Spatiochromatic statistics of natural scenes: first- and second-order information and their correlational structure

Aaron P. Johnson, Frederick A. A. Kingdom, and Curtis L. Baker, Jr.

McGill Vision Research Unit, Department of Ophthalmology, 687 Pine Avenue West, Room H4-14, Montréal, Québec, Canada H3A 1A1

Received January 24, 2005; revised manuscript received May 4, 2005; accepted May 16, 2005

Spatial filters that mimic receptive fields of visual cortex neurons provide an efficient representation of achromatic image structure, but the extension of this idea to chromatic information is at an early stage. Relatively few studies have looked at the statistical relationships between the modeled responses to natural scenes of the luminance (LUM), red–green (RG), and blue–yellow (BY) postreceptoral channels of the primate visual system. Here we consider the correlations among these channel responses in terms of pixel, first-order, and second-order information. First-order linear filtering was implemented by convolving the cosine-windowed images with oriented Gabor functions, whose gains were scaled to give equal amplitude response across spatial frequency to random fractal images. Second-order filtering was implemented via a filter–rectify–filter cascade, with Gabor functions for both first- and second-stage filters. Both signed and unsigned filter responses were obtained across a range of filter parameters (spatial frequency, 2–64 cycles/image; orientation, 0–135°). The filter responses to the LUM channel images were larger than those for either RG or BY channel images. Cross correlations between the first-order channel responses and between the first- and second-order channel responses were measured. Results showed that the unsigned correlations between first-order channel responses were higher than expected on the basis of previous studies and that first-order channel responses were highly correlated with LUM, but not with RG or BY, second-order responses. These findings imply that course-scale color information correlates well with course-scale changes of fine-scale texture. © 2005 Optical Society of America

OCIS codes: 100.2960, 330.1720, 330.4060, 330.4270, 330.6110.

1. INTRODUCTION

The role of color vision in the analysis of image structure has been a topic of long-standing interest in the biological and computer vision science community (e.g., see recent review by Regan[1]). At issue is the extent to which color vision provides information about the spatiotemporal structure of the visual world, rather than just information about the color, or hue, of its surfaces. There are a number of approaches to this issue, and one that is increasingly influential involves the analysis of the statistical information in the chromatic content of natural scenes and relating this information to the known behavioral and physiological properties of the color vision system.[2–5] This approach extends studies examining the statistical content of achromatic natural images,[6–10] and is predicated on the widely held belief that the visual system has evolved to code the useful information in the visual environment in an optimally efficient manner.[2,6,10–18]

In daylight vision the retinal image is transduced by three photoreceptors, the L (long-wavelength-sensitive), M (middle-wavelength-sensitive), and S (short-wavelength-sensitive) cones. The cone outputs are then combined into three postreceptoral channels: a luminance-sensitive channel that sums the outputs of the L and M cones, here referred to as the LUM channel; a chromatic channel that differences the outputs of the L and M cones, commonly referred to as the RG channel; and a chromatic channel that differences the outputs of the S with the sum of the L and M cones, commonly referred to as the BY channel. These channels have been revealed through psychophysics[19–23] and may have a neurophysiological substrate in the early stages of the visual pathway.[24–26] Psychophysical studies have shown that chromatic mechanisms are spatially tuned for both orientation and spatial frequency,[20,27,28] supporting the idea that the channels have an early cortical representation.

What information about the visual environment do these channels convey, and how similar is the information conveyed in the three channels? Two approaches have been adopted in previous studies to address these questions. The first approach has been to examine the *differences* between the statistical properties of model LUM, RG, and BY responses to natural scenes—for example, the differences in the amplitude spectra slopes among the three channels.[3,4] The second approach has been to consider the *relationships* between the LUM, RG, and BY responses to natural scenes by looking at the similarities in channel responses to a given image via cross-correlation analysis.[2,5]

It has been recognized that one function of the postreceptoral channels is to decorrelate the cone signals, that is, to remove information in their outputs that is redundant by virtue of the close overlap in the cones' spectral sensitivities.[2,13] Ruderman *et al.*[2] and Fine *et al.*[5] found that the average cross correlation between the RG, BY, and LUM channel responses to a set of natural images

was close to zero: for example, there was no consistent tendency toward red pixels being bright rather than dark, or bluish pixels being greenish rather than reddish. However, the nonrelatedness of the LUM, RG, and BY channel outputs to natural scenes revealed by cross correlating their point responses raises the question of whether correlations exist at the level of higher-order structure. Fine *et al.* showed that if one measured the *difference* between neighboring pixels within each of the three channels and discarded the sign of the difference, the channel responses were positively correlated in inverse proportion to the distance between pixels. The reason for positive correlations in the unsigned pairwise pixel relationships is straightforward: objects in natural scenes invariably differ in both lightness and color, that is, in both intensive and spectral reflectance, and as a result simultaneous changes in both of these dimensions will occur at most object borders. We might not expect these correlations to be very high, however, because of the presence of shading and shadows, which are primarily luminance-defined features.[29–32] In addition, even through coincident changes in color and luminance occur within the image, for example, at the boundaries of objects, the magnitudes of those changes may be uncorrelated.

The absence of significant changes in color at most shadow borders will inevitably reduce the overall estimates of channel cross correlations in unsigned pixel differences. Nevertheless, the magnitude of the correlations measured by Fine *et al.*[5] seem surprisingly low, especially given that their test images were selected to be devoid of deep shadows: at 6 pixels separation, average unsigned pixel-difference cross correlations were 0.24 for LUM-with-RG, 0.22 for LUM-with-BY, and 0.17 for RG-with-BY—an overall mean cross correlation of 0.21. (6 pixels corresponded to the cited value of 18 arcmin for this result. The images were 128×128, so 6 pixels also corresponded to 4.7% of image width.) Moreover, these values were only slightly larger in absolute magnitude than the "signed" pixel-difference cross correlations, also surprising because in preserving the sign of each pixel difference, one would expect the resultant cross correlations, now negative as well as positive, to yield a net correlation of zero. One aim of the present study is to measure the between-channel correlations in pairwise structure for a much larger set of images, using an approach based on biologically realistic bandpass-filter responses rather than differences of adjacent pixel responses. To capture the spatially coincident nature of the attribute transitions, we have correlated the absolute, i.e., unsigned, as well as the signed filter responses.

What of between-channel correlations in image structure that are of higher order than those defined by pixel differences? An abundance of psychophysical[33–37] and single unit neurophysiology[38–40] supports the existence of visual cortex neurons sensitive to changes in dimensions other than luminance or color—for example, variations in contrast, element size, or orientation—here collectively termed "texture variations." Such texture variations, and the visual mechanisms that are believed to detect them, have typically been designated as second-order in the biological vision literature. Note that in a context of image statistics, such stimuli might be considered fourth-order, because the detection of *differences* between textures requires the comparison of at least four image points. However, following common usage we will refer to stimuli that vary in luminance or chromaticity (and that minimally require the comparison of two image points) as first-order and stimuli that vary in texture (and that require the comparison of four image points), as second-order. Analyses of unfiltered single points in the image (statistically speaking, first-order) will be referred to as pixel analyses.

