

# Surface segmentation based on the luminance and color statistics of natural scenes

Ione Fine

Department of Psychology, Box 0109, 9500 Gilman Drive, University of California, San Diego,  
San Diego, California 92093-0109

Donald I. A. MacLeod

Department of Psychology, 0109, 9500 Gilman Drive, University of California, San Diego,  
San Diego, California 92093-0109

Geoffrey M. Boynton

The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 93037-1099

Received January 29, 2002; revised manuscript received September 12, 2002; accepted September 25, 2002

The luminance and color of surfaces in natural scenes are relatively independent under certain linear transformations, with the luminance of a surface providing little information about the color of that surface, and vice versa. However, *differences* in luminance between two locations in a natural scene remain strongly associated with *differences* in color. We used the statistics of the spatiochromatic structure of natural scenes as the priors for a Bayesian model that decides whether or not two points within an image fall on the same surface. This model provides a biologically plausible algorithm for surface segmentation that models observer segmentations well. © 2003 Optical Society of America

OCIS codes: 330.1690, 330.1720, 330.4060.

## 1. INTRODUCTION

A good deal of the variation in spectral power reflected off natural objects under natural illuminants can be represented by using three orthogonal axes that represent variation roughly along luminance, red–green, and blue–yellow directions in color space,<sup>1,2</sup> and these variations are very roughly independent.<sup>3</sup> Though it has been suggested that human visual systems might represent luminance and color (both red–green and blue–yellow) information independently,<sup>3,4</sup> the extent to which the human visual system represents luminance and color information within independent cardinal axes remains unclear.<sup>5–9</sup> This may be partly because even when the absolute luminance of a surface provides very little information about the absolute color of that surface, *differences* in luminance between two pixels are quite likely to be associated with *differences* in color, and vice versa. Differences in luminance and color are strongly interdependent, though not linearly correlated.<sup>9</sup> Two pixels that fall on the same surface are likely to have the same luminance and color, and two pixels that fall on different surfaces are likely to differ in both luminance and color.

In this paper we examine these statistics governing *differences* in luminance and color between neighboring pixels in natural scenes. We find, first, that for pixels of a given separation ( $r$ ), differences in luminance ( $\delta_l$ ) and color ( $\delta_{rg}$  and  $\delta_{by}$ ) are not independent: A change in luminance predicts a change in color, and vice versa. We also show that differences in luminance and color between nearby pixels can be modeled by assuming that the

probability of belonging to the same surface,  $p(\text{same}|r)$ , decreases (roughly according to an exponential function) with the distance between the two pixels. A Bayesian model based on these statistics segments natural scenes similarly to human observers.

## 2. METHODS

We used a set of 12 images of natural scenes, the ones recorded by Ruderman *et al.*<sup>3</sup> As described in their paper, the images were taken with an Electrim EDC-100TE camera. Light reaching the imaging CCD display was passed through a variable interference filter with a wavelength ( $\lambda$ ) range of 400–740 nm and a half-bandwidth that was typically 15 nm at each wavelength. To create each image, 43 successive data frames were taken at 7–8-nm intervals from 403–719 nm. Images were collected from a variety of natural environments such as temperate woodland, subtropical rain forest, and mangrove swamp. In the corner of each scene (in a region of the image excluded from our data set) was placed a pair of small rectangular white and black reflectance standards with known reflectance functions. 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. Each of the data frames was then calibrated by using the values of the small black and white standards within the frame. When this procedure resulted in negative intensity values, the minimum uncalibrated pixel within the

data frame was used in place of the dark standard. No attempt was made to correct for local variations in illumination. It should be noted that this data set contained very few deep shadows. These images are available at <ftp://ftp.sloan.salk.edu/pub/ruderman/hyperspectral>.

The three cone responses to each pixel were derived by using  $\sum_{\lambda} Q(\lambda)R(\lambda)I(\lambda)$ , where  $Q(\lambda)$  is the Stockman–MacLeod–Johnson spectral sensitivity of the given cone type,<sup>10</sup>  $R(\lambda)$  is the measured spectral reflectance and  $I(\lambda)$  is the standard illuminant D65 (meant to mimic a daylight spectrum); the sum is over the wavelengths represented in the spectral data. Each pixel in the image was therefore represented by three numbers representing the excitations of each of the three human cones at that pixel's position in space. We then transformed the data into Ruderman–Cronin–Chiao<sup>3</sup> co-ordinates by converting each cone output into a logarithmic signal (base 10) and subtracting the logarithmic mean. These logarithmic outputs ( $L$ ,  $M$ , and  $S$ ) were then transformed:

$$l = \frac{1}{\sqrt{3}}(L + M + S), \quad rg = \frac{1}{\sqrt{2}}(L - M),$$

$$by = \frac{1}{\sqrt{6}}(L + M - 2S). \quad (1)$$

The resulting axes, as shown by other authors,<sup>9,11–13</sup> are similar to the three orthogonal principal axes obtained by principal components analysis on this signal,<sup>2</sup> and are similar to the luminance, red–green and blue–yellow opponent mechanisms that have been characterized psychophysically and physiologically.<sup>3,4</sup> Correlations among these three measures are very small. For our scenes, the correlation coefficient between luminance ( $l$ ) and red–green ( $rg$ ) values was  $-0.0845$ , between luminance and blue–yellow ( $by$ ) values was  $-0.0391$ , and between red–green and blue–yellow values was  $0.0947$ . A scatter plot of a thousand randomly chosen pixels projected onto the luminance and red–green axes is shown in Fig. 1A.

Though other transformations can convert cone signals into relatively decorrelated opponent axes, one advantage of the Ruderman–Cronin–Chiao coordinates is that the (biologically plausible) logarithmic transformation of the data points distributes points along the axes relatively uniformly. Without the logarithmic transformation, data tend to be clustered near the zeros of the axes. Besides their compatibility with psychophysical and physiological data, these properties of decorrelation and even distribution of data points along the axes ensure that the statistical properties described in this paper are not due to first-order correlations between the luminance and color of individual pixels or to the luminance and color of individual pixels being clustered near gray. However, our choice of this particular coordinate space was not critical to our results.

