to temporal variations, so the remaining temporal variations belong to the stationary term. Thus, perceptually important information associated with this component, such as flicker, photometric effects of motion, and motion edge effects, can be extracted more easily because of the removal of the velocity field component from high temporal frequencies. This argument does not hold in the periphery, however, because tracking only ensures reduction of the velocity field component in the region of tracking.

(*b*) *Retinal pathways and eccentricity dependent architecture are matched to the tracked image*

The second stage of the coder (figure 1) are the M and P pathways which operate on the tracked image, $I_r(\boldsymbol{u},t)$. These pathways and the underlying single cell units from which they are made have received considerable attention. Because of their significance in the present context, they are briefly reviewed here. The spatiotemporal filter properties of the M and P pathways are based on single cell properties of phasic and tonic cells from the retina and M and P cells from the LGN (Marrocco *et al.* 1982; Kaplan & Shapley 1982; Hicks *et al.* 1982; Derrington & Lennie 1984; Blakemore & Vital–Durand 1986; Crook *et al.* 1988; Lee *et al.* 1989b; Purpura *et al.* 1990). P (tonic) cells respond well to low temporal frequencies (below 5 Hz), whereas M (phasic) cells attenuate these frequencies. The spatial resolution of the P pathway is about three times higher than the M pathway at all eccentricities. This is due to receptive field center size and spatial sampling rates of the respective arrays (Merigan 1989; Merigan *et al.* 1991). The main characteristics of the pathways are summarized in table 2. The numbers in table 2 represent averages within the respective pathways rather than the response of any particular cell since there are large deviations among cells even in the same pathway (Hicks *et al.* 1982; Marrocco *et al.* 1982; Derrington & Lennie 1984).

Figure 8 illustrates the spatiotemporal transfer function of the M and P pathways inferred from the specifications in table 2. To obtain these responses, we fitted a frequency transfer function with the form of a spatial and temporal difference of Gaussians (Rohaly & Buchsbaum 1988; Rohaly 1988) and selected constants so as to meet the spatial and temporal frequency slopes and peaks in table 2.

$$\mathrm{RF}(|\boldsymbol{k}|,f,e) = [C_1 \mathrm{e}^{-(\pi r_c(e)|\boldsymbol{k}|)^2} - S_1 \mathrm{e}^{-(\pi r_s(e)|\boldsymbol{k}|)^2}][C_2 \mathrm{e}^{-(T_c f)^2} - S_2 \mathrm{e}^{-(T_s f)^2}], \quad (8)$$

where e is eccentricity, $r_c(e)$, $r_s(e)$ are the centre and surround sizes for receptive fields, and $\boldsymbol{k}$ and f are the spatial and temporal frequencies, respectively. T_c, T_s are temporal constants selected so as to provide peak temporal response at a specified frequency, and C_1, C_2, S_1, S_2 were selected so as to provide a specified response at low temporal and spatial frequencies. The

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 2. *Spatial and temporal characteristics of the M and P pathways*

	M pathway	P pathway
spatial structure	centre-surround (relatively powerful surround)	centre surround (surrounds often have little power)
spatial resolution	one-third that of P cells (decreases with eccentricity)	three times that of M cells (decreases with eccentricity)
foveal spatial resolution	13 cycles per degree	40 cycles per degree
numbers of cells	10% of cells	80% of cells
spatial sampling rate	one-third that of P cells	three times that of M cells (80 samples per degree at fovea)
temporal frequency-peak	20 Hz: large variance between individual cells	10 Hz: large variance between individual cells
response at low frequencies	highly attenuated: phasic response to a step increase in light intensity	partially attenuated: tonic or sustained response to a step increase in light intensity
high frequency cutoff	up to 80 Hz	20–40 Hz
contrast sensitivity	high: eight times higher than P cells	low: eight times lower than M cells
speed (latency of response to visual stimulation)	fast: latency is about 24 ms at LGN for visual stimulation. Large variance	slow: latency is about 28 ms at LGN for visual stimulation. Large variance
suggested roles	carries information about quickly moving images with a low degree of spatial detail, such as flicker	carries slowly moving images with a high degree of spatial detail

centre and surround sizes for the receptive fields, $r_c(e)$ and $r_s(e)$, are assumed to increase in size with the inverse of the cortical magnification factor (Sakitt & Barlow 1982).

