Statistics of cone responses to natural images: implications for visual coding

Daniel L. Ruderman

Sloan Center for Theoretical Neuroscience, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037

Thomas W. Cronin and Chuan-Chin Chiao

Department of Biological Sciences, University of Maryland–Baltimore County, Baltimore, Maryland 21228

Received December 11, 1997; revised manuscript received April 7, 1998; accepted April 9, 1998

We gathered hyperspectral images of natural, foliage-dominated scenes and converted them to human cone quantal catches to characterize the second-order redundancy present within the retinal photoreceptor array under natural conditions. The data are expressed most simply in a logarithmic response space, wherein an orthogonal decorrelation robustly produces three principal axes, one corresponding to simple changes in radiance and two that are reminiscent of the blue–yellow and red–green chromatic-opponent mechanisms found in the primate visual system. Further inclusion of spatial stimulus dimensions demonstrates a complete spatial decorrelation of these three cone-space axes in natural cone responses. © 1998 Optical Society of America [S0740-3232(98)02408-9]

OCIS codes: 0120.0280, 330.4060, 330.1720, 330.5310, 120.5630.

1. BACKGROUND

Theories of efficient coding in the visual system are of much current interest.[1–5] They all include the notion that the visual system is optimally designed to process a certain class of images—those that are found in the natural environment. The study of natural images has thus ensued,[6–11] with the ultimate goal of relating their properties to the function of the visual system. In the present study we focus our attention on natural images as represented in the first stage of retinal processing, the photoreceptor layer.

In this work we measure the spectral distributions of light present in natural images by using a hyperspectral camera,[12–15] which provides a complete spectrum at each pixel. We derive human cone responses at each spatial location from the spectra, and from these we gather cone response statistics for analysis. This approach is related to that of Webster and Mollon,[14] with the key difference that whereas they contrast the differences between various images, we study the ensemble statistics as averaged over images.

Our results are qualitatively similar to those of Buchsbaum and Gottschalk,[16] who sought to understand theoretically, by using model stimuli, how the visual system might decorrelate natural cone signals through an orthogonal linear transformation. They found that under certain conditions this can be achieved through a transformation to a luminancelike channel and a pair of blue–yellow and red–green opponent channels. We will provide a related treatment that demonstrates decorrelation by chromatic-opponent processing through the analysis of data from the natural image itself rather than data from a model.

In the present work we primarily focus on characterizing the cone response statistics themselves and not on predicting how higher levels in visual systems might best encode images. Although our results apply to other theoretical treatments of visual system function that include the important limitations caused by noise (e.g., Refs. 4, 8, and 17), we here consider primarily the redundancies present within the photoreceptor array and how they might be eliminated.

2. METHODS

Spectral images were captured by using an Electrim EDC-1000TE camera with a resolution of 192×165 (horizontal $\times$ vertical) 8-bit pixels. Light reaching the imaging CCD array was passed through an Optical Coating Laboratory, Inc., semicircular, variable-interference filter with a wavelength range of 400–740 nm and a half-bandpass typically of 15 nm at each wavelength. Wavelengths were selected under control of a portable computer by using a stepping motor. In each scanned image 43 successive images were taken of each scene at 7–8-nm intervals from 403 to 719 nm. A 16-mm lens (typically operated at *f*/4 plus or minus one stop) was used to image the scene onto the CCD, producing an angular resolution of $0.047 \text{ deg} \times 0.055 \text{ deg}$ per pixel (horizontal $\times$ vertical). Comparisons among images collected at different wavelengths showed no evidence of systematic magnification or registration errors within the resolution of the system. No corrections for optical or CCD-element spatial filtering were made.

We attempted to select a diversity of typical foliage-dominated scenes. Therefore images were collected in several locations, including the vicinity of Baltimore, Maryland (temperate woodland), and Brisbane, Australia (sclerophyll forest, subtropical rain forest, and mangrove swamp). Selected scenes contained numerous natural

objects, including leaf foliage, bark, rocks, herbs, streams, bare soil, etc. In one corner of each imaged scene, a pair of small, square reflectance standards was placed. Spectralon 100% diffuse reflectance material (Labsphere) was used as the white standard, and a nominally 3% spectrally flat diffuse reflector (MacBeth) was used as the black standard. The actual reflectances of these objects may have varied somewhat from that of the scene as a whole, depending on their local illumination, but they served as comparisons for the overall range of reflectances in the scene itself.

In each imaged scene, exposure was individually determined at each wavelength to adjust the brightness of the white standard to a level of 220–235 (of a possible 255). In other words, the computer noted the value of the standard's brightness and adjusted the exposure time accordingly to the correct level. Exposures were typically 250–5000 ms in length, and the time required for a complete scan was typically approximately 3–5 min. As each data frame was collected, the brightness levels of the white standard and the black standard, as well as the exposure time for each frame, were stored in a separate status file.

Once the series of 43 images was collected, a completely opaque aperture was automatically placed in front of the camera, and a series of images was collected with no illumination to determine the level of dark noise for each exposure series under the conditions of each individual data set. The image series was then corrected by subtracting each dark frame from each corresponding data frame in succession. Each of the images was calibrated by using the values of the small black and white standards measured therein. In the event of negative resulting intensity values, the minimum uncalibrated image pixel was used in place of the dark standard. The result is a series of 43 images, each containing a reflectance map of the scene at a given wavelength. A uniform-illumination spectrum throughout the image is assumed in these measurements. In fact, the images that we gathered did not, in general, contain deep shadows, and no attempt was made to correct for local variations in illumination.

We collected images of 12 such natural scenes and further analyzed the central 128×128-pixel region. Each of the ($128 \times 128 \times 12 = 196{,}608$) pixels was converted to three theoretical cone responses as $\Sigma_\lambda Q(\lambda)R(\lambda)I(\lambda)$, where $Q(\lambda)$ is the Stockman–MacLeod–Johnson cone fundamental[18] for the given cone type, $R(\lambda)$ is the measured image reflectance data, $I(\lambda)$ is the standard illuminant D65 (which is meant to mimic a daylight spectrum[19]), and the sum is over wavelengths represented in the spectrum. Our results depend only very weakly on the choice of illuminant, so long as it is broadband. This procedure provides the cone response data $L(\mathbf{x})$, $M(\mathbf{x})$, and $S(\mathbf{x})$, proportional to the number of quanta absorbed in an L, M, or S cone at spatial location $\mathbf{x}$ within the image. The raw reflectance data for the 12 images are available through anonymous ftp at ftp://ftp.sloan.salk.edu/pub/ruderman/hyperspectral/.

3. PHOTORECEPTOR RESPONSES AT ONE POINT

Each of our image pixels contains three numbers: L, M, and S. Two scatterplots of the data as projected onto the $L-M$ plane and the $L-S$ plane are illustrated in Fig. 1 (data in the $M-S$ plane look similar to those in the $L-S$ plane because of the large correlations present between L and M signals).

