

Psychophysical estimates of the number of spectral-reflectance basis functions needed to reproduce natural scenes

Sérgio M. C. Nascimento

Department of Physics, Gualtar Campus, University of Minho, 4710-057 Braga, Portugal

David H. Foster and Kinjiro Amano

Visual and Computational Neuroscience Group, Faculty of Life Sciences, University of Manchester, Manchester M60 1QD, UK

Received July 8, 2004; revised manuscript received December 6, 2004; accepted December 8, 2004

Theoretical analyses of spectral reflectances of natural surfaces suggest that their perceived colors can be well reproduced by approximations comprising combinations of three or four spectral basis functions. The aim of the present work was to assess psychophysically the number of basis functions necessary to reproduce entire natural outdoor scenes. Hyperspectral images of 20 such scenes were each subjected to a principal component analysis and then reproduced with a variable number of basis functions. The quality of the color approximation under daylight illumination was quantified theoretically in CIELAB space and psychophysically by spatial and temporal two-alternative forced-choice measurements in which the original and the approximated images were compared on a calibrated color monitor. Although five basis functions produced on average unit error in CIELAB space, original images were visually indistinguishable from their approximations only if there were at least eight basis functions. The combination of the spectral diversity of the natural world and the observed levels of color discrimination suggest that estimates of the minimum number of basis functions necessary to reproduce natural scenes may need to be revised upward. © 2005 Optical Society of America

OCIS codes: 330.0330, 330.1720, 330.1730, 330.5510.

1. INTRODUCTION

The physical causes of surface color are generally well understood,¹ and part of this understanding is that surface spectral reflectances are sufficiently constrained that they may be adequately approximated by low-dimensional models,^{2–5} as indeed may daylight illuminant spectra.^{6–8} The dimensionality of the model refers to the number of basis functions that in linear combinations generate the set of reflectance or illuminant spectra. The particular value of this number is important in the generation of compact descriptions of the data⁹ and for certain approaches to color perception^{10,11} where a correlate of surface spectral reflectance is to be extracted, allowing color constancy under illuminant changes.

Computational studies based on collections of surface spectral reflectances not weighted by their frequencies of occurrence in nature have suggested that the minimum number of basis functions needed to represent reflectance spectra varies from five to eight,^{3–5} but when spectra are weighted by their frequencies of occurrence in some natural scenes this number may fall to three.¹²

When the goodness of fit of an approximation is modified to take into account the nonuniform spectral sensitivity of the eye,⁹ the number of basis functions is reduced by ≈ 1 .³ Thus in an ideal-observer analysis¹³ three or four basis functions were found to be enough to reproduce visually indistinguishable reflectance spectra. For natural daylight spectra, three basis functions^{6,7} have been as-

sumed sufficient for good reproduction, although more may be necessary in some circumstances.¹⁴

Curiously, all these estimates of the minimum number of basis functions needed to represent reflection spectra were obtained by theoretical methods. Apart from studies by Oxtoby and co-workers with Munsell surfaces,^{15–17} the question of how many basis functions are actually required seems not to have been addressed empirically.

In this study psychophysical methods were used to assess the adequacy of linear approximations of reflectance spectra of surfaces in natural images, which were obtained with a hyperspectral imaging system and displayed on a calibrated color monitor. In accordance with standard practice,^{3,4,12,13,18–23} principal component analysis (PCA) was used to obtain the basis functions, and, for ease of comparison with previous studies, approximations were not modified to accommodate the nonuniform spectral sensitivity of the eye. The use of PCA rather than any other decomposition procedure seems not too critical, as is made clear later. It was found that natural scenes illuminated by daylight require at least eight PCA basis functions for their approximations to be visually indistinguishable from an original.

2. METHODS

A. Images

Hyperspectral images were captured with a progressive-scanning monochrome digital camera (Pulnix TM-1010,



Fig. 1. Reproduction of an original scene under D_{65} and approximations based on 1, 2, and 4 spectral basis functions.

