Color constancy in natural scenes explained by global image statistics

DAVID H. FOSTER,[1] KINJIRO AMANO,[1] AND SÉRGIO M. C. NASCIMENTO[2]

[1]Sensing, Imaging, and Signal Processing Group, School of Electrical and Electronic Engineering, University of Manchester, Manchester, UK
[2]Department of Physics, Gualtar Campus, University of Minho, Braga, Portugal

Abstract

To what extent do observers' judgments of surface color with natural scenes depend on global image statistics? To address this question, a psychophysical experiment was performed in which images of natural scenes under two successive daylights were presented on a computer-controlled high-resolution color monitor. Observers reported whether there was a change in reflectance of a test surface in the scene. The scenes were obtained with a hyperspectral imaging system and included variously trees, shrubs, grasses, ferns, flowers, rocks, and buildings. Discrimination performance, quantified on a scale of 0 to 1 with a color-constancy index, varied from 0.69 to 0.97 over 21 scenes and two illuminant changes, from a correlated color temperature of 25,000 K to 6700 K and from 4000 K to 6700 K. The best account of these effects was provided by receptor-based rather than colorimetric properties of the images. Thus, in a linear regression, 43% of the variance in constancy index was explained by the log of the mean relative deviation in spatial cone-excitation ratios evaluated globally across the two images of a scene. A further 20% was explained by including the mean chroma of the first image and its difference from that of the second image and a further 7% by the mean difference in hue. Together, all four global color properties accounted for 70% of the variance and provided a good fit to the effects of scene and of illuminant change on color constancy, and, additionally, of changing test-surface position. By contrast, a spatial-frequency analysis of the images showed that the gradient of the luminance amplitude spectrum accounted for only 5% of the variance.

Keywords: Natural scenes, Color constancy, Image statistics, Spatial cone-excitation ratios, Spatial-frequency analysis

Introduction

The ability of human observers to make accurate judgments about the colors of surfaces under different colored lights depends on many factors. Predicting the accuracy of such judgments, that is, the degree of color constancy is difficult, especially when the surfaces are part of natural scenes containing complex spatial variations in spectral reflectance. The problem might, however, be made more tractable by taking a statistical approach in which the color properties of images as a whole are considered rather than just the particular features of the surface being judged and its local context. From psychophysical experiments with simpler geometric displays, the global properties of average scene hue, saturation, and the variation in these quantities over the field of view might all be relevant factors (e.g., Webster & Mollon, 1995; Brown & MacLeod, 1997; Kulikowski et al., 2001; Wachtler et al., 2001; Brenner et al., 2003). But there are few data in the literature

describing surface-color judgments in natural scenes that might provide the basis for such an analysis.

To address this problem, a psychophysical experiment was undertaken to measure surface-color matching with images of natural vegetated and non-vegetated scenes under different illuminants characteristic of the sun and sky at different times of the day (Judd et al., 1964; Wyszecki & Stiles, 1982). The images were generated from hyperspectral data, to allow the accurate and independent control of illuminant and reflectance spectra, and they were viewed on a high-resolution color monitor driven by a 30-bit RGB color-graphics computer system. An operational approach to the color-matching task was adopted (Craven & Foster, 1992; Foster, 2003) in which observers reported in each experimental trial whether a test surface in the scene had changed in its reflecting properties during the change in daylight. The spectral reflectance of the test surface was varied randomly from trial to trial, and observers' ability to detect that variation across successive images of the scene was used to quantify their color constancy (Foster & Nascimento, 1994, Appendix 1; Foster et al., 2003). As anticipated, observers' performance varied markedly with the scene.

To then determine how well this variation in performance could be explained by global image statistics, a linear regression analysis

---

Address correspondence and reprint requests to: Professor D.H. Foster, Sensing, Imaging, and Signal Processing Group, School of Electrical and Electronic Engineering, Moffat Building, University of Manchester, Sackville Street, Manchester M60 1QD, UK. E-mail: d.h.foster@manchester.ac.uk

was performed using a range of colorimetric and receptor-based properties of the images. The most successful explanatory factor was the mean deviation in spatial ratios of cone excitations due to light reflected from pairs of surfaces evaluated over the scene under two successive illuminants. In conjunction with other global statistics, namely, the mean chroma of the image of the scene under the first illuminant, its difference from the mean chroma of the image of the scene under the second illuminant, and the mean difference in hue, it was possible to explain 70% of the performance variation, the rest being attributed to local image properties and to individual observer variation.

A separate control experiment was undertaken in which the position of the test surface in the scene was changed. The resulting change in performance was limited, and its variation with scene could also be explained by these global color properties. To test the role of purely spatial global properties, the luminance distribution in each image was subjected to a spatial-frequency analysis. The gradient of the amplitude spectrum accounted for only 5% of the performance variation.

Materials and methods

Stimuli and procedure

The natural scenes used as stimuli were drawn from the Minho region of Portugal, which has a temperate climate and a variety of land covers. Twenty-one close-up and distant hyperspectral images of scenes were acquired. These comprised the main vegetated and non-vegetated land-cover classes (UNESCO, 1973; Federal Geographic Data Committee, 1997), including woodland, shrubland, herbaceous vegetation (e.g. grasses, ferns, and flowers), barren land (e.g., rock), cultivated land (fields, also farm outbuildings), and urban (residential and commercial buildings). Images of eight example scenes are shown in Fig. 1A to Fig. 1H (and a further eight in Foster et al., 2004). For the present purposes, the set of natural scenes did not have to be an exhaustive representation of the land-cover classes, merely sufficiently varied to produce a useful range of experimental performance levels. The fact that the main findings from the analysis of performance were stable under repeated resampling of the 21 scenes suggests that the set was indeed large enough, and moreover, it contained no or few outliers. Nevertheless, it remains a finite sample from a potentially infinite population.

Each scene included a gray or colored sphere in the field of view that provided the experimental test surface (indicated by arrows in Fig. 1A to Fig 1G), except for three distant scenes in which a uniform surface (e.g., a roof or wall, Fig. 1H) was used instead (introducing a gray sphere into the scene has been used previously to measure illumination with an RGB camera (Ciurea & Funt, 2003)). A larger image of the test sphere in Fig. 1F is shown in Fig. 2. Scenes were recorded under a cloudless sky with the sun behind the camera, or occasionally recorded under uniform cloud. Any scenes containing visible light sources, including the sky, were excluded and, as far as possible, also those containing water, glass, and other materials producing specular reflections.