$$r_c(e) = r_c(0)(1+0.33e),\ r_s(e) = r_s(0)(1+0.33e). \qquad (9)$$

The spatiotemporal frequency the velocity responses of M and P pathways can now be discussed in the context of the properties of images before and after tracking. Before tracking, the image spatiotemporal

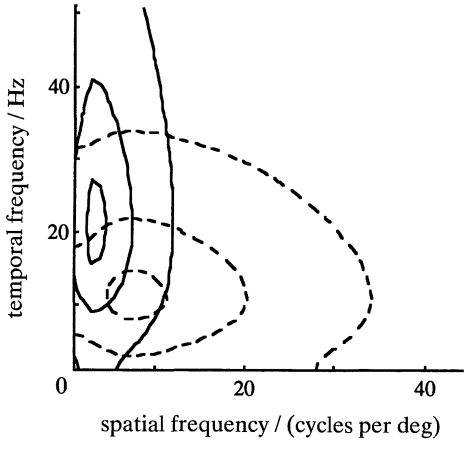


Figure 8. The spatiotemporal transfer function of the M and P pathways calculated from equation (8) with constants chosen to match details in table 2 (solid lines, M pathway; dashed lines, P pathway). Contours are at 0.1, 0.5, and 0.9 of the maximum value in the respective pathway. The P pathway is tuned to higher spatial frequencies and lower temporal frequencies than the M pathway, though there is considerable overlap.

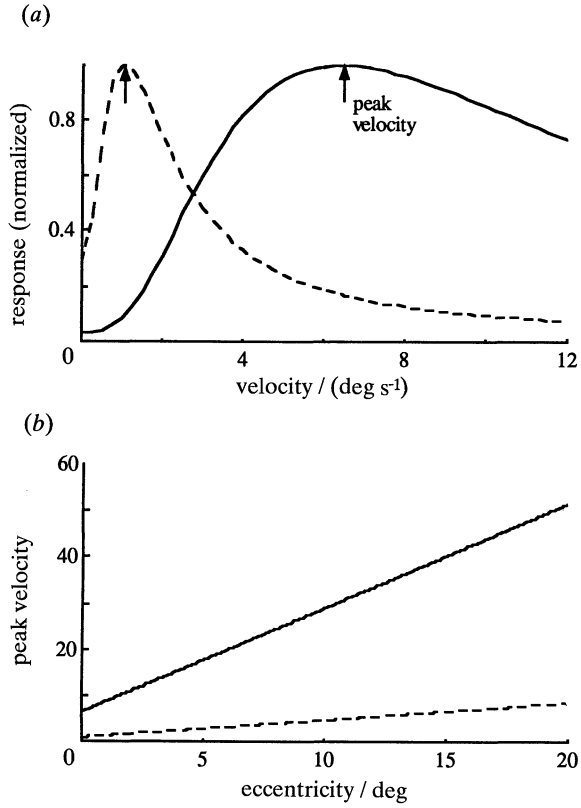


Figure 9. (*a*) The simulated response of M and P pathways at the fovea to a translating white noise stimulus as a function of velocity (solid line, M pathway; dashed line, P pathway). This is equivalent to integrating the frequency response over lines of constant velocity. The normalized peak response of the M and P pathways (arrows) is about 7 deg s^{-1} and 1 deg s^{-1}, respectively. (*b*) Peak velocity of the two pathways as a function of eccentricity. The peak velocity of the M pathway increases with eccentricity at a greater rate than the P pathway.

spectrum is broadly distributed across frequency space, which results in large responses in both the M and P pathways. However, after tracking, the signal energy becomes concentrated at low temporal frequencies, and the P pathway which is more sensitive in the low temporal frequency region, will respond to a greater degree than the M pathway. Figure 9*a* illustrates the response of the M and P pathways to a translating image, found by integrating the spatiotemporal transfer function (equation 8) over lines of constant velocity in frequency space. Figure 9*b*, computed using equations 8 and 9, illustrates the velocity of peak response as a function of eccentricity for the two pathways. At the fovea, the M and P pathways are predicted to have a peak response to images moving at velocities of 7 deg s^{-1} and 1 deg s^{-1}, respectively. However, the velocity of peak response increases with eccentricity, with the peak velocity increasing faster for the M than the P pathway. This change in peak velocity with eccentricity is a result of the increase in receptive field size (or decrease in spatial scale) described by equation 9.