Whereas first-order information in either luminance or color can be extracted by a single stage of linear filtering, most models of second-order processing use a filter–rectify–filter (FRF) cascade.[33,41,42] In this scheme the first-stage filter, which is typically bandpass in spatial frequency and orientation, detects fine-grain luminance or chromatic detail. The second-stage filter, which is also bandpass in orientation and spatial frequency (but with a lower center spatial frequency than its first-stage counterpart), detects large-scale variations in textural properties. This model is supported by both neurophysiology[43] and human psychophysics,[37,44–47] in which both the first- and second-stage filters are bandpass for spatial frequency and orientation. The intervening nonlinearity or "rectification" stage of FRF acts to demodulate the early-stage filtered signal. This has a consequence of making available to the second-stage filter the coarse-scale regional differences in the pattern of first-stage responses. Most FRF models implement the nonlinearity with a full-wave rectification,[41,42] although any nonlinearity will suffice[48,49] provided it contains even-order components.[50]

By modeling neurons that are spatially linear (i.e., that detect first-order properties) or operate in a FRF cascade (i.e., that detect second-order properties), it is possible to study the correlations between the lower- and higher-order responses to natural scenes[51,52] and by extension between the lower- and higher-order statistics of the LUM, RG, and BY images of natural scenes.

Here we use simulated neural images of the LUM, RG and BY postreceptoral channels to study their correlational relationships in response to a set of natural color images. We achieve this by simulating the responses of cortical neuronlike filters for linear first-order (linear, i.e., luminance and color-modulated) and nonlinear second-order (i.e., texture and contrast-modulated) information. We show that while correlations between unfiltered images sum to near zero, images filtered with either the linear or the FRF cascade show correlations that vary with spatial frequency. Finally, we consider how the results might relate to the behavioral and physiological properties of the primate visual system.

2. METHODS

A. Image Set

Eighty natural images, representing a variety of natural environments (forest, lake, mountains, meadow, flowers, and fruit), taken under a number of different illumination conditions (sunny and cloudy) and at a number of distances (0.5–1000 m), were obtained from the McGill Calibrated Colour Image database.[53] These images were acquired with a digital still camera (Nikon Coolpix 5700) and stored as uncompressed RGB TIFF files of dimension

1920×2560 pixels and 24 bits deep (256 levels for each R, G, and B image). All images were acquired using the camera's smallest aperture setting (f/11.4), which minimized within-image differences in focus for images with large depth fields.

B. RGB to LMS Conversion

The images were first gamma-corrected following measurement of the luminance response function of the cameras' RGB sensors. Each RGB digital image was converted to an LMS image, that is, an image modeled as the responses of the human L, M, and S cones. We provide a brief explanation of the calibration procedure here; more detailed documentation can be found on the database website.[53]

The luminance response function of each RGB sensor was measured as follows. A set of six gray Munsell papers were illuminated by an incandescent light driven by a constant-DC power supply and were photographed. The luminance of the light reflected from each paper was measured with a Topcon SR-1 spectroradiometer. The average R, G, and B pixel values were plotted against the corresponding measured luminances and fitted with gamma functions. These functions were used to gamma-correct the sensor outputs. In a separate laboratory setting, using a set of nine gray Munsell papers illuminated with a different light source and measured with a different photometer, we confirmed the linearity of the gamma-corrected camera outputs.[53]

The spectral sensitivities of the three camera sensors were measured by taking photographs of a white target made of cyanoacrylate powder through a series of narrowband optical interference filters spanning the range from 400 to 700 nm at 10 nm intervals. The resulting spectral sensitivity functions were then gamma-corrected and normalized to produce equal responses to a flat-spectrum light. The RGB images were then converted into human LMS cone images via a conventional linear matrix transformation whose weights were derived from the Smith and Pokorny[54] human cone spectral sensitivity functions and the measured spectral sensitivity functions of the camera RGB sensors.

C. LUM, RG, and BY Images

In the color vision literature, the postreceptoral channels are conventionally defined in terms of cone contrasts: ΔL/L, ΔM/M, and ΔS/S. For example, the RG channel is often modeled as ΔL/L−ΔM/M.[22,55–57] The denominator in each cone contrast term represents the level of cone adaptation and is typically calculated as the cone response to the luminance of the background upon which the stimulus is presented. These definitions are arguably appropriate for psychophysical experiments using simple stimuli such as gratings or patches presented briefly and at low contrast, where one can assume that the adaptation state of the cones is set by the background.

However, cone contrast definitions are arguably inappropriate for natural scenes where cone adaptation levels are set locally within the image and almost certainly vary across its extent. In particular, the RG and BY components of natural scenes modeled using cone contrast definitions with fixed denominators will pick up local changes in pure luminance, such as most shadows. This is because equal-ratio changes in the three cone responses at pure-luminance borders will translate into discontinuities in the cone-contrast-defined chromatic channels unless the luminance borders are symmetric around the assumed adaptation level. Considerations such as these have led us to use the following pixel-based definitions of color-opponent responses, following Parraga *et al.*[3,4] and Olmos and Kingdom[32]:

$$\begin{aligned} \mathrm{LUM} &= \mathrm{L} + \mathrm{M}, \\ \mathrm{RG} &= (\mathrm{L} - \mathrm{M})/(\mathrm{L} + \mathrm{M}), \\ \mathrm{BY} &= \mathrm{S} - \mathrm{LUM}/\mathrm{S} + \mathrm{LUM}. \end{aligned} \tag{1}$$

These channel images were cropped to 512 by 512 pixels and were windowed by using a cosine taper to reduce edge artifacts (see Fig. 1).

It should be pointed out that our model of the postreceptoral channels does not embody many important physiological properties of primate vision. First, we have not taken into account the differences in spatial resolu-

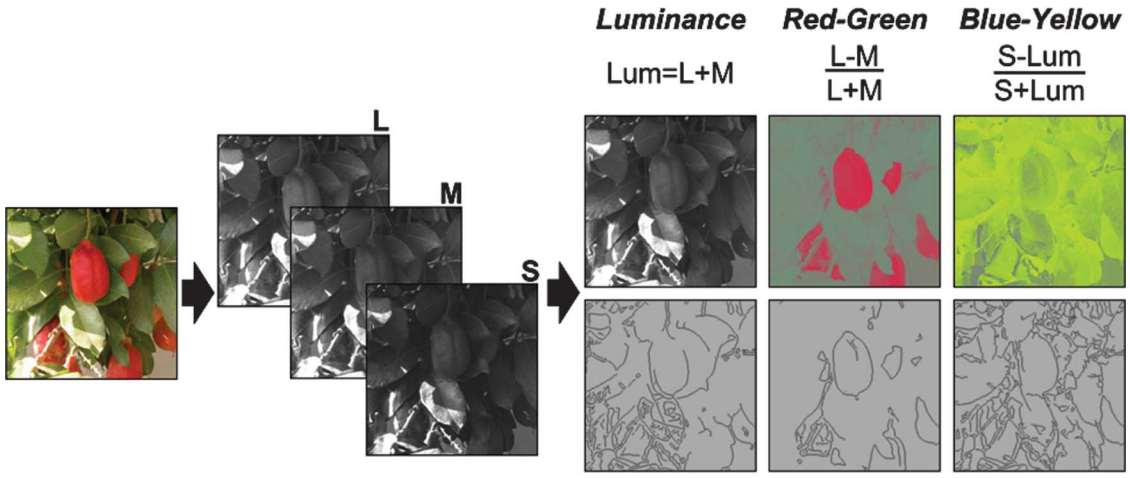

Fig. 1. Channel image representations. A typical RGB database image is first converted into L-, M-, and S-cone-sampled images, which are then combined to produce LUM, chromatic RG, and chromatic BY channel images. Note that channel values have been scaled to maximize available range. We also include the edge maps of each channel image to show the relationships among structural features.

