Research Article Vol. 42, No. 5 / May 2025 / Journal of the Optical Society of America A B133

Journal of the Optical Society of America A

OPTICS, IMAGE SCIENCE, AND VISION

Frequency of mismatching surface colors in the wild

David H. Foster

Department of Electrical & Electronic Engineering, University of Manchester, Manchester M13 9PL, UK (d.h.foster@manchester.ac.uk)

Received 3 July 2024; revised 19 November 2024; accepted 20 December 2024; posted 20 December 2024; published 31 January 2025

Colored surfaces may appear to match in one viewing condition but not in another, usually because of a change in illumination. The aim of this computational study was to estimate the frequency of mismatching outdoors under natural, uncontrolled, illumination changes, unlike the purely spectral changes in studies of illuminant metamerism. Data were taken from hyperspectral radiance images acquired at intervals of 1 min to more than 4 h. For pairs of randomly chosen surfaces in a scene, the relative frequency of their appearing initially the same and different later was around 10^{-4} to 10^{-3}, depending on color difference. However, if they already appeared the same, the relative frequency was higher, around 6% to over 60%, much higher than for illuminant metamerism, suggesting that real-world lighting changes may well impair surface identification by color.

https://doi.org/10.1364/JOSAA.534385

1. INTRODUCTION

If two surfaces in a scene look the same at one moment, will they look different later? The answer is probably no, providing that the only change in viewing conditions is in the spectral power distribution of the illumination. Yet this proviso, which has been common in studies of color appearance, is not always realistic with outdoor scenes, where natural, uncontrolled, illumination changes are more uneven, complex, and unpredictable than changes in spectra alone.

The phenomenon of two surfaces appearing to have the same color under one illuminant with a certain spectrum and different under another is termed illuminant metamerism [1,2]. Its frequency of occurrence outdoors has been estimated in a previous computational analysis of scenes whose effective spectral reflectances [3] at each point were known. Under a simulated change in illuminant from a daylight with a correlated color temperature (CCT) of 25,000 K to one of 4000 K [4,5], characteristic, respectively, of light from the north or polar sky and light from the setting Sun, the relative frequencies of metameric pairs, as a proportion of all pairs of matching surfaces in a scene, ranged from around 10^{-2} to 10^{-1}, averaged across 50 such scenes [6].

These frequencies, which are high enough to affect visual inferences about material identity, are conditional in that the pairs are assumed initially to match, or, more precisely, to have color differences less than some threshold for discrimination. When pairs were chosen without constraint, the resulting unconditional relative frequencies were much lower, ranging from around 10^{-6} to 10^{-4}, for the same illuminant change [6]. Similar estimates have been recorded with large ensembles of individual object reflectance spectra [7].

For clarity, the analyses of empirical spectra should be distinguished from those of hypothetical spectra that are metameric [4,8–10] but which may not represent spectra in the real world [11,12], where reflecting properties are complicated by typically non-Lambertian behavior.

By contrast with changes in illuminant spectra, illumination changes outdoors are more like a redistribution of light over a scene [13,14]. Changing solar elevation and fluctuating atmospheric transmission produce both large-scale and local changes in direct and indirect illumination [15,16], with little change in color cast during most of the day. In fact, the CCT of daylight is largely constant ([17], Fig. 5) ([18], Fig. 7). Only for short periods at dawn or dusk is there a pronounced spectral shift away from or toward longer wavelengths [19].

Figure 1 shows an example of the effect of solar elevation on the color appearance of natural non-Lambertian surfaces at 14:33, 15:13, and 16:12. The pairs of square patches are enlarged copies of the sample areas indicated in the images.

Changes in the upper pair are transient. They first match, then mismatch, and then match again as the cast shadow moves over the coplanar areas. Changes in the lower pair are progressive. They first match, then mismatch, and then mismatch further as the angle of incidence of the solar beam changes on the non-coplanar areas. There are also small changes in the spectrum of the direct beam, which affect neither matches nor mismatches. With larger, heterogeneous scenes, the differences in appearance are more varied still, as illustrated later. These

Fig. 1. Effect of solar elevation on the color appearance of pairs of surface areas. The sRGB images were rendered from hyperspectral radiance images acquired at 14:33, 15:13, and 16:12 [13]. The square patches are enlarged copies of the indicated areas averaged over approximately 1 mm^2 of the rendered image. The four locations are identical in all three images. The correlated color temperatures of the direct illumination on the scene were, from left to right, 5615 K, 5526 K, and 5345 K.

changes in appearance are clearly different from those produced by changes in illuminant spectra alone.

The aim of this study was to estimate computationally the frequency of mismatching surface colors in outdoor scenes undergoing natural illumination changes, over short and extended time intervals. Both conditional and unconditional relative frequencies were found to be much higher than with simulated changes in daylight spectra.

2. MATERIALS AND METHODS

All computations were performed in the MATLAB computing environment (Version R2023b, The MathWorks, Inc., Natick, MA).

A. Hyperspectral Radiance Image Pairs

Hyperspectral radiance images were taken from sources listed in the data availability statement. Technical accounts of the images and their acquisition with a wavelength-scanning camera and subsequent processing have been given elsewhere [6,14,20]. The hyperspectral images were acquired at different times during the day from 18 outdoor scenes in the Minho region of Portugal. The scenes are illustrated by the sRGB images in Fig. 2 rendered from the hyperspectral data.