The P pathway is thought to carry information about slowly moving images with a high degree of spatial detail (Schiller *et al.* 1990; Merigan 1989; Merigan *et al.* 1991). Figure 9*a* illustrates that the P pathway will respond better to low velocity images. The P pathway matches the properties of the velocity field component in the tracked region, and will carry the maximum amount of spatiotemporal information about this component. The M pathway is thought to carry information about quickly moving images with a low degree of spatial detail (Schiller *et al.* 1990; Merigan 1989; Merigan & Maunsell 1990). Figure 9*a* illustrates that the M pathway is tuned to higher velocities than the P pathway. However, large image velocities will only rarely arise in the tracked region. When large velocities do arise, it is during tracking errors which occur for unpredictable motion, and for cases such as transparent motion when there are two velocity field components in the same spatial region. Because tracking is generally quite accurate for motion of 'real world' stimuli (Steinman *et al.* 1990), the M pathway can be expected to carry only a small fraction of the velocity field component of image information in the region of tracking (the fovea). However, the stationary component is broadly distributed across spatiotemporal frequency space (figure 2*b*), and contains a significant amount of energy in the region covered by the M pathway. Therefore, in addition to carrying (infrequent) high velocity images, another role for the M pathway at the fovea could be to carry the stationary component of time varying images.

The change of the velocity tuning of the M and P pathways with eccentricity (figure 9*b*) is consistent with the change in the velocity distribution of tracked images with eccentricity. The peak velocity of the two pathways increase with eccentricity, though at different rates, so a larger average velocity and larger range of velocities is covered with increasing eccentricity. This can be compared with the velocity distribution after tracking (figure 4*b*). At the fixation point (fovea), the image has a narrowly distributed velocity distribution and a small temporal bandwidth. With increasing eccentricity, the velocity distribution broadens (figure 4*b*), the average velocity reaching the retina increases (figure 5), and the range of velocities increases. The broader image velocity distribution in the periphery means that information is lost due to temporal blur because of the limited temporal frequency bandwidth of retinal pathways. This decreases the average spatial frequency limit of the peripheral retinal image. Because of this, larger receptive fields can be utilized in the periphery without significant loss of information.

Psychophysical evidence also shows a gradual change in motion perception between the fovea and the periphery of the visual field. The fovea is sensitive to a lower range of velocities than the periphery and essentially becomes blind when this velocity range is exceeded (van de Grind *et al.* 1986; Baker & Braddick 1985). As eccentricity increases, the visual system is better able to discriminate images with a higher average velocity, and over a larger range of velocities (McKee & Nakayama 1984). This is consistent with the change in velocity distribution and average velocity with eccentricity (see figures 4 and 5) which results from tracking.

Hughes (1977) argues that receptor packing matches the change in velocity across the retina for the case of an observer moving through a scene (ego-motion). In some ways, this paper can be viewed as a generalization of Hughes (1977) original arguments, by showing that an eccentricity dependent velocity distribution results for any scene rather than the special case of ego-motion, as long as the observer continually tracks with eye movements. This paper diverges from Hughes by matching the retinal velocity distribution to the velocity sensitivity of the M and P pathways, rather than to the change in receptor packing. However, receptive field size and velocity sensitivity are linked so both arguments are complementary.

5. CONCLUSION

We examined the spatiotemporal spectrum and other attributes of natural time varying images in the context of efficient coding in the early visual system. The image is modeled as a combination of a velocity field component and a stationary component which have markedly different spatiotemporal spectra. Tracking, as implemented with smooth pursuit eye movements, decreases the average velocity and the variability of velocities reaching the fovea (tracked region). The result is a spectrum with minimal temporal bandwidth and variability in the tracked region, but which broadens with increasing eccentricity. Tracking does not affect the stationary component, which remains broadly distributed across temporal frequency space.

An efficient coding strategy will be influenced by tracking because it changes the image spectrum. In the tracked region, the spectrum has minimal temporal bandwidth and variability. This enables efficient coding of the image with fixed time invariant path-

ways as found in the retina. The reduction in temporal bandwidth ensures that minimal information will be lost due to motion blur in the tracked region. The stationary component of time varying images is emphasized in the tracked region, enabling temporal information not attributed to translational motion to be analyzed effectively. Finally, since the average velocity of the image increases with eccentricity from the tracked region, an efficient coding strategy should reflect this change with a corresponding change in velocity tuning with eccentricity.

The results suggest that the M and P pathways are matched to the tracked image. Both the M and P pathways are tuned to low image velocities at the fovea, where the image has consistently low velocities because of tracking. However, the M pathway, with the broader temporal frequency response, will respond better to the temporal changes of the stationary component. The M and P pathways are tuned to higher velocities and a broader range of velocities with increasing eccentricity from the fovea. This is matched to the change of image velocity after tracking, in which both the average velocity and range of velocities increase.