The data show two expected trends. First, signals between all pairs of cone type are strongly correlated: The data lie near the diagonals. This occurs because overall fluctuations in light intensity will tend to increase all cone responses simultaneously. Second, the correlation between the L and M photoreceptors is much larger than that between the L and S photoreceptors. This is primarily due to the large overlap of the L- and M-cone spectral sensitivities.

More surprising is the great deal of skew evident in the receptor response distributions. Since the mean of each response has been normalized to unity, the average of the distributions is at the coordinates (1, 1), which appears toward the lower left corner of the data clouds. The distributions are thus highly asymmetrical. Essentially, an identical amount of information to that present throughout the upper right area of the plots is compressed into the restricted space at the lower left.

A. Choosing a Coordinate System

Before proceeding with the analysis, we considered two compelling reasons for expressing our data in an im-

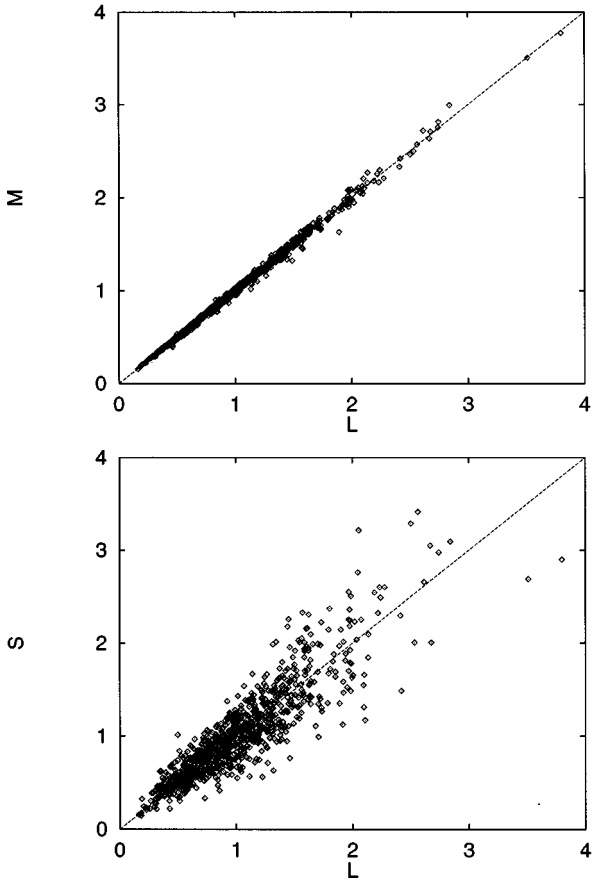

Fig. 1. Scatterplots of (top) L-versus-M and (bottom) L-versus-S data from 1000 pixels chosen at random from the image data set. The distributions show a high degree of correlation and asymmetry. Values are scaled so that the mean along each axis is 1.

proved space. First, the data show marked lack of symmetry. In other words, the intrinsic scale of a small volume of space changes systematically with location; many more data points occupy a given area near the origin than the same area further from it. Spreading out the data near the origin and compressing those away from the origin would tend to reduce this effect. Second, the space of relative photon catches necessarily includes only positive numbers, restricting the data to one octant of the space. Negative-going fluctuations from the mean thus have finite extent, whereas positive-going ones can have infinite extent. This lack of symmetry is inherent in the space rather than in the data.

A logarithmic transform will serve to improve greatly the coordinate space. First, the data will no longer be confined to only positive values. Second, the squeezing of data near the origin and its long-tailed extension at large values will both be remedied. We soon show that this improvement in symmetry is quite dramatic. Although a logarithmic transformation can be justified solely on the basis of an improvement of the data's appearance, it can also be related to results from psychophysics. The Weber–Fechner law states that uniform logarithmic changes in stimulus intensity tend to be equally detectable (see, for example, Ref. 20). Thus in a logarithmic space the perceptual noise level is uniform throughout. This law, however, does not hold universally (especially for small stimuli), so its role as a motivating factor for a logarithmic transformation is somewhat less than that of simply improving the data's symmetry.

The coordinate transformation is as follows. Each of the three quantal catches of a given image is converted to a logarithmic signal (base 10), whose mean is subtracted away:

$$\begin{aligned} \mathcal{L} &= \log L - \langle \log L \rangle, \\ \mathcal{M} &= \log M - \langle \log M \rangle, \\ \mathcal{S} &= \log S - \langle \log S \rangle. \end{aligned} \tag{1}$$

Here $\mathcal{L}$ is the new logarithmic signal for L, and $\langle \log L \rangle$ is the mean of $\log L$ over the image. By subtracting the mean, we are able to assess the responses without regard to an overall illumination level. This is analogous to a von Kries adaptation procedure, where each cone's set of responses is normalized independently.[19] Statistics of log-transformed natural images have been previously analyzed,[9,10] and are known to have second-order statistics similar to those of linear images.[21]

B. Rotation to Decorrelated Axes

In fixing the axes of the new logarithmic space, we considered only orthogonal transformations, rather than a more general linear transformation. This preserves the space's metric, which now has a fundamental meaning in terms of Weber thresholds. In a space whose axes have arbitrary scales, a more general transformation could equally well be applied.

Specifically, we took advantage of the benefits of decorrelation, which roughly means that an oblong data set is aligned along the coordinate system's axes. Data that are spherically distributed or that contain complicated structure are not so simply described, and decorrelation implies merely that there is no linear trend present in the data as plotted against any pair of axes. That is, if x_i is the ith coordinate of a data point where the axes are not correlated, then

$$\langle x_i x_j \rangle = \sigma_i^2 \delta_{ij}, \tag{2}$$

where $\langle \cdot \rangle$ is an ensemble average and σ_i is the standard deviation of the data as projected onto the ith axis. Such an orthogonal transformation is a standard data analysis technique known as principal-components analysis.[22]

This analysis also offers us some conceptual advantages. For instance, if the data occupy a compact cloud, without holes, then it may be reasonable to model them as a linear combination of uncorrelated sources. So in the original space whose axes are y_i, we can express any data point as

$$y_i = \sum_{j=1}^{N} A_{ij} x_j, \tag{3}$$

where A is the mixing matrix and N is the dimensionality of the space. The columns of the A matrix then have meaning as the basic directions of fluctuation within the data. If the data instead occupy the space sparsely, then higher-order correlations are implied between the x's, making the intuition less complete. Our data are of the former character, as we show in Subsection 3.C.