In each trial of the experiment, two images of a particular scene were presented in the same position in sequence on a computer-controlled color monitor, each for 1 s, with no interval (a design that yields higher levels of color constancy than side-by-side simultaneous presentation; see Foster et al. (2001*a*)). The images differed in the global illuminant on the scene, which was first a

spatially uniform daylight of correlated color temperature 25,000 K and then one of 6700 K; or first one of 4000 K and then one of 6700 K. During the global illuminant change, the spectral reflectance of the test surface in the second image also changed, by a random amount (see Fig. 2 for examples of illuminant and reflectance changes with detail of Fig. 1F). The observer's task was to decide whether the test surface in the successive images was the same or different; that is, whether an illuminant change alone or an illuminant change accompanied by a change in the spectral reflectance of the test surface had occurred (Craven & Foster, 1992). Responses were made with mouse buttons connected to a computer. Observers were allowed to move their eyes freely. At the beginning of the experimental session, the experimenter indicated the identity of the test surface to the observer verbally and by pointing, and gave a demonstration of illuminant changes and varying sizes of reflectance changes.

Although 21 scenes were available with the 25,000 K illuminant first, this number was reduced to 18 with the 4000 K illuminant first, owing to limits on the gamut of colors displayable by the monitor. The images were viewed binocularly at 100 cm and subtended approx. $18° \times 14°$ visual angle. Depending on the scene, the subtense of the test surface varied from 0.3° to 5.6°, with median 0.7°, interquartile interval 0.5°.

Over scenes, maximum pixel luminance varied from 8 to 33 cd m^{-2}, and minimum pixel luminance from 0 to 1 cd m^{-2} (actual black level of the display was approx. 0.004 cd m^{-2}). The experiment took place in a darkened room. The monitor was surrounded by an illuminated neutral surface with reflected luminance approx. 0.5 cd m^{-2} and was viewed by the observer within a black non-reflecting tunnel. Evidence offered elsewhere (Baraas et al., 2006), suggests that rods did not contribute to discrimination performance. Observers each performed no less than 325 (5 blocks $\times$ 65) trials per scene. Details of the design and randomization are given later (see *Illuminant and reflectance variation*). Each experimental session took about 1 h, and observers participated in no more than two experimental sessions per day, with at least a 1-h gap between the two.

In the control experiment on the effect of changing test-surface position, a subset of 6 scenes was selected yielding mid-range levels of color constancy; the test surface was inserted in a different position in the scene; and the foregoing measurements repeated.

Scene acquisition

The hyperspectral imaging system used to record the scenes for this study was based on a low-noise Peltier-cooled digital camera, which provided a spatial resolution of 1344×1024 pixels (Hamamatsu, model C4742-95-12ER, Hamamatsu Photonics K.K., Hamamatsu, Japan) with a fast tunable liquid-crystal filter (VariSpec, model VS-VIS2-10-HC-35-SQ, Cambridge Research & Instrumentation, Inc., Woburn, MA) mounted in front of the lens, together with an infrared blocking filter (Foster et al., 2004). Focal length was typically set to 75 mm and aperture to f/16 or f/22 to achieve a large depth of focus. The line-spread function of the system was close to Gaussian with standard deviation approximately 1.3 pixels at 550 nm. The intensity response at each pixel, recorded with 12-bit precision, was linear over the entire dynamic range. The peak-transmission wavelength was varied in 10-nm steps over 400–720 nm. The bandwidth (FWHM) was 10 nm at 550 nm, decreasing to 7 nm at 400 nm and increasing to 16 nm at 720 nm.
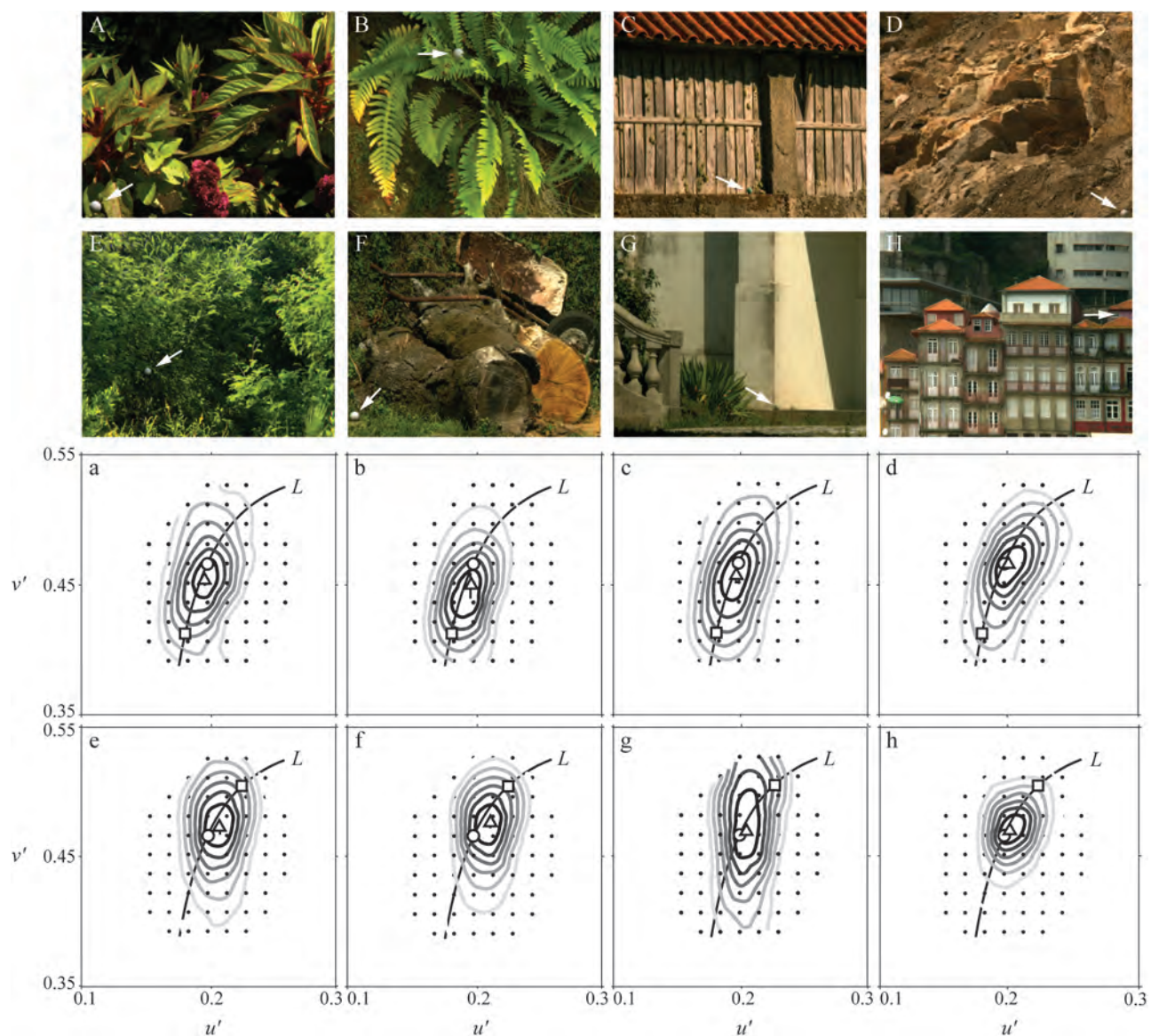
Fig. 1. Example scenes and corresponding plots of surface-color judgments. The images A–H subtended approx. $17° \times 14°$ visual angle in the experiment and each contained a test surface, either a small sphere (A–G) or part of a building (H), indicated by arrows. In the corresponding contour plots a–h, the relative frequency of observers' "illuminant-change" responses to a change in illuminant on the scene and variable test-surface reflectance is shown in the CIE 1976 (u',v') chromaticity diagram as a function of the chromaticity of the reflectance change: the darker the contour, the higher the frequency. The square symbols show the position of the first illuminant (daylight with correlated color temperature 25,000 K in a–d, 4000 K in e–h); the circles the second illuminant (6700 K); and the triangles the mode (and where large enough the bars show ±1 SE), from which the color-constancy index was derived. With perfect constancy, the triangles and circles are coincident. The line marked L is the daylight locus.