Pulnix America Inc., Sunnyvale, Calif.) with a CCD array of 1024×1024 pixels that, with an attached lens-filter system, produced an angular resolution of ≈ 1 arcmin per pixel. The intensity response at each pixel was recorded with 10-bit precision. Mounted in front of the lens was a tunable birefringent filter (VariSpec, model VS-VIS2-10HC-35-SQ, Cambridge Research and Instrumentation, Inc., Boston, Mass.) whose wavelength of peak transmission could be electronically tuned over the range 400–720 nm with a FWHM transmission of 10 nm at 500 nm. Scenes were sampled at intervals of 10 nm over this wavelength range. Each complete set of 33 spectral images took between 5 and 15 s to acquire, depending on the light levels on the scene.

After correction for dark noise and transmission effects of the lens-filter system, the reflectance spectrum at each pixel element was estimated by normalizing the nominal radiance spectrum at that point against that of a neutral standard, a small gray object placed in the scene and having a flat reflectance spectrum (further details of the apparatus and image processing are given in Nascimento *et al.*²⁴). The illumination was assumed to be spatially uniform, and those scenes failing this condition were excluded. For the purposes of computation and display, all image data were sampled to one third of the maximum resolution.

Twenty scenes were used in this study: ten of noncultivated rural scenes in the Minho region of Portugal containing rocks, trees, foliage, grass, and earth, and ten urban scenes containing buildings and room contents.²⁴ All images were collected during the summers of 1999 and 2000 under daylight between midmorning and midafternoon. Scenes were illuminated by direct sunlight in clear or nearly clear sky.

The set of reflectance spectra for each individual scene was subjected to PCA, from which eight approximations were synthesized based on 1, 2, ..., 8 principal components, each defined about the mean spectrum.^{25,26} This upper limit of eight on the number of components tested was based on the calculation that eight basis functions accounted for more than 99% of the variance in the spectra and on the previous finding that as many as eight components (defined about zero) are needed to achieve a low-error reconstruction of the Munsell spectra.⁴

Each original scene and its approximations were assumed to be illuminated by a common illuminant, namely CIE Standard Illuminant D_{65} , which was close to that illuminating the scene during image acquisition. The resulting images were displayed on a calibrated color moni-

tor. Only images for which more than 99% of the pixels could be displayed within the gamut of the monitor after luminance scaling were used. The pixels out of gamut were not clustered together and were displayed by clipping to the closest RGB values for both original images and approximations.

Each set of principal components was derived for each individual scene rather than for all 20 scenes as a whole in order to relate scene content more easily to the minimum number of basis functions needed to represent it. Basing the PCA on all 20 scenes instead would have produced higher minimum numbers of components than those reported here, as would including the 1% of pixels out of gamut and relaxing the constraint that scenes should have almost uniform illumination.

As expected, images reconstructed with just one component were monochrome and as the number of components increased, the original chromatic content was restored (Fig. 1).

B. Psychophysics

Images were displayed on a 17-in., RGB color monitor with flat screen (Trinitron, model GDM-F400T9; Sony Corp., Tokyo, Japan) controlled by a computer raster-graphics card providing 24 bits per pixel in true-color mode (VSG 2/5; Cambridge Research Systems, Rochester, UK). Screen resolution was 800×600 pixels and refresh rate was ≈ 80 Hz. A telespectoradiometer (SpectraColorimeter, PR-650; Photo Research Inc., Chatsworth, Calif.) whose calibration could be traced to the UK National Physical Laboratory was used to regularly calibrate the display system. Images subtended 3–4 deg visual angle. They had an average luminance of 8–15 cd m^{-2} , depending on the scene, and were displayed in a darkened room.

