Network: Computation in Neural Systems
June/September 2005; 16(2/3): 175–190

Taylor & Francis
Taylor & Francis Group

Is the early visual system optimised to be energy efficient?

BENJAMIN T. VINCENT, ROLAND J. BADDELEY, TOM TROSCIANKO, & IAIN D. GILCHRIST

Department of Experimental Psychology, University of Bristol, Bristol, UK

(Received 2 December 2004; revised 13 April 2005; accepted 10 June 2005)

Abstract
This paper demonstrates that a representation which balances natural image encoding with metabolic energy efficiency shows many similarities to the neural organisation observed in the early visual system. A simple linear model was constructed that learned receptive fields by optimally balancing information coding with metabolic expense for an entire visual field in a 2-stage visual system. The input to the model consists of a space variant retinal array of photoreceptors. Natural images were then encoded through a bottleneck such as the retinal ganglion cells that form the optic nerve. The natural images represented by the activity of retinal ganglion cells were then encoded by many more 'cortical' cells in a divergent representation. Qualitatively, the system learnt by optimising information coding and energy expenditure and matched (1) the centre surround organisation of retinal ganglion cells; (2) the Gabor-like organisation of cortical simple cells; (3) higher densities of receptive fields in the fovea decreasing in the periphery; (4) smaller receptive fields in the fovea increasing in size in the periphery; (5) spacing ratios of retinal cells; and (6) aspect ratios of cortical receptive fields. Quantitatively, however, there are small but significant discrepancies between density slopes which may be accounted for by taking optic blur and fixation induced image statistics into account. In addition, the model cortical receptive fields are more broadly tuned than biological cortical neurons; this may be accounted for by the computational limitation of modelling a relatively low number of neurons. This paper shows that retinal receptive field properties can be understood in terms of balancing coding with *synaptic* energy expenditure and cortical receptive fields with *firing rate* energy expenditure, and provides a sound biological explanation of why 'sparse' distributions are beneficial.

Keywords: *Retina, cortex, energy efficiency, coding, optimal*

Introduction

After many decades of experimental work, the visual system is now very well characterised, giving us a good idea about *what* happens in the early stages of visual processing. However, at times it is unclear as to *why* certain features of the visual system are as they are, and so theoretical work has attempted to address this issue. One way to do this is to assume that aspects of the visual system are optimised in some way; predictions are then made based on this optimisation and are compared to physiological data.

One particularly successful approach at explaining the receptive field organisation of neurons in the primary visual cortex has been sparse coding (Olshausen & Field 1996). In this approach, it is hypothesised that these neurons are optimsed to encode natural images whilst

Correspondence: B. T. Vincent, Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol, BS8 1TN, UK. Tel: 0117 9546942. Fax: 0117 9288588. E-mail: ben.vincent@bris.ac.uk

ISSN: 0954-898X print / ISSN 1361-6536 online
DOI: 10.1080/09548980500290047

having a zero-peaked distribution (one type of sparsity) of firing rates whereby the majority of neurons are inactive, but a few have high firing rates. This type of distribution can also be considered as energy efficient in terms of minimising the average firing rate of neurons (Baddeley 1996). In fact, many monotonically increasing cost functions tend to promote zero-peaked distributions.

This concept of metabolic energy efficiency was extended in our previous work, which considered the role of synaptic energy usage in encoding natural images (Vincent & Baddeley 2003). In the same way that monotonically increasing cost functions of firing rates in Olshausen & Field (1996) leads to zero-peaked firing rate distributions, the linear synaptic cost function leads to a zero peaked distribution of synaptic connection strengths. Encoding natural images under a metabolically based synaptic energy budget showed: (1) substantial metabolic saving can be made whilst maintaining information encoding capacity; (2) the optimal receptive fields had close quantitative similarity to the classic centre-surround type as observed in retinal ganglion cells; (3) the change from bandpass to lowpass characteristics in different light levels can be explained; (4) that details of the mosaic of receptive fields can be explained. These results, combined with the metabolic interpretation of sparse coding, suggest that energy efficient coding could be an important explanation of why many features of the visual system are as they are.

In this paper, the energy efficient coding hypothesis is extended in two important ways. Firstly, instead of using small patches of natural images, we use an entire (albeit scaled down) visual field. This visual field is sampled by a space variant retina like distribution of photoreceptors, with a sampling density that is high in the fovea and decreases with retinal eccentricity. This allows the expansion of the energy efficient explanation of local receptive field properties (Vincent & Baddeley 2003) to more global properties of the visual field. Examples of the global properties include how the optimal receptive field size changes across retinal eccentricity and how the optimal sampling densities change across the visual field. Secondly, energy efficient coding is extended from retinal image coding to investigate how cortical receptive fields that are energy efficient in terms of firing rates can be constructed from synaptically efficient centre-surround receptive fields.

In order to test whether the energy efficient coding hypothesis can account for local and global properties of retinal and cortical receptive fields, a simple linear model is constructed (see Figure 1). This model allows the calculation of receptive fields that are optimised to encode natural image information under a metabolic cost scheme where the cost of synaptic activity and firing rates can be fully manipulated. Here, we do not fully explore different combinations of costs in the retinal and cortical stages but this is the focus of future work.

In the retinal stage, a cost that increases in proportion to synaptic activity is imposed and from the results of previous work (Vincent & Baddeley 2003), we predict that the optimal organisation will be centre surround.

In the cortical stage, a cost proportional to the firing rate of neurons is imposed and from sparse coding, and our own previous unpublished investigations, we predict that a Gabor-like organisation will be optimal. The intuitive explanation of why a zero-peaked distribution of firing rates leads to oriented bar detectors is more established. The natural image input can be seen as a mixture of some ‘fundamental causes’ of that image and the oriented bars are simply an estimate of what these causes are given the assumption that any given image is caused by only few of all possible causes (Olshausen & Field 1996; Bell & Sejnowski 1997).

In addition to testing these predictions of the optimal spatial structure the model will allow us to test if (1) the same spatial organisations are optimal with a space variant input array; (2) how the receptive fields are distributed across the visual field; and (3) if cortical receptive fields can be learnt from a set of centre surround ‘building blocks’.

