Separation of Luminance and Chromatic Information by Hebb-Stent Rules

William McIlhagga
McGill Vision Research Unit
687 Pine Ave. West, H4-14 Montréal, Canada H3A-1A1

Graeme Cole
Dept. of Mathematics and Physical Sciences
Murdoch University, Western Australia

Abstract

The P cell pathway in primates is involved in both luminance and chromatic perception. This correlates well with P cell receptive fields, which are both spatially and chromatically opponent. Since, however, the luminance and chromatic channels found in psychophysics are independent, the mixed luminance and chromatic information in P cell signals must be demultiplexed in cortex.

We have examined the ability of an unsupervised neural network to demultiplex P cell signals, using realistic visual inputs. Digitised images, corrected to be statistically similar to retinal images, were sampled by a simulated retinal mosaic, and filtered by difference-of-gaussians P cell receptive fields. The simulated P cell signals were used as inputs to a network designed to maximise unit responses while minimising the correlation between units. After a period of training, we evaluated the receptive fields formed in the network. The neurons clearly fell into two categories. The first were those sensitive to changes in intensity in the retinal image; that is, luminance selective units. The second were those sensitive to a colour difference in the retinal image; that is, chromatically selective units.

1. Introduction

Light that falls on the retina is absorbed by a mosaic of photoreceptors. These are of three types (called L, M and S cones), sensitive to different regions of the spectrum. Our colour sensations do not depend directly on the cone responses; instead, they result from combinations of cone responses. The simplest and most widely accepted model says that human colour vision is mediated by three channels: a "luminance" channel, which is a sum of L and M cone responses; a "red/green" channel, which is a difference of L and M cone responses; and a "blue/yellow" channel, which compares the response of S cones with that of L and M cones[1,2] (Fig. 1a). The luminance channel is spatially bandpass, preferring medium spatial frequencies (2-5 cycles per degree). The colour channels are spatially low pass.[3]

These separate channels are not present in the retina or lateral geniculate nucleus (lgn). In the retina there are two distinct classes of ganglion cell: those with small cell bodies, which feed to the P layers of the lgn (and which we will call retinal P cells), and those with large cell bodies which feed to the M layers of the lgn (which we will call retinal M cells). Both classes of ganglion cell typically have

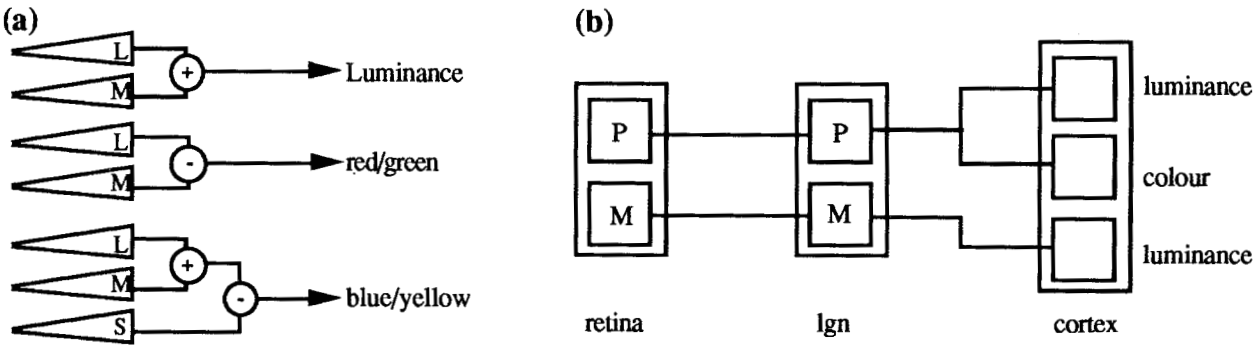


Figure 1. The colour and luminance channels from the point of view of (a) psychophysics (b) physiology.

an antagonistic centre/surround receptive field. Near fovea, the centre of a retinal P cell receptive field is so small that only one cone can be contained inside it; the surround, however, can contain many cones. Opponency between the centre cone (either L, M or S) and the cones in the surround (a more varied sample of cone types), yields a degree of colour opponency reminiscent of the red/green or blue/yellow channels.[4] However, the spatial opponency of the receptive field means that a retinal P cell is still responsive to changes in luminance. The centres of M cell receptive fields are large enough to contain many cones, so the degree of colour opponency in these cells is much less than in the P cells. The retinal P and M cells feed to the P and M layers of the lgn, where the receptive fields are essentially unchanged.[5] The P and M layers of the lgn then feed to different layers of the visual cortex (Fig. 1b). At this stage, presumably, the psychophysical channels come into existence; there have been reports of cortical cells which could be substrates of these channels.[6]

One important component of the schema in Fig. 1b is the separation of the P cell pathway into separate luminance and chromatic channels in cortex. (This organisation has been confirmed by lesion studies of the lgn.[7]) Somehow, the cortical cells must separate out the combination of luminance and colour information that exists in the P cell signals to produce the two channels. If the P cells were marked anatomically according to the kind of cone at their centre (by, for example, a characteristic membrane protein), this separation could be "hard-wired" in the cortex, and implement a simple demultiplexing scheme.[8] There is, however, reason to suspect that the P cells cannot be marked this way, because the P cells themselves cannot distinguish between the different cone types. In this case, separation of P cell signals into luminance and chromatic channels must occur by some kind of adaptive process.