Each of the 12 images was  $128 \times 128$  pixels, with each pixel roughly representing  $3 \times 3$  min of visual angle. Each picture therefore represented approximately  $6.4 \times 6.4$  deg of visual angle. As illustrated in Fig. 1B, we randomly sampled pairs of pixels from all of the 12 images. For each pair of sampled pixels we then computed

the difference along luminance ( $\delta_l$ ), red–green ( $\delta_{rg}$ ), and blue–yellow ( $\delta_{by}$ ) axes between the two pixels. The radius of separation ( $r$ ) between the pair of pixels varied between  $\sim 3$  and  $\sim 150$  min of visual angle. We sampled 24,000 randomly chosen pairs of pixels for each radius separation, sampling evenly from each of the 12 images. The distribution of  $\delta_{rg}$  is much tighter than for  $\delta_l$  or  $\delta_{by}$ .<sup>14</sup> To allow comparison of the shapes of the distributions, the sampled values of  $\delta_l$ ,  $\delta_{rg}$ , and  $\delta_{by}$  were independently scaled so that 90% of the samples lay between  $-1$  and  $1$ .

### 3. LUMINANCE AND COLOR DIFFERENCES ARE NOT INDEPENDENT

Although, as described above, correlations among luminance, red–green and blue–yellow axes are relatively small, *differences* between pairs of pixels for these directions in color space do not show the same independence. This interdependence does not, however, take the form of a linear correlation between the differences. Figure 1C shows a scatterplot of differences along luminance and red–green axes for 1000 randomly chosen pairs of pixels separated by 18 min, showing that differences along these two axes are relatively uncorrelated.

Even though differences in luminance between two pixels are associated with differences in color, the sign and the magnitude of these changes are relatively uncorrelated. As a result, correlation coefficients between differences in luminance and color across pixels, though significant, are not large.<sup>9,14</sup> For example, the correlation coefficients between color differences for pixels separated by 18 min were  $-0.1922$  between luminance and red–green axes,  $-0.0737$  between luminance and blue–yellow axes, and  $0.1783$  between red–green and blue–yellow axes. Correlation coefficients between absolute values of luminance and color differences were only slightly larger:  $0.2358$ ,  $0.2243$  and  $0.1733$ , respectively. When independent components analysis<sup>15</sup> is applied to hyperspectral, LMS,<sup>16</sup> or RGB images,<sup>17</sup> most of the resulting basis functions represent information along separate luminance, blue–yellow, and red–green opponent axes. Independent components analysis does not naturally discover the particular type of higher-order nonindependence conditional upon taking the *differences* in luminance and color between pixels that is described in this paper.

In contrast, examining the empirically sampled joint density function for luminance and color differences demonstrates a noticeable failure of independence. For pixel separations between 3 and 150 min we estimated (through sampling) the joint probability density function (pdf),  $p(\delta_l, \delta_{rg}, \delta_{by})$  for luminance and color differences between pairs of pixels. The joint pdf for each radius of pixel separation can be thought of as a three-dimensional cube with axes  $\delta_l$ ,  $\delta_{rg}$ , and  $\delta_{by}$ , filled with the probability values for those particular values of  $\delta_l$ ,  $\delta_{rg}$ , and  $\delta_{by}$ . Figure 1D shows a slice through this cube in the plane  $\delta_l = 0$  for a pixel separation of 18 min. As expected, the joint pdf is peaked at 0 along both  $\delta_{rg}$  and  $\delta_{by}$  axes; small differences along the red–green axis are associated with small differences along the blue–yellow axis. Slices through the cube in the planes  $\delta_{rg} = 0$  and  $\delta_{by} = 0$  look