The analysis is straightforward and gives approximately the following three orthonormal principal axes:

$$\begin{aligned} \hat{l} &= \frac{1}{\sqrt{3}} (\hat{\mathcal{L}} + \hat{\mathcal{M}} + \hat{\mathcal{S}}), \\ \hat{\alpha} &= \frac{1}{\sqrt{6}} (\hat{\mathcal{L}} + \hat{\mathcal{M}} - 2\hat{\mathcal{S}}), \\ \hat{\beta} &= \frac{1}{\sqrt{2}} (\hat{\mathcal{L}} - \hat{\mathcal{M}}), \end{aligned} \tag{4}$$

where $\hat{\mathcal{L}}$, $\hat{\mathcal{M}}$, and $\hat{\mathcal{S}}$ are the unit direction vectors in the logarithmic cone response space. The actual coefficient values correspond very closely to these integer values and are shown in Table 1. Note that these axes are precisely those suggested as a convenient set by Flanagan *et al.*[23] (in an appendix). The standard deviations of the data along each of the three axes are $\sigma_l = 0.353$, $\sigma_\alpha = 0.0732$, and $\sigma_\beta = 0.00745$, a ratio of 47:9.8:1. These resulting principal axes and standard deviation ratios are largely insensitive to image rescaling, as has been found previously for other statistics of natural images.[10,14]

C. Interpretation of the Result

The three axes that we find have surprisingly simple forms: They sample the axes in the original cone space in integer ratios. Furthermore, their particular directions have easily interpreted meanings. The l axis, for instance, measures equal logarithmic fluctuations in all three cone catches, such as would occur when the scene's radiance changes solely in magnitude. This direction is commonly referred to as achromatic and is the type of fluctuation that dominates natural images: The light

Table 1. Three Principal Axes for Image Pixels Expressed in the ($\mathcal{L}$, $\mathcal{M}$, $\mathcal{S}$) Basis and the Standard Deviations of the Data As Projected onto These Three Axes

Vector[a]	$\mathcal{L}$	$\mathcal{M}$	$\mathcal{S}$	σ
$\sqrt{3}l$	1.004	1.005	0.991	0.353
$\sqrt{6}\alpha$	1.014	0.968	−2.009	0.0732
$\sqrt{2}\beta$	0.993	−1.007	0.016	0.00745

[a] These vectors correspond very closely to those of Eqs. (4).

level can vary widely from areas of bright illumination to those in shadow. Not surprisingly, it is the direction of largest variance.

The two other axes are also interesting. Both combine the cone signals in an opponent fashion. The α direction, with the larger variance, opposes an average of the $\mathcal{L}$ and $\mathcal{M}$ signals against $\mathcal{S}$. This is reminiscent of the blue–yellow opponent mechanism well known in both the physiology[24–27] and the psychophysics[23,28–30] literature, where short-wavelength signals are opposed to long-wavelength ones. However, these chromatic mechanisms are not uniquely defined and often depend on the type of experiment performed; see Refs. 31 and 32 for reviews. Finally, the β direction simply opposes the L- and M-cone signals, acting as a red–green channel, which is the other fundamental primate chromatic mechanism. Projecting these principal axes onto the isoluminant plane ($\mathcal{L} + \mathcal{M}$ constant) gives the principal axes $\mathcal{S}$ and $(\mathcal{L} - \mathcal{M})/\sqrt{2}$, showing that isoluminant fluctuations in S-cone responses are decorrelated with red–green fluctuations. Note that the standard deviation of the red–green axis is only 0.00745 log units, which corresponds to a rms fluctuation of approximately 1.7%. L- and M-cone responses are very highly correlated indeed.

These measurements are all relative to the mean values subtracted from the cone responses at the start [see Eqs. (1)]. Means corresponding to the chromatic channels take average values over the 12 images of $\langle\langle\alpha\rangle\rangle = 0.342$ and $\langle\langle\beta\rangle\rangle = 0.044$. A uniformly reflecting patch under illuminant D65 gives corresponding values of $\alpha = 0.178$ and $\beta = 0.041$. Thus our average foliage scenes are, not surprisingly, much more yellowish and somewhat more greenish than this nominally gray reflector. Of more importance, however, are their standard deviations, which are $\sigma_{\langle\alpha\rangle} = 0.059$ and $\sigma_{\langle\beta\rangle} = 0.0059$. These are comparable with the size of fluctuations within individual images. Under our fixed but artificial illuminant, the "gray-world" hypothesis—in which the average chromaticity of scenes is expected to be nearly constant—appears to fail. The von Kries type of adaptation that we have employed has thus been important to our comparison of chromatic fluctuations between different images.

The data are shown plotted in the logarithmic space in Fig. 2. Note first that the distributions are both highly elongated and compact in each of the three scatterplots. The orthogonal transformation has aligned the data's long dimensions with the coordinate axes. Also, since the data are compact and dense, considering them a linear superposition of independent—rather than just decorrelated—sources is not a bad approximation (this will not apply to some other statistics of natural images, which are known not to be produced through linear superposition).

Note that the data are distributed far more symmetrically in this new space than in the original, linear cone response space. One implication of this symmetry is that the choice of logarithmic coordinates has allowed us to achieve a degree of independence higher than the second order that we originally demanded. To demonstrate this, we have plotted the data in the original linear space in Fig. 3. Here again an orthogonal decorrelation results in an achromatic axis and two chromatic-opponent axes that are very nearly those of Eqs. (4), except acting on L, M, and S of the original space. The figure shows a scatterplot of the data in the $l-\alpha$ plane. Although the data are aligned to these axes, a symmetry along the l axis is en-

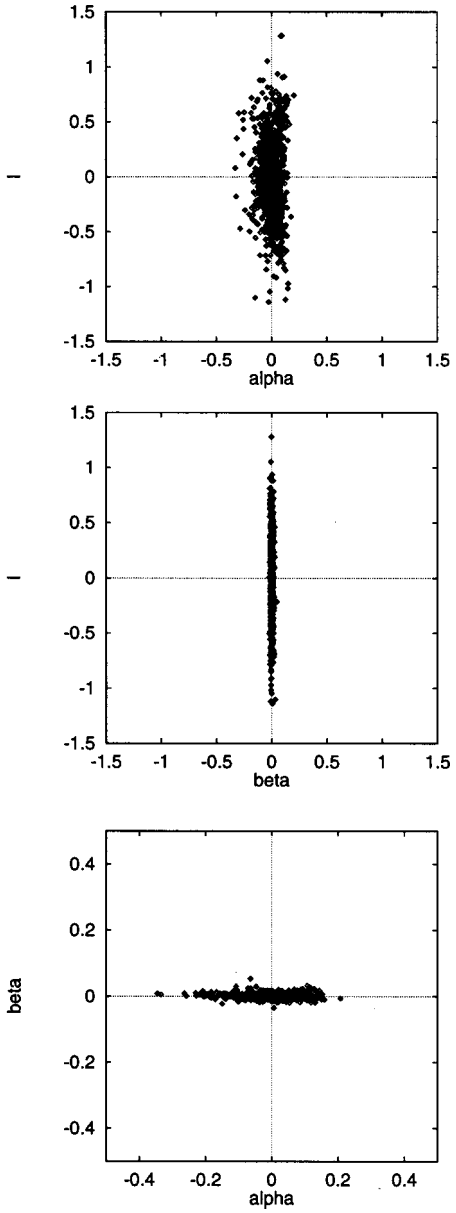

Fig. 2. Scatterplot of 1000 data points randomly selected as projected onto principal-axis pairs: (top) α versus l, (middle) β versus l, (bottom) α versus β. Note that the axes within each plot have the same scale, but the scales may change between plots.