Immediately after acquisition, the spectrum of light reflected from a small neutral (Munsell N5 or N7; see details later) reference surface in the scene was recorded with a telespectroradiometer (SpectraColorimeter, PR-650, Photo Research Inc., Chatsworth, CA), the calibration of which was traceable to the National Physical Laboratory. Images were corrected for dark noise, spatial nonuniformities (mainly off-axis vignetting), stray light, and any wavelength-dependent variations in magnification or translation (registration). The effective spectral reflectance at each pixel was then estimated by normalizing the corrected signal against that obtained from the reference surface. Further details are given elsewhere (Nascimento et al., 2002; Foster et al., 2004, 2006).

For each scene, a second hyperspectral image was also recorded with several spheres placed at different points in the field of view. The spheres were covered in Munsell N5 or N7 matt emulsion paint (VeriVide Ltd, Leicester, UK), and, depending on the scene, their diameters varied from 5 mm to 300 mm. The hyperspectral image of one of these spheres was subsequently inserted into the original hyperspectral image to provide the test surface (Fig. 1A to Fig. 1G). The location of the test surface varied
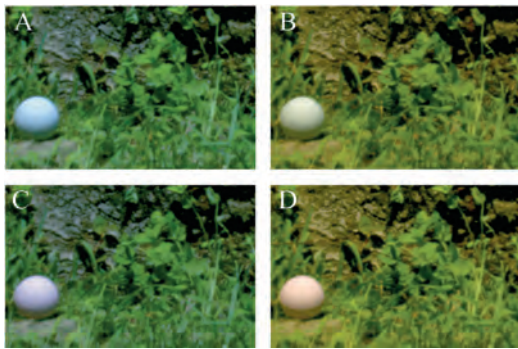
Fig. 2. Examples of illuminant and reflectance changes for detail of scene F of Fig. 1. A and B: a gray sphere in scene under daylight of correlated color temperature 25,000 K and 6700 K, respectively; C and D, a reddish sphere in scene under the same two illuminants. The sequences A to B and C to D both illustrate illuminant changes; the sequence A to D illustrates an illuminant change with a reflectance change.

from near the edge of the image to near the center, chosen partly to accommodate physical constraints and partly to avoid nearby similarly colored surfaces.

Display system and calibration

Stimuli were produced on the screen of a 21-inch RGB CRT color monitor (Trinitron Color Graphic Display, model GDM-F500R, Sony Corp., Tokyo, Japan), with spatial resolution 1600×1200 pixels, controlled by a color-graphics workstation (Fuel V12, Silicon Graphics, Inc., Mountain View, CA) whose 10-bit digital-to-analog converters provided an intensity resolution of 1024 levels on each of the red, green, and blue guns. Each image of approx. 1344×1024 pixels appeared in the central approx. 85% of the displayable area of the screen. A calibrated telespectroradiometer (SpectraColorimeter, PR-650, Photo Research Inc., Chatsworth, CA) and photometer (LMT, L1003, Lichtmesstechnik GmbH, Berlin, Germany) were used to monitor and calibrate the display system. Calibration data included the phosphor coordinates and voltage-intensity look-up tables for the three guns. The monitor was allowed 1 hour to warm up before use.

Images were prepared off-line. For each scene and color of test surface, a radiance image for a particular global scene illumination was obtained by multiplying the effective scene spectral reflectance derived from the hyperspectral data by the global illuminant spectrum (technical details in Foster et al., 2006). The spectral reflectance at any pixel producing out-of-gamut values on the monitor was iteratively affine transformed towards neutral while preserving luminance (i.e., desaturating the pixel) until it was in gamut for all illuminants. The mean proportion of pixels affected was 3% in scenes without flowers, but almost all these pixels were dark (99% had luminance <5% of maximum). Two close-up scenes of high-chroma flowers had 29% of pixels affected, but most of these were also dark (95% with luminance <5% of maximum). Images were saved in 48-bit RGB PNG format. At run time, they were converted to 10-bit-per-channel format and displayed on the monitor under real-time control with in-house software written in C and C++ with OpenGL. Screen refresh rate was approx. 60 Hz.