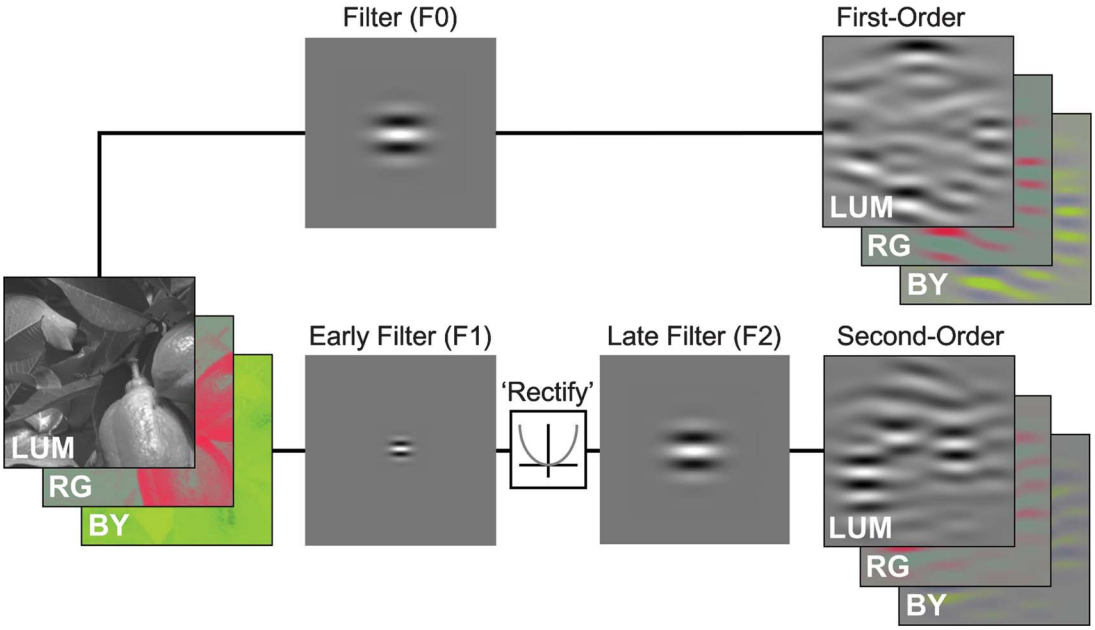

Fig. 2. Example of first-order model (top) and FRF cascade (bottom). The first-order model convolves each channel image with a single Gabor filter to detect luminance/chromatic variations. The FRF (second-order) model convolves each channel image with a high-spatial-frequency Gabor filter (F1) whose square rectified response is then convolved with a second Gabor (F2) having a lower spatial frequency to detect luminance/chromatic texture variations. In this example, the first-order (F0) and second-order (F2) filters (shown at magnified scale) have the same orientation, phase, and spatial frequency.

tion between the LUM, RG, and BY channels, in particular the poor resolution of the BY system.[58–60] Second, we have not modeled the effects of contrast gain controls that are known to occur at moderate and high levels of channel contrast.[61] Our channels must therefore be considered as idealized and our results therefore an approximation to channel performance.

D. Filter Convolution

The analysis of the first-order characteristics within each channel image was based on convolution with a Gabor filter (F0) designed to approximately model the spatial receptive field properties of area V1 neurons (Fig. 2, top). Two-dimensional Gabor functions[10,62] were generated by the following equations:

$$g_{\lambda\theta\phi\sigma}(x,y) = A\exp[(x^2+y^2)/2\sigma^2]\cos[2\pi(x/\lambda)+\Phi],$$

$$x = x\cos\theta + y\sin\theta,$$

$$y = -x\sin\theta + y\cos\theta, \qquad (2)$$

where λ is the spatial wavelength (reciprocal spatial frequency, in pixels), Φ signifies the phase offset of the sinusoidal carrier, σ specifies the width of the Gaussian envelope, and θ is the orientation of the Gabor. The filter gain ($A=1/\sigma^2$) was chosen to ensure equal amplitude responses across spatial frequency for a random fractal image.[10] A set of 48 Gabor filters were selected to cover the full range of orientations (θ=0°, 45°, 90°, 135°) and phases (Φ=0°, 90°), as well as a range of wavelengths (λ =8, 16, 32, 64, 128, 256 pixels), which correspond to spatial frequencies of 64, 32, 16, 8, 4, 2 cycles per image (cpi).

The second-order characteristics within each channel were analyzed with a FRF cascade, using Gabor spatial filters as in Eq. (2). The response image of the first-stage filter (F1) was subjected to a static nonlinearity, and then convolved with a second-stage filter (F2) (Fig. 2, bottom). The nonlinearity was an expansive power function (square law) defined by

$$[\text{Input of F2}] = [\text{Output of F1}]^2. \qquad (3)$$

The Gabor filters encompassed the full range of orientations as for first-order analysis, with the first- and second-stage filters having equal orientation. The wavelength (λ_{F1}) of the first-stage filter ranged from 8 to 64 pixels (spatial frequency 64–8 cycles per image), and λ_{F2} for the second-stage filter ranged from 32 to 256 pixels (spatial frequency 16–2 cpi). The first- and second-stage responses were averaged over filter phase, as previous studies have shown phase to have no significant effect on natural image responses over an ensemble of images.[49] To exclude biologically invalid combinations, the first-stage filter was constrained to higher spatial frequencies than the second-stage filter (i.e, $sf_{F2} < (sf_{F1})/2$).[43,49,52] Therefore the possible number of filter combinations within the FRF cascade is 96. This gives a total possible number of correlations between the first- and second-order filters of 4608 correlations per image.