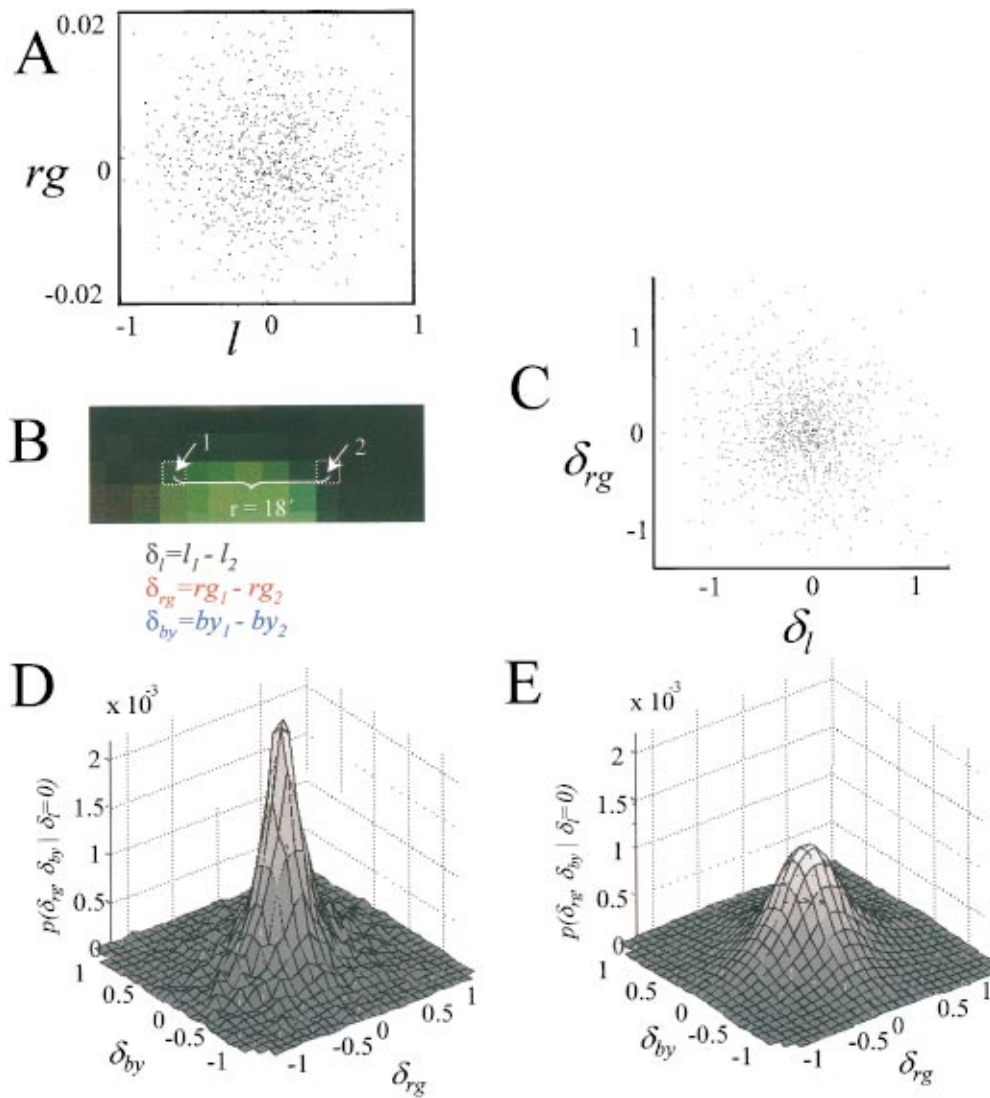


Fig. 1. A, Scatterplot of  $l$  versus  $rg$ . B, small patch from a natural image (for an example of a full image, see Fig. 5A). Two pixels, separated by a radius of 18 min, are shown. C, Scatterplot of  $\delta_l$  versus  $\delta_{rg}$ . D, Joint pdf of differences along red–green and blue–yellow axes for pixels separated by 18 min. E, Joint pdf of differences along red–green and blue–yellow axes for pixels separated by 18 min assuming independence.

very similar. Figure 1E shows how the  $\delta_l = 0$  slice would look if  $\delta_l$ ,  $\delta_{rg}$ , and  $\delta_{by}$  were independent, calculated by using their marginal probabilities, i.e., assuming that  $p(\delta_l, \delta_{rg}, \delta_{by}) = p(\delta_l) \cdot p(\delta_{rg}) \cdot p(\delta_{by})$ . The joint density functions obtained assuming independence are far less peaked near zero than the real joint pdf of Fig. 1D. This deviation from independence of differences in luminance and color across pixels is captured by the mutual information between (for instance)  $\delta_l$  and  $\delta_{rg}$ ,  $I(\delta_l; \delta_{rg})$ . This is the expected value of the binary logarithm of the ratio  $p(\delta_l, \delta_{rg})/p(\delta_l) \cdot p(\delta_{rg})$ , or, equivalently, Kullback–Leibler divergence between the joint density and the product of the marginal densities. For a pixel separation of 18 arc min the mutual information between  $\delta_l$  and  $\delta_{rg}$  was  $I(\delta_l; \delta_{rg}) = 0.154$ ; likewise,  $I(\delta_{rg}; \delta_{by}) = 0.138$ , and  $I(\delta_l; \delta_{by}) = 0.131$ . The mutual information between luminance and color was not strongly effected by pixel separation: For a pixel separation of 3 min,  $I(\delta_l; \delta_{rg}) = 0.182$ ; likewise,  $I(\delta_{rg}; \delta_{by})$

$= 0.085$ , and  $I(\delta_l; \delta_{by}) = 0.151$ , and for a pixel separation of 150 min,  $I(\delta_l; \delta_{rg}) = 0.106$ ,  $I(\delta_{rg}; \delta_{by}) = 0.126$ , and  $I(\delta_l; \delta_{by}) = 0.100$ .

#### 4. LUMINANCE AND COLOR DIFFERENCES AS A FUNCTION OF THE SEPARATION BETWEEN PIXELS

The solid curves in Figs. 2A–2C show the marginal pdfs for differences in luminance and color between pixels separated by 3 and 18 min and between pixels in different images, obtained by sampling randomly from all the natural images in our set. The  $x$  axis in each graph represents  $\delta_l$  (black text),  $\delta_{rg}$  (red text), or  $\delta_{by}$  (blue text), and the  $y$  axis represents the probability of  $\delta_l$ ,  $\delta_{rg}$ , or  $\delta_{by}$ .

In Figs. 2A and 2B the curves are “peaky,” or kurtotic, and centered on zero; i.e., small differences in luminance or color are more common than they would be in a Gauss-

ian distribution. The kurtoses of these pdfs are inversely related to the separation between the two pixels, as shown in Fig. 2D, which plots kurtosis as a function of pixel separation. We excluded the 0.5% tails of each distribution from our estimate of kurtosis, since these tails added a large amount of noise to the estimates, and the binning procedure used in our other analyses also limits the effect of these extreme outliers. The kurtosis is roughly similar for  $\delta_l$ ,  $\delta_{rg}$ , and  $\delta_{by}$ , with the exception that the  $\delta_{rg}$  distribution retains some kurtosis for large separations, even in comparisons between pixels from different images (Fig. 2C); this may be an artifact resulting from the limited gray-level resolution of the CCD camera, which affects the  $\delta_{rg}$  distribution selectively owing to its smaller dispersion.

The distance-dependent kurtosis of Fig. 2 is not unexpected. When pairs of pixels are close to each other they are more likely to fall on the *same* surface; consequently, differences in luminance and color between them are likely to be small. When pairs of pixels fall on different surfaces, larger differences in both luminance and color are more likely. Suppose that any pair of sampled pixels falls into one of two categories: Either the two pixels belong to the *same* surface (*same*) or they belong to different

surfaces (*diff*). As a basis for an initial analysis we assumed that the pdf for pixels falling on the *same* surface could be approximated by the pdf for adjacent pixels (separated by  $\sim 3$  arc min):

$$p(\delta_l, \delta_{rg}, \delta_{by}|same) \approx p(\delta_l, \delta_{rg}, \delta_{by}|r = 3'), \quad (2)$$

and that the pdf for pixels falling on different (*diff*) surfaces could be approximated by choosing pairs of pixels from different images. Sampled pixel pairs are drawn from the population of pixel pairs belonging to the same surface with probability  $p(same|r)$ , or from the population of pixel pairs belonging to different surfaces with probability  $1 - p(same|r)$ , where  $p(same|r)$  depends on the spatial separation of the pixels. Using the given approximations, the pdf for pairs of pixels, for any given separation  $r$ , could then be modeled as follows:

$$p(\delta_l, \delta_{rg}, \delta_{by}|r) = p(same|r)p(\delta_l, \delta_{rg}, \delta_{by}|same) + \{1 - p(same|r)\}p(\delta_l, \delta_{rg}, \delta_{by}|diff). \quad (3)$$