In this paper, we will explore the possibility that the luminance and colour pathways in cortex that arise from the P cell pathway are a result of a learning process. The outline of the paper is as follows. First we will analyse the local statistical structure of real images, using principal components methods. We find that for colour images, the local principal components correspond well to what we would call luminance and chromatic channels. Next, we develop a simple, biologically plausible, neural network which approximates the extraction of principal components from its inputs. When this network is run using real images as inputs, the units in the network develop receptive fields which correspond to a selection of colour and luminance principal components. The receptive fields are similar to those found in area V1 of primate cortex.

2. Principal Components

2.1 Definition

Principal components analysis represents a data vector $\mathbf{z}$ (a random variable) as a sum of vectors which are in some sense typical of the distribution of $\mathbf{z}$. There are a number of derivations of principal components. Here, we will define the first principal component of $\mathbf{z}$ as the unit vector π which minimises

$$\left\langle \left\| \mathbf{z} - \pi(\mathbf{z}' . \pi) \right\|^2 \right\rangle \tag{1}$$

where <.> is the expected value over the ensemble of possible data vectors, and $\mathbf{z}'$ is the transpose of z. In this definition, the principal component $\mathbf{z}\pi$ is the pattern which, when scaled by $\mathbf{z}'\pi$, most closely matches the data vector $\mathbf{z}$. Expanding yields

$$\begin{aligned}
\left\langle \left\| \mathbf{z} - \pi(\mathbf{z}' . \pi) \right\|^2 \right\rangle &= \left\langle \left(\mathbf{z}' - \pi'(\mathbf{z}' . \pi)\right) . \left(\mathbf{z} - \pi(\mathbf{z}' . \pi)\right) \right\rangle \\
&= \left\langle \mathbf{z}' . \mathbf{z} - 2(\mathbf{z}' . \pi)^2 + \pi' . \pi (\mathbf{z}' . \pi)^2 \right\rangle \\
&= \left\langle \mathbf{z}' . \mathbf{z} - (\mathbf{z}' . \pi)^2 \right\rangle \quad \text{since} \quad \pi' . \pi = 1 \\
&= \langle \mathbf{z}' . \mathbf{z} \rangle - \pi' \langle \mathbf{z} . \mathbf{z}' \rangle \pi
\end{aligned}$$

which is minimised when π is the eigenvector of $\langle \mathbf{z}\,\mathbf{z}' \rangle$ with largest eigenvalue. The second principal component of $\mathbf{z}$ is the principal component of the residual $\mathbf{z} - \pi.(\mathbf{z}'\ \pi)$; and so on for the third, fourth, fifth principal components. Each successive principal component accounts for less and less of the variation in $\mathbf{z}$. If we write the i-th principal component as π^i, the data vector $\mathbf{z}$ can be expanded as a weighted sum of these components:

$$\mathbf{z} = \sum_i \pi^i (\mathbf{z}'\ \pi^i) \tag{2}$$

The number of principal components in the summation is the same as the rank of $\langle \mathbf{z}\ \mathbf{z}' \rangle$, but if the summation is reduced to include only the first k principal components say, with $k < \text{rank}(\langle \mathbf{z}\,\mathbf{z}' \rangle)$, then the equation holds true in a least-squares sense, because each principal component is orthogonal to the others.

2.2 Local Principal Components

In image analysis, $\mathbf{z}$ is a vector of pixel values: $\mathbf{z}_x$ is taken to be the pixel value at position x. The definition (1) above will yield the global principal components of an ensemble of images, of which $\mathbf{z}$ is one example; but the global principal components are a cumbersome way of describing the image, and like global Fourier analysis lack information about local features. A better approach is to look at local or windowed principal components (in an analogy to the Gabor filter). Choose a window function, and represent it by an image vector $\mathbf{w}$ where $\mathbf{w}_x$ is the value of the window at position x. Letting $\mathbf{W} = \text{diag}(\mathbf{w})$, the windowed image is then $\mathbf{Wz}$. To cover the image with these local principal components, we need a series of windows W_k such that $\sum W_k = 1$. Since natural images are statistically translation invariant, we should select each W_k to be a translation of a standard window (e.g. a gaussian). Then the local principal components within each window W_k are just translations of the local principal components under the standard window.

There are a few possible definitions of local principal components (LPC). We shall define a local principal component, using window W, as the unit vector π that minimises

$$\left\langle \left\| \mathbf{Wz} - \pi(\mathbf{z}'\,\mathbf{W}\pi) \right\|^2 \right\rangle \tag{3}$$

Since this merely substitutes $\mathbf{Wz}$ for $\mathbf{z}$ in (1), the principal component π is the eigenvector of $\mathbf{W}\langle \mathbf{z}\,\mathbf{z}' \rangle\mathbf{W}$ with largest eigenvalue. Note that if we want to identify the local

principal components with a neural receptive field or with a filter kernel, the receptive field weights are $W\pi$ not just π, since it is the former that is applied to the image data **z**.

The data vector **z** can be expanded into a weighted sum of local principal components, in much the same way as it can be expanded into a sum of global principal components in equation (2). If $\pi_k^1, \pi_k^2, \pi_k^3, \ldots \pi_k^n$ are the successive local principal components of the weighted image $W_k\mathbf{z}$, then the original image **z** can be expanded as

$$\mathbf{z} = \sum_k W_k\mathbf{z} = \sum_{k,i} \pi_k^i (\mathbf{z}' W_k \pi_k^i)$$

Unlike global principal components, however, if each windowed image $W_k\mathbf{z}$ is not completely expanded into it's components, the equation does not hold in a least-squares sense, because the principal components of adjacent windows are not orthogonal.