The hyperspectral images were grouped into two sets. The short-interval set consisted of just two images from each of the 18 scenes in Fig. 2, with intervals of about 1–15 min [14,20,21], determined by the imaging conditions. The extended-interval set consisted of multiple images from each of the four scenes in the top row of Fig. 2 with intervals of about 1 min to over 4 h [13]. Differences in time of day thus took the place of differences in CCT used with simulated changes in daylight spectra [6]. Each image had dimensions of 1344 pixels × 1024 pixels and spectral range of 400, 410, … , 720 nm. The shortest interval of about 1 min between images was set by the process of image acquisition [3]. The reported intervals are rounded to whole minutes.

Standard MATLAB routines were used to register each of the images over wavelength by uniform scaling and translation to compensate for variations in optical image size [3]. Because of their different acquisition times, the two members of each image pair were also registered with respect to each other by translation to compensate for any residual differences in optical alignment. Goodness of alignment was determined to within 0.1 pixels by maximizing the mutual information between the images, estimated with an offset form of the Kozachenko–Leonenko estimator [22,23], which converges more rapidly and accurately than the original estimator [24].

For some scenes, padding artifacts a few pixels wide were visible at the edges of the images and were subsequently trimmed. Images were calibrated for spectral radiance against independent spectral radiance data recorded from one or more neutral reference surfaces embedded in each scene or introduced into the field of view [3].

The angular subtense of each scene at the hyperspectral camera was approximately 6.9° × 5.3°. Since images were not averaged over successive acquisitions to reduce noise, each was downsampled by spatial averaging over 2 × 2 pixels. Each pixel therefore represented an image spectral radiance integrated over an elementary area subtending approximately 0.6 × 0.6 arc min, possibly containing a mixture of several distinct spectral reflectances at some finer scale, unresolved by either the instrument or the eye [13].

B. Natural Illumination Changes

With much larger scenes than in Fig. 1, changes in solar elevation can affect surface appearance differently at different depths. Figure 3 shows sRGB images rendered from a hyperspectral radiance image of the scene in Fig. 2(d) at 16:07 and 17:11 and a grayscale plot of the absolute differences in total radiance at each point of the scene. Spectral radiances were scaled linearly to match physically over the approximately uniform sky region, where fluctuations in the intensity of the direct light, of the

Fig. 2. Color images of the 18 scenes used in this study. Each of the scenes provided single pairs of hyperspectral radiance images separated by short intervals of about 1–15 min [14,20,21]. Each of the four scenes in the top row also provided multiple pairs of hyperspectral images separated by extended intervals of about 1 min to 4.6 h [13]. Adapted from Ref. [14] (Fig. 2) under CC BY 4.0.

order of 0.25%, were smaller than in light reflected from surfaces [14] and of the same order as independent estimates from pyrheliometer recordings [25,26].

Differences in the reflected light are distributed over the scene but tend to increase as the distance of the reflecting surface decreases from the top of the grayscale image to the bottom. The ratios of the mean and maximum absolute differences in radiance relative to the mean radiance were about 9% and 340%, respectively (the grayscale plot was clipped to 99% of maximum). Similar though smaller differences were found with intervals of the order of minutes.

The number of scenes in Fig. 2 with these large-scale illumination changes was roughly the same as the number with more local changes.

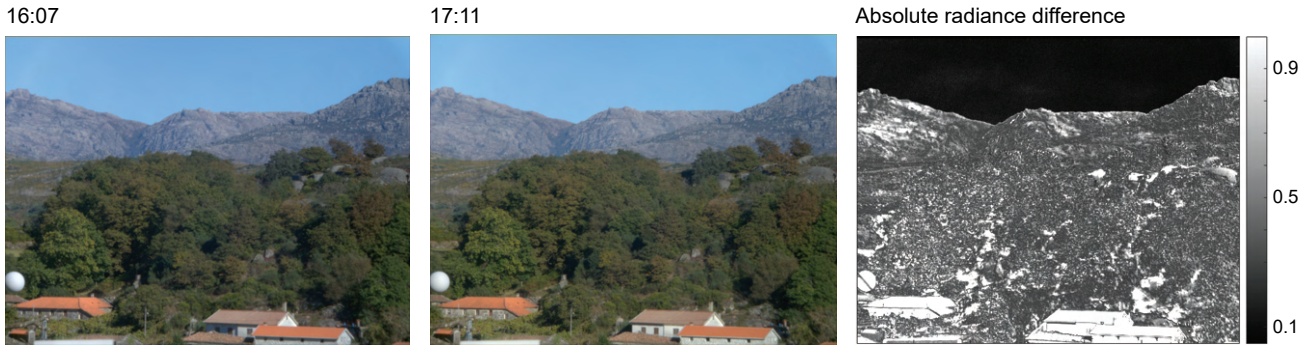


Fig. 3. Changes in scene radiance over 64 min. The left and middle panels show sRGB images rendered from hyperspectral radiance images of the scene in Fig. 2(d) acquired at 16:07 and 17:11, respectively. The right panel shows a grayscale plot of the absolute difference in total radiance at each point clipped to 99% of the maximum. In the bottom left of the plot, the sphere has an oblique dark line across it marking the transition between radiance differences of opposite sign but the same magnitude. The CCTs of the direct illumination on the scene at 16:07 and 17:11 were 5538 K and 5149 K, respectively.