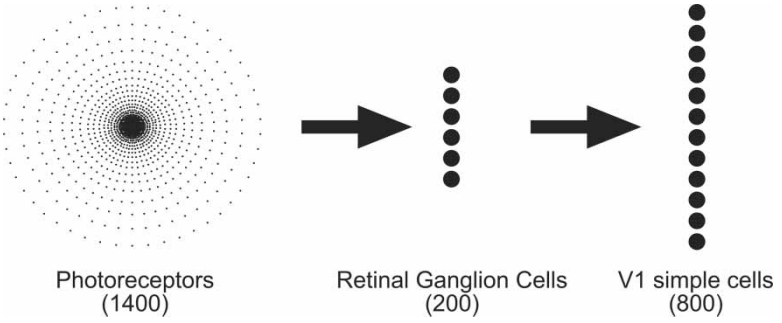

Figure 1. The basic structure of the model. The input is a space variant array of 1400 photoreceptors which sample natural images. The first stage, representation of the visual input which converges to 200 retinal ganglion cells. The second stage is a divergent representation of 800 cortical neurons which goes some way in reflecting the number of cells in the cortex but is only 800 due to computational limitations. All neurons are fully interconnected to all units in the previous layer, and it is the pattern of connections from a given unit to those in the previous layer which defines that units receptive field.

Methods

Space variant visual field

The prevailing method of presenting images in image coding approaches is to use a space invariant, regular grid of pixels. In this paper we take the space variant approach in order to simulate a basic retina (see Bolduc & Levine (1998)). A log-polar model is used which calculates a set of positions of photoreceptors, (see Figure 1 (left)). A uniformly spaced set of 50 radial angles are determined and a set of distances are calculated where $\mathbf{d} = 160e^{\mathbf{n}}$, where $\mathbf{n}$ is a vector of values increasing from -30 to 0 in steps of 0.15. This defines a set of photoreceptor locations in log-polar coordinates which are then converted from polar co-ordinates to cartesian and then rounded such that photoreceptor locations correspond to a specific pixel location. After this rounding, duplicate sets of points with identical distances were removed. The final set of photoreceptors (with 50 radial values and 28 distance values) can then sample from a standard image with a randomly chosen fixation position within the image, yielding the space variant retinal representation. This procedure results in a simple model of the photoreceptor distribution but resembles physiology in that the density decreases sharply with distance from the fovea (see Figure 7). We assume photoreceptor size is constant (1 pixel) across the whole retina for simplicity. This has implications for the very centre of the fovea as only a single photoreceptor can occupy a single pixel location. This results in a small (7 pixel radius) region of maximal (uniform) photoreceptor density in the very centre of the fovea. Outside of this region, however, the photoreceptor spacing decreases with eccentricity. This assumed uniform size in photoreceptors is not too gross an assumption as physiological measurements of the New World monkey show that the cone inner segment diameter is approximately constant across the retina but with a slight decrease in size (factor of 2) in the central 5° (da Costa & Hokoç 2000).

Natural image stimuli

The initial images, from which the space variant images are derived, are based upon a dataset of natural spatio-chromatic images taken in a Ugandan rainforest (Troscianko et al. 2003). This dataset of 244 images is calibrated so that the three channels correspond to L, M and

S cone types (Párraga et al. 2002), then luminance was calculated as L+M. From this, 204 samples are taken from each image from random positions, subject to the photoreceptor array being inside the image boundary, resulting in a dataset of 49,776 space variant natural image samples. The photoreceptor array spans a diameter of 320 pixels and the source natural images had a resolution of 1600 by 1200 pixels. The mean and variance of each photoreceptor (over the entire dataset) was normalized to minimise any sampling artifacts, for example the top half of the images were brighter on average than the bottom half due to presence of the sky.

The basic model

In order to calculate the optimal filters to encode natural images under a synaptic energy constraint, a standard linear model is used (Baldi & Hornik 1995). This model consists of a set of inputs $\mathbf{x}$ that are fully connected by synapses $\mathbf{W}$ to a set of output units $\mathbf{y}$. The images are presented to the inputs and propagate through the weights (receptive fields) to result in a set of output unit activations. The output activity (analogous to firing rates) essentially encodes the inputs. Then the network is 'tested' to ensure that it has accurately encoded the inputs. This is done by predicting the input from the set of output activations and measuring the difference between the input and the predicted input. This results in an error signal which is then used to update the receptive fields by a learning rule so as to more accurately encode the inputs.

This linear model is used once for the first (retinal) stage and a second time for the second (cortical) stage. The retinal stage receives luminance input from the photoreceptor array and passes its output, as firing rates, onto the cortical stage.

Enforcing synaptic and firing rate efficiency

We define a cost function such that optimisation of the model results in receptive fields that optimally encode inputs subject to the costs of synaptic activity and firing rates. We start by defining the cost function of the model

$$E = \left\langle \frac{1}{2}\mathbf{e}^2 \right\rangle + \alpha\Omega_{\text{synapse}} + \beta\Omega_{\text{rate}} \tag{1}$$

where $\langle \frac{1}{2}\mathbf{e}^2 \rangle$ is the encoding error over the whole input ensemble, $\mathbf{e} = \mathbf{x} - \mathbf{W}'\mathbf{y}$, Ω_{synapse} is the cost of synaptic activity and Ω_{rate} is the cost of firing rates. The relative importance of the two costs are set with α and β, this way investigation of optimal coding under a range of different conditions is possible. If $\alpha = 0$ and $\beta > 0$, then firing rate efficiency is important and oriented bar type filters are predicted to emerge. Conversely, if $\alpha > 0$ and $\beta = 0$, synaptic efficiency is important and centre surround type filters are predicted to emerge.

The synaptic energy efficiency is dealt with slightly differently than in Vincent and Baddeley (2003). Previously a 'hard' constraint was imposed where a given synaptic energy budget was defined and the model could not exceed it. As this hard constraint was applied to individual receptive fields, it promoted similarity amongst them in terms of size for example. Here a 'soft' weight constraint is used where the importance of synaptic efficiency (α) is defined relative to the other costs (encoding information and firing rates). This soft constraint allows synaptic resources to be placed anywhere amongst the receptive fields and so it allows them to be of different sizes.