The performance of observers in discriminating between an original image and its approximation with a variable number of components was tested in two different two-interval experiments: In one the images for comparison were presented simultaneously; in the other, they were presented sequentially (Fig. 2). In the two-interval simultaneous procedure [Fig. 2(a)] pairs of images were presented side-by-side, separated by 0.75 deg, in each of two temporal intervals. In one temporal interval, the pair consisted of two identical original images; in the other temporal interval, the pair consisted of the original and an approximation reconstructed from a specific number of components. Each temporal interval lasted for 5 s with a 1-s interval separating them when the screen was dark. The total stimulus duration was therefore 11 s. In the

two-interval sequential procedure [Fig. 2(b)] the images in each pair were presented sequentially for 1 s each in the same position on the screen, with an interval between the pair also of 1 s. The total stimulus duration was therefore 5 s. In both procedures, the observer had to indicate the interval, first or second, containing the approximation. Observers were not instructed to fixate specific locations.

The use of the indistinguishability of an approximation from the original is a stringent criterion. Other possible criteria, such as perceived realism, would require fewer basis functions, but would make the results of this work more difficult to compare with previous theoretical and computational predictions.

In each experimental session twenty scenes with eight approximations of each scene were tested. In each trial the number of components in the approximation and the spatial or temporal ordering of the approximation (left or right, and first, second, third, or fourth interval) were chosen randomly. The ordering of scenes was balanced over sessions. The simultaneous procedure and the sequential procedure were used in different sessions. Each observer performed at least 15 trials for each of the eight approximations of any individual scene.

Three observers, JL, FS, and PP, participated in the experiment, and all were unaware of its purpose. Each had normal color vision as assessed with the Farnsworth–Munsell 100-Hue test and Ishihara plates, and each had normal Snellen acuity.

3. RESULTS

To quantify the colorimetric quality of the approximations and provide a comparison with previous data,⁵ an average color difference was calculated between each original image and its approximation in terms of the usual Euclidean distance in CIELAB space, with white point a perfect reflecting diffuser under the same CIE Standard Illuminant D_{65} used to illuminate the scenes for display purposes. Thus if pixel i of the original image had CIELAB coordinates (L_i^*, a_i^*, b_i^*) and the approximation had CIELAB coordinates $(\hat{L}_i^*, \hat{a}_i^*, \hat{b}_i^*)$, then the average color difference, $\overline{\Delta E_{ab}^*}$ say, over the N pixels of the image is given by

$$\overline{\Delta E_{ab}^*} = \frac{1}{N} \sum_{i=1}^N [(L_i^* - \hat{L}_i^*)^2 + (a_i^* - \hat{a}_i^*)^2 + (b_i^* - \hat{b}_i^*)^2]^{1/2}.$$

Figure 3 shows $\overline{\Delta E_{ab}^*}$ as a function of the number n of components used in each approximation calculated over the ten rural scenes (squares) and over the ten urban

scenes (triangles), and Table 1 shows the corresponding numerical values. The color difference decreased as n increased, reaching unity at $n=6$. Rural and urban scenes produced closely similar color differences. Even at $n=8$, however, $\overline{\Delta E_{ab}^*}$ was a little larger than 0.5, a level similar to that reported by Vrhel *et al.*⁵ The proportion of the pixels with $\Delta E_{ab}^* > 3$, corresponding to clearly discriminable colors, is also shown in Table 1. Even with the best approximation, 2%–3% of the images were on average distinguishable by this measure.

Consider now the psychophysical data. Figure 4 shows as an example one observer's discrimination of a rural scene from its approximations with the sequential procedure. The symbols show percent-correct discrimination based on 15 trials per level as a function of the number n of components in each approximation. With three or fewer components, this observer was able to discriminate the original image from the approximation in almost all trials; but as n increased, discrimination performance decreased, reaching chance levels, that is, 50%, near $n=8$.