Routine monitoring of the display system tested whether errors in the displayed CIE (x, y, Y) coordinates of a white test patch were <0.005 in (x, y) and <5% in Y (<10% at low light levels). Tests of image fidelity used images from the experiments. Thus, 35 separate measurements were made with patches of width >20 pixels and approximately constant chromaticity (usually the test surface) or edited to have exactly constant chromaticity, with values in the CIE 1976 chromaticity diagram of $0.167 \leq u' \leq 0.286$ and $0.383 \leq v' \leq 0.541$ at different positions on the screen. Errors were ≤ 0.002 in (u', v'), and <10% in Y. For patches of this size, chromatic errors were therefore less than 15% of the 0.015 grid spacing in the (u', v') plane used to sample observers' responses (small solid points in graphs of Fig. 1a to Fig. 1h). For much smaller patches (width <8 pixels, i.e., <11 arcmin) surrounded by pixels of markedly different color, chromatic errors about twice this size were recorded with the aid of a 2-mm aperture mask fixed to the monitor screen. Because images were presented sequentially in the same position on the screen, position-dependent chromatic errors in each pair of images were the same.

Illuminant and reflectance variation

The ordering of scenes and global illuminant changes was chosen randomly but fixed in each experimental session. The reflectance of the test surface in the first image was manipulated independently of the global illuminant: five different initial test-surface colors were tested in five separate blocks. In each block, the spectral reflectance of the test surface in the second image varied randomly, from trial to trial, in one of 65 ways (all randomization was without replacement). This variation was achieved by a computational device as follows. Suppose that the initial spectral reflectance of the test surface was $R(\lambda; x, y)$ at wavelength λ and position (x, y) and that the global illuminant spectrum was $E(\lambda)$, so that the color signal at the eye was $R(\lambda; x, y)E(\lambda)$. With a change in spectral reflectance to $R'(\lambda; x, y)$, say, the color signal becomes $R'(\lambda; x, y)E(\lambda)$; but the same color signal can be achieved with the original reflectance $R(\lambda; x, y)$ by replacing $E(\lambda)$ locally by a different daylight $E'(\lambda)$ such that $R'(\lambda; x, y)E(\lambda) = R(\lambda; x, y)E'(\lambda)$; the change in reflectance $R'(\lambda; x, y)/R(\lambda; x, y) = E'(\lambda)/E(\lambda)$. Varying the chromaticity of this local illuminant is closely related to varying the chromaticity of the test surface, although the representation of changes in spectral reflectances $R'(\lambda; x, y)/R(\lambda; x, y)$ in terms of changes in local illuminants $E'(\lambda)/E(\lambda)$ has the advantage of a natural colorimetric parameterization that is independent of the initial spectral reflectance of the test surface, so that averages may be calculated over stimuli (see Foster et al., 2001a). These local illuminants were constructed from a linear combination of the daylight spectral basis functions (Judd et al., 1964) whose corresponding chromaticities were drawn from the gamut in the (u', v') diagram consisting of the 65 locations shown by the small solid points in the plots in Fig. 1a to Fig. 1h, with spacing 0.015. The same technique was used to produce the five different initial test-surface spectra, whose corresponding (u', v') chromaticities were shifted from the original neutral Munsell N5 or N7 by $(0.015, 0)$, $(0, 0.015)$, $(-0.015, 0)$, $(0, -0.015)$, and $(0, 0)$.

Observers

Twelve observers (5 male, 7 female), aged 17–30 years, took part in the experiment with the 25,000 K illuminant first, and a subset

of eight (3 male, 5 female) with the 4000 K illuminant first (except for one scene where seven observers were available). All observers had normal or corrected-to-normal visual acuity and normal color vision as assessed with Ishihara pseudoisochromatic plates, the Farnsworth-Munsell 100-Hue test, and Rayleigh and Moreland anomaloscopy. The experiments were conducted in accordance with principles embodied in the Declaration of Helsinki (Code of Ethics of the World Medical Association), and they were approved by the Research Ethics Committee of the University of Manchester. All observers were unaware of the purpose of the experiment. Seven observers participated in the control on changing test-surface location.

Analysis

For each scene, the relative frequency of "illuminant-change" responses, pooled over observers, was calculated as a function of the chromaticity of the local illuminant in the (u', v') chromaticity diagram. The frequency plots were smoothed by a two-dimensional non-parametric locally weighted quadratic regression ("loess" Cleveland & Devlin, 1988), and contour plots derived as shown in Fig. 1a to Fig 1h (cf. Bramwell & Hurlbert, 1996, who used a two-dimensional Gaussian model; Foster et al., 2003). Each contour represents a constant relative frequency: the darker the contour, the higher the frequency; differences between contours represent approx. 0.10–0.15 differences in frequency. The position of the maximum of each frequency distribution was obtained numerically from the loess analysis (shown by the triangles in Fig. 1a to Fig. 1h). If the observer had perfect color constancy, that position would coincide with the position of the second illuminant (circles). To summarize the error in the surface-color judgment (i.e., the bias) a standard color-constancy index (Arend et al., 1991) was then derived. That is, if a is the distance between the positions of the maximum (triangle) and the 6700 K illuminant (circle) and b the distance between the positions of the 25,000 K or 4000 K illuminant (square) and 6700 K illuminant (circle), then the constancy index is $1 - a/b$. Perfect constancy corresponds to an index of unity and perfect inconstancy corresponds to an index of 0, where the response peak coincides with the first global illuminant. The standard error (SE) of this index was estimated with a bootstrap procedure, based on 1000 replications, with resampling over observers (Efron & Tibshirani, 1993).

The constancy indices for each scene and illuminant change were assessed against possible global image properties in a linear regression analysis, the indices weighted by their estimated SEs. Global properties were here defined as those functions of the whole image that did not depend on the properties of the test surface, in particular, its spatial location. As already indicated, the properties considered were of two kinds: one was colorimetric, based on CIELAB lightness L^*, hue h_{ab}, and chroma C_{ab}^* (which correlates with colorfulness as a proportion of the brightness of a similarly illuminated area that appears white; see e.g., Fairchild (2005)); the other was receptor-based, involving simple combinations of excitations in long-, medium-, and short-wavelength-sensitive cones (L, M, and S), calculated as in Foster et al. (2004). As a result of previous work, one of these receptoral properties included the spatial ratio of cone excitations between pairs of points in the image (Foster & Nascimento, 1994; Nascimento & Foster, 1997), although here evaluated over all surfaces rather than just between the test surface and other surfaces or averages over surfaces in the scene (Amano & Foster, 2004), possibly in some

nonlinear form (Lucassen & Walraven, 1993, 2005). Differences in ratios across images were calculated in the following way (Nascimento & Foster, 1997). If $\boldsymbol{r}_{i,j} = (r_{ij}^L, r_{ij}^M, r_{ij}^S)$ is the triplet of cone-excitation ratios for L, M, and S cones obtained at each pair of distinct pixels i, j in an image, and $|\boldsymbol{r}_{i,j}|$ represents the Euclidean norm $[(r_{ij}^L)^2 + (r_{ij}^M)^2 + (r_{ij}^S)^2]^{1/2}$, then the mean relative deviation between the images of a scene under first and second illuminants, (1) and (2), was defined by $\mathrm{MRD}(\boldsymbol{r}) = \mathrm{E}[|\boldsymbol{r}(1) - \boldsymbol{r}(2)|/\min\{|\boldsymbol{r}(1)|, |\boldsymbol{r}(2)|\}]$, where E represents the average over pairs of pixels i, j (see Table 1). In practice, to avoid the effects of correlations due to the 1.3-pixel line-spread function, only alternate pixels in the images were used in the calculation, giving a total of typically $(1344 \times 1024)/4 = 344,064$ pixels.