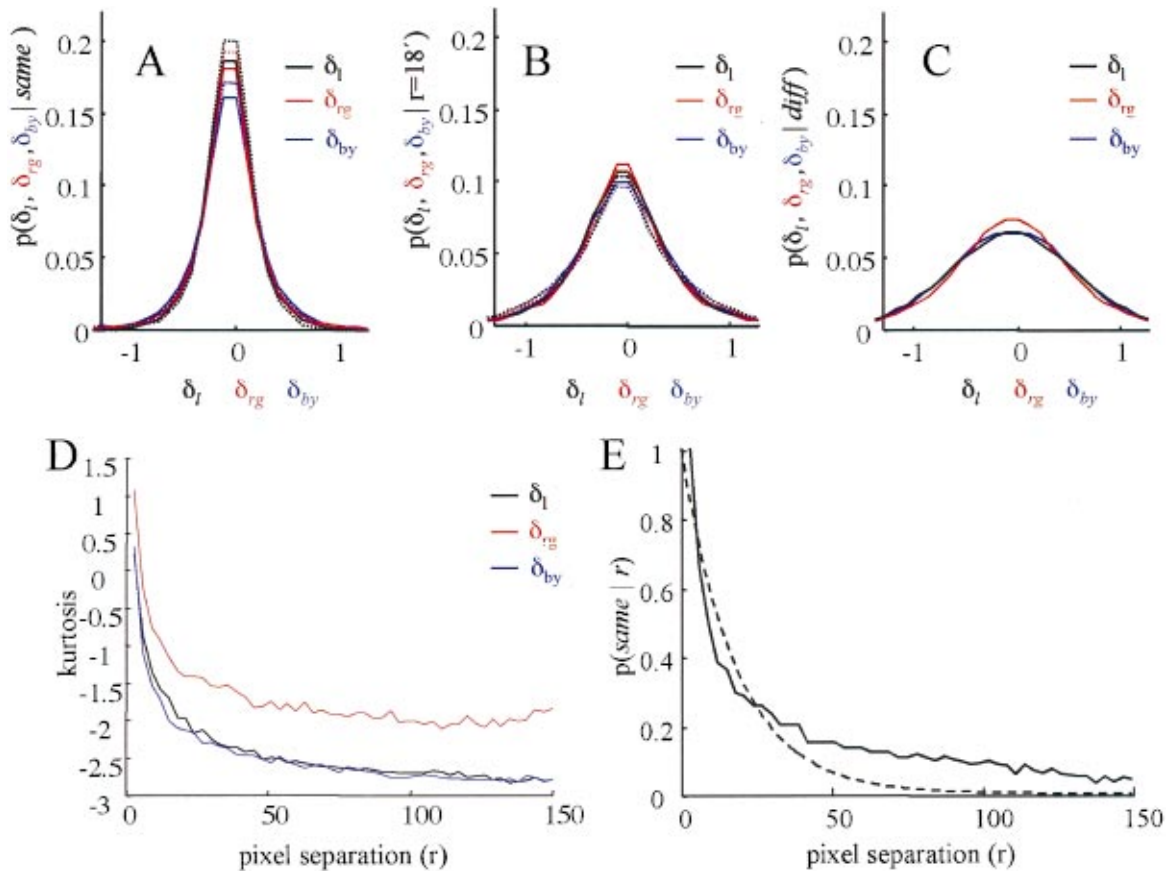


Fig. 2. A, Solid curves: pdf for differences in luminance and chromaticity between two pixels separated by 3 min of visual angle. Dashed curves: modeled pdf for differences in luminance and chromaticity between two pixels falling on the same surface. Pdf's for luminance differences are shown in black, pdf's for red-green chromatic differences are shown in red, and pdf's for blue-yellow chromatic differences are shown in blue. B, Sampled and modeled pdfs for differences in luminance and color between two pixels separated by 18 min. C, Sampled pdf for differences in luminance and color between two pixels on different surfaces. D, Kurtosis of pdf's for differences in luminance and color between pixels as a function of pixel separation. For comparison, the standard normal distribution has a kurtosis of 0. E, Estimates of  $p(same|r)$  as a function of pixel separation for a simple model based on the assumption that adjacent pixels always belonged to the same surface (solid curves) and an exponential fit (dashed curves).

The best-fitting value of  $p(\text{same}|r)$  was found, for each pixel separation in turn, by minimizing the rms difference between the values of  $p(\delta_l, \delta_{rg}, \delta_{by})$  obtained from Eq. (3) and the observed values. The solid curve in Fig. 2E shows this best-fitting value of  $p(\text{same}|r)$  as a function of the  $r$ , the separation between pixel pairs. Next, as a test of the assumptions of the model, estimated pdfs for differences in color (along red–green and blue–yellow dimensions separately) were constructed by weighting the pdfs for pixels falling on the same and different surfaces using the values of  $p(\text{same}|r)$  based on luminance alone. Fits were good for all pixel separations: For each separation the probability distributions for luminance, red–green and blue–yellow differences could all be reasonably well modeled by using a single parameter  $p(\text{same}|r)$ . The dashed curves in Fig. 2B show modeled pdfs for luminance and color for a pixel separation of 18 min each based on the value of  $p(\text{same}|r)$  constrained by luminance alone.

On the border between surfaces, adjacent pixels do not fall on the same surface, so our assumption that adjacent pixels belong to the same surface is clearly somewhat inadequate. To deal with this, we added the assumption that the probability of belonging to the same surface can be described by an exponential function, where, as the distance between two pixels increases by a fixed distance, the probability of their belonging to the same surface decreases by a fixed percentage; i.e.,  $p(\text{same}|r) = \exp(-r/r_0)$ . The advantage of using this simple equation (as compared with, for instance a power function, which could have a form invariant with viewing distance) is that as  $r$  approaches 0,  $p(\text{same}|r)$  approaches 1, as it should, since pixels infinitely close to each other belong to the same surface; and as  $r$  approaches infinity,  $p(\text{same}|r)$  approaches 0 (though the fit at large separations could clearly be improved). We used an iterative procedure to find the best fitting space constant  $r_0 = 19.5$  arc min. The dashed curve in Fig. 2E shows this new estimate of  $p(\text{same}|r)$ . We then reestimated the pdfs for the same surfaces, using Eq. (3) but replacing our initial estimate of  $p(\text{same}|r)$  with the exponential function. As shown by the dashed curves in Fig. 2A, these revised pdfs were only slightly different from those obtained by sampling adjacent pixels. Note that both of these approximations assume that the distribution of luminance and chromatic differences for pixels falling on the same surface remains independent of the spacing between pixels,  $p(\text{same}|r, d) = p(\text{same}|r)$ .