To implement these soft constraints with linear cost functions, we set the synapse and firing rate costs as follows:

$$\Omega_{\text{synapse}} = \Sigma|\mathbf{W}| \tag{2}$$

$$\Omega_{\text{rate}} = \Sigma|\mathbf{y}| \tag{3}$$

where $|\bullet|$ indicates the absolute value.

The model is optimised to find the receptive fields $\mathbf{W}$ such that it minimises the error function E. The optimal set of outputs $\mathbf{y}$ are chosen to minimise E using gradient descent then the receptive fields are updated using gradient descent. This procedure of calculating output activations and updating weights is iterated with a decreasing learning rate until E is minimised which is assumed to happen when the RMS change in $\mathbf{W}$ over 500 iterations decreases below a threshold of 0.001.

In the first stage, synapses incur a cost but firing rates are metabolically 'free' and this is implemented by setting $\alpha = 0.1$ and $\beta = 0$. In the second stage, this is swapped such that firing rates incur a cost and synapses are 'free' by setting $\alpha = 0$ and $\beta = 0.1$.

Analysis

In order to qualitatively compare the models' receptive fields with physiological data, they were fitted with paramatised mathematical functions. We choose to use the popular difference of Gaussians to fit center surround receptive fields and an elliptical, oriented Gabor model to fit the bar-type receptive fields (see the Introduction and Discussion for an explanation of why center surround and bar detector type receptive fields emerge).

For the difference of Gaussian model, the receptive field sensitivity at retinal location (x, y) is

$$\text{DOG}(x, y) = k_c \exp\left[-\left(d^2/2r_c^2\right)\right] - k_s \exp\left[-\left(d^2/2r_s^2\right)\right] \tag{4}$$

where

$$d = \sqrt{(x - c_x)^2 + (y - c_y)^2} \tag{5}$$

and the six parameters are: location of receptive field centre (c_x, c_y), centre radius (r_c) and sensitivity (k_c) and corresponding surround radius (r_s) and sensitivity (k_s). The radii correspond to the standard deviations of the Gaussians. A set of 12 models were fitted to each receptive field with each model having a range of different initial parameters.

For the elliptical, oriented Gabor model the receptive fields sensitivity at retinal receptive field location is (x, y) (Movellan 2005).

$$\text{Gabor}(x, y) = \cos(2\pi(ux + vy) + P).K \exp(-\pi(a^2 X^2 + b^2 Y^2)) \tag{6}$$

where

$$X = (x - c_x)\cos(\theta) + (y - c_y)\sin\theta \tag{7}$$

$$Y = -(x - c_x)\sin\theta + (y - c_y)\cos\theta \tag{8}$$

$$u = F\cos\theta \tag{9}$$

$$v = F\sin\theta \tag{10}$$

This Gabor model has 8 parameters, K scales the magnitude of the Gaussian envelope, (a, b) scales the two axis of the Gaussian envelope which is oriented at the angle θ in radians. (c_x, x_y) determines the centre position, P is the phase of the sinusoid and F^{-1} is its spatial frequency (in cycles/pixel).

Approximately 300 Gabor models were fitted to each cortical receptive field, each starting with different initial parameters that span the range of sensible values. Then, the Gabor fit which explained most of the receptive fields variance (highest R^2 value) was chosen as the final fit.

From the final sets of Gabor fits, we calculated the half-magnitude orientation bandwidth.

$$\theta_{1/2} \approx 2 \tan^{-1} \left(\frac{b\left(\sqrt{\frac{\log 2}{\pi}}\right)}{F} \right) \tag{11}$$

Each cortical receptive field was classified either as lowpass or bandpass; those whose response to a mean image intensity (i.e., sensitivity at zero spatial frequency) was less than half the peak response were classed as bandpass. 41% of cells satisfied this criteria and their spatial frequency bandwidths were calculated in octaves as

$$F_{1/2} = \log_2(F_{\max}/F_{\min}) \tag{12}$$

where $F_{\min}$ and $F_{\max}$ are the lowest and highest, respectively, spatial frequencies resulting in half the peak response.

Results

Retinal ganglion results

Our previous work established that centre surround organisation is the optimal spatial arrangement to encode natural images under a synaptic energy budget (Vincent & Baddeley 2003). The present study replicates this finding (equation 1, $\alpha = 0.1$, $\beta = 0$) and shows that the result is robust under the new space variant approach (see Figure 2).

To further investigate the centre surround receptive fields in the model and to compare this to the physiology, the raw receptive field data were fitted with difference of Gaussian models (see Methods). It was found that this difference of Gaussian model was effective at describing the data, with a median R^2 of 0.75 (7 out of the 200 receptive fields had $R^2 < 0.5$ and were removed from further analysis).

Figure 3 shows qualitatively that the receptive fields in the fovea are small but increase in size out toward the periphery. Figure 4a shows that this increase in receptive field size is linear with retinal eccentricity such that the entire visual field is tiled by receptive fields. To test this quantitatively, the spacing ratio is calculated; this is the distance between nearest neighbour receptive field centres normalised by the standard deviation of the centre radii (Devries & Baylor 1997). A spacing ratio of 1 means that the 1σ profiles of two cells just touch and corresponds to equal sampling of visual space. In physiology, mean spacing ratio for different ganglion cell classes range from 1.18 to 1.23 (see Devries & Baylor 1997, their Table III). Our data, however, shows spacing ratios change as a function of retinal eccentricity ranging from ~ 0.5 to ~ 3; see Figure 4b. Thus, the fovea is oversampled due to receptive field overlap and the periphery is undersampled. Whether this trend exists in the physiology remains to be seen.