To estimate more precisely the number of components for which a scene and its approximation become indistinguishable, a smooth curve was fitted to the data by locally weighted logistic regression.²⁷ The bandwidth of the local fitting procedure was determined by cross validation, constrained to exclude nonmonotonic fits. For two criterion levels of performance of 75% and 55% correct, threshold numbers of components of $n_{75\%}$ and $n_{55\%}$, respectively,

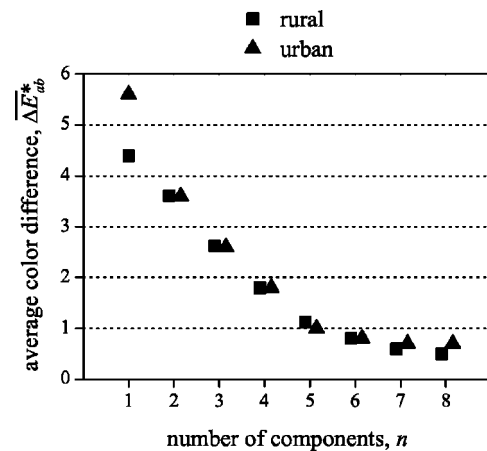


Fig. 3. Average CIELAB color difference $\overline{\Delta E_{ab}^*}$ between original and PCA-approximated images as a function of the number n of components in each approximation, calculated over ten rural scenes (squares) and over ten urban scenes (triangles). Overlapping points have been slightly displaced horizontally.

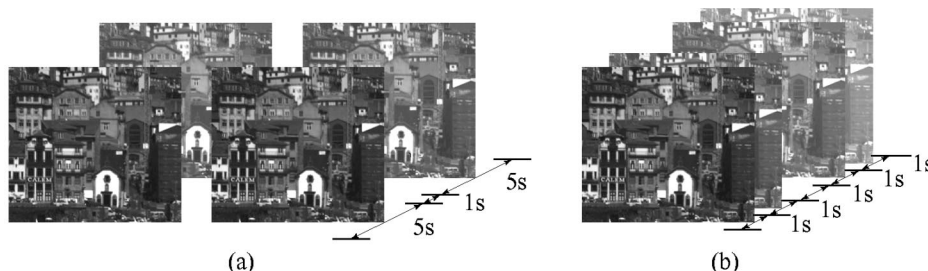


Fig. 2. Simultaneous and sequential procedures for stimulus presentation.

Table 1. Quality of Colorimetric Approximations as a Function of Number of Components

Colorimetric Quantities	Number of Components							
	1	2	3	4	5	6	7	8
Rural scenes ^a								
$\overline{\Delta E_{ab}^*}$	6.0	4.7	3.7	2.6	1.7	1.1	0.9	0.7
$P(\Delta E_{ab}^* > 3)$	0.7	0.6	0.4	0.3	0.1	0.06	0.03	0.02
Urban scenes ^a								
$\overline{\Delta E_{ab}^*}$	7.5	4.7	3.6	2.4	1.5	1.1	0.9	0.8
$P(\Delta E_{ab}^* > 3)$	0.7	0.5	0.4	0.2	0.1	0.07	0.04	0.03

^aThe entries $\overline{\Delta E_{ab}^*}$ represent the mean CIELAB color difference across scenes and $P(\Delta E_{ab}^* > 3)$ the proportion of pixels where the CIELAB color difference is greater than 3. Data are based on ten rural scenes and ten urban scenes.

were obtained from the fitted curve (as illustrated in Fig. 4). The level of 55% was a compromise between the stability of the threshold estimate and the closeness of the criterion to the true chance level of 50%.