In conclusion, the visual system combines smooth pursuit tracking with specialized pathways and an eccentricity dependent retinal architecture to efficiently code time varying images.

We thank Horace Barlow and Peter Sterling for their many comments and suggestions, and Andrew B. Watson for his aid in collecting the image sequences. This work was supported by AFOSR grant 91–0082 and the NASA graduate student fellowship program.

REFERENCES

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

Baker, C.L. & Braddick, O.J. 1985 Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Res.* **25**, 803–812.

Barlow, H.B. 1961 Possible principles underlying the transformation of sensory messages. In *Sensory communication* (ed. W. A. Rosenblith), pp. 217–234. MIT Press.

Barlow, H.B. 1981 The Ferrier Lecture: critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. Lond.* B**212**, 1–34.

Barnes, G.R. & Lawson, J.F. 1989 Head-free pursuit in the human of a visual target moving in a pseudo-random manner. *J. Physiol., Lond.* **410**, 137–155.

Blakemore, C. & Vital-Durand, F. 1986 Organization and post-natal development of the monkey's lateral geniculate nucleus. *J. Physiol., Lond.* **380**, 453–491.

Buchsbaum, G. & Gottschalk, A. 1983 Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond.* B**220**, 89–113.

Crook, J.M., Lange-Malecki, B., Lee, B.B. & Valberg, A. 1988 Visual resolution of macaque retinal ganglion cells. *J. Physiol., Lond.* **396**, 205–224.

Derrico, J.B. & Buchsbaum G. 1991 A computational model of spatiochromatic image coding in early vision. *J. visual commun. Image Repres.* **2**, 31–38.

Derrington, A.M. & Lennie, P. 1984 Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J. Physiol., Lond.* **357**, 219–240.

Eckert, M.P., Buchsbaum, G. & Watson, A.B. 1992 The separability of spatiotemporal spectra of image sequences. *IEEE Trans. Pattern Anal. Machine Intell.* **PAMI-14**, 1210–1213.

Field, D. 1987 Relations between the statistics of natural images and the response properties of cortical cells. *J. opt. Soc. Am.* A**4**, 2379–2394.

Flipse, J.P., Wildt, G.J. v.d., Rodenburg, M., Keemink, C.J. & Knol, P.G.M. 1988 Contrast sensitivity for oscillating sine wave gratings during ocular fixation and pursuit. *Vision Res.* **28**, 819–826.

Girod, B. 1987 The efficiency of motion-compensating prediction for hybrid coding of visual sequences. *IEEE J. Selected Areas Commun.* **5**, 1140–1154.

van de Grind, W.A., Koenderink, J.J. & van Doorn, A.J. 1986 The distribution of human motion detector properties in the monocular visual field. *Vision Res.* **26**, 797–810.

Heeger, D.J. 1987 Model for the extraction of image flow. *J. opt. Soc. Am.* A**4**, 1455–1471.

Hicks, T.P., Lee, B.B. & Vidyasagar, T.R. 1982 The responses of cells in macaque lateral geniculate nucleus to sinusoidal gratings. *J. Physiol., Lond.* **337**, 183–200.

Hughes, A. 1977 The topography of vision in mammals of contrasting life style: comparative optics and retinal organization. In *Handbook of sensory physiology. The visual system in vertebrates* (ed. F. Crescitelli), pp. 613–756. Berlin: Springer-Verlag.

Jain, J.R. & Jain, A.K. 1981 Displacement measurement and its application in interframe image coding. *IEEE Trans. Commun.* **COM-29**, 1799–1808.

Jayant, N.S. & Noll, P. 1984 *Digital coding of waveforms: principles and applications to speech and video.* Englewood Cliffs, New Jersey: Prentice-Hall.

Kaplan, E. & Shapley, R. 1982 X and Y cells in the lateral geniculate nucleus of macaque monkeys. *J. Physiol., Lond.* **330**, 125–143.

Kassam, S.A. & Poor, H.V. 1983 Robust signal processing for communication systems. *IEEE Commun. Mag.* 20–28.

Kassam, S.A. & Poor, H.V. 1985 Robust techniques for signal processing. *Proc. IEEE* **73**, 433–481.

Kelly, D.H. 1979 Motion and vision II: stabilized spatio-temporal threshold surface. *J. opt. Soc. Am.* A**69**, 1340–1349.