E. Statistical Analysis of Filter Responses via Cross Correlation

Correlation coefficients were calculated between different first-order channel responses (e.g., LUM with RG) and between first- and second-order channel responses (e.g., first-order BY with second-order LUM or BY-with-LUM channels). The signed correlation (R) between filter response a and b was calculated as

$$R = \frac{\sum (a_{xy} b_{xy})}{\left[\sum (a_{xy}^2) \times \sum (b_{xy}^2)\right]^{0.5}}, \quad (4)$$

where a_{xy} and b_{xy} are the pixels at location x,y of the responses being compared. For the unsigned correlation, the absolute values of the filter responses (i.e., the full-wave rectified responses) were compared, which prevents cancellation between the positive and negative spatially coincident changes.[49] Correlations were calculated on

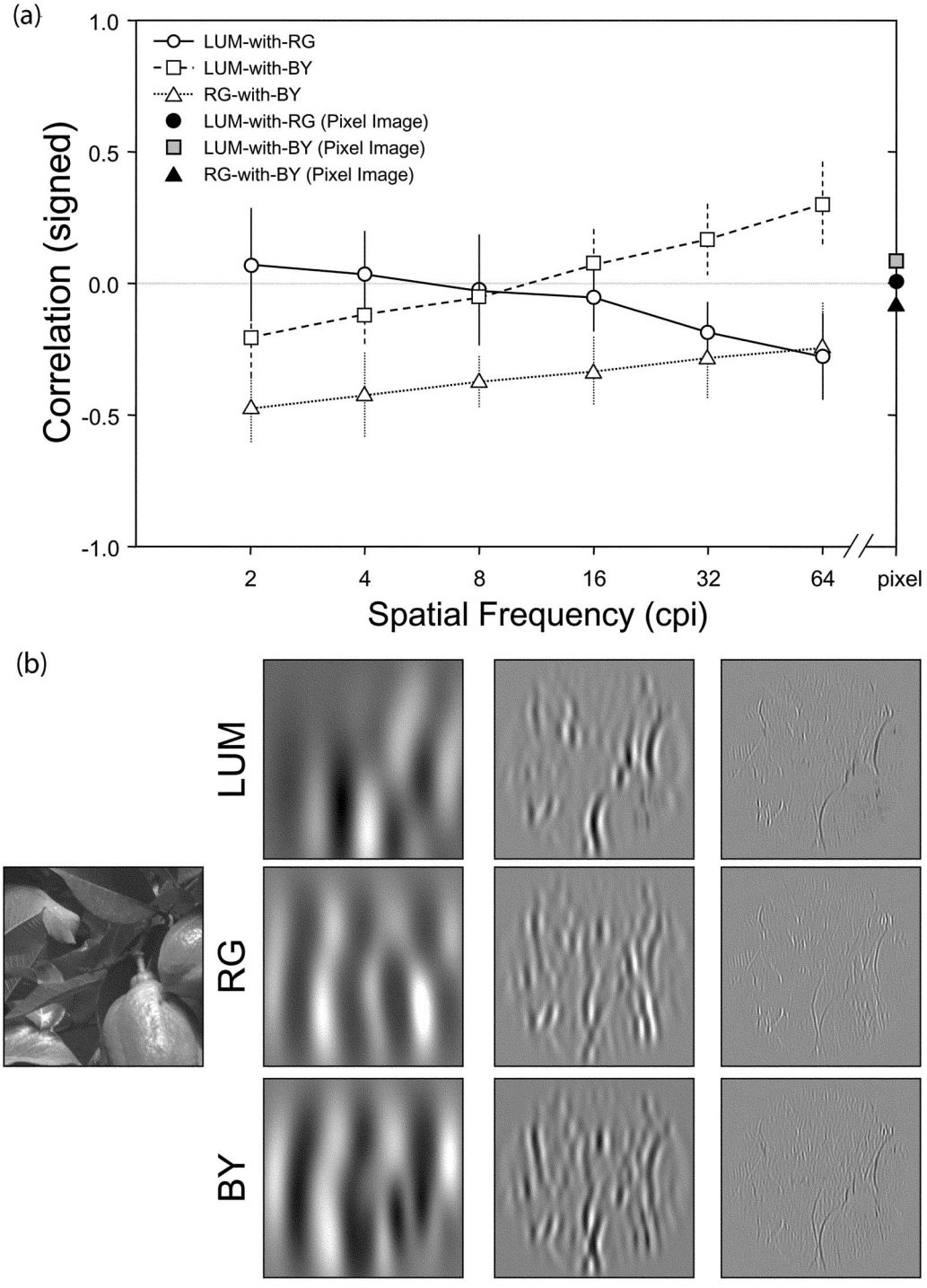

Fig. 3. Correlation between signed first-order channel responses as a function of spatial frequency (cycles per image), averaged over filter orientation and even/odd phase from an ensemble of natural scenes ($n=80$). (a) Circles correspond to correlations between LUM-with-RG channels, squares to LUM-with-BY, and triangles to RG-with-BY. For comparison, the solid symbols show the corresponding correlations for the raw pixels. Error bars denote the standard error across images and filter responses. (b) Sample of LUM, RG, and BY filter responses from a typical image ranging from low spatial frequency (left) to high spatial frequency (right), with the filter orientation of 0° (i.e., vertical) and odd-symmetric phase (90°).

each pair of filter responses and then averaged over the ensemble. The standard error was calculated from the standard deviation of the ensemble correlations.

3. RESULTS

A. Correlations between First-Order Filtered Channel Images

Signed correlations preserve the directionality of response and therefore range from −1 to +1. However, the sign of the correlation at any particular location in the image depends on the formula used for the chromatic channel. For example, defining the RG channel as (M−L)/(L+M), or "Green–Red," rather than (L−M)/(L+M), or "Red–Green," would reverse the sign of all RG correlations. Although the direction of the signed correlation is not significant in itself, it is important when considering the underlying physical cause of the correlation (see Discussion).

Figure 3(a) shows pixel correlations (the rightmost points) and signed first-order filter correlations as a function of filter spatial frequency, with filter responses averaged over all other filter parameters. The pixel correlations are all close to zero, as found in previous studies.[2,5] The signed filter correlations, however, are in many cases significantly different from zero, in particular at high spatial frequencies for the LUM-with-RG channels and LUM-with-BY channel correlations, and at all spatial frequencies for the RG-with-BY correlations.

The unsigned filter correlations, which can range from 0 to 1, are shown in Fig. 4. Remember that for these correlations, discontinuities such as bright-red to dark-green and dark-red to bright-green contribute positively to the overall correlation between LUM and RG. The average correlation across all channel combinations and all images is 0.42 (±0.034 S.E.). The results are similar at mid and high spatial frequencies (s.f. ≥ 8 cpi), although they show differences at lower spatial frequencies (s.f. ≤ 4 cpi). There were no significant differences in the results between different filter orientations and between even and odd filter phases (results not shown).