For each observer, thresholds were thereby obtained from data pooled over rural scenes and data pooled over urban scenes. Standard errors (SEs) for each class of scenes were estimated from a bootstrap resampling with replacement over scenes and based on 1000 replications rather than by binomial resampling (see Ref. 28). Figure 5 shows the results for simultaneous and sequential experimental procedures. Thresholds were systematically lower for rural scenes than for urban scenes, but there was no difference in the effects of simultaneous and sequential procedures. Taken over subjects and procedures, the mean threshold (± 1 SE) for a criterion level of 75% was 5.1 ± 0.2 for rural scenes and 6.4 ± 0.3 for urban scenes, and, for a criterion level of 55%, it was 8.0 ± 0.3 for rural scenes and 8.5 ± 0.4 for urban scenes. Notice that these SE estimates refer to the variability of scenes, not observers.

4. DISCUSSION

Contrary to the implications of some theoretical analyses, approximations of spectral reflectances of natural surfaces need more than just three or four spectral basis functions to be acceptable visually. In fact, as the experiments here have shown, at least eight basis functions seem necessary for the colors of daylight-illuminated natural scenes to be indistinguishable from their approximations.

What is the origin of this disparity? Two factors, at least, are critical. First, there is the problem of sampling density. Previous theoretical estimates of mean or maximum CIELAB color differences ΔE_{ab}^* with synthetic or natural spectra have not taken into account their frequencies of occurrence in nature (one previous frequency-based analysis¹²—of spectra from forests and coral reefs—used proportion of variance explained rather than colorimetric estimates of perceived color difference). The fact that even with eight components in the approximations the mean values of ΔE_{ab}^* were still larger than 0.5, and that ≈ 2 –3% of the images had $\Delta E_{ab}^* > 3$ (Fig. 3 and Table 1) suggest that the approximations may still have been inadequate visually.

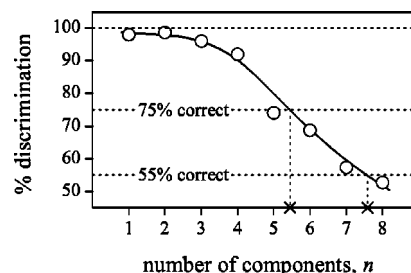


Fig. 4. Discriminability of original and PCA-approximated images by one observer. Percent-correct discrimination based on 15 trials per level is plotted (circles) as a function of the number n of components in each approximation. The smooth curve is a locally weighted logistic regression. Chance level is 50%. For two criterion levels of performance of 75% and 55%, threshold numbers of components of $n_{75\%}$ and $n_{55\%}$ are indicated by crosses on the abscissa. Images were of a rural scene, the stimulus presentation was sequential, and the observer was JL.

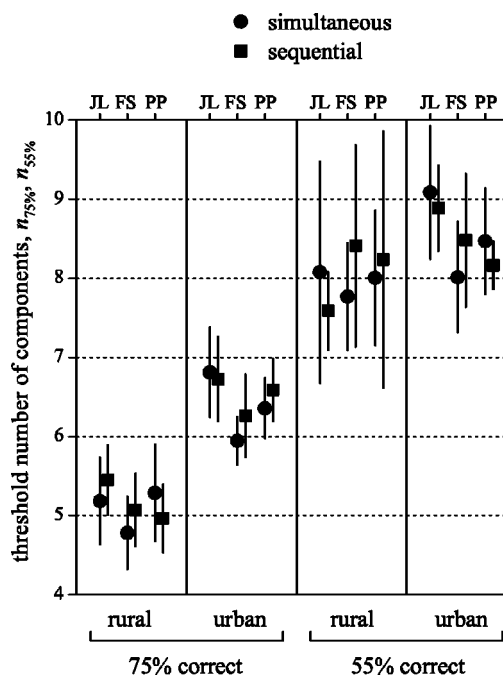


Fig. 5. Threshold numbers of components for discriminating original and PCA-approximated images. For two criterion levels of performance of 75% and 55% correct, thresholds $n_{75\%}$ and $n_{55\%}$ are shown for ten rural and ten urban scenes, three observers, and simultaneous and sequential stimulus presentations. Error bars represent ± 1 SE estimated from a bootstrap based on 1000 replications with resampling over scenes.