C. Resolution Limits on Frequency Estimates

For convenience, the following summary of the system calibration measurements is extracted from two previous accounts [6,27].

The line-spread function of the imaging system was almost exactly Gaussian with standard deviation (SD) of about 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 nominal peak-transmission wavelength differed by less than 1 nm from the actual peak-transmission wavelength, and the bandwidth (FWHM) ranged from 7 nm at 400 nm to 16 nm at 720 nm. A system measurement of the spectral reflectances of a GretagMacbeth ColorChecker had a root mean square error over wavelength of 0.25% [27].

Notwithstanding these measurements, the estimates of the frequency of mismatching are limited not by the intensity resolution at each pixel but by the frequency with which two pixels match at one instant and mismatch by some multiple of discrimination threshold at another. With a downsampled image size of, say, $M = 650 \times 500 = 325{,}000$ pixels, the lowest frequency of resolvable pairs is 1 in $M(M-1)/2$, i.e., $\sim 2 \times 10^{-11}$. For computational reasons, the number of pairs was limited to 5×10^8, which corresponds to a lowest resolvable frequency of 2×10^{-9}, several orders of magnitude lower than the frequencies reported here.

D. Spectral Change Controls

To compare mismatches under natural illumination changes with mismatches due to illuminant metamerism, images were generated from each scene to simulate purely spectral illumination changes. Thus, one member of each pair of radiance images (Section 2.A) was converted into an effective spectral reflectance image by dividing it pointwise by the spectrum of the direct illumination recorded from a reference surface in the scene [3]. The reflectance image was next converted into two radiance images by multiplying it pointwise by daylight spectra [5] with CCTs of 25,000 K and 4000 K, the largest spectral change used previously [6]. Estimates of the frequency of metamerism under this illuminant change were then obtained exactly as with natural illumination changes. Estimates from a previous analysis [6] were not used since the sets of scenes and uniform color spaces were different.

The foregoing procedure is an approximation and contingent on the measurement of direct illumination, but it ensures that individual pairs of surfaces drawn from the scene are treated equally. The presence of shadows complicates the representation of reflectances, though with little effect on frequency estimates ([6], Appendix A and Section 3.F).

E. Uniform Color Spaces

The spectral radiance at each image point was mapped into tristimulus values and then into a color space where discriminability could be quantified for a standardized observer and viewing conditions [3]. Two color spaces were used with different colorimetric properties. The first was the approximately uniform color space CAM02-UCS [5], which has coordinates (J', a'_M, b'_M), where J' is a correlate of lightness, a'_M of redness–greenness, and b'_M of yellowness–blueness (these chromatic axis names are not used attributively). Given differences $\Delta J'$, $\Delta a'_M$, $\Delta b'_M$ between two sets of coordinates, the total color difference [2] was evaluated with the Euclidean norm, $\Delta E = [(\Delta J')^2 + (\Delta a'_M)^2 + (\Delta b'_M)^2]^{1/2}$. When color differences were assessed chromatically, i.e., without the lightness term, the norm was adjusted to $\Delta E = [(\Delta a'_M)^2 + (\Delta b'_M)^2]^{1/2}$.

The second color space was the popular albeit less uniform CIELAB space [5], which has corresponding coordinates L^*, a^*, b^*, with total color differences also evaluated with the Euclidean norm. To preserve its interpretation, a more uniform color-difference formula such as CIEDE2000 [28] was not applied, but a chromatic adaptation transform CMCCAT2000 [29] was introduced because CIELAB does not accommodate illuminants very different from average daylight [5,30]. With both color spaces, the observer was assumed to be fully adapted to the ambient illumination, which was recorded from a barium sulfate plug or a neutral (Munsell N7) sphere or flat reference surface placed in the scene [6,14,20].