The major problem with LPC is its efficiency in representing an image. The sampling rate of each of the local principal components is tied to the sampling rate of the window function W_k: that is, the number of copies of W needed to cover the image **z**. Conceivably, some of the local principal components could be sampled at a lower rate than this, whereas others may represent some image characteristics that are better sampled at a higher rate. It would be interesting to develop a theory which combines the optimality properties of principal components and the sampling flexibility of multi-rate or pyramid representations of an image,[9] though this is outside the scope of the present work.

3. Colour Principal Components

3.1 Multisensor Images

A colour (or multisensor) image is a set of images formed by integrating the incident light over a series of spectral sensitivity curves. An example is the multi-image produced by an RGB digitizer. A multi-image can be represented by a set of image data vectors $\{\mathbf{z}^1, \mathbf{z}^2, \ldots \mathbf{z}^r\}$. A multi-image can be analysed for local principal components in exactly the same way as a single image, by analysing the vector formed by concatenating the multi-image components i.e. by analysing the principal components of the vector $\mathbf{z} = (\mathbf{z}^{1'}, \mathbf{z}^{2'}, \ldots \mathbf{z}^{r'})'$. The windowing function is also repeated as often as necessary to cover all image components; that is W = diag(**w,w,w,... w**).

A more revealing approach is to decorrelate the images on a pixel-by-pixel basis,[10] before computing the local principal components. The pixel values of all the images at point x form a vector $(z_x^1, z_x^2, \ldots z_x^r) = p_x$ say. We can compute the principal components of p_x over all positions x. These are the eigenvectors of the matrix $M_{ij} = \langle p_x^i . p_x^i \rangle$, where the expectation is over all positions x. Let $\mathbf{b}_q$ be the q-th such eigenvector. Then we can form the image

$$\mathbf{d}^q = \sum \mathbf{b}_{qi}\mathbf{z}^i, \text{ where } \mathbf{b}_{qi} \text{ is the i-th element of } \mathbf{b}_q.$$

The set $\{\mathbf{d}^1, \mathbf{d}^2, \ldots \mathbf{d}^r\}$ is another multi-image, but one with zero pixel-wise correlation. If the cross-correlation between two images in the original set $\langle \mathbf{z}_x^i . \mathbf{z}_w^i \rangle$ is factorable into a spatial covariance $c_{x,w}$ (common to all images) and a sensor covariance c_{ij} (unique to the particular pair i, j of images) i.e. $\langle \mathbf{z}_x^i . \mathbf{z}_w^j \rangle = c_{x,w} . c_{i,j}$, the new multi-image consists of a set of completely independent images, which may be analysed for their spatial principal components separately.

An example of this kind of analysis is shown in Figure 2. We analysed the red and green channels of a digitised RGB image. For simplicity, we did not perform a full 2D analysis, but instead a 1D analysis of the scanlines of the image. The image was first high-pass filtered to set the mean pixel value to zero (a useful step in principal components analysis), by subtracting a gaussian-blurred image (using a blur a of 4 pixels) from the original image. We built up a small (16 by 16) cross-correlation matrix by sampling 16 pixel segments from randomly selected rows. The cross-correlation matrix (or covariance matrix) of the red and green channels is shown in Fig 2a. After pixel-wise decorrelation, the cross-correlation matrix looked like Fig. 2b. The original channels were almost, but not quite, factorable into a sum and a difference component. A local principal components analysis was performed on the pixel decorrelated multi-image, using a gaussian window with σ = 4 pixels. The first five components $W\pi^1$, $W\pi^2$ ... $W\pi^5$ are shown in Figure 2c.

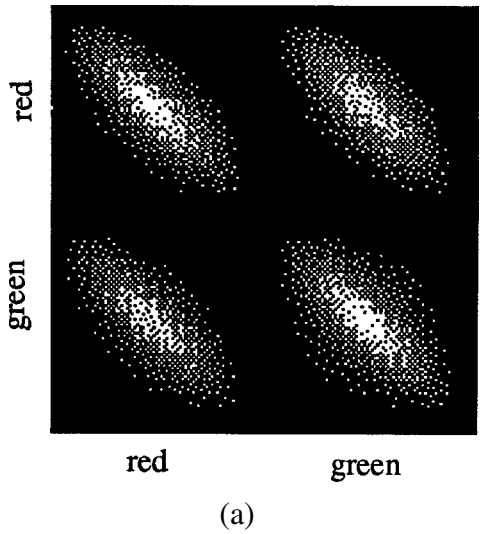
3.2 Retinal Images

The multi-image of interest in this paper is one produced by integrating the incident light with the sensitivity curves of the human L, M and S cone photoreceptors. We shall use the notation $\{\mathbf{z}^L, \mathbf{z}^M, \mathbf{z}^S\}$ to indicate the L, M, and S cone images: $\mathbf{z}^L{}_x$ being the quantal catch of an L cone situated at point x, etc. Each image is also scaled to have a mean of 1 (von Kries adaptation). The retinal image differs from $\{\mathbf{z}^L, \mathbf{z}^M, \mathbf{z}^S\}$ in that only one cone is present at each retinal position. If we let L_x, M_x, and S_x be the characteristic functions of each cone distribution in the retina (i.e. L_x =1 if point x contains an L cone, 0 otherwise) then the retinal image **r** is