Second, there is the potential confounding effect of spatial structure on color discrimination. The salience or otherwise of poorly approximated spectral reflectances in natural scenes can only be determined empirically. Numerically small errors may assume a much larger visual significance when present in real scenes. The actual nature of those scenes had only a modest effect here, in that 0.5–1.3 more components were needed with urban scenes than with rural scenes to elicit similar levels of discrimination. This increase may be attributable to urban scenes being chromatically richer and more complex than rural scenes.

Although not tested here, an effect of scene size might also be expected. In normal viewing, visual fields can be very large, and the small residuum of clearly distinguishable approximations with eight components obtained with the 3–4 deg viewing angle used here might be expected to rise markedly with unrestricted viewing angles.

The particular procedure by which the scenes were presented for comparison, that is, across space or across time, proved immaterial.

There is, however, a general issue concerning the relevance of decompositions of spectra by PCA. It seems that PCA is not itself special, in that experiments like those reported here but with Mondrian arrays of Munsell surfaces¹⁷ have shown that there is little difference in the minimum number of components whether determined by PCA, by independent component analysis, or by artificial neural networks, although other methods of obtaining basis functions, such as nonnegative matrix factorization,²⁹ may give different results. More important, as Vrhel *et al.*⁵ have emphasized, PCA does not minimize the color difference ΔE_{ab}^* . A set of basis functions specifically matched to some color-difference metric might yield a lower minimum.

As a final point, recall that each set of components was here derived for each individual scene. If PCA had been applied to the pooled sets of 10 rural scenes and 10 urban scenes, or indeed to all 20, then it is very likely that the minimum number would have been larger still. In general, for images of natural scenes to be reproduced by combinations of basis functions so that they are visually indistinguishable from their originals, more than eight such functions may well be necessary in any application.

ACKNOWLEDGMENTS

This work was supported by the Fundação para a Ciência e Tecnologia (grant POSI/SRI/40212/2001), by the Centro de Física da Universidade do Minho, Braga, Portugal, and by the Engineering and Physical Sciences Research Council, UK. Parts of this work were reported at the Annual Meeting of the Association for Research in Vision and Ophthalmology, 2001, Fort Lauderdale, Florida,³⁰ and at the European Conference on Visual Perception, 2000, Groningen, The Netherlands.³¹

Corresponding author Sérgio Nascimento can be reached by email at smcn@fisica.uminho.pt, by phone at 351-253-604320, or by fax at 351-253-678981.