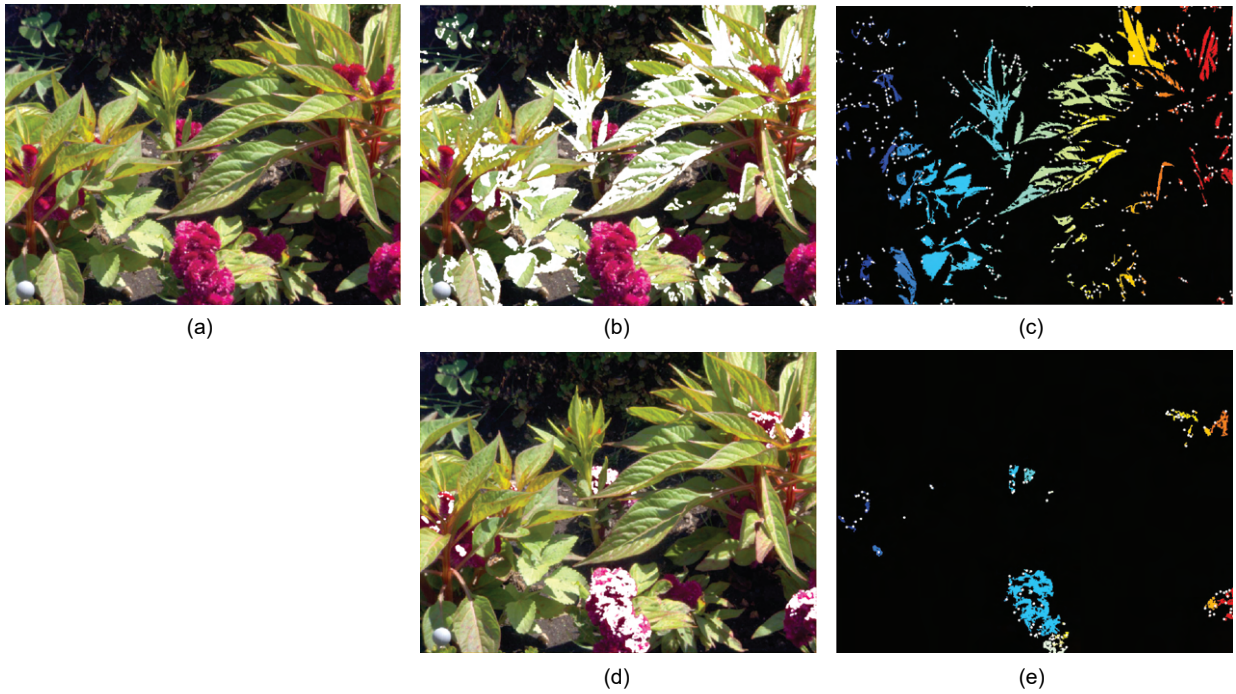


Fig. 4. Sampling from connected components. Panel (a) shows an sRGB reproduction of the scene in Fig. 2(i). Panels (b) and (d) show, respectively, points in white belonging to CAM02-UCS [5] color intervals (30, −15, 15) to (45, 0, 30), selective for yellow–green surfaces, and (30, 30, 0) to (45, 45, 15), selective for red surfaces. Panels (c) and (e) show the corresponding connected components in false color and the locations of single representative points in white.

F. Threshold Color Differences

Reference threshold color differences ΔE^{thr} were specified for each color space. These thresholds were used to decide whether two surfaces were indistinguishable under one illumination and distinguishable under another, and, if so, by what multiple of the threshold. In practice, values depend on whether surfaces are adjacent or not [31] and on the presence of other surfaces in the scene [32]. These configurations were not tested individually but sampled randomly, thereby allowing for their natural frequencies of occurrence.

For CIELAB space, a discrimination threshold of 1.0 has often been assumed [33–35], though for detecting color differences in scene images, empirical thresholds of about 2.2 may be more appropriate [32,36]. Both were tested. Higher thresholds may be used in acceptability measures [37] and categorization tasks [38]. Whatever the choice, expressing the mismatch as a multiple of a threshold color difference reduces the indeterminacy of near-threshold values.

Despite there being no unique mapping of CIELAB thresholds to CAM02-UCS thresholds [39], a previous empirical approximation over 50 hyperspectral images of outdoor scenes [40] suggested a scaling factor of about 0.7. The two reference thresholds of 1.0 and 2.2 for CIELAB space were therefore transformed to 0.7 and 1.5, respectively, for CAM02-UCS.

G. Scene Sampling and Connected Surface Areas

Only the spectral properties of the light reflected at each point in a scene were considered [14], not the local spatial attributes, such as texture, shape, location, and proximity to other surfaces, which, unlike the illumination, were taken as constant. A total of 5×10^8 unique pairs of points were drawn randomly from each scene [6]. This pointwise sampling was indifferent to what physically defined the contents of each scene [41], since that knowledge was unavailable receptorally.

That said, a confound could have arisen with surface areas where sample points were adjacent to or connected to other sample points with similar reflecting properties and local illumination.

To test the effect of what is a low-level articulation of the scene, known to be important in other contexts [42–47], the random sampling was repeated in a stratified way with progressively wider colorimetric tolerances on points in spatially connected components from which just one representative point was chosen.

The procedure was as follows. For a given CAM02-UCS color interval width defining the tolerance, the range of coordinates $(J', a'_{\text{M}}, b'_{\text{M}})$ was divided into a finite number of non-overlapping intervals, and each point of the image was assigned to one of those intervals. For example, for a coordinate interval of width 15, the range of J' values was divided into intervals from 0 to 15, 15 to 30, and so on, and similarly for a'_{M} and b'_{M}. The color coordinates of a particular point, say (38, −4, 17), would accordingly belong to the interval from (30, −15, 15) to (45, 0, 30). Each subset of image points with coordinates in the same color interval was partitioned into spatially connected components. The J', a'_{M}, b'_{M} values in each connected component were therefore all within 15 of each other. Notice that the J', a'_{M}, b'_{M} values were not themselves quantized.

Samples still consisted of 5×10^8 distinct pairs of points but with no more than one point from each connected component, rather than being drawn randomly from the scene as a whole.