Notice that colorimetric and receptor-based properties were used as explanatory factors, rather than experimental variables such as scene illuminant, because they represent the information available to the observer in the color signal. Although the distinction between colorimetric and receptor-based properties is not intrinsic, for each may be expressed in terms of others (e.g., chroma expressed as a function of cone excitations), they have different interpretations (Walsh, 1999; Smithson, 2005). More important is the stability of the linear regression, which requires that explanatory factors should not be highly correlated (Draper & Smith, 1998). To this end, combinations of factors that were linearly dependent were explicitly excluded from the analysis.

Results and comment

From the frequency plots of "illuminant-change" responses, color-constancy indices were obtained from the 21 scenes under a change in daylight from a correlated color temperature of 25,000 K to 6700 K and from the 18 scenes under a change in daylight from a correlated color temperature of 4000 K to 6700 K. For the eight example scenes in Fig. 1, indices for scenes A to D under illuminant changes of 25,000 K to 6700 K were 0.77, 0.69, 0.81, and 0.94, respectively, (plots a to d) and for scenes E to H under illuminant changes of 4000 K to 6700 K were 0.75, 0.65, 0.90, and 0.88, respectively (plots e to h). Very high indices are not, however, special to non-vegetated scenes (Fig. 1D); for example, with a close-up of a yellow lily (see Foster et al. (2004), Fig. 1, top right) the color-constancy index was 0.97 with an illuminant change of 25,000 K to 6700 K.

To explain this variation with scene and illuminant change, the regression analysis referred to in Methods was applied to the list of image statistics in Table 1. As an example of how a particular image statistic can account for the variation, Fig. 3 shows color-constancy index plotted against the log of the mean relative deviation in spatial cone-excitation ratios for each scene and illuminant change. A log transformation was used to accommodate the extrema in these ratios, and the axis has been reversed so that the level of constancy generally improves as the difference in cone-excitation ratios across the two illuminants decreases. The proportion R^2 of variance accounted for in this regression was 43%, corresponding to a product moment correlation coefficient of 0.66, which is statistically highly significant ($t = -5.3$, 2-tailed $P < 0.00001$).

The explanatory power of each image statistic was summarized by this quantity R^2, with its estimated SE based on a bootstrap with resampling over scenes and illuminant changes (Efron & Tibshirani, 1993). The global statistics in Table 1 are listed in ascending order of R^2, and consist of the mean (denoted by E), SD, and mean
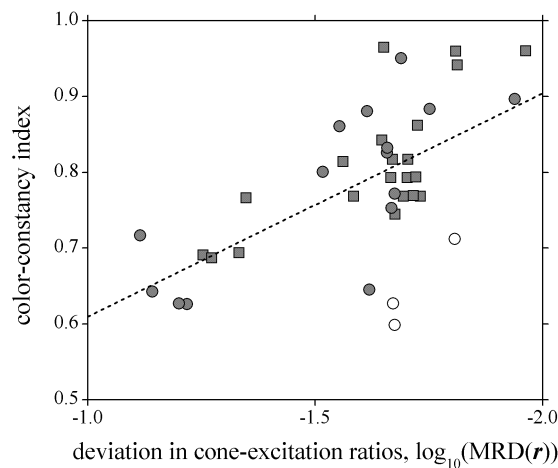
Fig. 3. Variation of surface-color judgments. Color-constancy index is plotted against the log of the mean relative deviation in spatial cone-excitation ratios for each scene under two illuminants. Filled squares are for 21 scenes with first illuminant a daylight with correlated color temperature 25,000 K and second illuminant 6700 K (data from 12 observers); filled circles are for 18 scenes with first illuminant 4000 K and second illuminant 6700 K (data from 7–8 observers); open circles are for images in the quartile with the highest mean chroma under the first illuminant or highest difference in mean chroma. The dotted line is an unweighted linear regression.

relative deviation (MRD) of basic colorimetric or receptor-based properties.

Combinations of statistics were formed additively. Higher-order moments, namely skewness and kurtosis, were found to offer no particular advantage over these quantities.

In general, colorimetric properties provided a limited explanation of the variance in color-constancy index over scenes and illuminants, at most 28% from the standard deviation of the chroma of the second image $SD(C_{ab}^*(2))$. By contrast, receptor-based properties were more successful, with log mean relative deviation in cone-excitation ratios $\log_{10}(MRD(r))$ accounting for most variance, namely 43%, as already noted. Increasing the number of explanatory properties from one to two or more increased R^2, but by progressively smaller amounts. Thus, with two properties, the factor in combination with log mean relative deviation in cone-excitation ratios giving the largest increase in R^2, from 43% to 54%, was mean chroma of the first image, $E(C_{ab}^*(1))$. Including the interaction of these two factors as a third term in the regression increased R^2 by only 1.7% and did not improve the fit significantly $(F(35,36) = 1.32, P > 0.2)$.

With three properties, the factor in combination with the previous two giving the largest increase in R^2, from 54% to 63%, was the mean difference in chroma between first and second images, $E(\Delta C_{ab}^*)$, equivalent to adding the chroma of the second image as an independent factor. As with two factors, including the pairwise interactions of three factors as three additional terms in the regression increased R^2 by a further 1.9%, and did not improve the fit significantly $(F(32,35) = 0.57, P > 0.5)$. Interactions were not considered further.

With four properties, the factor in combination with the previous three giving the largest increase in R^2, from 63% to 70% (corresponding to a multiple correlation coefficient of 0.84), was

the mean difference in hue between first and second images $E(\Delta h_{ab})$. Although here added step-by-step, the same four factors proved optimal in an unconstrained fit, that is, without imposing the results of the previous fits with one, two, and three factors.