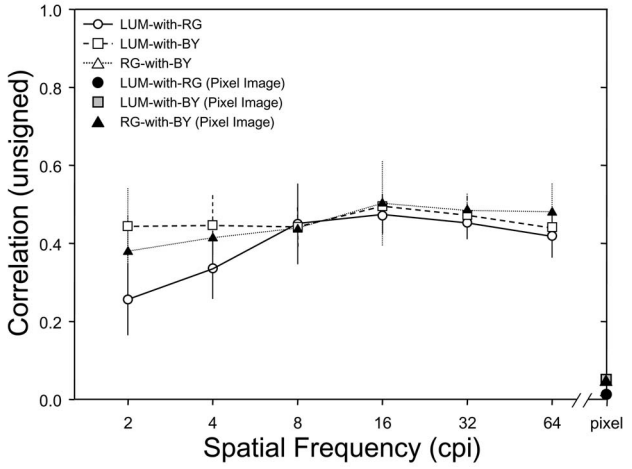

Fig. 4. Correlations between unsigned first-order channel responses as a function of spatial frequency (cycles per image), averaged over filter orientation and phase for the ensemble of images ($n=80$). Responses were full-wave rectified to remove the directionality of response. Circles correspond to correlations between LUM and RG channels, squares to LUM and BY, and triangles to RG and BY. Error bars denote the standard error across images and filter responses.

B. Correlation between First- and Second-Order Channels

Here we consider the correlations between first-order-filtered and second-order-filtered LUM, RG, and BY images. Again, the filter orientations have no effect as long as both filters (i.e., first-order linear filter and second-stage filter of FRF) have the same orientation; otherwise, there was a reduction in correlation, consistent with previous results for achromatic images.[49,51,52]

The correlations are plotted in Fig. 5 as a function of the spatial frequency ratio between responses of the first-order filter (F0) and the second-stage filter of the FRF cascade (F2). The signed correlations, shown in Fig. 5(a), are quite small: they range between 0 and 0.15 (±0.03 S.E.) and are similar to those obtained previously for purely achromatic images of natural scenes.[49] The correlations for other channel combinations (i.e., chromatic second-order) are effectively the same and so will not be shown here.

The results for the corresponding unsigned correlations are shown in Fig. 5(b). The correlations for the LUM images (circles) display a peak at a spatial frequency ratio (F0/F2) of 2:1, in keeping with our previous findings for achromatic images,[49] though if anything the dependence here is more peaked, perhaps owing to the larger number of images analyzed. The maximum correlation was 0.62, with the average correlation being 0.38 (±0.023 S.E.).

The first-order RG and BY images are also well correlated with the second-order LUM images and show a similar though less pronounced peak around F0/F2=2:1. The maximum correlation for RG-with-LUM is 0.53, with an average 0.38 (±0.01 S.E.), and the maximum correlation for BY-with-LUM is 0.45, with an average 0.36 (±0.009 S.E.). In all three plots, the correlations decline rapidly beyond the 2:1 ratio. In contrast, correlations between any of the first-order filtered images (LUM, RG, or BY) with second-order RG [Fig. 5(c)] or BY images [Fig. 5(d)] show little dependence on spatial frequency ratio and a slightly lower average of 0.33 (±0.003 S.E.). This would suggest that for certain second-to-first-order spatial frequency relationships, first-order information is more strongly correlated with second-order luminance than with second-order chromatic information.

4. DISCUSSION

To summarize:

1. Average pixel correlations between the raw channel images are all close to zero.
2. Signed first-order correlations are in many cases significantly nonzero, indicating systematic tendencies of some colors to be lighter or darker than others.
3. Unsigned first-order correlations are significantly positive.
4. Correlations between signed first- and second-order filter responses are close to zero.

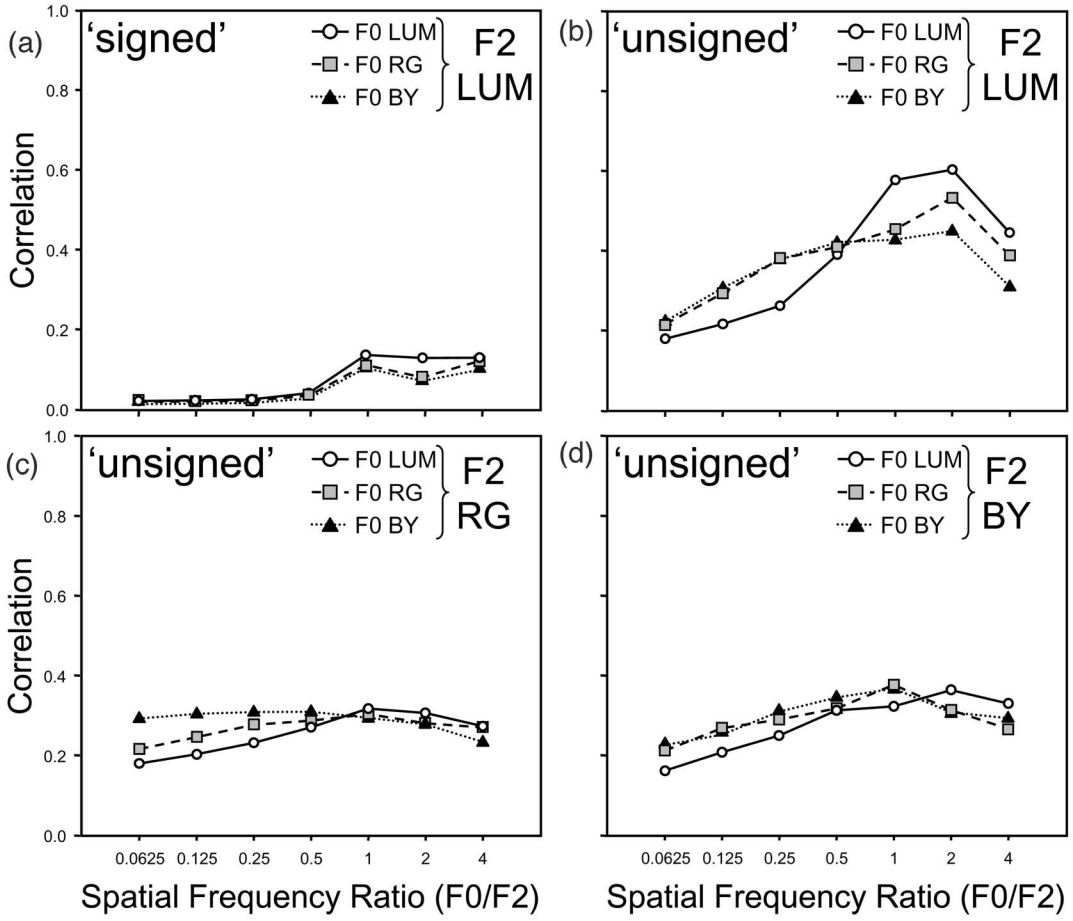

Fig. 5. Correlation between first-order (F0) and second-order (F2) responses as a function of spatial frequency ratio, averaged across the ensemble of natural scenes ($n=80$) and averaged across other filter parameters satisfying the constraint that the early-stage (F1) to late-stage (F2) of the FRF ratio equals 8:1 or 4:1 (optimal ratios of FRF[49] with similar response): (a) signed correlation of all three first-order channels (LUM, RG, BY) with second-order LUM; (b) unsigned correlation for all three channels with second-order LUM; (c) unsigned correlation with second-order RG; (d) unsigned correlation with second-order BY. Note that first-order chromatic channels correlate more strongly with second-order luminance channel than with either of the second-order chromatic channels.