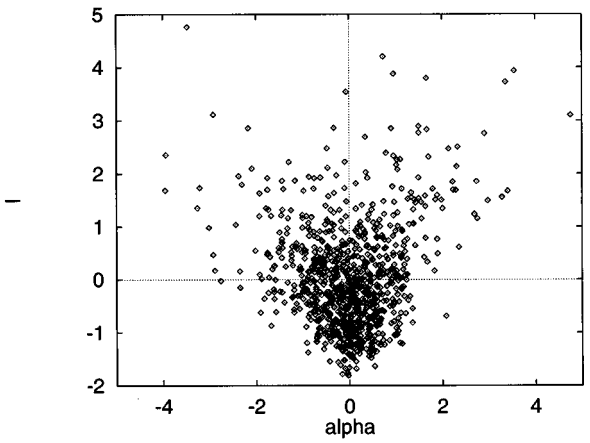

Fig. 3. 1000 datapoints selected at random as decorrelated in the linear (L, M, S) space, projected along the l and α axes (roughly equal to those of the logarithmic space). The standard deviation along the α axis is approximately proportional to l (data not shown), representing a third-order correlation in the data. These data have been normalized to unit variance along both axes.

tirely lacking. As l gets larger, so does the variance, or spread, of the data along the α axis. This higher-order dependence—between the variance of one variable and the value of another (a third-order correlation)—is absent in the logarithmic data. Since the logarithmic coordinate system seems naturally symmetrical, in a simple sense these coordinates are more fundamental than the originals. In Fig. 4 we further show the marginal distributions of the data as plotted along the l, α, and β axes in the logarithmic and original linear spaces. The figure demonstrates the much more nearly Gaussian and symmetrical distributions found in the logarithmic space.

In summary, orthogonally decorrelating our cone response data in a logarithmic space provides the following:

1. A set of three principal axes that encode fluctuations along an achromatic dimension, a yellow–blue opponent direction, and a red–green opponent direction.
2. A representation of the data that is compact and symmetrical.
3. An automatic decorrelation to higher than second order, unlike what we find in the original linear space.

4. INCLUSION OF SPATIAL DIMENSIONS

The preceding development has given us insight into the statistical structure of cone responses when viewing one location within an image. To extend our understanding to the spatial structure of cone responses across the retina, we now consider cone responses at different spatial locations in our statistics.

In such an analysis the dimensionality of the signal space increases as three times the number of pixels. A high-dimensional space requires vast amounts of data for statistical validity, so we limit ourselves to the relatively modest 27 dimensions of a 3×3-pixel lattice. By way of analysis we simply find the principal axes of such patches in our data set (nearly 2×10^5 of them) in the logarithmic coordinate system.

A. Orthogonal Decorrelation of Spatiochromatic Signals

As in Section 3, the result of the decorrelation is an orthogonal set of axes and the associated variances of the data along each of these axes. It is impossible to visualize fully either the data or the axes in this 27-dimensional space, so the presentation represents the principal axes as color-coded 3×3 image patches. They are shown in Fig. 5, where gray levels denote fluctuations in the achromatic, or l, direction, blue–yellow coloration denotes the α direction, and red–green coloration denotes the β direc-

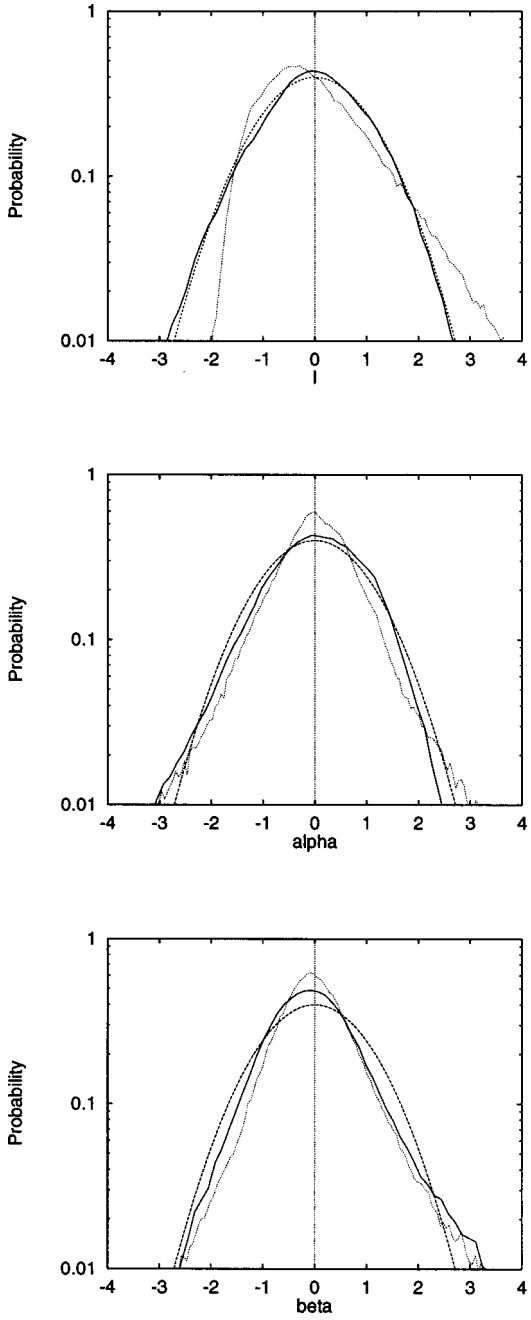

Fig. 4. Marginal distributions of our data along the three principal axes: (top) l axis, (middle) α axis, (bottom) β axis. Logarithmic space data (solid curves) and linear space data (dotted curves) are shown together for comparison. The dashed curves are unit-variance Gaussian distributions (all data have been rescaled to unit variance).