Table 2 shows the coefficients of the four factors in the optimal fit, each significantly different from zero. Their correlations ranged from 0.07 to 0.23. Overall, they provided good fits to the effects of scene and illuminant: adding a fifth property increased R^2 by 3% at most, and just failed to improve the fit significantly $(F(33,34) = 3.80, P = 0.06)$.

The coefficient for log mean relative deviation in Table 2 is negative (i.e., constancy improved as mean relative deviation decreased) and also negative for the mean chroma of the first image and for the mean difference in hue between images (i.e., constancy worsened as each increased). The influence of chroma is indicated in Fig. 3, where the points marked by open circles fall in the quartile of scenes with the highest mean chroma under the first illuminant or highest difference in mean chroma.

Test-surface size and position

The test-surface size varied in visual angle by a factor of about 18 over scenes, but it had no detectable effect on color-constancy index: the proportion R^2 of variance accounted for was 1%; the slope of the regression was -0.01 with SE 0.05. For the control experiment in which the position of the test surface was changed, there was a modest change in color-constancy index: the mean absolute difference in values across scenes was 0.14 (cf. the range in Fig. 3). Log mean relative deviation in cone-excitation ratios accounted for 38% of variance in this difference, a proportion which rose to 58% with the addition of the mean difference in hue between first and second images, although the improvement in the fit was not significant $(F(3,4) = 1.37, P = 0.3)$.

Spatial statistics

Although colorimetric and receptor-based descriptions of natural scenes were the properties of interest here, it is possible that spatial properties alone might influence color constancy (e.g. Courtney et al., 1995; Jenness & Shevell, 1995; Zaidi et al., 1997; Brenner & Cornelissen, 1998; Wachtler et al., 2001; Zaidi, 2001; Werner, 2003; Hurlbert & Wolf, 2004). A useful spatial statistic for natural images is the spatial power or amplitude spectrum, which is a second-order statistic. In general, the amplitude of the spectrum falls off as the reciprocal of the spatial frequency (Field, 1987). Both second- and higher-order statistics are important in determining spatial discrimination performance (e.g., Knill et al., 1990; Thomson & Foster, 1997; Párraga et al., 2005)

To test whether spatial statistics might be relevant to the present analysis, the discrete 2-dimensional Fourier transform of the luminance distribution in each scene under a daylight of correlated color temperature 6700 K was calculated and the log of the absolute value of the amplitude plotted against log spatial frequency averaged over horizontal and vertical directions (results not shown here). On these log–log plots, the amplitude spectra were well described by linear regressions, with the correlation coefficient varying from 0.91 to 0.98 over the 21 scenes. The gradient varied from -1.5 to -1.0 (cf. Knill et al., 1990; Tolhurst et al., 1992; Thomson & Foster, 1997) but explained little of the variation in color-constancy index: the proportion R^2 of variance accounted for was 5%, not significantly different from zero $(P = 0.15)$.

Table 1. *Global image properties in ascending order of proportion of variance R^2 in color-constancy index explained by a linear regression on the corresponding statistic*

Global property	Statistic	R^2 (%)	SE (%)
Mean hue difference	$E(\Delta h_{ab})$	0.0	5.3
Standard deviation of hue image 1	$SD(h_{ab}(1))$	0.1	10.9
Standard deviation of lightness image 1	$SD(L^*(1))$	0.3	5.5
Standard deviation of lightness image 2	$SD(L^*(2))$	0.4	5.5
Mean hue image 1	$E(h_{ab}(1))$	0.4	10.5
Mean lightness image 1	$E(L^*(1))$	0.5	7.9
Standard deviation of hue difference	$SD(\Delta h_{ab})$	2.2	8.3
Standard deviation of hue image 2	$SD(h_{ab}(2))$	3.1	10.8
Standard deviation of chroma difference	$SD(\Delta C^*_{ab})$	6.2	8.9
Standard deviation of color difference	$SD(\Delta E^*_{ab})$	7.6	9.5
Mean lightness difference	$E(\Delta L^*)$	13.6	12.3
Standard deviation of cone excitations image 1	$SD(q(1))$	17.1	13.9
Standard deviation of lightness difference	$SD(\Delta L^*)$	17.2	10.6
Mean chroma difference	$E(\Delta C^*_{ab})$	22.8	13.7
Mean chroma image 1	$E(C^*_{ab}(1))$	23.5	13.9
Standard deviation of chroma image 1	$SD(C^*_{ab}(1))$	26.7	11.8
Standard deviation of chroma image 2	$SD(C^*_{ab}(2))$	27.5	12.1
Standard deviation of difference in cone excitations	$SD(\Delta q)$	29.7	13.8
Mean relative deviation in cone-excitation ratios	$\log_{10}(MRD(r))$	43.2	14.5
Mean relative deviation in cone-excitation ratios and mean chroma image 1	$\log_{10}(MRD(r)) + E(C^*_{ab}(1))$	53.9	11.5
Mean relative deviation in cone-excitation ratios, mean chroma image 1, and mean chroma difference	$\log_{10}(MRD(r)) + E(C^*_{ab}(1)) + E(\Delta C^*_{ab})$	63.3	10.8
Mean relative deviation in cone-excitation ratios, mean chroma image 1, mean chroma difference, and mean hue difference	$\log_{10}(MRD(r)) + E(C^*_{ab}(1)) + E(\Delta C^*_{ab}) + E(\Delta h_{ab})$	70.5	8.0

Statistics were based on one or both images of a scene under two successive illuminants. Thus, for each variable X, the mean $E(X)$ and standard deviation $SD(X)$ were evaluated over the N pixels of the image; e.g. if lightness L^* at pixel i was L^*_i, then $E(L^*) = \Sigma_i L^*_i/N$. Differences ΔX between variables were taken between corresponding pixels of the first and second images (1) and (2), so $\Delta X_i = X(1)_i - X(2)_i$. The variable q is the triplet of L, M, S cone excitations (q^L_i, q^M_i, q^S_i) obtained at each pixel i, and the variable r is the triplet of cone-excitation ratios (r^L_i, r^M_i, r^S_i) obtained at each pair of distinct pixels i, j; e.g., $r^L_{ij} = q^L_i/q^L_j$, with $q^L_j > 0$. The mean relative deviation $MRD(q)$ of q was given by $E[|\Delta q|/\min\{|q(1)|, |q(2)|\}]$, where $|q|$ is the Euclidean norm. Values of CIELAB lightness L^*, chroma C^*_{ab}, and hue h_{ab} were calculated from the adaptation model CMCCAT2000 (see e.g. Li et al., 2002) and colorimetric differences from CIEDE2000 (Luo et al., 2001), so $\Delta E^*_{ab} = \Delta E_{00}$. The coefficients of the additive model shown in the last row are given in Table 2. The estimated SE of R^2 was obtained from a bootstrap (Efron & Tibshirani, 1993).