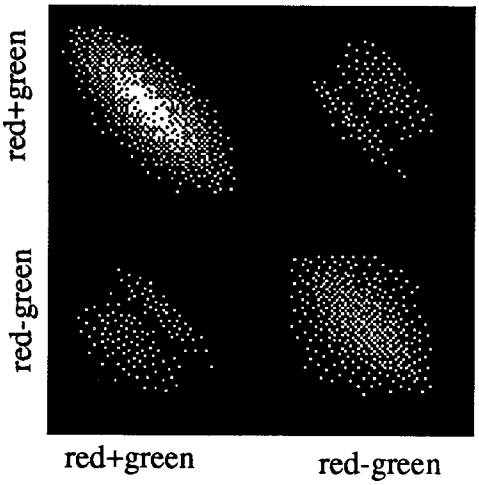
$$\mathbf{r}_x = L_x\mathbf{z}^L{}_x + M_x\mathbf{z}^M{}_x + S_x\mathbf{z}^S{}_x$$

One can, as before, compute the local principal components of **r**, but their functional interpretation is different from a single-spectrum image. In the retinal image **r**, colour information is encoded as a variation in cone response with position, since it is position that determines the cone type. For example, a red colour will cause L cones to respond slightly more than M cones, and so cause a spatial "ripple" in the retinal image r with the same pattern as $L_x - M_x$. The colour ripple has a high spatial frequency, and so it is enhanced by high-pass filtering. This presumably is why the retinal P cells have such small receptive fields.

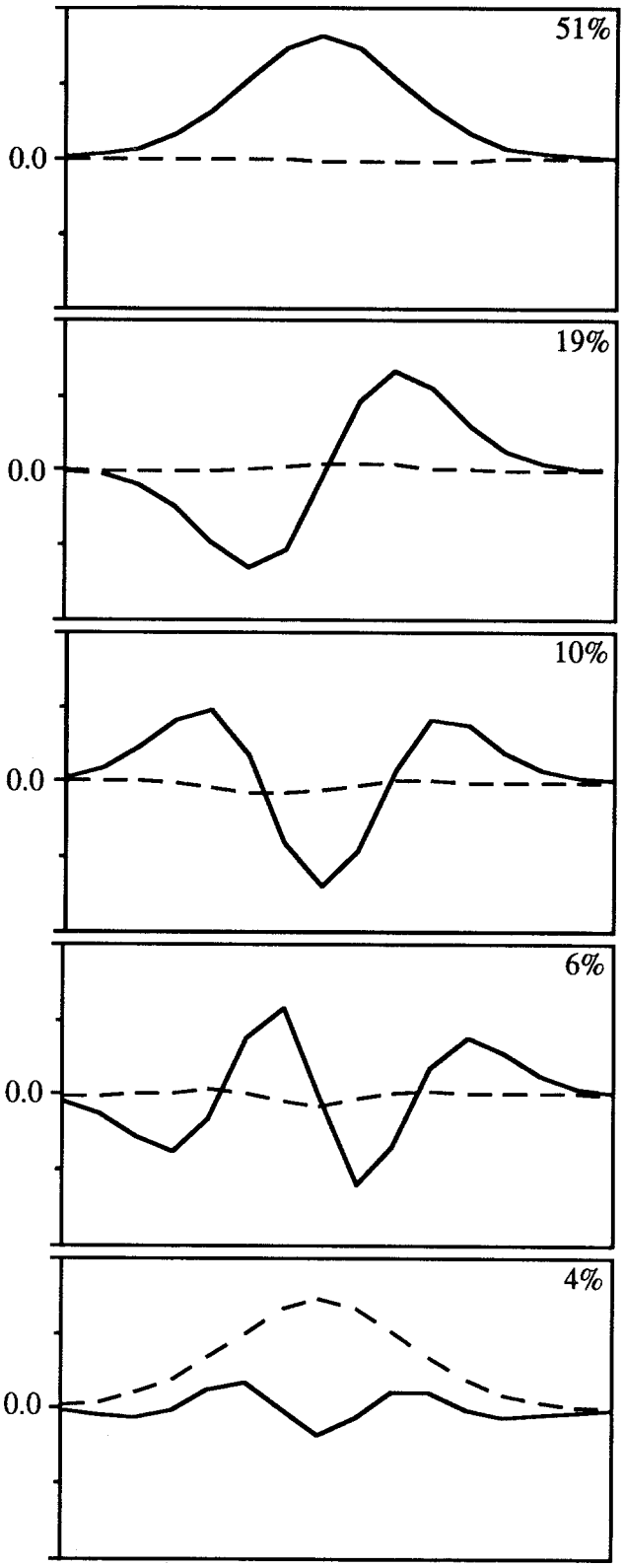
The other critical difference between the principal components of an ordinary image and the those of die retinal image is that the latter are not translation invariant, because the distribution of cones in the retina is not translation invariant. Thus, the principal components of a retinal image cannot be interpreted as convolutional filter kernels, though they may still be interpreted as neural receptive fields. Principal components from a simulated retina are shown in Figure 3. Only $\mathbf{z}^L$ and $\mathbf{z}^M$ were used to build the retinal image. We approximated $\mathbf{z}^L$, $\mathbf{z}^M$ by a linear combination of the red and green channels of the image used in Figure 2. The linear combination was selected to increase the pixel-



(a)



(b)



(c)

Figure 2. (a) Spatial Covariance matrix of the red and green channels of a natural image. (b) Covariance matrix of the same image after decorrelating into two channels, red+green and red-green. (c) Local Principal components from the covariance matrix b. Solid line indicates the component due to red+green; dashed lines are the component part from red-green channel. Numbers in upper left corner give the percentage of variance accounted for by the component.