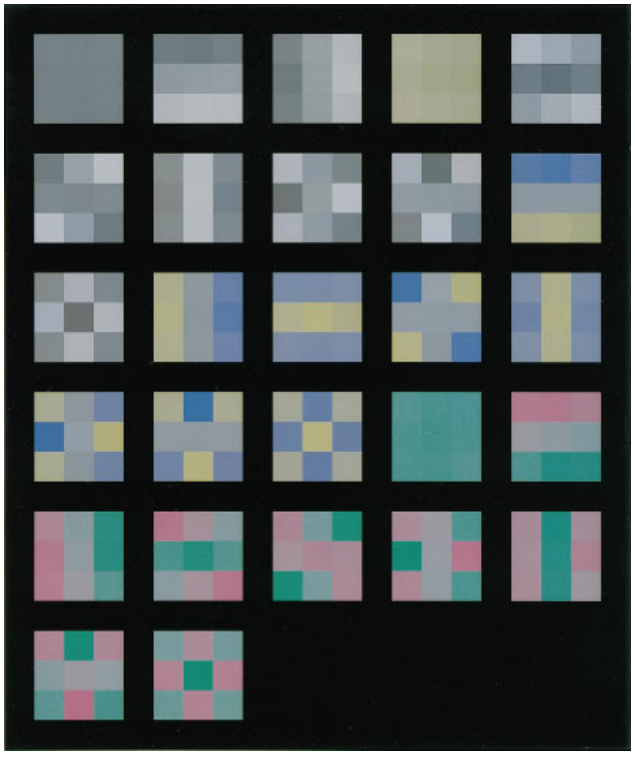

Fig. 5. Principal axes of 3 × 3-pixel chromatic patches arranged in order of decreasing eigenvalue from left to right and top to bottom. The color values (R, G, and B) for each pixel in this image were determined from the $\mathcal{L}$, $\mathcal{M}$, and $\mathcal{S}$ coordinates of each element through a direct linear correspondence [e.g., $R = \lfloor 128(\mathcal{L} + 1) \rfloor$, giving a range of 0–255]. For instance, a component that extracts only L-cone responses would contain pixels that range from reddish ($R = 255$; $G,B = 128$) to cyan ($R = 0$; $G,B = 128$).

tion (see the figure caption for details). The 27 principal axes are presented from left to right and top to bottom in order of decreasing projected signal variance.

These axes show striking regularities. First, although nothing in the analytical approach prevented it, no pixels within these patches appear as anything other than the primary gray, blue–yellow, or red–green colors described above. Purple or orange do not appear at any location, for instance. In other words, in these new principal axes no mixing of the previous l, α, and β axes is found. Furthermore, within any one patch there are *only* colors corresponding to a single one of these axes (or medium gray, which represents the origin). This is emphasized in Fig. 6, which shows the squares (variances) of the projections of each of the axes along the three principal directions. In each case nearly all of the projection is along a single cone-space axis. Thus the result for single pixels holds even when the space is greatly expanded to include 27 total dimensions. The l, α, and β axes remain fundamental.

The overall structure of the principal axes can be summarized as follows. For each of the principal directions in cone space, there are nine associated spatial dimensions. These dimensions are decorrelated through a nearly stereotypical set of principal axes, which are repeated for each cone-space direction. The spatial axes are largely symmetrical, and as the eigenvalue (variance) falls within any one class (i.e., l, α, or β), the spatial patterns take on more high-frequency detail. This is in fact what we expect from a translation-invariant ensemble such as natural images: Principal-components analysis is achieved through a Fourier transform. (Since we have a finite number of pixels, this statement is only approximate.) The fact that higher-spatial-frequency components have smaller variances corresponds well with the ubiquitous finding that power spectra of natural image ensembles are decreasing functions of spatial frequency (e.g., Refs. 7–9 and 11). Note also that, in general, the overall order of the variances decreases from l to α to β, just as in the single-pixel case.

B. Understanding the Result

What does this statistical structure reveal? Simply put, the three spatial signals $l(\mathbf{x})$, $\alpha(\mathbf{x})$, and $\beta(\mathbf{x})$, with $\mathbf{x}$ the spatial position, are entirely decorrelated with one another. This is true not only at a single pixel but at neighboring and presumably distant pixels as well. When moving from three to 27 dimensions, any spatial structure could have arisen, but in fact the simplest possible

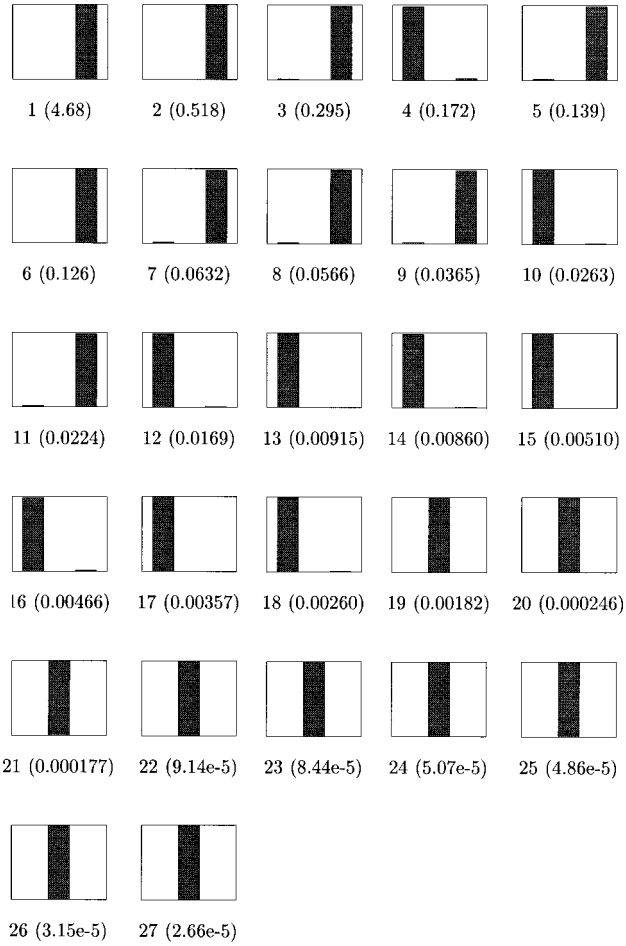

Fig. 6. Histograms showing the fraction of energy within each principal component, which represents each of the three principal axes (l at left, α at center, β at right). Each bar can range from 0 to 1 in value, a value of 1 meaning that all the energy resides along that single principal axis. The histograms show that all of the 27 components essentially lie along only a single axis. The numbers below the histograms are the component number and its eigenvalue (variance) in parentheses.

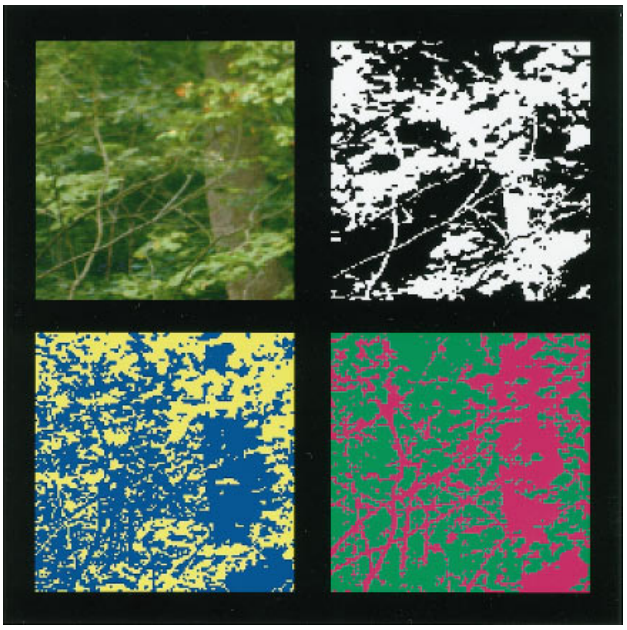

Fig. 7. Image from our data set (upper left) and its thresholded projection onto the three principal axes: l (upper right), α (lower left), β (lower right). The spatial patterns in each of these three projections are completely uncorrelated with one another.

one did. Natural cone responses can be thought of (at least to second order) as arising from three uncorrelated spatial processes that are superposed. This lack of correlation between spatial and chromatic dimensions has been previously indicated by van Hateren[33] (in the appendix) and Webster and Mollon.[14] It can be illustrated by presenting a natural image that has been dissected into these three fundamental components.