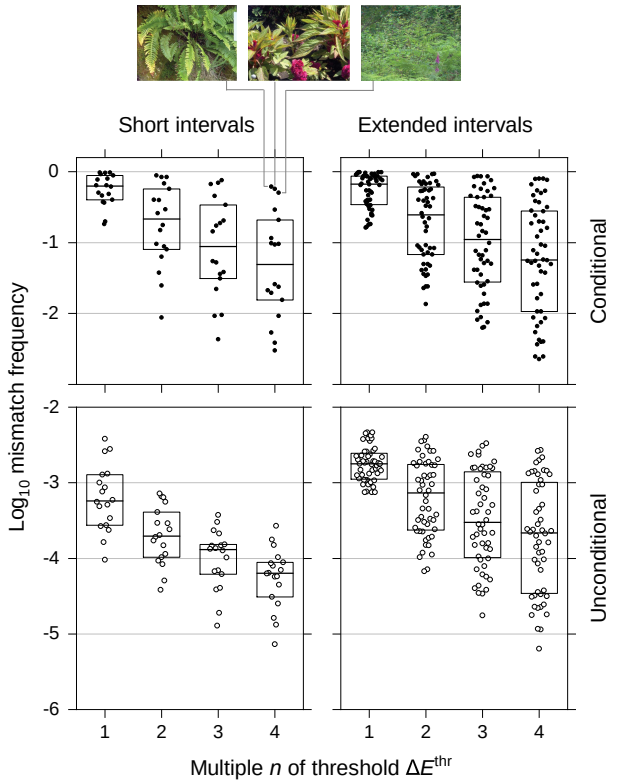


Fig. 5. Frequency of mismatching at integer multiples of threshold. The logarithm of the conditional and unconditional relative frequency of mismatching is plotted against the criterion multiple n of the CAM02-UCS [5] threshold $\Delta E^{\text{thr}} = 0.7$. The boxes placed over the data points mark quartile values. The panels on the left are for the 18 scenes in Fig. 2 with intervals between images of about 1–15 min. The panels on the right are for the four scenes in the top row of Fig. 2 with intervals between images extending from about 1 min to over 4 h. Data points have been jittered horizontally for clarity.

The one point was chosen to be representative of the component in having the median color coordinates, although random selection from the component produced similar outcomes.

Frequencies of mismatching were obtained with interval widths of 2.5, 5.0, … , 20. As the interval width increased, the number of sample points decreased until a width was reached where the number of pairs available reached the minimum of 5×10^8.

Figure 4 shows an example of connected components for the scene in Fig. 2(i) with CAM02-UCS color coordinates drawn from two of 82 color intervals, each of width 15. Figure 4(a) shows an sRGB image rendered from the hyperspectral radiance image of the scene. Figure 4(b) shows all the points in white that belong to the color interval (30, −15, 15) to (45, 0, 30), which is selective for yellow–green surfaces. Figure 4(c) shows in false color the connected components from those yellow–green surfaces, and in white the locations of the single points in each component used for frequency estimates. Figures 4(d) and 4(e) show corresponding data for the color interval (30, 30, 0) to (45, 45, 15), selective for red surfaces.

The number of color intervals and their representative midpoints should not be confused with the number of scene colors that, in other circumstances, might be judged by an observer as empirically relevant, for example, in describing a scene [48,49] or categorizing its colors [50,51].

H. Relative Frequency and Discrimination Threshold

The frequency of mismatches in a scene undergoing a change in the illumination was estimated by taking a random sample of pairs of surfaces and counting the number of pairs for which color differences were subthreshold before the change and suprathreshold by a certain multiple n after the change. An alternative approach modeled on the CIE special metamerism index: change in illuminant [5] does not lend itself to this application since it requires the initial color difference to be zero (or corrected to zero) to avoid the risk of a confound [52].

In more detail, for a given scene imaged at times t_1 and t_2, say, suppose that N distinct pairs of points are drawn randomly from the scene (Section 2.G) and suppose that N_0 of these pairs have color differences ΔE less than the chosen threshold ΔE^{thr} at time t_1, where $N_0 > 0$. Suppose further that N_1 of these pairs have color differences ΔE greater than n times ΔE^{thr} at time t_2, where $n = 1, \ldots, 4$ specifies the criterion multiple. The unconditional relative frequency of mismatching in the scene is then N_1/N, and the conditional relative frequency is N_1/N_0. These definitions are the same as in an earlier analysis of illuminant metamerism [6].

As a side note, estimated median frequencies were similar when the order of images was reversed, that is, when N_0 of pairs had color differences ΔE less than ΔE^{thr} at time t_2 and N_1 of those pairs had color differences ΔE greater than n times ΔE^{thr} at time t_1.

I. Statistical Quantities

The main reporting statistics are median relative frequencies across scenes and image intervals, according to context. Medians were used in preference to means [6] to allow for occasional extreme values with small samples. Logarithmic scales (base 10) were used to stabilize variance and linearize dependencies across relative frequencies. Uncertainties in estimates were quantified with 95% confidence intervals (CIs) estimated by Efron's BCa bootstrap method [53] with 1000 bootstrap replications across scenes and intervals.

3. RESULTS AND COMMENT

A. Relative Frequencies for Short and Extended Time Intervals

Relative frequencies of mismatching varied with the scene, with the interval between images, and with the criterion multiple of the discrimination threshold. Figure 5 shows column scatter plots of the logarithm of the relative frequency against the criterion multiple n. The panels on the left are for all 18 scenes in Fig. 2 with time intervals between the images of each scene of about 1–15 min. The panels on the right are for the four scenes in the top row of Fig. 2 with intervals between images of each scene extending from about 1 min to over 4 h. The CAM02-UCS reference threshold ΔE^{thr} specifying whether two surfaces were indistinguishable was 0.7.