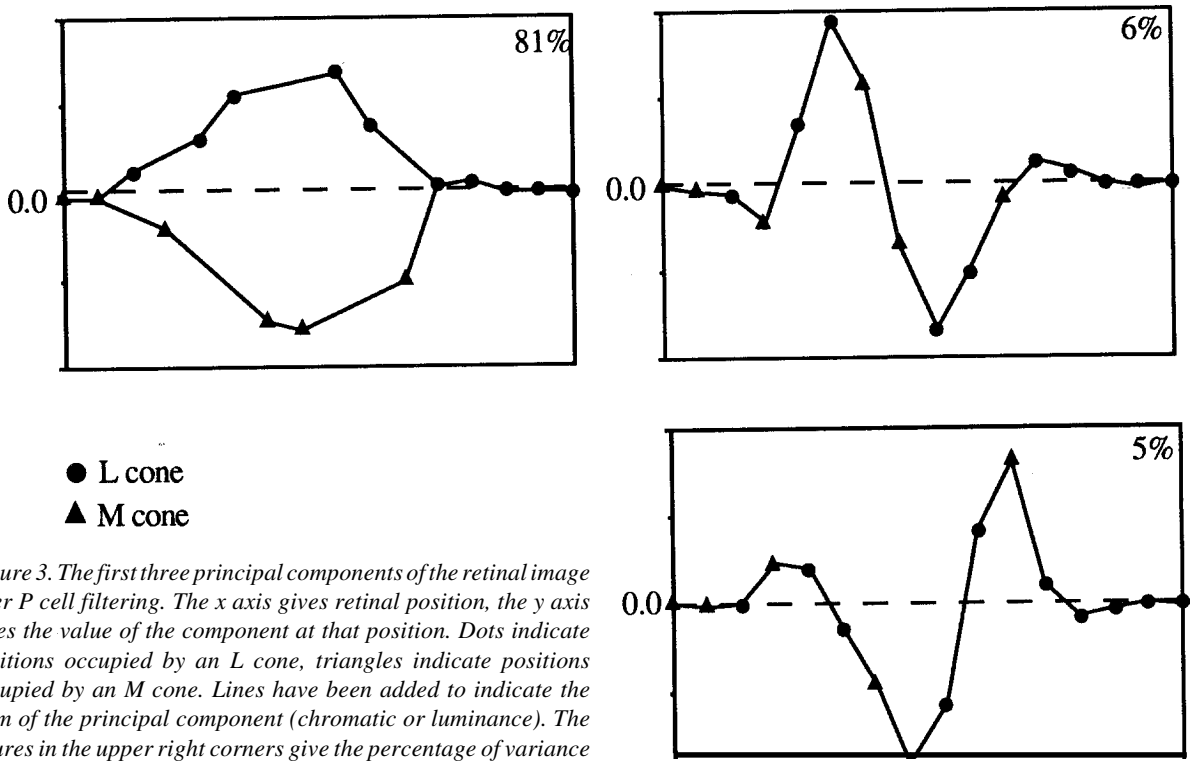


Figure 3. The first three principal components of the retinal image after P cell filtering. The x axis gives retinal position, the y axis gives the value of the component at that position. Dots indicate positions occupied by an L cone, triangles indicate positions occupied by an M cone. Lines have been added to indicate the form of the principal component (chromatic or luminance). The figures in the upper right corners give the percentage of variance accounted for by each component.

wise correlation of $\mathbf{z}^L$ and $\mathbf{z}^M$ to 98%.[11] Because natural images are almost factorable, and because natural reflectances can be approximated by a 3 element basis set[12] so that—with a fixed illuminant—any two trichromatic spaces can be linked by a linear transform, we believe that the fake $\{\mathbf{z}^L, \mathbf{z}^M\}$ is statistically very similar to the real thing.

Each of $\mathbf{z}^L$ and $\mathbf{z}^M$ was sampled by a 16-pixel "retina" of randomly selected L and M cones, with on average 2 L cones for each M. The retinal image was then filtered with a difference-of-gaussians receptive field simulating a retinal P cell. Various sizes of the centre were used, from σ = 0.5 pixels to 1.5 pixels. The surround σ was always 1.6 times larger. Below is an example of the retinal local principal components (using the same windowing function as in Figure 2) for σ = 0.7 pixels. The first principal component is a colour component, differencing the L and M cone responses. It is the dominant component for σ up to 1.2 pixels, but disappears when σ = 1.5 pixels. Note also that the other principal components in Figure 3 represent fine detailed luminance variation; Thus, if the schema shown in Figure lb is correct, the P cell pathway must extract more than just the first principal component of its retinal signal.

4. Neural Nets

4.1 Neural Net Algorithm

There is a strong connection between principal components analysis and Hebbian learning in neural networks. A model neuron generally accepts an input vector **z**, which is weighted by a vector of synaptic strengths **s** to produce a neural response n = **w.z**. A Hebbian neuron changes the weight vector **s** after each input by an amount $\Delta\mathbf{s}$ proportional to n**z**. Provided **s** is normalised, it will tend towards the first principal component of the ensemble of inputs **z**.[13]

The basic Hebbian rule can be extended to extract all principal components of an input, using a number of neurons interconnected in various clever ways.[14,15,16] While all these algorithms are effective in extracting principal components, all of them have one or more unbiological features. Either the wiring scheme required is not known to exist, or too many inhibitory interneurons are needed (only about 10% of visual cortical neurons are inhibitory), or the number of interconnections in the network rises as the square of the number of neurons, which could overwhelm the brain if the network is reasonably large. In the remainder of this section, we shall motivate a more biologically plausible network algorithm which is able to extract multiple principal components.