Figure 7 shows an image patch from one of our scenes together with its threshold projections onto the l, α, and β axes. For simplicity they are shown in white–black, yellow–blue, and red–green, respectively, where the threshold for each has been set to zero and the first named color of each pair corresponds to positive fluctuations. That these signals are decorrelated is reflected in the fact that the color in any one of the three thresholded images does not predict the color at the same location in either of the two others. Even though many of the most distinct changes occur across such well-defined locations as branch borders, the direction of the achromatic or brightness change, for instance, does not tell us how either blue–yellow or red–green will change at the same border. Note also that the structure in the chromatic channels seems to better betray the locations of object borders than does the achromatic image; changes in the latter are often dominated by variations in the illuminant rather than in reflectivity.

Finally, it should be noted that although the image values in the (l, α, β) axes are uncorrelated, this does not mean that they are entirely independent. They simply do not linearly predict one another. Clearly, object borders locate large absolute changes in all three values but not in predictable directions. A simple nonlinear transformation, such as squaring the differences across object borders, will show much more marked correlations. Large changes in one coordinate across location may predict large changes in the other coordinates, but this does not imply that they are linearly correlated.

C. Three Spatial Processes

A close inspection of Fig. 5 reveals that within any one cone-space class the sequence of spatial patterns of the principal axes in order of decreasing eigenvalue is identical. This suggests a very simple mechanism at work: Each principal axis is the product of a cone-space vector and a spatial vector. As such, each would take the form

$$p_{mn}(a, i) = c_m(a)s_n(i), \tag{5}$$

where $p_{mn}(a, i)$ is the principal axis associated with indices m and n (m ranging from 1 to 3 and n ranging from 1 to 9). The variables a and i are cone-space and spatial dimensions, respectively. Finally, the vectors c_m and s_n are the principal cone-space axes (l, α, or β) and the spatial frequency (Fourier-like) patterns, respectively. This form represents the overall spatiochromatic signal as the product of two uncorrelated signals, one spatial and the other chromatic.

If the above is true, then the variances associated with each axis should also take the form of a product:

$$\sigma^2_{mn} = \sigma^2_c(m)\sigma^2_s(n), \tag{6}$$

where σ^2_{mn} is the variance along the principal axis indexed by m and n, $\sigma^2_c(m)$ is the variance along the mth principal axis of the cone-space process, and $\sigma^2_s(n)$ is the variance along the nth principal axis of the spatial process.

We have strong reason to believe the form of Eq. (6). It can be shown that scale invariance of the image data set combined with translation invariance requires it.[34] To test the hypothesis, we examined the variances of the three cone-space channels as a function of the spatial pattern number (from 1 to 9). Figure 8 displays these on a semilogarithmic scale, so that for the hypothesis to be true the curves should be shiftable vertically into one another. The α and β chromatic axes obey the relationship

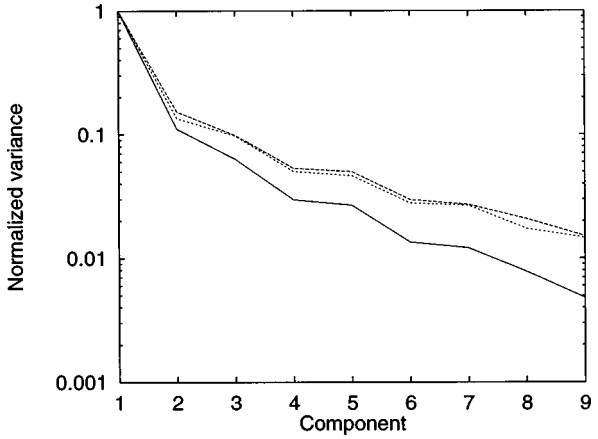

Fig. 8. Normalized variances along the principal axes in each of the three principal cone-space directions: l (solid curve), α (dashed curve), β (dotted curve). The order of the nine axes is as they appear in Fig. 5. The α- and β-eigenvalue spectra are nearly identical. The spectrum of l eigenvalues, which does not compare relative photon catches, differs significantly.

very well. But this is not so for the l axis, which shows a greater falloff with pattern number (spatial frequency). A similar circumstance is seen from the spatial power spectra of our images as projected onto the three principal cone-space axes (data not shown), where the two chromatic axes have an identical spectral shape and the l-axis spectrum decreases more steeply at high spatial frequency (they are more alike at the lower frequencies not represented in our 3×3 patches).[35] We suspect this to be a consequence of spatial filtering within the optical system, which manages to cancel out when logarithmic images are differenced (as in the chromatic axes).

5. DISCUSSION

Our main result is that there are three special directions in logarithmic cone response space along which natural response data are robustly decorrelated. Interestingly, these directions have a simple meaning: One varies in response to illumination changes (such as when leaves on a tree all face slightly different orientations relative to the incident sunlight), and the two others embody a blue–yellow and a red–green opponency. A further examination of spatial dimensions reveals that these three cone-space directions are entirely spatially uncorrelated with one another in natural, foliage-dominated scenes such as ours.

Could these results follow directly from the spectral tuning curves of the photoreceptors, with little or no influence of the spectra of natural images? To find out, we followed the approach of Buchsbaum and Gottschalk,[16] assuming an average radiance spectrum $I(\lambda)$ that is flat (i.e., equal power per unit wavelength) and that has uncorrelated fluctuations [i.e., $\langle \delta I(\lambda) \delta I(\lambda') \rangle = A \delta(\lambda - \lambda')$, with A being some constant]. Thus all the correlations present in the resulting cone responses will be due solely to their spectral overlaps. If the signal is taken as the fractional changes in cone response relative to its mean, then an orthogonal decorrelation yields the three channels and variances shown in Table 2. These are qualitatively similar to the principal axes found for logarithmically transformed data from real images. One axis responds to same-sign fluctuations in all cone types, one opposes the S cone to the average of the L and M cones, and one opposes L and M cones with little S-cone influence. Thus, in a qualitative sense, the three principal directions that we find can be predicted from the spectral tunings of the receptors (assuming well-behaved statistics for the incident spectrum), consistent with the general findings of Buchsbaum and Gottschalk.[16]

Nevertheless, the two sets of results have significant quantitative differences. First, the coefficients of Table 2 are not as nearly integer; the deviations are now larger. More importantly, the relative variances of the three channels are quite different; in particular, the two largest have nearly equal standard deviation (Table 2, last column), in contrast to the fivefold ratio obtained by using natural images. Furthermore, none of the higher-order statistics, such as skew and the marginal histograms of Fig. 4, are available in this simple approach. The use of natural images is necessary both to understand how each channel may contribute to the overall spectral information in the scene and to learn how spatial and spectral distributions are organized in the visual world.