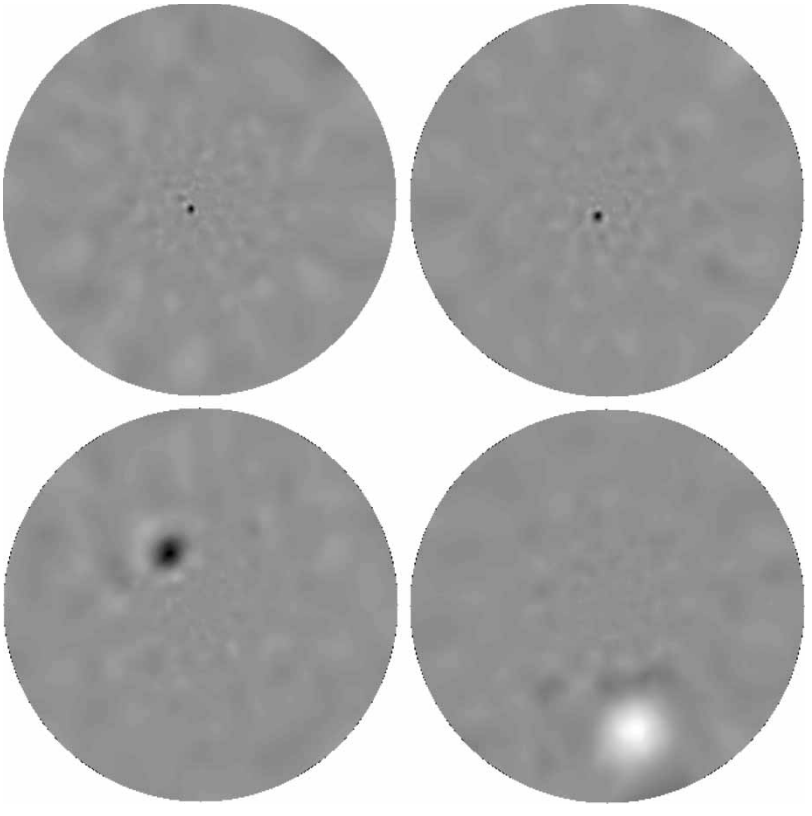

Figure 2. Some retinal ganglion receptive fields. All show clear centre surround organisation despite some residual but very low sensitivity to photoreceptors in other areas of the visual field. In order to display as an image, the raw data of sensitivities at each photoreceptor location was interpolated to get receptive field sensitivities for a uniform grid.

Figure 4c shows the total number of synaptic connections that each centre surround neuron has with the photoreceptors. A synaptic connection was classed as present if the magnitude of that connection was greater than 5% of the peak connection magnitude for a given retinal neuron. The plot shows that, contrary to the physiology, there are more photoreceptors per ganglion cell in the fovea than in the periphery. We have already seen in Figure 4a that receptive fields are in fact smaller in the fovea, despite more connections, but this discrepancy is accounted for by the closer receptor spacing in the fovea.

Figure 4d shows that synapse related metabolic cost of retinal receptive fields is high in the fovea but decreases as a power law with eccentricity, so retinal ganglion cells in the fovea cost more energy than peripheral ones, by approximately a factor of 2.

Cortical results

The cortical stage calculated the optimal receptive fields which balance natural image encoding with costs associated with firing rates (Equation 1, $\alpha = 0$, $\beta = 0.1$). Unlike sparse coding and independent component analysis, the cortical receptive fields are built up from combinations of centre surround receptive fields of the retinal layer of the model. Figure 5 shows resemblance to the well known Gabor-like oriented bar detectors. These plots show

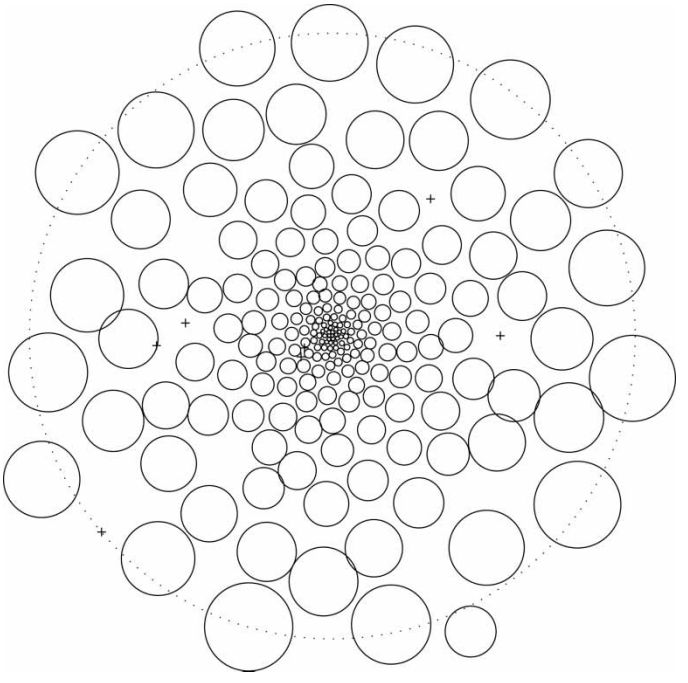

Figure 3. The receptive field centres over the space variant retina. The size of receptive field centres and their positions are shown by circles, derived from difference of Gaussian fits and the total visual field is shown by the large dotted circle. Fits that explain less then 50% of the variance ($R^2 < 0.5$) are indicated by the + sign. Note that the radius of the circles correspond to the standard deviation of the receptive field centre sizes thus are smaller than the total spatial extent of the receptive field.

differences to the types of receptive fields seen from sparse coding and ICA in that they appear much more broadly tuned.

To investigate the cortical receptive fields and to compare them to physiological measures and other models of simple cells, all receptive fields were parameterised with an elliptical Gabor model. These Gabor models were good fits to the data, having a median R^2 value of 0.71 (80 of 800 fits had $R^2 < 0.5$ and were excluded from further analysis).