Laughlin, S. 1983 Matching coding to scenes to enhance efficiency. In *Physical and biological processing of images* (ed. O. J. Braddick & A. Sleigh), pp. 42–52. Berlin: Springer-Verlag.

Lee, B.B., Martin, P.R. & Valberg, A. 1989*a* Amplitude and phase of responses of macaque retinal ganglion cells to flickering stimuli. *J. Physiol., Lond.* **414**, 245–263.

Lee, B.B., Martin, P.R. & Valberg, A. 1989*b* Sensitivity of macaque retinal ganglion cells to chromatic and luminance flicker. *J. Physiol., Lond.* **414**, 223–243.

Lisberger, S.G., Morris, E.J. & Tychsen, L. 1987 Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *A. Rev. Neurosci.* **10**, 97–129.

Marrocco, R.T., McClurkin, J.W. & Young, R.A. 1982 Spatial summation and conduction latency classification of cells of the lateral geniculate nucleus of macaques. *J. Neurosci.* **2**, 1275–1291.

McKee, S.P. & Nakayama, K. 1984 The detection of motion in the peripheral visual field. *Vision Res.* **24**, 25–32.

Merigan, W.G. 1986 Spatio-temporal vision of macaques with severe loss of P retinal ganglion cells. *Vision Res.* **26**, 1751–1761.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Merigan, W.H. 1989 Assessing the role of parallel pathways in primates. In *Seeing contours and colour* (ed. J. J. Kulikowski, C. M. Dickinson & I. J. Murray). Pergamon Press.

Merigan, W.H. & Maunsell, J.H.R. 1990 Macaque vision after magnocellular lateral geniculate lesions. *Vis. Neurosci.* **5**, 347–352.

Merigan, W.H., Katz, L.M. & Maunsell, J.H.R. 1991 The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys. *J. Neurosci.* **11**, 994–1001.

Miller, J.W. & Ludvigh, E. 1962 The effect of relative motion on visual acuity. *Surv. Opthalm.* **7**, 83–116.

Murphy, B.J. 1978 Pattern thresholds for moving and stationary gratings during smooth eye movement. *Vision Res.* **18**, 521–530.

Pentland, A. 1991 Photometric motion. *IEEE Pattern Anal. Machine Intell.* **13**, 879–890.

Purpura, K., Tranchina, D., Kaplan, E. & Shapley, R.M. 1990 Light adaptation in the primate retina: analysis of changes in gain and dynamics of monkey retinal ganglion cells. *Vis. Neurosci.* **4**, 75–93.

Rohaly, A.M. & Buchsbaum, G. 1988 Inference of global spatiochromatic mechanisms from contrast sensitivity functions. *J. opt. Soc. Am.* A**5**, 572–576.

Rohaly, A.M. 1988 A global multidimensional model of human visual contrast sensitivity. Ph.D. thesis, Department of Bioengineering, University of Pennsylvania.

van Santen, J.P.H. & Sperling, G. 1985 Elaborated Reichardt detectors. *J. opt. Soc. Am. A* **2**, 300–321.

Sakitt, B. & Barlow, H.B. 1982 A model for the economical coding of the visual images in cerebral cortex. *Biol. Cybern.* **43**, 97–108

Schiller, P.H., Logothetis, N.K. & Charles, E.R. 1990 role of the color-opponent and broad-band channels in vision. *Vis. Neurosci.* **5**, 321–346.

Shapley, R.M. & Perry, V.H 1986 Cat and monkey retinal ganglion cells and their visual functional roles. *Trends Neurosci.* 229–235.

Snyder, A.W., Laughlin, S.B. & Stavenga, D.G. 1977 Information capacity of eyes. *Vision Res.* **17**, 1163–1175.

Srinivisan, M.V., Laughlin, S.B. & Dubs, A. 1982 Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond.* B**216**, 427–459.

Stark, L., Vossius, G. & Young, L.R. 1962 Predictive control of eye tracking movements. *IRE Trans. Human Factors Electr.* 52–57.

Steinman, R.M., Kowler, E. & Collewijn, H. 1990 New directions for oculomotor research. *Vision Res.* **30**, 1845–1864.

Tsukamoto, Y., Smith, R.G. & Sterling, P. 1990 'Collective coding' of correlated cone signals in the retinal ganglion cells. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1860–1864.