Across short and extended time intervals, median conditional relative frequencies ranged from about 0.06 to 0.63, i.e., 6% to

Table 1. Median Relative Frequencies of Mismatching Pairs under Natural Illumination Changes with CAM02-UCS Threshold $\Delta E^{thr} = 0.7$[a]

Estimate Type	Time Intervals	Criterion Multiple n of Threshold 1	4
Conditional	18 short	6.3×10^{-1} (4.3×10^{-1}, 8.5×10^{-1})	6.0×10^{-2} (1.8×10^{-2}, 1.9×10^{-1})
	4 extended	6.7×10^{-1} (4.3×10^{-1}, 8.0×10^{-1})	5.7×10^{-2} (3.8×10^{-2}, 1.5×10^{-1})
Unconditional	18 short	5.7×10^{-4} (3.1×10^{-4}, 1.0×10^{-3})	6.4×10^{-5} (3.8×10^{-5}, 8.6×10^{-5})
	4 extended	1.8×10^{-3} (1.5×10^{-3}, 2.0×10^{-3})	2.2×10^{-4} (1.1×10^{-4}, 3.6×10^{-4})

[a]Entries show median values of relative frequencies and 95% confidence intervals in parentheses. Other details, including time intervals, as for Fig. 5.

Table 2. Relative Frequencies of Mismatching Pairs under Natural Illumination Changes with CAM02-UCS Threshold $\Delta E^{thr} = 0.7$[a]

Scene Label	Scene Name	Time Interval, min	Criterion Multiple n of Threshold: Conditional Estimate 1	2	3	4	Unconditional Estimate 1	2	3	4
a	Nogueiró	3	6.4×10^{-1}	1.9×10^{-1}	6.6×10^{-2}	2.5×10^{-2}	2.3×10^{-3}	7.0×10^{-4}	2.4×10^{-4}	9.2×10^{-5}
b	Gualtar	3	1.9×10^{-1}	4.6×10^{-3}	1.2×10^{-3}	7.6×10^{-4}	6.7×10^{-4}	1.6×10^{-5}	4.4×10^{-6}	2.7×10^{-6}
c	Sete Fontes	1	5.6×10^{-1}	1.2×10^{-1}	4.0×10^{-2}	1.8×10^{-2}	1.1×10^{-3}	2.4×10^{-4}	7.9×10^{-5}	3.6×10^{-5}
d	Levada	2	1.8×10^{-1}	2.4×10^{-2}	9.2×10^{-3}	3.7×10^{-3}	7.0×10^{-4}	9.5×10^{-5}	3.7×10^{-5}	1.5×10^{-5}
e	Souto Farm Barn	1	4.6×10^{-1}	1.0×10^{-1}	2.9×10^{-2}	6.6×10^{-3}	3.1×10^{-3}	6.9×10^{-4}	1.9×10^{-4}	4.5×10^{-5}
f	Ruivães Fern	3	9.8×10^{-1}	9.0×10^{-1}	7.8×10^{-1}	6.5×10^{-1}	1.6×10^{-4}	1.4×10^{-4}	1.2×10^{-4}	1.0×10^{-4}
g	Tibães Garden	4	8.8×10^{-1}	5.6×10^{-1}	3.3×10^{-1}	2.1×10^{-1}	2.3×10^{-4}	1.5×10^{-4}	8.7×10^{-5}	5.5×10^{-5}
h	Ribeira Hotel	2	3.7×10^{-1}	3.8×10^{-2}	9.2×10^{-3}	5.5×10^{-3}	4.7×10^{-4}	4.9×10^{-5}	1.2×10^{-5}	7.1×10^{-6}
i	Bom Jesus Red Flower	8	9.6×10^{-1}	8.4×10^{-1}	6.9×10^{-1}	5.6×10^{-1}	9.5×10^{-5}	8.3×10^{-5}	6.8×10^{-5}	5.6×10^{-5}
j	Tibães Corridor	5	7.6×10^{-1}	3.8×10^{-1}	1.8×10^{-1}	7.7×10^{-2}	5.5×10^{-4}	2.8×10^{-4}	1.3×10^{-4}	5.6×10^{-5}
k	Ribeira Old Tower	2	4.6×10^{-1}	1.5×10^{-1}	5.5×10^{-2}	1.9×10^{-2}	1.5×10^{-3}	4.7×10^{-4}	1.8×10^{-4}	6.1×10^{-5}
l	Santuário Sameiro	6	5.5×10^{-1}	2.4×10^{-1}	1.4×10^{-1}	8.9×10^{-2}	2.1×10^{-4}	9.1×10^{-5}	5.2×10^{-5}	3.3×10^{-5}
m	Ruivães Cottage	3	9.2×10^{-1}	6.7×10^{-1}	4.6×10^{-1}	2.9×10^{-1}	2.7×10^{-4}	2.0×10^{-4}	1.3×10^{-4}	8.6×10^{-5}
n	Gualtar Villa	15	3.9×10^{-1}	8.0×10^{-2}	4.8×10^{-2}	2.8×10^{-2}	2.3×10^{-3}	4.6×10^{-4}	2.7×10^{-4}	1.6×10^{-4}
o	Ribeira Houses Shrubs	3	4.9×10^{-1}	1.2×10^{-1}	5.8×10^{-2}	3.5×10^{-2}	3.8×10^{-4}	9.5×10^{-5}	4.5×10^{-5}	2.7×10^{-5}
p	Vila Verde	7	6.5×10^{-1}	2.7×10^{-1}	1.6×10^{-1}	1.1×10^{-1}	5.5×10^{-4}	2.3×10^{-4}	1.3×10^{-4}	9.5×10^{-5}
q	Ribeira Houses Concrete	6	7.9×10^{-1}	3.8×10^{-1}	1.7×10^{-1}	8.3×10^{-2}	1.3×10^{-3}	6.1×10^{-4}	2.7×10^{-4}	1.3×10^{-4}
r	Ruivães Ferns	10	9.7×10^{-1}	8.4×10^{-1}	6.7×10^{-1}	5.1×10^{-1}	3.4×10^{-4}	3.0×10^{-4}	2.4×10^{-4}	1.8×10^{-4}