### 5. BAYESIAN SEGMENTATION BASED ON COLOR

The statistics above give us the priors needed to develop a Bayesian model that can predict whether two pixels belong to the same surface given the distance between them and the difference in luminance and color between them. The probability that two pixels belong to the same surface given their separation is multiplied by the probability of those luminance and color differences given that the two pixels belong to the same surface normalized by the probability of those luminance and color differences:

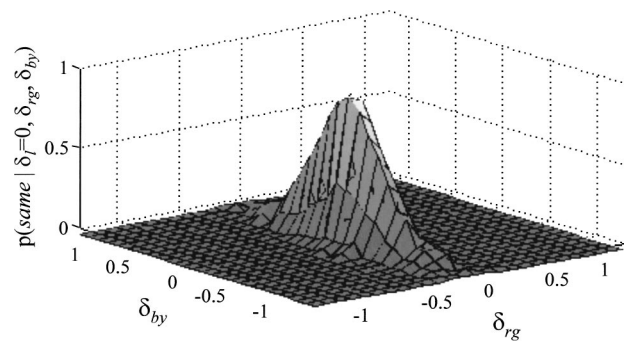


Fig. 3. Probability, computed from the model, that two pixels separated by 18 min belong to the same surface, as a function of the difference in red–green and blue–yellow chromaticity between the two pixels.

$$\begin{aligned}
 & p(\text{same}|\delta_l, \delta_{rg}, \delta_{by}, r) \\
 &= p(\text{same}|r)p(\delta_l, \delta_{rg}, \delta_{by}|\text{same})/p(\delta_l, \delta_{rg}, \delta_{by}|r). \tag{4}
 \end{aligned}$$

The model has no free parameters. We obtained  $p(\delta_l, \delta_{rg}, \delta_{by}|\text{same})$  from the approximation described in Eq. (2). We used the values of  $p(\text{same}|r)$  shown in Fig. 2, and we obtained  $p(\delta_l, \delta_{rg}, \delta_{by})$  by directly sampling images, as shown in Fig. 1B.

We calculated a three-dimensional probability cube for every pixel separation between 3 and 150 min. For example, Fig. 3 shows a slice through the cube obtained for  $r = 18$  min in the plane  $\delta_l = 0$ . The ordinate is the probability [from Eq. (4)] that two pixels 18 min apart belong to the same surface, given the difference in luminance and color between them. One common difficulty with using Bayesian models to predict human behavior is that estimating observers’ priors often requires *ad hoc* assumptions or choosing those priors that best predict observers’ performance. In both cases there is an issue of circularity, where it has to be assumed (rather unrealistically) that the human visual system has some sort of innate access to these priors. In the case of our model,  $p(\delta_l, \delta_{rg}, \delta_{by}|\text{same})$  and  $p(\text{same}|r)$  are based on the statistics of the natural environment, without any need for *ad hoc* assumptions. This makes our model a good test case for determining whether human performance matches that predicted by Bayesian statistics.

### 6. SEGMENTATION

We compared our model with segmentations made by two observers. Observers were presented with 36 image patches (3 patches from each of the 12 images) subtending  $105 \times 105$  min (corresponding to  $35 \times 35$  pixels). These patches were randomly chosen from the natural image set used to fix the model parameters. The choice of patches was constrained to be nonoverlapping and not to extend over the boundary of the image. Image patches were then converted from cone space to red–green–blue space on a calibrated monitor. Observers performed the segmentations on image patches scaled to subtend 10.2

deg of visual angle while referring to a smaller, correctly scaled image patch (the full image was not available to the observer). In addition, both the model and observers were presented with 18 image patches from 6 uncalibrated digital photographs of man-made environments.

The central pixel of each image patch acted as a reference pixel. Observers categorized every other pixel in the image patch as “very unlikely to belong to the same

surface as the reference pixel,” “somewhat likely,” “likely,” or “very likely.” Observers were given instructions to consider pixels as falling on the same surface if they appeared to belong to the same type of “stuff,” regardless of whether they were in spatial contiguous regions within the image patch. Each pixel was given a  $p(\text{same})$  value of 0, 0.33, 0.67 or 1 depending on how it had been categorized. This process was repeated four times, with a ran-