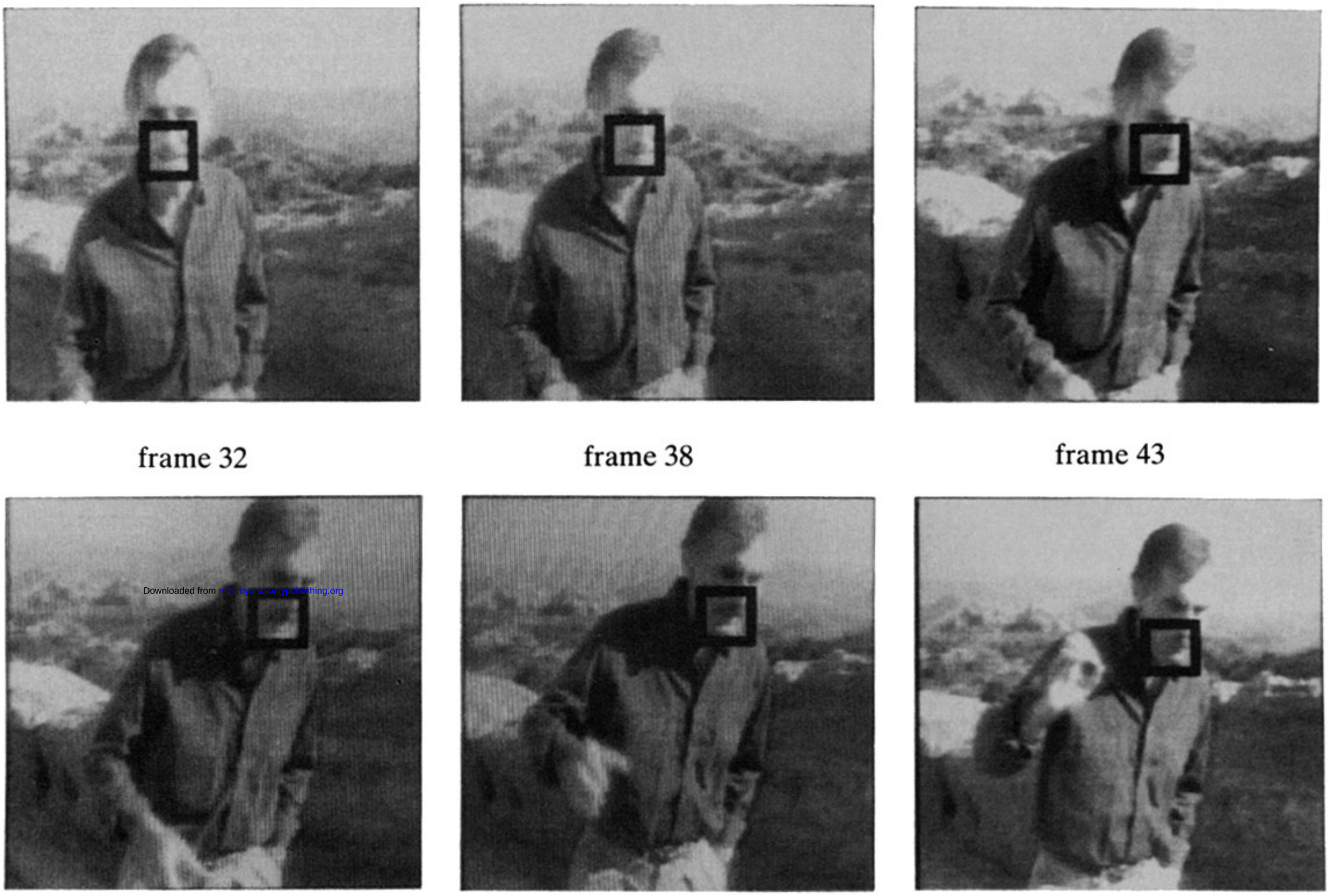
Watson, A.B. 1983 A look at motion in the frequency domain. NASA Tech. Report Tech. Memo. 84352.

Watson, A.B. & Ahumada, A.J. 1985 Model of human visual motion sensing. *J. opt. Soc. Am.* A**2**, 322–341.

Watson, A.B. 1990 Perceptual-components architecture for digital video. *J. opt. Soc. Am.* A**7**, 1943–1954.

Received 4 March 1992; accepted 19 October 1992

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



gure 3. A typical image sequence over which the spatiotemporal statistics were analysed (sequence IJ12426). The ack box indicates a region which was selected for tracking. The graph illustrates the magnitude of the velocity of e tracked region as a function of time.