General discussion

With the variety and complexity of natural scenes, it seems unlikely that any single image property would provide a useful predictor of surface-color judgments under different illuminants. Yet, as the present analysis has shown, it is possible to explain 43% of the variance in color-constancy index by the mean relative deviation in spatial cone-excitation ratios across images of natural scenes under successive illuminants: in short, the smaller the deviation, the better the constancy. With the addition of other global image properties, a further 20% of the variance could be explained by the mean chroma of the scene under the first illuminant and its difference from the mean chroma of the scene under the second illuminant, and a further 7% by the difference in mean

Table 2. *Values of four most important global image statistics accounting for variation in color-constancy index with scene and illuminant*

Global property	Statistic	Value	SE	P
(intercept)	1	0.34	0.11	0.004
Mean relative deviation in cone-excitation ratios	$\log_{10}(MRD(r))$	-0.326	0.062	0.00001
Mean chroma image 1	$E(C^*_{ab}(1))$	-0.0042	0.0013	0.003
Mean chroma difference	$E(\Delta C^*_{ab})$	0.046	0.012	0.0006
Mean hue difference	$E(\Delta h_{ab})$	-0.0037	0.0013	0.007

The estimated value and standard error (SE) of each coefficient in the linear regression is shown with corresponding P-value for a 2-tailed t-test against the hypothesis that the coefficient is zero (d.f. = 34).

hue. Taken together, all four factors accounted for 70% of the variance and provided a good fit to the effects of scene and of illuminant change on color constancy.

Why should deviations in spatial cone-excitation ratios have such a strong effect on surface-color judgments? It has been noted elsewhere that, in general, such ratios, which can also be calculated across post-receptoral combinations (Zaidi et al., 1997) and spatial averages of cone signals (Amano & Foster, 2004), are almost invariant under changes in illuminant on scenes of Mondrian-like patterns of Munsell papers (Foster & Nascimento, 1994) and on natural scenes (Nascimento et al., 2002). Even so, they are not exactly invariant, and when deviations occur they are interpreted by observers, incorrectly, as evidence of reflectance changes rather than of an illuminant change (Nascimento & Foster, 1997). This sensitivity to changes in cone-excitation ratios may underlie observers' judgments of transparency with overlapping surfaces (Westland & Ripamonti, 2000), the spatially parallel detection of violations in color constancy in single and multiple targets (Foster et al., 2001b), and asymmetric color matching with center-surround geometry (Tiplitz Blackwell & Buchsbaum, 1988; Amano & Foster, 2004). In the present analysis, it was the deviations in ratios evaluated over all the surfaces in the scene that helped explain the variation in judging test-surface color. Given observers' misinterpretation of these deviations, it is perhaps not surprising that their occurrence over the whole field affected performance.

For a given level of deviation in cone-excitation ratios, the worsening of surface-color judgments in scenes that under the first illuminant had high chroma may have been due to a reduction in receptor response range. On theoretical grounds, highly chromatic scenes have been linked to poor color constancy, either through the increased variance of spatial cone-excitation ratios (Nascimento et al., 2004) or through effects involving chromatic-adaptation transforms (Morovič & Morovič, 2005). The improvement in surface-color judgments with increasing chromatic difference between the scenes under the two illuminants (equivalent to increasing the chroma of the second image) is harder to interpret. Although statistically significant, its effect was similar in magnitude to decreasing the hue difference between the scenes under the two illuminants.

Global image statistics do not of course account for all of the variation in observers' performance. In addition to individual differences, there are scene-specific effects involving remote elements in the field of view as well as local effects of test-surface surround noted earlier (e.g., Shevell & Wei, 1998; Kraft & Brainard, 1999; Wachtler et al., 2001; Brenner et al., 2003). Although differences in spatial amplitude spectra did not influence performance across scenes, a more comprehensive analysis might attempt to include these local and remote effects and other properties of the test surface, including its position. Nevertheless, it is interesting that global statistics account for so much of the variation in performance, and, moreover, that the effects of changing test-surface position can be interpreted in terms of the same explanatory factors.

Acknowledgments

References

AMANO, K. & FOSTER, D.H. (2004). Colour constancy under simultaneous changes in surface position and illuminant. *Proceedings of the Royal Society of London Series B—Biological Sciences* **271**, 2319–2326.

AREND, L.E., JR., REEVES, A., SCHIRILLO, J. & GOLDSTEIN, R. (1991). Simultaneous color constancy: Papers with diverse Munsell values. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* **8**, 661–672.

BARAAS, R., FOSTER, D.H., AMANO, K. & NASCIMENTO, S.M.C. (2006). Anomalous trichromats' judgments of surface color in natural scenes under different daylight. *Visual Neuroscience* **23** (#23-329)■■–■■.

BRAMWELL, D.I. & HURLBERT, A.C. (1996). Measurements of colour constancy by using a forced-choice matching technique. *Perception* **25**, 229–241.

BRENNER, E. & CORNELISSEN, F.W. (1998). When is a background equivalent? Sparse chromatic context revisited. *Vision Research* **38**, 1789–1793.

BRENNER, E., RUIZ, J.S., HERRÁIZ, E.M., CORNELISSEN, F.W. & SMEETS, J.B.J. (2003). Chromatic induction and the layout of colours within a complex scene. *Vision Research* **43**, 1413–1421.

BROWN, R.O. & MACLEOD, D.I.A. (1997). Color appearance depends on the variance of surround colors. *Current Biology* **7**, 844–849.

CIUREA, F. & FUNT, B. (2003). A large image database for color constancy research. In *Eleventh Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, pp. 160–164. Scottsdale, AZ: Society for Imaging Science and Technology.

CLEVELAND, W.S. & DEVLIN, S.J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* **83**, 596–610.

COURTNEY, S.M., FINKEL, L.H. & BUCHSBAUM, G. (1995). Network simulations of retinal and cortical contributions to color constancy. *Vision Research* **35**, 413–434.

CRAVEN, B.J. & FOSTER, D.H. (1992). An operational approach to colour constancy. *Vision Research* **32**, 1359–1366.