Although it is tempting to take these results as predictions of how neurons in the visual pathway should optimally respond to images, we do not advocate so basic a task as orthogonal decorrelation for any stage of visual processing. Rather, we have set out simply to characterize the correlational relationships present in the signal responses themselves. A direct comparison with visual processing is made difficult. Most current models of opponent processing are linear, and there is little evidence for a truly logarithmic transformation within the visual pathway.

Furthermore, chromatic and spatial processing in primate vision does not neatly separate information into mutually decorrelated pathways. Although the primate visual pathway is roughly divided into a luminance, or magnocellular, pathway and a color, or parvocellular, pathway containing chromatic opponent neurons, they are not completely segregated. For example, parvocellular lateral geniculate nucleus neurons that are red–green opponent at low spatial frequencies show marked luminance responses at high spatial frequencies.[32] This basic character of result has been predicted by the theory of Atick *et al.*,[36] who include the effects of noise and the desire for equalized channel variance among neurons (see also the approach of Derrico and Buchsbaum[37]). Although our data are not meant to be predictive of neural function in and of themselves, they may be straightforwardly applied to theories of visual system design requiring second-order statistics of natural images.

In comparing our results with those of Buchsbaum and Gottschalk,[16] note first that ours derive from natural image data rather than the spectral white-noise model that they employed (although the later results of Moorhead showed that their transformation is similar to those derived from actual data[38]). Buchsbaum and Gottschalk showed that one generically expects to find chromatic opponency when orthogonally decorrelating cone responses. Moorhead[38] and Webster and Mollon[14] demonstrated that the precise form of the result depends strongly on which particular image is analyzed. Ours, applied to an entire ensemble of logarithmic data, is a simple transformation with near-integer values and two chromatic-opponent directions.

It is not entirely clear why one should be restricted to an orthogonal, rather than a more general, linear trans-

Table 2. Three Principal Axes in Linear Cone Response Space Assuming a Spectral Distribution That Is Equal Energy on Average and Has White-Noise Fluctuations That Are Independent at Each Wavelength[a]

Vector	L	M	S	σ
$\sqrt{3}l$	0.999	1.111	0.876	1.10
$\sqrt{6}\alpha$	0.902	0.854	−2.111	0.965
$\sqrt{2}\beta$	1.032	−0.966	0.050	0.0218

[a] Each cone-space axis (L, M, and S) has first been rescaled by its mean before decorrelation. Note the similarity to Table 1, the result for natural image data in a logarithmic space.

formation. In our case logarithmic space has a specific meaning in terms of fractional signal fluctuations. Further, maintaining this space's metric through the transformation requires it to be orthogonal. Buchsbaum and Gottschalk's restriction seems arbitrary, unless the particular gains used in their L-, M-, and S-cone spectral sensitivities have some special physiological meaning. In fact, their receptor primaries are those of Vos and Walraven,[39] where the gain factors are chosen to represent relative contributions to the psychophysical luminance function. Although this is relevant from the psychological viewpoint, it is not necessarily fundamental to the cone responses themselves. It would be interesting to see the result of a more general linear decorrelation, which would have a continuum of solutions rather than just one. We take this as a broader lesson: When applying principal-components analysis to a data set, one should have sufficient reason (such as a desire to preserve the metric) for excluding the larger class of nonorthonormal solutions.

In the logarithmic space it is also unclear whether uniform axis scaling is appropriate. It is known, for instance, that the Weber fraction for S cones is larger than that for L and M cones.[40] It might be sensible, therefore, to squeeze the $\mathcal{S}$ axis to reflect its ability to encode fewer levels of information, a refinement that would no doubt alter our results. This approach is taken by Webster and Mollon,[14] who scale their axes relative to the detection thresholds. We deliberately avoided consideration of noise as an undue complication to our treatment, since in estimating noise levels it is necessary to assume a particular receptoral integration time. Furthermore, the space that we used has a basic meaning in terms of fractional fluctuations in photon capture, which offers an intuitive and physical understanding. A more physiological approach to the data would include an accurate treatment of photoreceptor noise. In any case the principal axes that we find depend crucially on our choice of uniform logarithmic axis scaling.

Another limitation of our approach is an unrealistic spatial model of photoreception. For example, the retina does not have equal numbers of the different cone types, with the L and M cones far outweighing the S cones, especially in the fovea.[41] Also, chromatic aberration prevents the short-wavelength signals from achieving the same contrast as that of long-wavelength ones. This will tend to reduce the contrast of the signals along our α axis, possibly altering the resulting principal axes. These are obvious directions for further exploration.

Webster and Mollon[14] found a large variation in cone response distributions to different images. We present only a single distribution, which is an average over many images, similar to Webster and Mollon's Fig. 3. Our individual images also show significant variability in their distributions as well (data not shown), but, when averaged, they produce one that is localized and continuous (Fig. 2). Two differences in the approaches limit our ability to make a comparison. First, we plot data in a logarithmic space, which is not linearly related to theirs; thus a straightforward comparison does not have a simple interpretation. Second, and perhaps more important, the two image ensembles studied are quite different. Whereas theirs includes panoramic scenes including sky, ours concentrates on foliage. We have replicated their finding that the largest chromatic fluctuations are along a blue–yellow axis. We expect this effect to be even more dominant in their images, as they include blue sky, which we have avoided. Although the question of which scene type to select for analysis remains open, it is clear that various choices will have much larger effects on chromatic statistics than on spatial ones, which seem to be nearly universal.[21]

Finally, these data inform us about typical cone response fluctuations. The L and M cones, for instance, differ in their photon capture rates by only a couple of percent over the spatial scales that we sampled (up to approximately 0.05 deg). Thus visual judgments based on this cone difference signal will be reliable only under conditions of high photoreceptor signal-to-noise ratios or with considerable spatiotemporal averaging. Nevertheless, these small $L-M$ signals may be critically significant to foraging primates,[42] which suggests why these cones have substantially lower Weber fractions than that of the S-cone class.[40]

ACKNOWLEDGMENTS

We thank H. B. Barlow, D. J. Field, J. H. van Hateren, D. Heeger, D. I. A. MacLeod, D. Osorio, and M. A. Webster for helpful comments, and J. D. Pettigrew for hosting a data collection visit to the University of Queensland. This work was supported in part by National Science Foundation grant IBN-9413357 (to T. W. Cronin) and by a postdoctoral fellowship from the Alfred P. Sloan Foundation (to D. L. Ruderman).