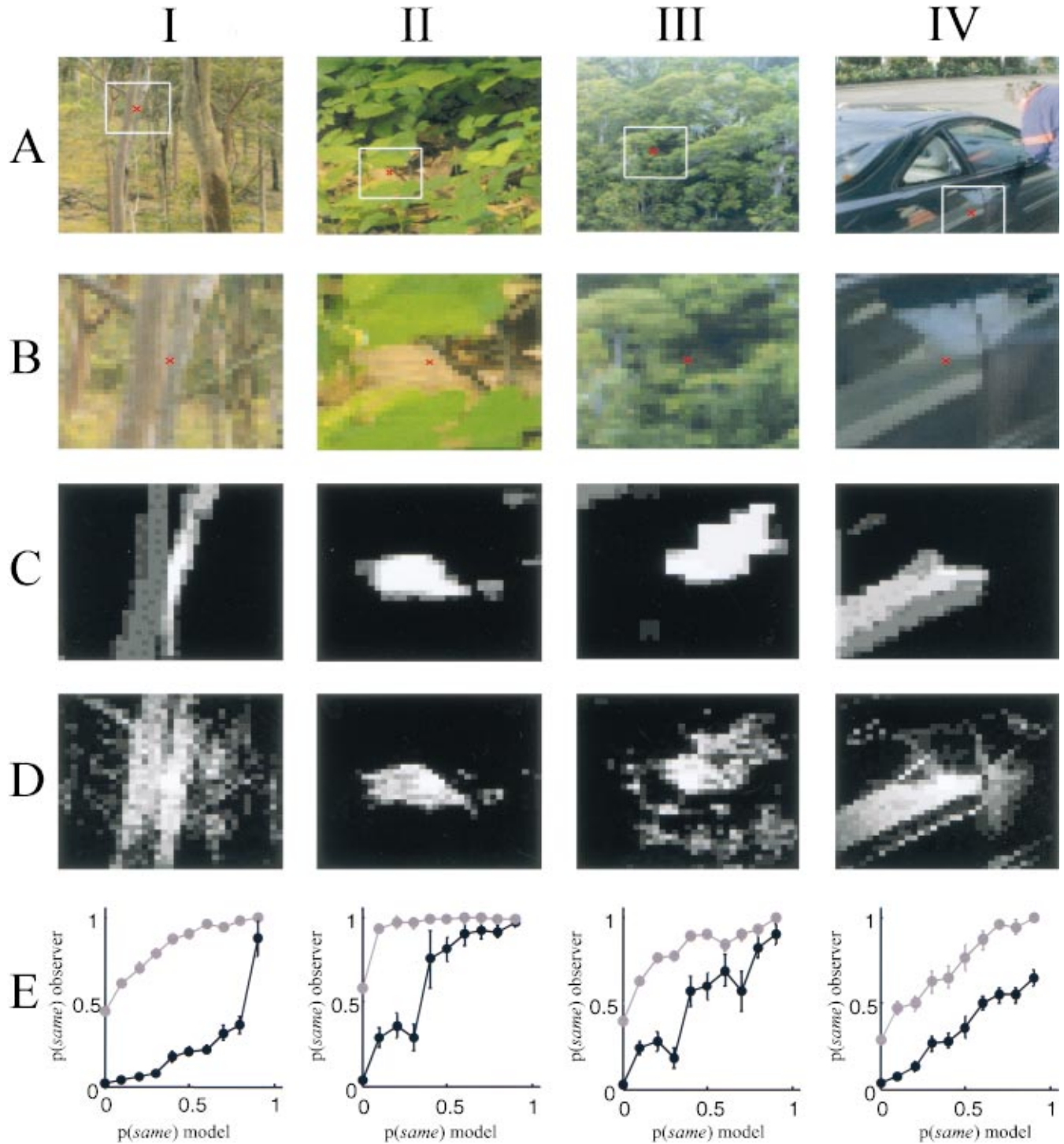


Fig. 4. A, Four of the scenes that were used, with a randomly chosen  $12 \times 12$  excised patch outlined. B, Image patches. C, Mean estimates from a single observer of the likelihood of each pixel belonging to the same surface as the central pixel. D, Estimates from our Bayesian model. E, Comparison of estimates of  $p(\text{same})$  made by two observers (IF, black symbols; SK, gray symbols) with estimates made by the model. Error bars subtend plus and minus one standard error.

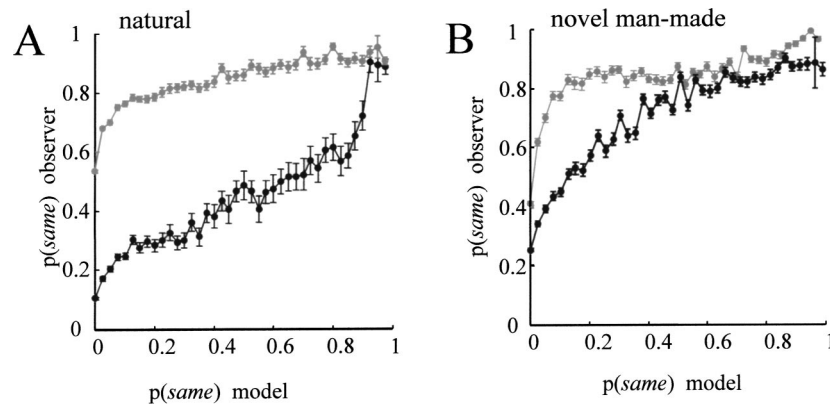


Fig. 5. A, Comparison of estimates of  $p(\text{same})$  made by observers with estimates of  $p(\text{same})$  made by the model averaged across all 36 sample patches of natural scenes. B, Comparison of estimates of  $p(\text{same})$  made by observers with estimates of  $p(\text{same})$  made by the model averaged across all 18 novel sample patches of man-made environments (IF, black symbols; SK, gray symbols). Error bars subtend plus and minus one standard error.

domized order of patch presentation. The mean of the four segmentations was considered to be a measure of  $p(\text{same})$  for each observer.

The images in columns I–III in Fig. 4 come from the natural scenes used in constructing our model. Typical examples are shown. Panels A show the full images, with the image patch outlined. Panels B show examples of the image patches that were segmented. In Panels C  $p(\text{same})$  for one observer (IF) are shown using gray scale, with lighter pixels representing larger values. Panels D represent  $p(\text{same})$  as estimated by the model for the same patch.

In column I the reference pixel happened to fall on an “object” (a tree trunk) so the observer and the algorithm picked out all “tree-trunkish stuff.” In column II the reference pixel fell on the background, so both observer and model picked out the background. Column III shows an image patch that contained relatively little discernible structure (though it was classified as segmentable by observers): such images were relatively common in our data set. With these unstructured images there was less correlation between model and observer estimates. In some of the natural scene image patches there was no discernible structure at all (e.g., the entire patch was a uniform color or was uniformly heavily textured) or the reference pixel fell on an edge. In those cases the observer categorized the entire image patch as unsegmentable, and that image patch was not included in the analysis.

For natural images, across all pixels, the correlation coefficient between estimates of  $p(\text{same})$  made by observer SK (naïve, gray symbols) and the model was 0.42, and the correlation coefficient between estimates of  $p(\text{same})$  made by observer IF (author, black symbols) and the model was 0.51. In comparison, the correlation coefficient between estimates of  $p(\text{same})$  made by observers SK and IF was 0.50. Variation between the two observers was as great as variation between the model and each observer for the natural images.

The image patch in column IV comes from the set of novel man-made environments. Despite our man-made environments containing almost no vegetation (three were taken indoors and three were taken outside in entirely urban settings), segmentation by the model was at

least as good as for the original set of natural scenes. Correlation between the model and observers was slightly higher for the novel man-made scenes: 0.54 between the model and observer SK and 0.66 between the model and observer IF. Correlation between the two observers was 0.75, higher than for the natural scenes and higher than the correlation between the model and observers. This is probably because the made-made scenes contained a large amount of contour information not available to the model.