5. Correlations between unsigned first-order (LUM, RG, or BY) and second-order LUM images peak at a first-to-second-order spatial frequency ratio of approximately 2:1.

6. Correlations between unsigned first-order (LUM, RG, or BY) and second-order RG and BY are less pronounced and largely independent of the first-to-second-order spatial frequency ratio.

The discussion will deal first with the pixel and signed first-order correlations, second with the unsigned first-order correlations, and finally with the first-with-second-order correlations.

A. Pixel and Signed First-Order Correlations

While some previous studies have found strong between-channel pixel correlations for particular subclasses of image, for example between BY and LUM channels for arid scenes[63,64] or between RG and LUM channels for reddish scenes,[65] our pixel-correlation results obtained from a large variety of images reinforce the conclusion that overall, the postreceptoral channels of the human visual system decorrelate the cone responses to natural scenes.[2,13]

On the other hand, our results with Gabor-filtered images reveal significant, albeit for the most part modest, between-channel correlations, with a pronounced spatial frequency dependence. The fact that none of these signed correlational relationships is observed in the pixel correlations reinforces the importance of looking at filter correlations across a range of spatial scales.

Inspection of individual channel images provides hints as to the possible cause of these correlations. For example, in Fig. 3(b), the low-frequency energy in the image is due primarily to the presence of the red fruit against the green foliage. Because the fruit is relatively bright and the foliage dark, there is a signed, positive correlation between LUM and RG for this image. The positive LUM-with-RG correlation found across the ensemble of images shown in Fig. 3(a) is probably due to the presence of similar object relationships throughout our image set.

The spatial-frequency-dependent pattern of correlation between LUM and BY [Fig. 3(a)] might arise from shadows. It is often stated that shadows tend to be bluish,[4,66] because they are relatively more illuminated by blue skylight. However, the reverse situation appears to hold for the shaded regions of dense foliage. Inspection of the BY

response in Fig. 1 and the other images in our set that contain foliage show that the shaded parts of the foliage are shifted slightly toward the yellow end of the BY spectrum, i.e., are less bluish. The reason for this may be that the top surface of many leaves have a specular reflectance component, and the daylight they reflect will have a higher proportion of short wavelengths. On the other hand, light emanating from the shaded regions of foliage will have undergone multiple absorptions and reflections and will tend to have a narrower spectrum centered around green. Hence the light emanating from the more shaded regions in foliage will stimulate more the yellow end of the BY mechanism. Thus at the higher spatial frequencies in dense foliage, the relatively yellowish fine-grained shadows might give a negative LUM-with-BY correlation; in contrast the blueish object shadows at low spatial frequencies would give a positive LUM-with-BY correlation.

Finally, the RG-with-BY filter correlations reveal that red tends to be associated with yellow, and green with blue, across a range of spatial frequencies. This might be because sky tends to stimulate the blue end of the BY and the green end of the RG channel, while the ground tends to stimulate the yellow end of the BY and the red end of the RG channel.

B. Unsigned First-Order Correlations

As pointed out in the Introduction, pixel-based correlation measures do not capture many of the visible structural similarities between channel images, and neither will signed filtered-channel correlations because their positive and negative correlations will often cancel when averaged. Unsigned filter correlations, however, do capture much of the visible structure because discontinuities that are coincident across channels will be picked up irrespective of their individual signs. When averaged across a range of filter spatial frequencies and orientations and across the three types of channel cross correlation, our unsigned cross-channel correlation was 0.42. This is more than double the value obtained by Fine *et al.*[5] for between-channel correlations of pixel differences at six pixel separations (see Introduction). Why the difference in results? The first possibility concerns the metric used to define the channels. Fine *et al.*[5] used Ruderman *et al.*[2] metrics for the RG, BY, and LUM channels, which were defined in terms of combinations of logarithmically transformed cone responses. This differs from the color metric we employ, particularly with regard to the LUM channel (see Methods). To test whether the difference in metric could account for the difference in mean channel cross correlation, we measured the mean channel cross correlation for a sample of ten images using both our metric and that employed by Ruderman *et al.*[2] The values obtained were 0.47 (±0.093 S.E.) for our metric and 0.49 (±0.075 S.E.) for Ruderman *et al.*'s metric, suggesting that the difference in metric was not the cause of the discrepancy.

The most likely cause of the discrepancy between our results and those of Fine *et al.* is the difference in method of measurement. We employed an array of Gabor filters, across a range of scales and orientations, which respond more specifically to the locally oriented structure of edges and contours in natural images than randomly chosen pairs of pixels, while their Gaussian envelopes will make them less sensitive to uncorrelated pixel noise. We cannot rule out, however, that the discrepancy could be due to our use of a different (and/or larger) set of images.

Recent neurophysiological findings bear upon the significant unsigned correlations we have observed between the chromatic and luminance components of images. Johnson *et al.*[67] found that of those cells in macaque V1 that responded to color contrast, most also responded to luminance contrast. Both simple cells, which are sensitive to contrast sign, as well as complex cells, which are insensitive to contrast sign, were found with these properties. Moreover, the spatial frequency and orientation tuning of these "color-luminance" cells was similar in both dimensions. Johnson *et al.*[67] suggested that the color-luminance cells would be especially suitable for detecting object boundaries, where color and luminance contrasts are spatially coincident. Note that the high correlation between unsigned channel responses does not imply that color-luminance cells should be mainly insensitive to contrast sign. Our reason for analyzing unsigned responses was to *reveal* the presence of significant amounts of coincident discontinuities between the chromatic and luminance channels that would otherwise tend to be lost due to cancellation between positive and negative correlations in the averaging process. More recently, Horwitz *et al.*[68] also found color-luminance cells in macaque V1; and, interestingly, of those sensitive to S-cone-isolating stimuli (i.e., the BY chromatic direction), many were sensitive to the sign of S-cone, but not luminance, contrast. Again, we have observed significant unsigned correlations between BY and LUM, and it is these coincident discontinuities in BY and LUM that would be expected to elicit strong responses in these cells.

Does the presence of significant first-order unsigned correlations imply a degree of redundancy in the structure of the LUM, RG, and BY components of natural images? The answer is yes, but it is a redundancy that is exploited by the human visual system. Chromatic borders, and those luminance borders that are aligned with them, generally arise from changes in surface reflectance, whereas luminance borders that are nonaligned with chromatic borders generally arise from shadows and shading.[29–32] Recent studies have indicated that the visual system has knowledge of these relationships and uses this knowledge to help segment the scene into its "intrinsic images,"[69] that is, into reflectance and illumination layers, and to facilitate tasks such as shape-from-shading and shadow identification.[66,70,71] In terms of the present study, unsigned LUM and chromatic (particularly RG; see below) discontinuities that are aligned and that therefore contribute to an overall positive cross orrelation are likely interpreted by the visual system as reflectance changes, whereas unsigned LUM discontinuities that are nonaligned and that therefore do not contribute to the cross correlation are likely interpreted as shadows and shading.