[a]Entries show relative frequencies for individual scenes identified by letter as in Fig. 2 and by public name for download at Ref. [54]. The time interval refers to the interval between successive images, which varies with the scene. Further details in Section 2.A.

63%, depending on the criterion multiple n of the threshold. Median unconditional relative frequencies were much lower, ranging from about 6×10^{-5} to 6×10^{-4}. Both conditional and unconditional frequencies declined steadily with increasing n (see also Ref. [6], Fig. 3). Median values are summarized in Table 1.

Relative frequencies for individual scenes are given in Table 2, though values depend on both scene content and the interval between images (Section 2.A). Examples of color mismatches are illustrated in Table 3.

Relative frequencies with the higher CAM02-UCS threshold ΔE^{thr} of 1.5 had a similar distribution and are not plotted. Conditional relative frequencies were lower, with median values ranging from about 1% to 40%. Unconditional relative frequencies were higher, consistent with the higher threshold for pairs to be discriminable. Median values are summarized in Table 4.

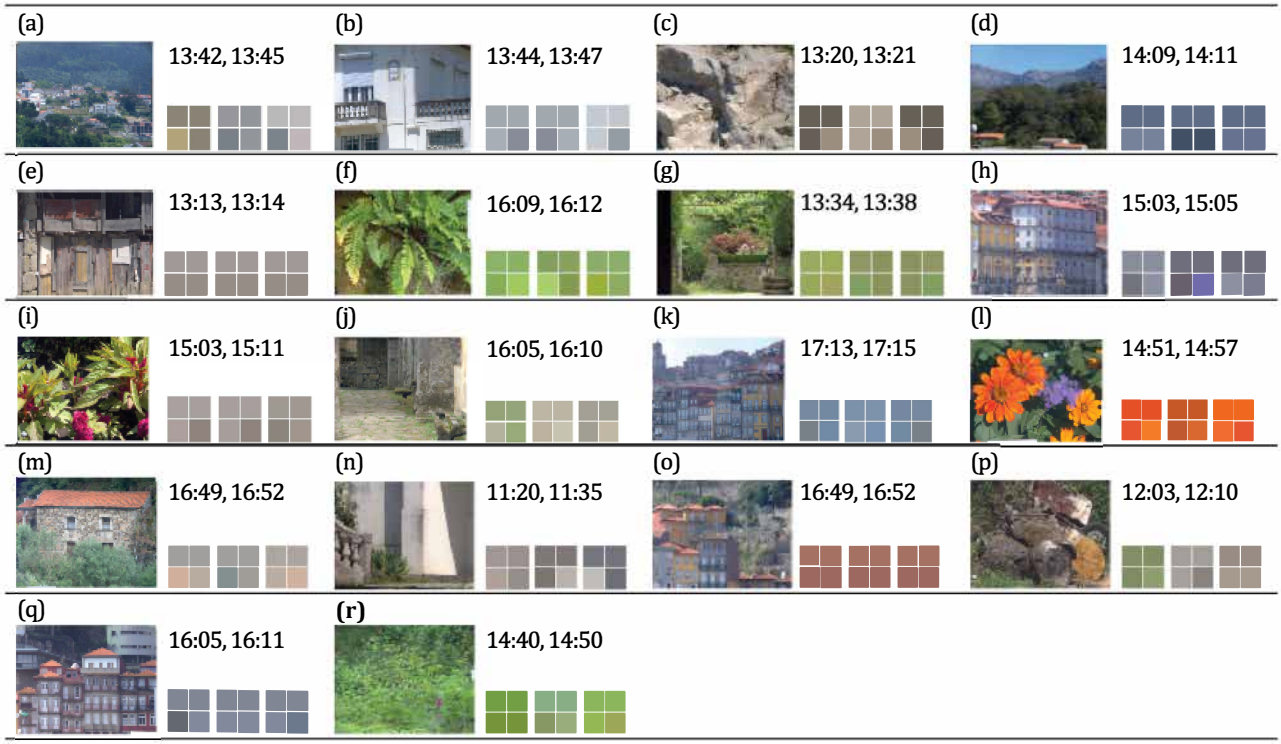
Similar or somewhat higher unconditional relative frequencies were obtained with CIELAB space and corresponding reference thresholds ΔE^{thr} of 1.0 and 2.2 (Section 2.F) and somewhat lower conditional relative frequencies. Median values are summarized in Table 5 and Table 6.