REFERENCES AND NOTES

1. F. Attneave, "Some informational aspects of visual perception," Psychol. Rev. **61**, 183–193 (1954).
2. H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, W. A. Rosenblith, ed. (MIT Press, Cambridge, Mass., 1961).
3. M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: a fresh view of inhibition in the retina," Proc. R. Soc. London, Ser. B **216**, 427–459 (1982).
4. J. J. Atick and N. Redlich, "Towards a theory of early visual processing," Neural Comput. **2**, 308–320 (1990).
5. B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature (London) **381**, 607–609 (1996).
6. G. J. Burton and I. R. Moorhead, "Color and spatial structure in natural scenes," Appl. Opt. **26**, 157–170 (1987).
7. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).
8. J. H. van Hateren, "Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation," J. Comp. Physiol. A **171**, 157–170 (1992).
9. D. L. Ruderman and W. Bialek, "Statistics of natural images: scaling in the woods," Phys. Rev. Lett. **73**, 814–817 (1994).
10. D. L. Ruderman, "The statistics of natural images," Network **5**, 517–548 (1994).
11. D. W. Dong and J. J. Atick, "Statistics of natural time-varying images," Network **6**, 345–358 (1995).
12. N. G. Nagle and D. Osorio, "The tuning of human photopigments may minimize red–green chromatic signals in natu-

ral conditions," Proc. R. Soc. London, Ser. B **252**, 209–213 (1993).
13. G. Brelstaff, A. Párraga, T. Troscianko, and D. Carr, "Hyperspectral camera system: acquisition and analysis," in *Geographic Information Systems, Photogrammetry, and Geological/Geophysical Remote Sensing*, J. B. Lurie, J. Pearson, and E. Zilioli, eds., Proc. SPIE **2587**, 150–159 (1995).
14. M. A. Webster and J. D. Mollon, "Adaptation and the color statistics of natural images," Vision Res. **37**, 3283–3298 (1997).
15. C. A. Párraga, G. Brelstaff, and T. Troscianko, "Color and luminance information in natural scenes," J. Opt. Soc. Am. A **15**, 563–569 (1998).
16. G. Buchsbaum and A. Gottschalk, "Trichromacy, opponent colour coding and optimum colour information transmission in the retina," Proc. R. Soc. London, Ser. B **220**, 89–113 (1983).
17. D. L. Ruderman, "Designing receptive fields for highest fidelity," Network **5**, 147–155 (1994).
18. A. Stockman, D. I. A. MacLeod, and N. E. Johnson, "Spectral sensitivities of the human cones," J. Opt. Soc. Am. A **10**, 2491–2521 (1993).
19. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (Wiley, New York, 1982).
20. D. R. J. Laming, *Sensory Analysis* (Academic, London, 1986).
21. D. L. Ruderman, "Origins of scaling in natural images," Vision Res. **37**, 3385–3398 (1997).
22. I. T. Jolliffe, *Principal Component Analysis* (Springer-Verlag, New York, 1986).
23. P. Flanagan, P. Cavanagh, and O. E. Favreau, "Independent orientation-selective mechanisms for the cardinal directions of colour space," Vision Res. **30**, 769–778 (1990).
24. F. M. DeMonasterio and P. Gouras, "Functional properties of ganglion cells of the rhesus monkey retina," J. Physiol. (London) **251**, 167–195 (1975).
25. F. M. DeMonasterio, P. Gouras, and D. J. Tolhurst, "Trichromatic colour opponency in ganglion cells of the rhesus monkey retina," J. Physiol. (London) **251**, 197–216 (1975).
26. A. M. Derrington, J. Krauskopf, and P. Lennie, "Chromatic mechanisms in lateral geniculate nucleus of macaque," J. Physiol. (London) **357**, 241–265 (1984).
27. R. C. Reid and R. M. Shapley, "Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus," Nature (London) **356**, 716–718 (1992).
28. E. Hering, *Outlines of a Theory of the Light Sense* (Harvard U. Press, Cambridge, Mass., 1964).
29. D. Jameson and L. M. Hurvich, "Some quantitative aspects of an opponent-colors theory. I. Chromatic responses and spectral saturation," J. Opt. Soc. Am. **45**, 546–552 (1955).
30. J. Krauskopf, D. R. Williams, and D. W. Heeley, "Cardinal directions of color space," Vision Res. **22**, 1123–1131 (1982).
31. M. A. Webster, "Human colour perception and its adaptation," Network **7**, 587–634 (1996).
32. P. Lennie and M. D'Zmura, "Mechanisms of color vision," CRC Crit. Rev. Clin. Neurobiol. **3**, 333–400 (1988).
33. J. H. van Hateren, "Spatial, temporal and spectral preprocessing for colour vision," Proc. R. Soc. London, Ser. B **251**, 61–68 (1993).
34. Translation invariance implies that correlations depend only on pixel separations. From scale invariance the correlation matrix in the cone subspace is not a function of the pixel separation (up to an overall multiplicative factor). Thus the correlation matrix is of the form $F(x)C_{ab}$, where x is the pixel separation vector and a and b are index cone-space directions. Such a correlation matrix can be diagonalized through a Fourier transform in the variable x and a separate diagonalization of the matrix C (yielding the l, α, and β directions). The associated eigenvalues will be a product of the F (spatial) eigenvalues and the C (cone-space) eigenvalues, as in Eq. (6). This is analogous to two decorrelated multiplicative processes, one in real space and the other in cone space.
35. This invariance of the spectral shape is consistent with the hypothesis that spatial statistics in natural images are dominated by the size distribution of objects within the scenes.[21]
36. J. J. Atick, Z. Li, and A. N. Redlich, "Understanding retinal color coding from first principles," Neural Comput. **4**, 559–572 (1992).
37. J. B. Derrico and G. Buchsbaum, "A computational model of spatiochromatic image coding in early vision," J. Visual Commun. Image Represent. **2**, 31–38 (1991).
38. I. R. Moorhead, "Human color vision and natural images," in *Colour in Information Technology and Information Displays* (Institution of Electronic and Radio Engineers, London, 1985).
39. J. J. Vos and P. L. Walraven, "On the derivation of the foveal receptor primaries," Vision Res. **11**, 799–818 (1971).
40. W. S. Stiles, "Color vision: the approach through increment threshold sensitivity," Proc. Natl. Acad. Sci. USA **45**, 100–114 (1959).
41. D. R. Williams, D. I. A. MacLeod, and M. M. Hayhoe, "Foveal tritanopia," Vision Res. **21**, 1341–1356 (1981).
42. D. Osorio and M. Vorobyev, "Colour vision as an adaptation to frugivory in primates," Proc. R. Soc. London, Ser. B **263**, 593–599 (1996).