DRAPER, N.R. & SMITH, H. (1998). *Applied Regression Analysis*, 3rd ed. New York: Wiley.

EFRON, B. & TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

FAIRCHILD, M.D. (2005). *Color Appearance Models*. Chichester: John Wiley & Sons, Ltd.

FEDERAL GEOGRAPHIC DATA COMMITTEE. (1997). National Vegetation Classification Standard. FGDC-STD-005. U.S. Geological Survey, Reston, Virginia.

FIELD, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* **4**, 2379–2394.

FOSTER, D.H. (2003). Does colour constancy exist? *Trends in Cognitive Sciences* **7**, 439–443.

FOSTER, D.H., AMANO, K. & NASCIMENTO, S.M.C. (2001a). Colour constancy from temporal cues: Better matches with less variability under fast illuminant changes. *Vision Research* **41**, 285–293.

FOSTER, D.H., AMANO, K. & NASCIMENTO, S.M.C. (2003). Tritanopic colour constancy under daylight changes? In *Normal & Defective Colour Vision*. ed. MOLLON, J.D., POKORNY, J. & KNOBLAUCH, K., pp. 218–224. Oxford, UK: Oxford University Press.

FOSTER, D.H., AMANO, K. & NASCIMENTO, S.M.C. (2006). Frequency of metamerism in natural scenes. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* (in press).

FOSTER, D.H. & NASCIMENTO, S.M.C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of London Series B—Biological Sciences* **257**, 115–121.

FOSTER, D.H., NASCIMENTO, S.M.C. & AMANO, K. (2004). Information limits on neural identification of colored surfaces in natural scenes. *Visual Neuroscience* **21**, 331–336.

FOSTER, D.H., NASCIMENTO, S.M.C., AMANO, K., AREND, L., LINNELL, K.J., NIEVES, J.L., PLET, S. & FOSTER, J.S. (2001b). Parallel detection of violations of color constancy. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8151–8156.

HURLBERT, A. & WOLF, K. (2004). Color contrast: A contributory mechanism to color constancy. *Progress in Brain Research* **144**, 147–160.

JENNESS, J.W. & SHEVELL, S.K. (1995). Color appearance with sparse chromatic context. *Vision Research* **35**, 797–805.

JUDD, D.B., MACADAM, D.L. & WYSZECKI, G. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America* **54**, 1031–1040.

KNILL, D.C., FIELD, D. & KERSTEN, D. (1990). Human discrimination of fractal images. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* **7**, 1113–1123.

KRAFT, J.M. & BRAINARD, D.H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 307–312.

KULIKOWSKI, J.J., STANIKUNAS, R., JURKUTAITIS, M., VAITKEVICIUS, H. & MURRAY, I.J. (2001). Colour and brightness shifts for isoluminant samples and backgrounds. *Color Research and Application* **26**, S205–S208.

LI, C.-J., LUO, M.R., RIGG, B. & HUNT, R.W.G. (2002). CMC 2000 chromatic adaptation transform: CMCCAT2000. *Color Research and Application* **27**, 49–58.

LUCASSEN, M.P. & WALRAVEN, J. (1993). Quantifying color constancy: Evidence for nonlinear processing of cone-specific contrast. *Vision Research* **33**, 739–757.

LUCASSEN, M.P. & WALRAVEN, J. (2005). Separate processing of chromatic and achromatic contrast in color constancy. *Color Research and Application* **30**, 172–185.

LUO, M.R., CUI, G. & RIGG, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research and Application* **26**, 340–350.

MOROVIČ, J. & MOROVIČ, P. (2005). Can highly chromatic stimuli have a low color inconstancy index? In *Thirteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, pp. 321–325. Scottsdale, AZ: Society for Imaging Science and Technology.

NASCIMENTO, S.M.C., DE ALMEIDA, V.M.N., FIADEIRO, P. T. & FOSTER, D. H. (2004). Minimum-variance cone-excitation ratios and the limits of relational color constancy. *Visual Neuroscience* **21**, 337–340.

NASCIMENTO, S.M.C., FERREIRA, F.P. & FOSTER, D.H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* **19**, 1484–1490.

NASCIMENTO, S.M.C. & FOSTER, D.H. (1997). Detecting natural changes of cone-excitation ratios in simple and complex coloured images. *Proceedings of the Royal Society of London Series B—Biological Sciences* **264**, 1395–1402.

PÁRRAGA, C.A., TROSCIANKO, T. & TOLHURST, D.J. (2005). The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model. *Vision Research* **45**, 3145–3168.

SHEVELL, S.K. & WEI, J. (1998). Chromatic induction: Border contrast or adaptation to surrounding light? *Vision Research* **38**, 1561–1566.

SMITHSON, H.E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society B—Biological Sciences* **360**, 1329–1346.

THOMSON, M.G.A. & FOSTER, D.H. (1997). Role of second- and third-order statistics in the discriminability of natural images. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* **14**, 2081–2090.

TIPLITZ BLACKWELL, K. & BUCHSBAUM, G. (1988). Quantitative studies of color constancy. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* **5**, 1772–1780.

TOLHURST, D.J., TADMOR, Y. & CHAO, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics* **12**, 229–232.

UNESCO (1973). International classification and mapping of vegetation. Paris, France: UNESCO Publishing.

WACHTLER, T., ALBRIGHT, T.D. & SEJNOWSKI, T.J. (2001). Nonlocal interactions in color perception: Nonlinear processing of chromatic signals from remote inducers. *Vision Research* **41**, 1535–1546.

WALSH, V. (1999). How does the cortex construct color? *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13594–13596.

WEBSTER, M.A. & MOLLON, J.D. (1995). Color constancy influenced by contrast adaptation. *Nature* **373**, 694–698.

WERNER, A. (2003). The spatial tuning of chromatic adaptation. *Vision Research* **43**, 1611–1623.

WESTLAND, S. & RIPAMONTI, C. (2000). Invariant cone-excitation ratios may predict transparency. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* **17**, 255–264.

WYSZECKI, G. & STILES, W.S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. New York: John Wiley & Sons.

ZAIDI, Q. (2001). Color constancy in a rough world. *Color Research and Application* **26**, S192–S200.

ZAIDI, Q., SPEHAR, B. & DEBONET, J. (1997). Color constancy in variegated scenes: Role of low-level mechanisms in discounting illumination changes. *Journal of the Optical Society of America A. Optics, Image Science, and Vision* **14**, 2608–2621.