REFERENCES

1. K. Nassau, *The Physics and Chemistry of Color. The Fifteen Causes of Color* (Wiley, New York, 1983).
2. J. Cohen, "Dependency of the spectral reflectance curves of the Munsell color chips," *Psychonomic Sci.* **1**, 369–370 (1964).
3. 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).
4. J. P. S. Parkkinen, J. Hallikainen, and T. Jaaskelainen, "Characteristic spectra of Munsell colors," *J. Opt. Soc. Am. A* **6**, 318–322 (1989).
5. M. J. Vrhel, R. Gershon, and L. S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Res. Appl.* **19**, 4–9 (1994).
6. D. B. Judd, D. L. MacAdam, and G. Wyszecki, "Spectral distribution of typical daylight as a function of correlated color temperature," *J. Opt. Soc. Am.* **54**, 1031–1040 (1964).
7. J. Romero, A. García-Beltrán, and J. Hernández-Andrés, "Linear bases for representation of natural and artificial illuminants," *J. Opt. Soc. Am. A* **14**, 1007–1014 (1997).
8. C. C. Chiao, D. Osorio, M. Vorobyev, and T. W. Cronin, "Characterization of natural illuminants in forests and the use of digital video data to reconstruct illuminant spectra," *J. Opt. Soc. Am. A* **17**, 1713–1721 (2000).
9. D. H. Marimont and B. A. Wandell, "Linear-models of surface and illuminant spectra," *J. Opt. Soc. Am. A* **9**, 1905–1913 (1992).
10. M. D'Zmura and G. Iverson, "Color constancy. III. General linear recovery of spectral descriptions for lights and surfaces," *J. Opt. Soc. Am. A* **11**, 2389–2400 (1994).
11. L. T. Maloney, "Physics-based approaches to modeling surface color perception," in *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe, eds. (Cambridge U. Press, Cambridge, UK, 1999), pp. 387–416.
12. C. C. Chiao, T. W. Cronin, and D. Osorio, "Color signals in natural scenes: characteristics of reflectance spectra and effects of natural illuminants," *J. Opt. Soc. Am. A* **17**, 218–224 (2000).
13. J. L. Dannemiller, "Spectral reflectance of natural objects: how many basis functions are necessary?" *J. Opt. Soc. Am. A* **9**, 507–515 (1992).
14. J. Hernández-Andrés, J. Romero, J. L. Nieves, and R. L. Lee, "Color and spectral analysis of daylight in southern Europe," *J. Opt. Soc. Am. A* **18**, 1325–1335 (2001).
15. E. K. Oxtoby, D. H. Foster, K. Amano, and S. M. C. Nascimento, "How many basis functions are needed to reproduce coloured patterns under illuminant changes?" *Perception* **S31**, 66 (2002).
16. E. K. Oxtoby, D. H. Foster, and R. C. Baraas, "How many spectral basis functions do red-green dichromats need to discriminate surface colours under different lights?" *Perception* **S32**, 147 (2003).
17. E. K. Oxtoby and D. H. Foster, "Perceptual limits on low-dimensional models of Munsell reflectance spectra" (to be published).
18. T. Jaaskelainen, J. Parkkinen, and S. Toyooka, "Vector-subspace model for color representation," *J. Opt. Soc. Am. A* **7**, 725–730 (1990).
19. M. H. Brill and G. West, "Chromatic adaptation and color constancy: a possible dichotomy," *Color Res. Appl.* **11**, 196–204 (1986).
20. L. T. Maloney and B. A. Wandell, "Color constancy: a method for recovering surface spectral reflectance," *J. Opt. Soc. Am. A* **3**, 29–33 (1986).
21. G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.* **310**, 1–26 (1980).
22. M. D'Zmura and P. Lennie, "Mechanisms of color constancy," *J. Opt. Soc. Am. A* **3**, 1662–1672 (1986).
23. M. J. Vrhel and H. J. Trussell, "Color correction using principal components," *Color Res. Appl.* **17**, 328–338 (1992).
24. S. M. C. Nascimento, F. P. Ferreira, and D. H. Foster, "Statistics of spatial cone-excitation ratios in natural scenes," *J. Opt. Soc. Am. A* **19**, 1484–1490 (2002).

25. Principal components of spectral data can be calculated about the mean spectrum or about the zero spectrum. For a discussion of the two methods, see Ref. 26.
26. M. H. Brill, "A non-PC look at principal components," *Color Res. Appl.* **28**, 69–71 (2003).
27. C. Loader, *Local Regression and Likelihood* (Springer, New York, 1999).
28. D. H. Foster and W. F. Bischof, "Thresholds from psychometric functions: superiority of bootstrap to incremental and probit variance estimators," *Psychol. Bull.* **109**, 152–159 (1991).
29. G. Buchsbaum and O. Bloch, "Color categories revealed by non-negative matrix factorization of Munsell color spectra," *Vision Res.* **42**, 559–563 (2002).
30. S. M. Nascimento, D. H. Foster, and K. Amano, "Reproduction of colors of natural scenes by low-dimensional models," *Invest. Ophthalmol. Visual Sci.* **42**, S720 (2001).
31. S. M. C. Nascimento and D. H. Foster, "Chromatic quality of natural scenes represented by low-dimensional approximations to reflectance functions," *Perception* **S29**, 72 (2000).