Consider a collection of neurons, labelled i = 1...m. Each neuron accepts an input vector **z** which it weights by a synaptic weight vector $\mathbf{s}_i$ (such that $||\mathbf{s}_i|| = 1$). Each neuron is inhibited by a feedback term f_i which comes from the other neurons in the network (the exact form of f_i will be specified later, but it is a weighted average of the responses of neurons). The response of the i-th neuron n_i is given by

$$n_i = h(\mathbf{s}_i.\mathbf{z} - f_i)$$

where h(.) is a half-wave rectification: h(x) = 0 if x < 0, otherwise h(x) = x. This is a simple integrate-and-fire neuron, which only fires when there is an excess of excitation over inhibition. We want the network responses $\{n_1, n_2, \ldots n_m\}$ to encode the inputs **z** as efficiently as possible. To maximise the information transfer, we would like to maximise $\sum n_i^2$. To minimise the redundancy of the encoding, we would like to minimise $\sum n_i f_i$, since f_i is a measure of the response of other neurons in the network. To optimise both simultaneously, we can maximise

$$P = \sum_i n_i^2 - \beta \sum_i n_i f_i$$

where β is a constant. We can do this by choosing an update rule where the increment $\Delta\mathbf{s}_k$ is proportional to $\partial P/\partial\mathbf{s}_k$. Following some simple algebra, we have

$$\frac{\partial P}{\partial \mathbf{s}_k} = (2n_k - \beta f_k)h'(\mathbf{s}_k\mathbf{z} - f_k)\mathbf{z} - \sum_i \left((2+\beta)n_i - \beta f_i\right)\frac{\partial f_i}{\partial \mathbf{s}_k}$$

where the derivative of the rectifier h'(.) is h'(x) = 0 if x ≤ 0, otherwise h'(x) = 1. Assuming that $\partial f_i/\partial\mathbf{s}_k$ is negligible, the update rule is

$$\Delta\mathbf{s}_k = \lambda(n_k - (\beta/2)f_k)\mathbf{z} \qquad \text{if } n_k > 0, \text{ otherwise } \Delta\mathbf{s}_k = 0.$$

where λ is the learning rate. The weight vectors $\mathbf{s}_k$ must be renormalised after every update. The rule works by comparing the neuron response n_k with the feedback term f_k, which quantifies the response of the rest of the network. When the response n_k exceeds the term $(\beta/2)f_k$, the weight vector $\mathbf{s}_k$ is reinforced moved closer to $\mathbf{z}$. When it does not, the weight vector is moved away from $\mathbf{z}$. Thus each neuron in effect competes with the rest of the network. Standard Hebbian learning only uses the first term in the update rule, $n_k\mathbf{z}$. The rule proposed above, since it allows for both reinforcement and attenuation of the weight vector, is a form of Hebb-Stent rule.

The rule was tested on a global principal components task. In this task, the feedback term f_k was simply the mean response of all neurons in the network, multiplied by a factor which was the regression of n_k on f_k. This can be implemented by adding a single inhibitory interneuron to the network, which sums the neural responses and feeds back on each neuron through a modifiable inhibitory synapse. The synaptic weight of this inhibitory synapse is the regression coefficient $\langle n_k f_k\rangle/\langle f_k^2\rangle$. With a value of $\beta = 2$, this network succeeded in finding all the principal components of the input vector, even when the variance ratio of highest to lowest component was 256:1. Clearly the algorithm, despite the approximations used to derive it, is effective in finding principal components, provided the value of f_k is selected appropriately.

4.2 Neural Nets and Local Principal Components

The aim of this study is to see if a simple unsupervised learning algorithm can demultiplex the luminance and chromatic information contained in P cell signals. From section 3.2, we have seen that luminance and colour information are contained in different local principal components of the retinal image, after filtering by P cells. Thus demultiplexing can be achieved by finding the local principal components. In section 4.1, we introduced a simple modification of Hebbian learning that can find many principal components of its inputs. In this section, we shall apply the algorithm to simulated P cell signals.

As previously, the retina was a ID structure, aligned with the scan-lines of a digitised RGB image. The L and M cone catches were computed as in section 3.2, except that this time the retina was 64 pixels long. The P cell receptive field had a centre σ of 0.7 pixels, though other widths were also tried with similar results. The neural network consisted of 64 neurons. Each was assigned a retinal position. The weight vector $\mathbf{s}_i$ of the i-th neuron was nonzero only within a small interval (9 pixels) around the neuron's retinal position. These synaptic weights were applied to simulated P cell responses from the retina to compute the neuron's response. The feedback term f_i for each neuron was computed from the local mean response of the network. The local mean response was computed by low-pass filtering the response array $\{n_1, n_2, n_3, \ldots n_{64}\}$ by a gaussian filter. Letting m_i be the local mean at position i, the value of f_i was set to

$$f_i = \mu_1 m_i + \mu_2(m_{i-4} + m_{i+4})$$

and μ_1 is the regression coefficient $\langle n_i m_i\rangle / \langle m_i^2\rangle$, and μ_2 similarly. This choice of f_i was motivated by three considerations. First, the response of the neuron n_i should be different from the responses of neurons close to it, hence f_i should contain a term due to m_i. Second, as a local representation of the image, the neuron response should be different from flanking neurons, hence f_i should contain a term due to m_{i-x} for some x. Finally, since we do not expect the near and far local means, m_i and $(m_{i-4} + m_{i+4})$, to be statistically related, each should have its own modifiable feedback synapse. A diagram of this network is shown in Figure 4.