Just as with the centre surround receptive fields, the cortical receptive field sizes increase with eccentricity. However, there is much more variability so a range of different receptive field sizes (calculated as the average of major and minor axes of the elliptical Gabor) exist at any eccentricity (see Figure 6a). The model aspect ratios show a close match to the distribution found in the physiology: see Figure 6b (physiological data from Parker and Hawken (1988) with the same assumption as in van Hateren and van der Schaaf (1998) that the width of the receptive field envelope is three times the width of the central lobe). Analysis of the orientation bandwidths (Figure 6c) is interesting the range of values in the physiology (data from Parker and Hawken (1988) their Figure 3a) is from 0° to around 100° but the model has a bimodal distribution of orientation bandwidths. Half seem to follow the distribution in physiology but the rest have a greater bandwidth based around 120°. The spatial frequency bandwidths of the model (shown in Figure 6d) are greater than that seen in physiology (data from de Valois et al. (1982) their Figure 5).

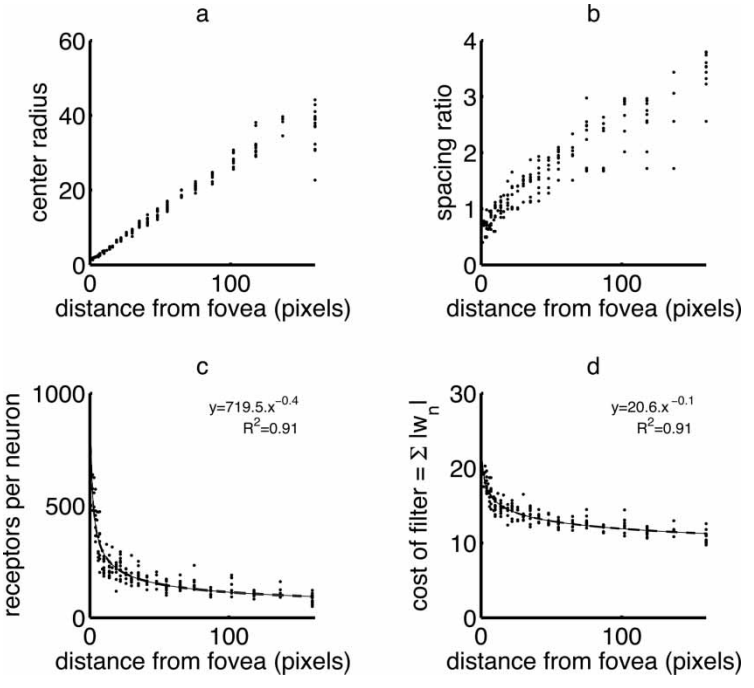

Figure 4. Analysis of retinal receptive field properties. The size of the central regions of receptive fields increase linearly with retinal eccentricity (a). The spacing ratio alters depending upon eccentricity, there being more overlap in the fovea and less overlap in the periphery (b). The connectivity of photoreceptors as a function of eccentricity is shown in terms of number of connections (above 5% peak sensitivity) and decreases as a power law (c). This does not mean that foveal receptive fields are larger as (a) shows an approximately linear increase in receptive field centre size with eccentricity. Foveal receptive fields are more synaptically costly, with cost decreasing exponentially with eccentricity (d). Power law fits are shown with standard error curves (dashed lines) obtained by bootstrap resampling.

Convergence

The photoreceptor distribution is defined *a priori* in the space variant model (see Methods), but the manner in which the visual information is represented is a property of this and balancing coding of the input statistics with metabolic energy expenditure. Under these conditions, there was an excellent qualitative match. Both the retinal and cortical layers place much more (and smaller) receptive fields in the fovea of the visual field whilst the periphery is represented by a much lower number of neurons. Figure 7 shows that both retinal ganglion cells and cortical cells are up to three orders of magnitude more dense in the fovea compared to the periphery. The inset of Figure 7 shows this much higher foveal density.

Examining the slopes of the power law fits to receptive field density as a function of eccentricity shows small but significant differences from physiology (see Table I). The slope of this power law fit gets slightly shallower from receptors to retinal to cortical cells, showing that for each successive stage, the fovea is relatively under-represented as compared to the periphery from the previous representation. This is the opposite trend to physiology, where even though the neural explanation is unclear, the consensus is that ‘cortical magnification’, where the cortex over-represents the fovea relative to the periphery, does exist (Wässle et al. 1989, 1990; Azzoprdi & Cowly 1993).

Table I. Power law fit parameters of the data shown in Figure 7. The (distance, density) data is fitted (in the log–log domain) with power laws such that density = multiple.distance^{exponent}. Standard deviations obtained by bootstrap resampling.

	Exponent	Multiple
Photoreceptors	-2.15 ± 0.01	81.93 ± 3.16
Centre surround	-1.81 ± 0.02	3.65 ± 0.22
Cortical	-1.77 ± 0.07	52.56 ± 12.1

Discussion

Receptive field predictions

The model presented learns receptive fields by finding the optimal balance between encoding natural images and metabolic expense. In the model retina, the synapses incur a cost and in the cortex, it is the firing rates that incur a cost. This was chosen because it is what previous results suggest is important (Olshausen & Field 1996; Vincent & Baddeley 2003), future work will focus on examining different cost combinations in different parts of the visual system.