C. First-with-Second-Order Correlations

The positive unsigned correlations between LUM first- and second-order responses supports the findings of our

previous study dealing with achromatic natural scenes.[52] Moreover, the 2:1 first-to-second-order spatial frequency ratio that produced the highest correlation (0.63) echoes analogous findings in neurophysiology[43] and psychophysics.[37,44–46] The results here, however, show that in natural scenes texture borders are often aligned with chromatic as well as with luminance borders.

The relationship between textural (i.e., second-order), luminance and chromatic variations in the retinal image of natural scenes is complex. Some textural variations, in particular those defined by changes in local orientation, can result from changes in surface orientation,[72,73] and these will sometimes be accompanied by luminance changes due to shading[74] but not by chromatic changes since these generally arise only from changes in surface reflectance. On the other hand, some texture borders arise from changes in surface material, and these will often be accompanied by changes in both luminance and color. On these grounds we might therefore expect the unsigned second-order LUM images to be more highly correlated with the unsigned first-order LUM images than with the unsigned first-order RG and BY images. Our results do indeed show such a difference in the peak-ratio first-order- to second-order-LUM correlations. These correlations are ordered LUM > RG > BY, though the differences are not large (0.62, 0.53, and 0.45, respectively). The lack of a peak-ratio correlation between first-order and second-order chromatic responses is consistent with the idea that at fine scales, chromatic information is too noisy to convey information about texture but instead provides information about the broad-scale structure within a scene. Conversely, the luminance channel does convey information about the fine-scale structure in the scene.

ACKNOWLEDGMENTS

This work is funded by a grant from the Natural Sciences and Engineering Research Council of Canada (OPG0001978) given to Curtis L. Baker, Jr., and a Canadian Institute of Health Research grant given to Frederick A. A. Kingdom (MOP-11554). We thank T. Troscianko and A. Parraga (University of Bristol, UK) for lending us their facilities for calibrating the cameras and for help with the calibration. We are grateful to A. Olmos at McGill Vision Research for providing us with the images and to M. Webster and the anonymous referees for helpful comments.

Corresponding author Aaron Johnson can be reached by e-mail at aaron.johnson1@mcgill.ca or by fax at 514-843-1691.

REFERENCES

1. D. Regan *Human Perception of Objects* (Sinauer, 2000).
2. D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," J. Opt. Soc. Am. A **15**, 2036–2045 (1998).
3. C. A. Parraga, G. Brelstaf, T. Troscianko, and I. R. Moorehead, "Color and luminance information in natural scenes," J. Opt. Soc. Am. A **15**, 563–569 (1998).
4. C. A. Parraga, T. Troscianko, and D. J. Tolhurst, "Spatiochromatic properties of natural images and human vision," Curr. Biol. **12**, 483–487 (2002).
5. I. Fine, D. A. MacLeod, and G. M. Boynton, "Surface segmentation based on the luminance and color statistics of natural scenes," J. Opt. Soc. Am. A **20**, 1283–1291 (2003).
6. D. J. Field, "Relations between the statistics of natural images and the response profiles of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).
7. D. J. Field, "What is the goal of sensory coding," Neural Comput. **6**, 559–601 (1994).
8. D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," Ophthalmic Physiol. Opt. **12**, 229–232 (1992).
9. A. van der Schaaf and J. H. van Hateren, "Modeling the power spectra of natural images: statistics and information," Vision Res. **36**, 2759–2770 (1996).
10. D. J. Field and N. Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," Vision Res. **37**, 3367–3384 (1997).
11. H. B. Barlow, "Single units and sensation: a neuron doctrine for perceptual psychology?" Perception **1**, 371–394 (1972).
12. H. B. Barlow, "Redundancy reduction revisited," Network Comput. Neural Syst. **12**, 241–253 (2001).
13. 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).
14. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" Vision Res. **37**, 3311–3325 (1997).
15. B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," Curr. Opin. Neurobiol. **14**, 481–487 (2004).
16. M. G. A. Thomson, "Visual coding and the phase structure of natural scenes," Network Comput. Neural Syst. **10**, 123–132 (1999).
17. M. G. A. Thomson, "Higher-order structure in natural scenes," J. Opt. Soc. Am. A **16**, 1549–1553 (1999).
18. M. G. A. Thomson, "Beats, kurtosis and visual coding," Network Comput. Neural Syst. **12**, 271–287 (2001).
19. J. Krauskopf, D. H. Williams, and D. W. Heeley, "Cardinal directions of color space," Vision Res. **22**, 1123–1131 (1982).
20. A. Bradley, E. Switkes, and K. DeValois, "Orientation and spatial frequency selectivity of adaptation to color and luminance gratings," Vision Res. **28**, 841–856 (1988).
21. K. R. Gegenfurtner and D. C. Kiper, "Contrast detection in luminance and chromatic noise," J. Opt. Soc. Am. A **9**, 1880–1888 (1992).
22. M. J. Sankeralli and K. T. Mullen, "Postreceptoral chromatic detection mechanisms revealed by noise masking in three-dimensional cone contrast space," J. Opt. Soc. Am. A **14**, 906–915 (1997).
23. K. T. Mullen and M. J. Sankeralli, "Evidence for the stochastic independence of the blue-yellow, red-green and luminance detection mechanisms revealed by subthreshold summation," Vision Res. **39**, 733–745 (1999).
24. M. Livingstone and D. H. Hubel, "Anatomy and physiology of a color system in the primate visual cortex," J. Neurosci. **4**, 309–356 (1984).
25. P. Lennie, "Color coding in cortex," in *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe, eds. (Cambridge Univ. Press, Cambridge, UK, 1999), pp. 235–247.
26. E. N. Johnson, M. J. Hawken, and Robert Shapley, "The spatial transformation of colour in the primary visual cortex of the macaque monkey," Nat. Neurosci. **4**, 409–416 (2001).
27. R. A. Humanski and H. R. Wilson, "Spatial-frequency adaptation: evidence for a multiple channel model of short-wavelength-sensitive cone spatial vision," Vision Res. **33**, 665–675 (1993).
28. K. T. Mullen and M. A. Losada, "The spatial tuning of color and luminance peripheral vision measured with notch filtered noise masking," Vision Res. **39**, 721–731 (1999).
29. J. M. Rubin and W. A. Richards, "Color vision and image