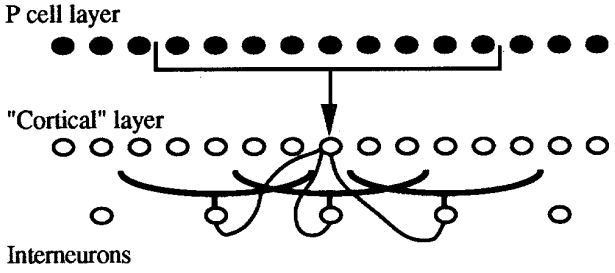


Figure 4. Diagram of the network. Each cell in the P cell layer "cortical" layer accepts inputs from 9 adjacent P cells. Interneurons average the responses of cortical units to compute the local mean response. These are used to feedback on each of the units in the network.

The network was run for 20000 iterations with $\beta = 2$ and a learning rate $\lambda = 1.0$. Afterwards, we examined the receptive fields developed by neurons in the network. Because of edge-effects at either end of the 64 pixel retina, only neurons in the centre 32 positions were looked at. The units could be categorised by inspection as luminance or chromatic sensitive. However, the segregation of the neurons into these two categories is more clearly displayed using a quantitative analysis. Each of the neural weight vectors s is a set of weights which are applied to P cell inputs coming from particular points on the retina. If the receptive field weight is predictable on the basis of the centre cone type (L or M) of the P cell at that point, the cell is probably colour sensitive. If the weight is not predictable on this basis, the cell is probably luminance sensitive. We computed the correlation coefficient R^2 between the receptive field weights and the cones at the corresponding retinal positions (using 0 to indicate an L cone, and 1 to indicate an M cone). The distribution of R^2 coefficients is shown in Figure 5a. Clearly, the neurons segregate into two groups: those with high values of R^2 are chromatically sensitive, while those with a low R^2 are luminance sensitive. Examples of three receptive fields (comparable to the principal components shown in figure 3) are shown in Figure 5, together with their R^2 values. Two of the receptive fields are luminance sensitive, and respond to changes in the luminance of the image. The third is a chromatic cell, which responds to a field of colour. These receptive fields are

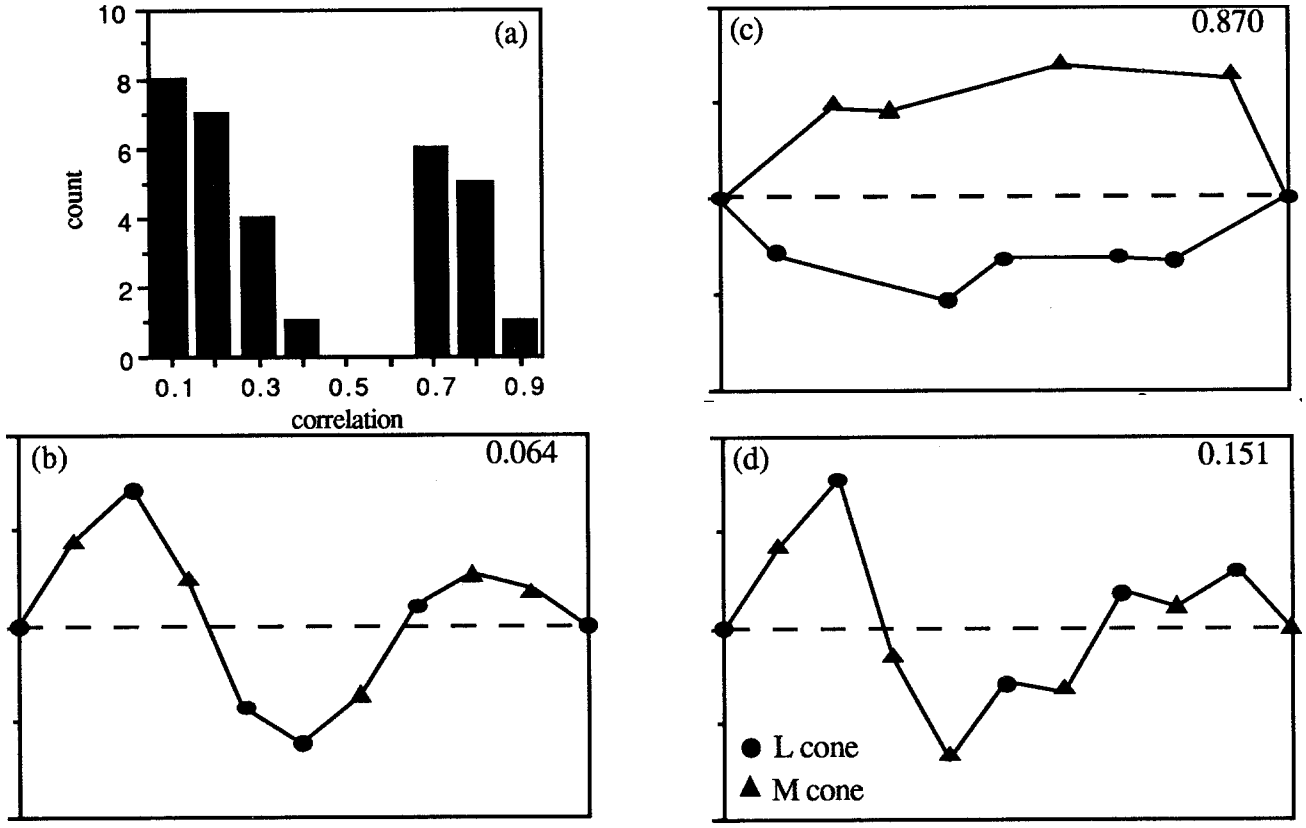


Figure 5. (a) The distribution of R^2 values for the neuron analysed. The neurons fall into two groups: one with small R^2 (less than 0.5) and those with large R^2 (over 0.6). These groups correspond to luminance and chromatic neurons. (b,c,d) Examples of the synaptic weights of neurons. (c) is a chromatic cell, (b,d) are luminance cells. The x axis gives the retinal position (relative to the neuron's position) and the y axis gives the synaptic weight. The number in the upper right corner of each frame is the neuron's R^2.