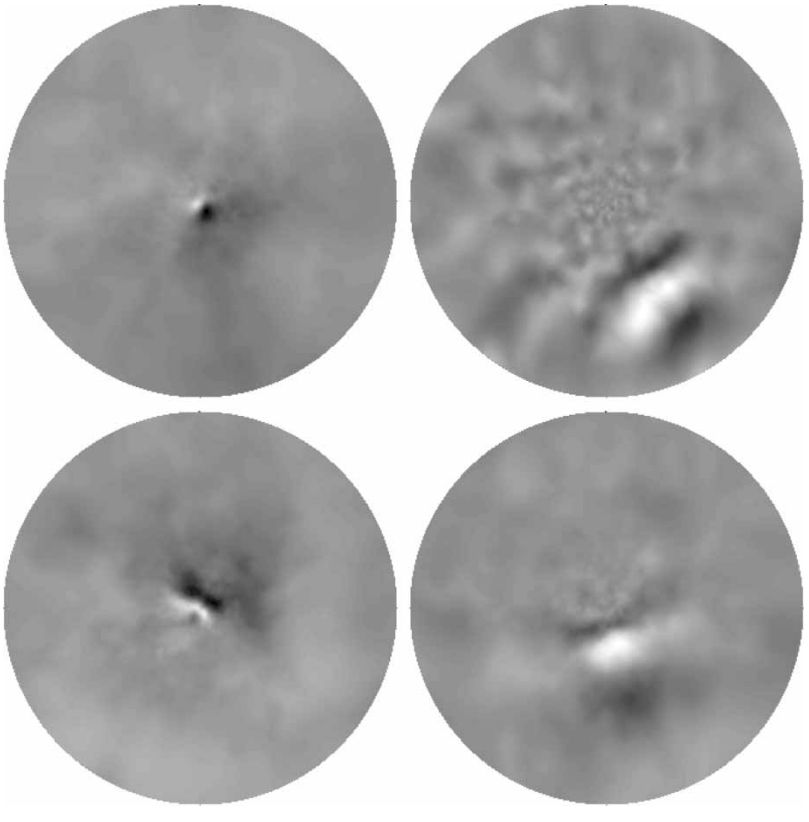

Figure 5. Some cortical Gabor-like receptive fields. The raw receptive field data was interpolated as in Figure 2 for plotting as an image.

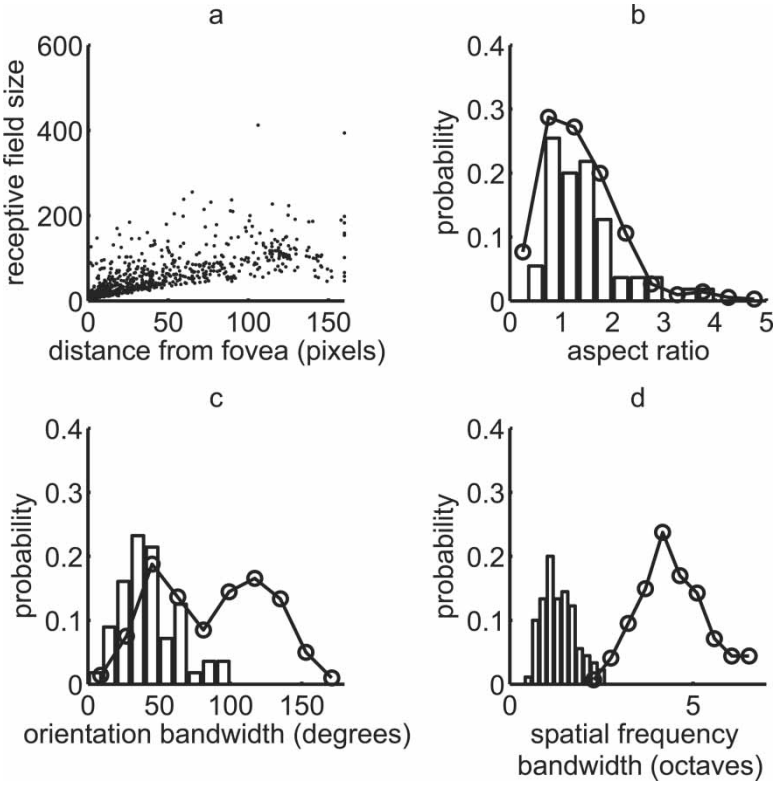

Figure 6. Analysis of cortical receptive field properties. Receptive field size (averages of major and minor axis) increase with retinal eccentricity (a). In (b), (c), (d) physiological data is shown in histograms and the model results are plotted as curves. The aspect ratio (length/width) of the model is a close match to the physiological data (b). The orientation and frequency bandwidths are shown in (c) and (d), respectively.

The retinal receptive fields have centre surround organisation, being very small in the fovea and increasing in size out toward the periphery. The spacing ratio is a similar value to that seen in physiology but the model predicts the spacing ratio may change depending on retinal eccentricity.

The model cortex includes a firing rate constraint where cost is proportional to firing rate which is equivalent to a cost function used in sparse coding (Olshausen & Field 1996; Harpur 1997). The cost functions are such that they promote a zero-peaked distribution of firing rates. The cortical receptive fields form Gabor-like bar detectors, and these also increase in size out toward the periphery. We find that the aspect ratios of model receptive fields fits the distribution found in physiology, but a number of other properties differ from physiology. Around half of the receptive fields have a similar orientation bandwidth as cortical neurons, but the remainder have a much greater orientation bandwidth. In addition, the frequency bandwidth of the model receptive fields is greater than cortical cells in biology. These two deviations show that the model cortical receptive fields are more broadly tuned than the results from sparse coding, ICA and the physiological data. We suggest two reasons why this may be so. Firstly, the spatial organisation of the cortical receptive fields is limited by being the weighted sum of the retinal centre-surround receptive fields. Although this is exactly the situation in biology, the low number of neurons in the model limits the spatial

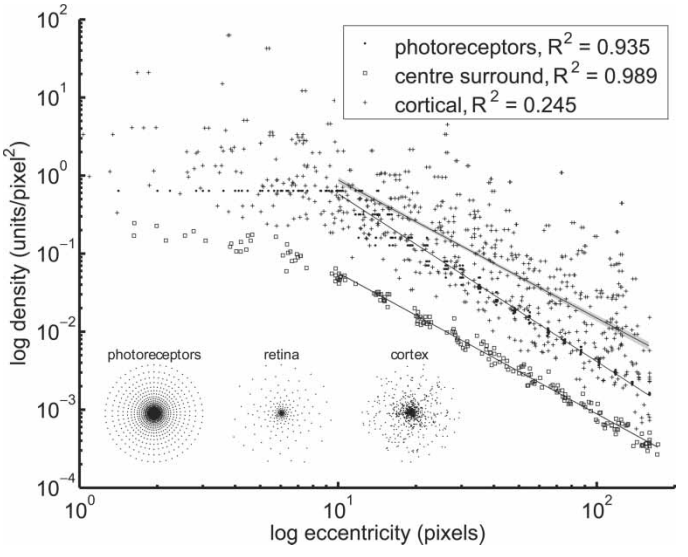

Figure 7. Analysis of magnification factors. Here, the densities of photoreceptors and centre-surround and cortical receptive fields are compared as a function of retinal eccentricity. The photoreceptors show a small region of uniform density within the central 10 pixels. Straight lines represent mean power law fits (from eccentricity 10 to 160) and shaded regions are standard deviations of the fits obtained by bootstrap resampling. Inset shows the spatial locations of the photoreceptors, retinal and cortical cells demonstrating the much higher densities in the fovea.

fineness, so more retinal receptive fields would allow for finer, more narrowly tuned, cortical receptive fields. Secondly, the number of cortical neurons was restricted to 800 due to limited computational resources. Because of this, the receptive fields are probably more broadly tuned in order to capture the highest-variance grosser features of the natural images. We predict that the cortical receptive fields would become more finely tuned and thus have lower orientation and frequency bandwidths given increased numbers of cortical neurons in the model.