Panels E in Fig. 4 compare estimates of  $p(\text{same})$  made by both observers to model estimates of  $p(\text{same})$ . For each image we divided pixels into ten bins based on model estimates of  $p(\text{same})$ . Again, there were different numbers of pixels in each bin, and we averaged observer estimates of  $p(\text{same})$  for all the pixels in each bin. These ten points are shown. As the model’s estimate of  $p(\text{same})$  increases, so do observers’ estimates of  $p(\text{same})$ .

Figures 5A and 5B compare estimates of  $p(\text{same})$  made by both observers to model estimates of  $p(\text{same})$  averaged across all natural images patches and across all 18 novel scenes. Because there were a large number of pixels, we were able to divide pixels into 40 bins based on model estimates of  $p(\text{same})$ . (There were again different numbers of pixels in each bin.) We then averaged observer estimates of  $p(\text{same})$  for all the pixels in each bin. These 40 points are shown in Fig. 5. As the model’s estimate of  $p(\text{same})$  increases, so do observers’ estimates of  $p(\text{same})$ , although one observer’s probabilities always exceed those of the model (see below).

Although observers were asked to estimate the probability that two pixels belonged to the same surface, it is likely that observers were in fact performing some sort of rating judgment (such as rating the similarity and color between the two pixels). The probability values assigned by observer IF were fairly close to those of the model. This could reflect experimenter bias, since although the observer had never seen the performance of the model on the particular test patches used, the observer was aware of the distribution of probability values typically assigned by the model. The  $p(\text{same})$  assignments of the naïve observer (SK) showed response compression, best fitted by a power function of the model probabilities with an expo-

ment of 0.1746 and a correlation coefficient of 0.94. SK's probability assignments therefore resembled rating judgments, which usually follow a power function,<sup>18</sup> more than genuine probability assignments. However, as noted above, the correlation coefficient between observer IF's estimates of  $p(\textit{same})$  and the model's estimate of  $p(\textit{same})$  was little better than the correlation coefficient between observer SK's estimates of  $p(\textit{same})$  and those of the model.

## 7. MODEL LIMITATIONS AND COMPARISON WITH OTHER MODELS

The performance of the model showed some dependence on spatial scale. Correlations between the model and the observer for the central region of our test patch (statistics were computed over the central  $10 \times 10$  pixel square of each image patch) were 0.55 for observer IF and 0.53 for observer SK. Performance for the outer corners of the test patch (statistics were computed over a  $5 \times 5$  pixel square in each of the four corners of each image patch) was worse, 0.31 for observer IF and 0.31 for observer SK. There is obviously information in the image patches that is used by observers that is not available to the model, and these other types of information seem to increase in importance as a function of the distance between two pixels.

One possible source of this spatial-scale dependency is gradual illumination changes over the scene. Our model is insensitive to many of the cues that observers may use to attribute a change in luminance and chromaticity to a gradual change in illumination rather than a change in surface. For example, our model is insensitive to whether the change in color and luminance between two pixels occurs gradually or abruptly in a step function. Our model also makes an assumption, inconsistent with the presence of gradual changes in illumination, that the distribution of luminance and chromatic differences for pixels on a common surface is independent of the spacing between the pixels. Another possible source of spatial-scale dependency is that object-level information may play more of a role when a large distance separates pairs of pixels. For example, our model considers two pixels of the same luminance and chromaticity as belonging to the same surface, regardless of the color and luminance of intervening pixels. Observers are likely to be sensitive to the luminance and color of intervening pixels in deciding whether two bright green pixels belong to the same leaf or two separate leaves.

Our model also does not consider the role of highlights, shadows, or shading that are due to surface slant. All these result mainly in a change in luminance, but they do generally also result in a small chromatic shift: For example, shadows cast by foliage tend to have a greenish hue. It should be noted that the natural image set that we used, like some others in the literature,<sup>19</sup> did not contain many deep shadows or highlights. This lack of shadows and highlights in our initial data set may have affected the statistics of our model slightly. However when testing our model, we made no attempt to avoid deep shadows in our novel man-made environment scenes. We noticed that when observers segmented

scenes containing deep shadows in either the natural or the man-made environment images, they tended to assign the shadow or highlight to a different surface, as did the model; see Fig. 4, column IV, for an example. It is plausible that under natural viewing conditions high-level object information, which was not generally present in our 2.5-deg image patches, interacts with color and luminance information in order to discount deep shadows and highlights.

Restricting the available information to luminance significantly impairs the ability of the model to imitate human judgments. For natural images the correlation coefficient between human estimates of  $p(\textit{same})$  and those from the Bayesian model dropped from 0.42 (SK) and 0.51 (IF) when both color and luminance information was used in the model to 0.35 and 0.38, respectively, when only luminance information was used in the model. For many scenes, removing color information made little difference; but in a few image patches, such as the one in Fig. 4, column II, where luminance differences within the scene were small but color differences were large, segmentation was much better when color information was included.

## 8. CONCLUSIONS

Various roles for color vision have been proposed, including finding or discriminating edible fruits and leaves,<sup>20,21</sup> facilitating scene and object recognition,<sup>22,23</sup> and improving search under certain conditions.<sup>24</sup> There has also been increasing interest within computer science in using color as a means of segmenting and identifying "meaningful regions" within a scene.<sup>25-28</sup>

This paper describes a biologically plausible segmentation model based on luminance and color differences within natural scenes. We found that color and luminance differences show a striking lack of independence: When two points fall on the same surface, differences in luminance and color between the two points tend to be small. Conversely, when two points fall on different surfaces, the distribution of luminance and color differences tend to be larger. These statistics for color and luminance differences are not easily captured by correlation statistics or by independent components analysis, though our Bayesian model captured these dependencies well. Demonstrations of a similar lack of independence have been made in other domains, such as motion and achromatic contrast,<sup>29,30</sup> and analogous approaches have been successful in modeling contour detection performance.<sup>31,32</sup>

We found that a Bayesian model (with no free parameters) based on color and luminance pairwise differences segmented images similarly to human observers, both with an image set based on natural scenes and with a very different novel image set containing both natural and man-made environments. The generalization of our results to novel scenes is consistent with the notion that observers may use a single set of priors based on an average distribution of luminance and color pairwise differences as a basis for image segmentation.