similar to those found by physiologists.[6] Neural nets which extract colour information from the P cell signal have been demonstrated previously.[17]

5. Conclusion

We have shown that a simple unsupervised neural network is capable of separating the signal from retinal P cells into luminance and chromatic components. The neural network implements a form of local principal components decomposition of the P cell signal. It is, in one important respect, superior to the LPC decomposition. The algorithm forces each neuron to be (relatively) uncorrelated with its neighbours. Under these circumstances, LPC components which only have to be represented at a low sampling rate will tend to be represented at a low rate in the network. Thus, even though the chromatic principal component in the case analysed in Figure 3 accounts for over 80% of the variance, less than half the units in the network (Fig. 5) are chromatic units. This is because the chromatic receptive field is effectively a low-pass filter, so its sampling rate can be correspondingly low.

The neural network approach also provides an interesting perspective on the evolution of a colour pathway. If the cortex is, indeed, a self-organising network, then the only prerequisites for a colour channel are the existence of at least some colour signal in the P cell inputs. If the P cells were already selected to have small receptive fields (for say the representation of fine detail in the retinal image) then they would be pre-adapted to convey a colour signal. Then, the only step needed to achieve colour vision is the mutation of one cone photopigment into a different kind (for example, the mutation of an L cone photopigment into an M-type).

There remain a number of deficiencies in the theory presented here. First, the LPC theory has no obvious way of controlling the sample rate of the components, nor of choosing the size of the window, or of dealing with multiple window sizes. All of these aspects are important, particularly when dealing with natural images, which have a scale-invariant statistical structure.[18] Second, the neural network theory does not provide us with a good way of choosing the exact form of the feedback term f_k to achieve any particular representation. Instead, we are currently forced to select it on various ad hoc grounds. Nevertheless, the algorithm is still quite effective in its performance.

6. Acknowledgments

W. McIlhagga is supported by the Human Frontiers Research Program.

7. References

1. R. M. Boynton, *Human Color Vision*, Holt Rinehart and Winston, New York, 1979.

2. G. R. Cole, T.J. Hine, and W. H. McIlhagga, "Detection mechanisms in L- M- and S-cone contrast space," *J. Opt. Soc. Am.* Vol. **A10** pp. 38-51, 1993.
3. K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *J. Physiol.* Vol. **359** pp. 381-400 1985.
4. P. Lennie, P. W. Haake, and D. R. Williams, "The design of chromatically opponent receptive fields", in *Computational Models of Visual Processing*, ed. M. S. Landy and J. A. Moyshon, pp 71-82, MIT Press, Cambridge, Mass 1991.
5. A.M. Derrington, J. Krauskopf, and P. Lennie, "Chromatic mechanisms in lateral geniculate nucleus of macaque," *J. Physiology* Vol. **357** pp 241-265.
6. M. Livingstone and D. H. Hubel, "Anatomy and physiology of a color system in the primate visual cortex," *J. Neuroscience*, Vol. **4** pp 309-356, 1984.
7. P. H. Schiller and N. K. Logothetis, "The color-opponent and broad-band channels of primate visual system," *Trends in Neurosciences*, Vol. **13** pp 392-398, 1990.
8. E. Martinez-Uriegas, "Spatiotemporal multiplexing of chromatic and achromatic information in human vision," *SPIE/ SPSE technical Conference on Human Vision*, Santa Clara, California, 1990.
9. P. Desarte, B. Macq, and D. T. M. Slock, "Signal-adapted multiresolution transform for image coding," *IEEE Trans. on Information Theory*, Vol. **32** pp. 897-904, 1992.
10. G. Buchsbaum and A. Gottschalk, "Trichromacy, opponent colours coding and optimum colour information transmission in the retina," *Proc. R. Soc. London* Vol **B220**, pp 89-113, 1983.
11. G. J. Burton and I. R. Moorehead, "Color and spatial structure in natural scenes," *Applied Optics* Vol. **26** pp 157-170, 1987.
12. L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Am.* Vol. **A3** pp 1673-1683, 1986.
13. E. Oja, "A simplified neuron model as a principal component analyser," *J. Math. Biol*. Vol. **15** pp. 267-273, 1982.
14. T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks* Vol. **2** pp. 459-479, 1989.
15. P. Foldiak, "Adaptive network for optimal linear feature extraction," *Proceedings IEEE/INNS International Joint Conference on Neural Networks*, IEEE Press, pp. 401-405, 1989.
16. J. Rubner and K. Schulten, "Development of feature detectors by self-organisation," *Biol. Cybern*. Vol. **62** pp. 193-199, 1990.
17. A. J. Ahumada, "Competitive learning and red-green opponency," *Invest. Ophthal. Vis. Sci.* (Supplement) Vol. **33** p 755 (1992)
18. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am*. Vol. **A4** pp 2379-2394, (1987).

published previously in SPIE, Vol. 2179, page 388