Convergence/magnification factors

In our previous work, we tested the predictions of synaptically efficient coding at various eccentricities by altering the convergence ratio of photoreceptors to ganglion cells (Vincent & Baddeley 2003). This showed that receptive field size decreases as the number of ganglion cells increase (i.e., more foveal locations), with the receptive fields consisting of direct, one-to-one connections where the number of ganglion cells is equal to the number of photoreceptors. These predictions are interesting and valid, but a stronger test of the predictive ability of energy efficient coding would be to measure the optimal receptive field placement over an entire space variant visual field as we do here.

Qualitatively, the model matches the physiology very well. Both in the retina and the cortex, the fovea is up to 3 orders of magnitude more densely sampled than in the periphery.

Quantitatively there were small but significant differences to physiology. A greater area in the cortex is devoted to foveal than peripheral regions. In the model, we observe a slight shallowing of the density function from photoreceptors to retinal to cortical cells. This means that the model cortex does have many more receptive fields in the fovea but

would require even more to reflect the physiological measurements. Although the locus of this over-representation is unclear, be it the retina (Wässle et al. 1989, 1990) or the cortex (Azzoprdi & Cowey 1993) it does exist.

An initial prediction would be that if the variance was identical across inputs, then the density of any subsequent stage would reflect the density of inputs. We did not find this, and we argue that it is the image statistics combined with the space variant sampling array which leads to this slight discrepancy. The autocorrelation function of natural images means that closer regions (i.e., the dense foveal photoreceptor sampling) will be more correlated than the less dense sampling of the periphery. This causes the model, which is concerned with information and thus variance, to place slightly more receptive fields in the periphery relative to the fovea. This is believed to be the cause of the slight relative over-representation of the periphery in each stage of the model.

The model predictions are highly dependent upon the input statistics, and even though we have gone to lengths to capture 'natural' image statistics, we cannot be sure that it truly reflects the statistics met by an organism in the natural world. We believe the following three factors could lead to different model predictions including the foveal over-representation as seen in the cortex:

- We did not take increased optical blurring in the periphery into account (Wang & Ciuffreda 2004). This may be relevant as it would broaden the autocorrelation function of peripheral photoreceptors more than those in the fovea. This would decrease the input variance of the peripheral photoreceptors relative to the fovea and thus induce more resources into the fovea.
- Fixation behaviour of humans is found to have a correlate with the presence of edges (Tatler et al. 2005). If this were taken into account then the input statistics would alter such that edges occur more in the fovea (fixation position) and more dense sampling of the cortex is likely.
- With independent component analysis it was found that taking the time domain into account resulted in much better fits to the physiology (van Hateren & Ruderman 1998). Extending the energy efficient coding model to the spatiotemporal domain would certainly alter the optimal receptive fields.

Comparison to sparse coding and independent components analysis approaches

Previous studies have shown that the oriented bar type receptive fields are optimal in terms of encoding natural images when a 'sparse' zero-peaked distribution is enforced (Olshausen & Field 1996). A zero-peaked firing rate distribution simply means that at any given time the majority of neurons are inactive or weakly active whilst a low proportion are strongly active. In addition, independent component analysis makes a very similar prediction: oriented bar type receptive fields are optimal (Bell & Sejnowski 1997; van Hateren & Ruderman 1998). These studies elude to a fairly robust understanding of simple cells in the primary visual cortex (or at least their linear characteristics see (Olshausen & Field 2004)), but there are certain unresolved issues of interpretation with both sparse coding and ICA.

A sparse code or a zero-peaked distribution of firing rates has (in itself) no intrinsic appeal over any other code and so some would question the justification of sparsity other than the fact that its optimal receptive fields are similar to simple-cells. ICA certainly is a principled approach (Hyvarinen & Hoyer 2001); various linear or non-linear processing steps may increase the independence of neural responses but true independence seems to be an

unreachable goal. Nevertheless, ICA is a very interesting approach and increasing independence does seem to be a sensible coding strategy.

From the results conducted thus far, we propose that 'energy efficient coding' provides a solid framework with which to understand the *why* of early visual processing. It shows specifically that a zero-peaked distribution of firing rates is desirable to minimise the metabolic requirements of encoding natural images (see here and Baddeley (1996)). Similarly, energy efficient coding shows that a zero-peaked distribution of synapses minimises costs associated with synaptic activity and results in realistic centre surround receptive fields (here and Vincent & Baddeley (2003)). The justification of metabolic energy efficiency is clear from the evolutionary and physiological viewpoints in addition to predicting many features of centre surround and cortical processing.

Conclusions

We have investigated the optimal receptive field organisation of a two-layer visual system which balances natural image coding with metabolic energy efficiency. Energy efficiency is obviously important biologically and goes a long way to explain why the early visual system is arranged how it is. Analysis of the optimal receptive fields shows that although there are some deviations from physiology, many properties of the early visual system are captured. Not only can the centre-surround and oriented-bar type organisation be explained, but also the distribution of receptive fields is captured fairly well.

The model had a limited number of neurons because of computational constraints, and we believe that this accounts for why the cortical receptive fields are more broadly tuned than real cortical neurons. Creating a larger scale model could result in more narrowly tuned receptive fields. In addition, we expect that some of the models foveal underrepresentation may be a result of not capturing the full statistics of fixated natural scenes. Thus, a greater understanding of how eye fixation positions affect the image statistics could lead to improved matches between the energy efficient coding approach and physiological observations.