Our model was very naturally implemented as a region-based segmentation model (though it could be easily extended to an edge-detector model). Region-based models perform very differently from edge detectors; for



example they are robust to occlusion. Our model, because it contains little information about spatial statistics, considers pixels as belonging to the same surface if they are similar in luminance and color, regardless of whether there is a boundary between them. It links disjointed regions that are separated by edge-based segmentation techniques. In some contexts this is a shortcoming, in others an advantage. It is easy to see how the information provided by our model might subserve processes like global feature-based attention and search, although the Bayesian prior as a function of separation would be less relevant under such circumstances.

Interestingly, our model predicts that segregation based on *differences* in color and luminance should be carried out by neurons that are sensitive to differences in both luminance and color but insensitive to the sign of those differences. Such chromatic complex cells are apparently common within many cortical areas; but because they are insensitive to the sign of chromatic contrast, they have not traditionally been defined as color cells.<sup>4,32</sup> These neurons may play an important role in processing the joint luminance and color structure of scenes.

## ACKNOWLEDGMENTS

Thanks to Ian Abramson, Marni Bartlett, Dirk Beer, Greg Horwitz, Robert Jacobs, Eero Simoncelli, and two anonymous reviewers. Also thanks to Thomas Cronin and Daniel Ruderman for the use of their spectroradiometric natural image data. Supported by National Institutes of Health grants EY-13149, EY12925, and EY-01711, National Science Foundation grant SBR-9870897, and a La Jolla Interdisciplinary Fellowship.

Corresponding author I. Fine can be reached by e-mail at fine@salk.edu.

## REFERENCES

1. G. Buchsbaum and A. Gottschalk, "Trichromacy, opponent colours coding and optimum colour information transmission in the retina," *Proc. R. Soc. London Sec. B* **220**, 89–113 (1983).
2. L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Am. A* **3**, 1673–1683 (1986).
3. J. Krauskopf, D. R. Williams, and D. W. Heeley, "Cardinal directions of color space," *Vision Res.* **22**, 1123–1131 (1982).
4. P. Lennie, J. Krauskopf, and G. Sclar, "Chromatic mechanisms in striate cortex of macaque," *J. Neurosci.* **10**, 649–669 (1990).
5. K. R. Gegenfurtner and D. C. Kiper, "Contrast detection in luminance and chromatic noise," *J. Opt. Soc. Am. A* **9**, 1880–1888 (1992).
6. D. C. Kiper, S. B. Fenstermaker, and K. R. Gegenfurtner, "Chromatic properties of neurons in macaque area V2," *Visual Neurosci.* **14**, 1061–1072 (1997).
7. A. Li and P. Lennie, "Mechanisms underlying segmentation of colored textures," *Vision Res.* **37**, 83–97 (1997).
8. M. A. Webster and J. D. Mollon, "Changes in colour appearance following post-receptor adaptation," *Nature* **349**, 235–238 (1991).
9. D. R. 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).
10. A. Stockman, D. I. MacLeod, and N. E. Johnson, "Spectral sensitivities of the human cones," *J. Opt. Soc. Am. A* **10**, 2491–2521 (1993).
11. 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).
12. D. L. Ruderman, "The statistics of natural images," *Network* **6**, 345–358 (1994).
13. M. A. Webster and J. D. Mollon, "Adaptation and the color statistics of natural images," *Vision Res.* **37**, 3283–3298 (1997).
14. T. von der Twer and D. I. MacLeod, "Optimal nonlinear codes for the perception of natural colours," *Network* **12**, 395–407 (2001).
15. A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Res.* **37**, 3327–3338 (1997).
16. T. Wachtler, T. W. Lee, and T. J. Sejnowski, "Chromatic structure of natural scenes," *J. Opt. Soc. Am. A* **18**, 65–77 (2001).
17. D. R. Taylor, L. H. Finkel, and G. Buchsbaum, "Color-opponent receptive fields derived from independent component analysis of natural images," *Vision Res.* **40**, 2671–2676 (2000).
18. S. S. Stevens, "Psychophysics of sensory function," *Am. Sci.* **48**, 226–252 (1960).
19. C. A. Parraga, G. Brelstaff, T. Troscianko, and I. Moorhead, "Color and luminance information in natural scenes," *J. Opt. Soc. Am. A* **15**, 563–569 (1998).
20. N. J. Dominy and P. W. Lucas, "Ecological importance of trichromatic vision to primates," *Nature* **410**, 363–366 (2001).
21. B. C. Regan, C. Julliot, B. Simmen, F. Vienot, P. Charles-Dominique, and J. D. Mollon, "Frugivory and colour vision in *Alouatta seniculus*, a trichromatic platyrrhine monkey," *Vision Res.* **38**, 3321–3327 (1998).
22. K. R. Gegenfurtner and J. Rieger, "Sensory and cognitive contributions of color to the recognition of natural scenes," *Curr. Biol.* **10**, 805–808 (2000).
23. L. H. Wurm, G. E. Legge, L. M. Isenberg, and A. Luebker, "Color improves object recognition in normal and low vision," *J. Exp. Psychol. Hum. Percept. Perform.* **19**, 899–911 (1993).
24. A. L. Nagy and R. R. Sanchez, "Chromaticity and luminance as coding dimensions in visual search," *Hum. Factors* **34**, 601–614 (1992).
25. M. Celenk and S. H. Smith, "Gross segmentation of color images of natural scenes for computer vision systems," *Appl. Artif. Intell.* **III**, 333–344 (1986).
26. J. Liu and Y. Yang, "Multi-resolution color image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 689–700 (1994).
27. R. Ohlander, K. E. Price, and R. Reddy, "Picture segmentation by a recursive region splitting method," *Comput. Graph. Image Process.* **8**, 313–323 (1978).
28. Y. I. Ohta, T. Kanade, and T. Sakai, "Color information for region segmentation," *Comput. Graph. Image Process.* **13**, 222–241 (1980).
29. E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
30. W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly, "Edge co-occurrence in natural images predicts contour grouping performance," *Vision Res.* **41**, 711–724 (2001).
31. M. Sigman, G. A. Cecchi, C. D. Gilbert, and M. O. Magnasco, "On a common circle: natural scenes and Gestalt rules," *Proc. Natl. Acad. Sci. USA* **98**, 1935–1940 (2001).
32. 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).