intensities: When are changes material?" Biol. Cybern. **45**, 215–226 (1982).
30. P. Cavanagh, "Vision at equiluminance," in *Vision and Visual Dysfunction Volume V: Limits of Vision*, J. J. Kulikowski, I. J. Murray, and V. Walsh, eds. (CRC Press, 1991), pp. 234–250.
31. M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," in *Advances in Neural Information Processing Systems*, (NIPS, MIT Press, 2003), Vol. 15, pp. 1343-1350.
32. A. Olmos and F. A. A. Kingdom, "A biologically-inspired algorithm for separating reflectance and shading images," Perception **33**, 1463–1473.
33. C. Chubb and G. Sperling, "Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception," J. Opt. Soc. Am. A **5**, 1986–2007 (1988).
34. T. Ledgeway and A. T. Smith, "Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision," Vision Res. **34**, 2727–2740 (1994).
35. A. T. Smith and T. Ledgeway, "Separate detection of moving luminance and contrast modulations: fact or artefact?" Vision Res. **37**, 45–62 (1997).
36. A. Sutter, G. Sperling, and C. Chubb, "Measuring the spatial frequency selectivity of second-order texture mechanisms," Vision Res. **35**, 915–924 (1995).
37. A. J. Schofield and M. A. Georgeson, "Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour," Vision Res. **39**, 2697–2716 (1999).
38. T. D. Albright, "Form-cue invariant motion processing in primate visual cortex," Science **255**, 1141–1143 (1992).
39. A. Chaudhuri and T. D. Albright, "Neuronal responses to edges defined by luminance vs. temporal texture in macaque area V1," Visual Neurosci. **14**, 949–962 (1997).
40. C. L. Baker, Jr., "Central neural mechanisms for detecting second-order motion" Curr. Opin. Neurobiol. **9**, 461–466 (1999).
41. C. Chubb and M. S. Landy, "Orthogonal distribution analysis: a new approach to the study of texture perception," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds. (MIT Press, 1991), pp. 291–301.
42. H. R. Wilson, V. P. Ferrera, and C. Y. O, "A psychophysically motivated model for two-dimensional motion perception," Visual Neurosci. **9**, 79–97 (1992).
43. I. Mareschal and C. L. Baker, Jr., "Temporal and spatial response to second-order stimuli in cat A18," J. Neurophysiol. **80**, 2811–2823 (1998).
44. N. Graham, A. Sutter, and C. Venkatesan, "Spatial-frequency- and orientation-selectivity of simple and complex channels in region segregation," Vision Res. **33**, 1893–1911 (1993).
45. F. A. A. Kingdom and D. R. T. Keeble, "On the mechanism for scale invariance in orientation-defined textures," Vision Res. **39**, 1477–1489 (1999).
46. N. Graham and S. Wolfson, "A note about preferred orientations at the first and second stages of complex (second-order) texture channels," J. Opt. Soc. Am. A **18**, 2273–2281 (2001).
47. N. Prins and F. A. A. Kingdom, "Orientation- and frequency-modulated textures at low depths of modulation are processed by off-orientation and off-frequency texture mechanisms," Vision Res. **42**, 705–713 (2002).
48. N. Graham and A. Sutter, "Spatial summation in simple (Fourier) and complex (non-Fourier) channels in texture segregation," Vision Res. **38**, 231–257 (1998).
49. A. P. Johnson and C. J. Baker, Jr., "First- and second-order information in natural images: a new view of what second-order sees," J. Opt. Soc. Am. A **21**, 913–925 (2004).
50. Y.-X. Zhou and C. L. Baker, Jr., "Spatial properties of envelope responses in area 17 and 18 of the cat," J. Neurophysiol. **75**, 1038–1050 (1996).
51. A. J. Schofield, "What does second-order vision see in an image?" Perception **29**, 1071–1086 (2000).
52. A. P. Johnson and C. J. Baker, Jr., "Sparse coding in first- and second-order filtered images," J. Vision **4**, 542 (2004).
53. A. Olmos and F. A. A. Kingdom, "McGill Calibrated Colour Image database," http://ego.psych.mcgill.ca/labs/mvr/database/database.html (2004).
54. C. Smith and J. Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500nm," Vision Res. **15**, 161–171 (1975).
55. C. Norlander and J. J. Koenderink, "Spatial and temporal discrimination ellipsoids in color space," J. Opt. Soc. Am. **73**, 1533–1543 (1983).
56. C. F. Stromeyer III, G. R. Cole, and R. E. Kronauer, "Second-site adaptation in the red-green chromatic pathways," Vision Res. **25**, 219–237 (1985).
57. G. R. Cole, T. Hine, and W. McIlhagga, "Detection mechanisms in L-, M-, and S-cone contrast space," J. Opt. Soc. Am. A **10**, 38–51 (1993).
58. K. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," J. Physiol. (London) **359**, 381–400 (1985).
59. N. Sekiguchi, D. R. Williams, and D. H. Brainard, "Aberration-free measurements of the visibility of isoluminant gratings," J. Opt. Soc. Am. A **10**, 2105–2117 (1993).
60. A. B. Poirson and B. A. Wandell, "The appearance of colored patterns: pattern-color separability," J. Opt. Soc. Am. A **12**, 2458–2471 (1993).
61. M. D'Zmura and B. Singer, "Contrast gain control," in *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe, eds. (Cambridge Univ. Press, Cambridge, UK, 1999), pp. 369–385.
62. J. G. Daugman, "Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," J. Opt. Soc. Am. A **2**, 1160–1169 (1985).
63. G. J. Burton and I. R. Moorhead, "Color and spatial structure in natural scenes," Appl. Opt. **26**, 157–170 (1987).
64. M. A. Webster and J. D. Mollon, "Adaptation and the color statistics of natural images," Vision Res. **37**, 3283–3298 (1997).
65. J. Golz and D. I. A. MacLeod, "Influence of scene statistics on colour constancy," Nature **415**, 637–640 (2002).
66. F. A. A. Kingdom, S. Rangwala, and K. Hamammji, "Chromatic properties of the colour-shading effect," Vision Res. **45**, 1425–1437 (2005).
67. E. N. Johnson, M. J. Hawken, and R. Shapley, "The spatial transformation of color in the primary visual cortex of the macaque monkey," Nat. Neurosci. **4**, 409–416 (2001).
68. G. D. Horwitz, E. J. Chichilnisky, and T. D. Albright, "Blue-yellow signals are enhanced by spatiotemporal luminance contrast in Macaque VI," J. Neurosci. **93**, 2263–2278 (2005).
69. H. G. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *Computer Vision Systems*, A. R. Hanson and E. M. Riseman, eds. (Academic, 1978), pp. 3–26.
70. F. A. A. Kingdom, "Colour brings relief to human vision," Nat. Neurosci. **6**, 641–644 (2003).
71. F. A. A. Kingdom, C. Beauce, and L. Hunter, "Colour vision brings clarity to shadows," Perception **33**, 907–914 (2004).
72. K. A. Stevens, "The line of curvature constraint and the interpretation of 3D shape from parallel surface contours," in *Natural Computation*, W. Richards, ed. (MIT Press, 1998), pp. 107–114.
73. D. Knill, "Contour into texture: information content of surface contours and texture flow," J. Opt. Soc. Am. A **18**, 12–35 (2001).
74. P. Mamassian and M. S. Landy, "Interaction of prior visual constraints," Vision Res. **41**, 2653–2668 (2001).