These results add to a growing body of work which recognises the importance of the brains biophysical nature in determining its structure and functioning (Laughlin & Sejnowski 2003). For example, the approach of minimum wiring volume can account for many *structural* aspects of the brain such as cortical folds (Essen 1997), topographic map organisation (Koulakov & Chklovski 2001) and neural branching patterns (Mitchison 1992; Chklovskii 2000). Concurrently, a number of labs have studied the impact of metabolic energy expenditure upon neural *function*. Energy considerations explain, for example, why it can be optimal that fewer neurons are active than is predicted from an information coding perspective alone (Levy & Baxter 1996) and is exemplified by visual cortical receptive fields (Olshausen & Field 1996; Harpur 1997). In addition, the patterns of spikes emitted from retinal ganglion cells can be viewed as optimal when noise and costs are considered (Balasubramanian & Berry 2002). The nervous system is one of the most energy hungry of all organs (Rolfe & Brown 1997) and a detailed characterisation of energy consumption in neural tissue (such as Attwell and Laughlin (2001)) is vital, providing empirical data upon which theoretical studies can be based.

In conclusion, we argue that the energy efficient coding approach moves the understanding of the early visual system forward as it provides a real biologically-based justification for sparse coding in the cortex and retinal ganglion cells.

Acknowledgments

This work was supported by EPSRC grant number GR/S47953/01(P). We wish to thank Alex Párraga, Ute Leonards, and J. Troscianko for the use of the Ugandan image dataset and George Lovell for constructive discussions.

References

Attwell D, Laughlin SB. 2001. An energy budget for signaling in the grey matter of the brain. J Cereb Blood Flow Metab 21:1775–1783.

Azzoprdi P, Cowey A. 1993. Preferential representation of the fovea in the primary visual cortex. Nature 361:719–721.

Baddeley R. 1996. Visual perception. An efficient code in V1. Nature 381:560–561.

Balasubramanian V, Berry MJ. 2002. A test of metabolically efficient coding in the retina. Network: Comput Neural Syst 13:531–552.

Baldi PF, Hornik K. 1995. Learning in linear neural networks: A survey. IEEE Transactions on Neural Networks 6:837–858.

Bell AJ, Sejnowski TJ. 1997. The 'independent components' of natural scenes are edge filters. Vision Res 37:3327–3338.

Bolduc M, Levine MD. 1998. A review of biologically motivated space-variant data reduction models for robotic vision. Comp Vision Image Understanding 69:170–184.

Chklovskii DB. 2000. Optimal sizes of dendritic and axonal arbors in a topographic projection. J Neurophysiol. 83:2113–3119.

da Costa BLSA, Hokoç JN. 2000. Photoreceptor topography of the retina in the new world monkey *Cebus apella*. Vision Res 40:2395–2409.

Devries SH, Baylor DA. 1997. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. J Neurophysiol 74:2048–2068.

de Valois RL, Albrecht DG, Thorell LG. 1982. Spatial frequency selectivity of cells in the macaque visual cortex. Vision Res 22:545–559.

Essen DCV. 1997. A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature 385:313–318.

Harpur GF. 1997. Low entropy coding with unsupervised neural networks. PhD thesis, Oxford University.

Hyvarinen A, Hoyer PO. 2001. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. Vision Res 14:2413–2413.

Koulakov AA, Chklovski DB. 2001. Orientation preference patters in mammalian visual cortex: A wire length minimization approach. Neuron 29:519–527.

Laughlin SB, Sejnowski TJ. 2003. Communication in neuronal networks. Science 301:1870–1874.

Levy WB, Baxter RA. 1996. Energy efficient neural codes. Neural Computation 8:531–543.

Mitchison G. 1992. Axonal trees and cortical architecture. TINS 15:122–126.

Movellan JR. Tutorial on Gabor filters. Available: http://mplab.ucsd.edu/tutorials/pdfs/Gabor.pdf.

Olshausen B, Field D. 1996. Emergence of simple-cell receptive fields properties by learning a sparse code for natural images. Nature 381:607–609.

Olshausen BA, Field DJ. 2004. What is the other 85% of V1 doing? In: van Hemmen JL, Sejnowslki TJ. Problems in Systems Neuroscience. Oxford: Oxford University Press.

Parker AJ, Hawken MJ. 1988. Two-dimensional spatial structure of receptive fields in monkey striate cortex. J Opt Soc Am A 5:598–605.

Párraga CA, Troscianko T, Tolhurst DJ. 2002. Spatiochromatic properties of natural images and human vision. Current Biol 12:483–487.

Tatler BW, Baddeley RJ, Gilchrist ID. 2005. Visual correlates of fixation selection: Effects of scale and time. Vision Res 45:643–659.

Rolfe DF, Brown GC. 1997. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. Physiol Rev 77:731–758.

Troscianko T, Párraga CA, Leonards U, Baddeley R, Troscianko J, Tolhurst DJ. 2003. Leaves, fruit, shadows and lighting in kibale forest, Uganda. Perception 32 (suppl.):51a.

van Hateren JH, Ruderman DL. 1998. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. Proc R Soc Lond B Biol Sci 265:2315–2320.

van Hateren JH, van der Schaaf A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. Proc R Soc Lond B Biol Sci 265:359–366.

Vincent B, Baddeley R. 2003. A role for energy efficiency in retinal processing. Vision Res 43:1283–1290.

Wang B, Ciuffreda KJ. 2004. Depth-of-focus of the human eye in the near retinal periphery. Vision Res 44:1115–1125.

Wässle H, Grünert U, Röhrenbeck J, Boycott B. 1989. Cortical magnification factor and the ganglion cell denisity of the primate retina. Nature 341:643–646.

Wässle H, Grünert U, Röhrenbeck J, Boycott B. 1990. Retinal ganglion cell density and cortical magnification factor in the primate. Vision Res 30:1897–1911.