B. Image Interval Duration

The interval between images is a convenient proxy for more direct measurements of illumination change and its progressive effects on mismatching. With short-interval image pairs, interval duration varied with the scene, and since only one interval was available (Section 2.A), its effects on the frequency of mismatching were confounded with scene content. With extended-interval image pairs, however, from 9 to 16 different intervals were available for each of the four scenes in the top row of Fig. 2.

Figure 6 shows for these image pairs, scatter plots of the logarithm of the relative frequency of mismatching against the logarithm of the image interval for criterion multiples n of the threshold from 1 to 4. Open and filled symbols denote unconditional and conditional frequencies, respectively. The slanting

Table 3. Examples of Matching and Mismatching Pairs of Surface Colors at Successive Instants[a]



[a]Each panel (a) to (r) shows a scene image and three 2 × 2 arrays of matching and mismatching pairs of surfaces at times t_1 and t_2. The top row of each 2 × 2 array shows the color appearance of a pair of surfaces in the scene at time t_1 and the bottom row their color appearance at time t_2. Scenes are labeled as in Fig. 2.

Table 4. Median Relative Frequencies of Mismatching Pairs under Natural Illumination Changes with CAM02-UCS Threshold $\Delta E^{thr} = 1.5$[a]

		Criterion Multiple n of Threshold	
Estimate Type	**Time Intervals**	**1**	**4**
Conditional	18 short	4.0×10^{-1} (2.7×10^{-1}, 6.5×10^{-1})	1.2×10^{-2} (4.4×10^{-3}, 4.0×10^{-2})
	4 extended	4.4×10^{-1} (3.0×10^{-1}, 5.8×10^{-1})	2.1×10^{-2} (1.4×10^{-2}, 4.4×10^{-2})
Unconditional	18 short	2.4×10^{-3} (1.7×10^{-3}, 3.8×10^{-3})	1.0×10^{-4} (5.1×10^{-5}, 2.0×10^{-4})
	4 extended	7.6×10^{-3} (5.1×10^{-3}, 1.0×10^{-2})	3.5×10^{-4} (2.1×10^{-4}, 8.6×10^{-4})

[a]Entries show median values of relative frequencies and 95% confidence intervals in parentheses. Other details as for Table 1.

Table 5. Median Relative Frequencies of Mismatching Pairs under Natural Illumination Changes with CIELAB Threshold $\Delta E^{thr} = 1.0$[a]

		Criterion Multiple n of Threshold	
Estimate Type	**Time Intervals**	**1**	**4**
Conditional	18 short	4.9×10^{-1} (3.7×10^{-1}, 6.5×10^{-1})	2.6×10^{-2} (7.3×10^{-3}, 7.2×10^{-2})
	4 extended	5.9×10^{-1} (4.5×10^{-1}, 7.1×10^{-1})	4.7×10^{-2} (3.0×10^{-2}, 7.2×10^{-2})
Unconditional	18 short	1.1×10^{-3} (8.3×10^{-4}, 2.2×10^{-3})	6.7×10^{-5} (3.6×10^{-5}, 1.1×10^{-4})
	4 extended	2.6×10^{-3} (2.0×10^{-3}, 3.2×10^{-3})	2.4×10^{-4} (1.2×10^{-4}, 4.5×10^{-4})

[a]Entries show median values of relative frequencies and 95% confidence intervals in parentheses. Other details as for Table 1.

lines are linear regression fits. High conditional frequencies appear compressed because of the 100% ceiling on values. The CAM02-UCS threshold ΔE^{thr} was 0.7.

The regression fits show an upward trend in conditional and unconditional relative frequencies with increasing interval. For example, for intervals of 1 min and 1 h, the conditional frequency increases from about 34% to 62% for $n = 1$ and from about 1% to 12% for $n = 4$. Estimates with 95% CIs are summarized in Table 7.

Estimates with the higher CAM02-UCS threshold ΔE^{thr} of 1.5 were lower but had a similar distribution, summarized in Table 8. Slopes of the regression lines are summarized in Table 9.

Table 6. Median Relative Frequencies of Mismatching Pairs under Natural Illumination Changes with CIELAB Threshold $\Delta E^{thr} = 2.2$[a]

		Criterion Multiple n of Threshold	
Estimate Type	**Time Intervals**	**1**	**4**
Conditional	18 short	3.1×10^{-1} (2.3×10^{-1}, 4.4×10^{-1})	6.4×10^{-3} (1.4×10^{-3}, 2.9×10^{-2})
	4 extended	3.7×10^{-1} (2.9×10^{-1}, 4.3×10^{-1})	1.3×10^{-2} (2.7×10^{-3}, 3.3×10^{-2})
Unconditional	18 short	3.3×10^{-3} (2.7×10^{-3}, 6.6×10^{-3})	8.9×10^{-5} (4.4×10^{-5}, 2.3×10^{-4})
	4 extended	1.0×10^{-2} (6.9×10^{-3}, 1.5×10^{-2})	3.6×10^{-4} (1.2×10^{-4}, 9.1×10^{-4})

[a]Entries show median values of relative frequencies and 95% confidence intervals in parentheses. Other details as for Table 1.

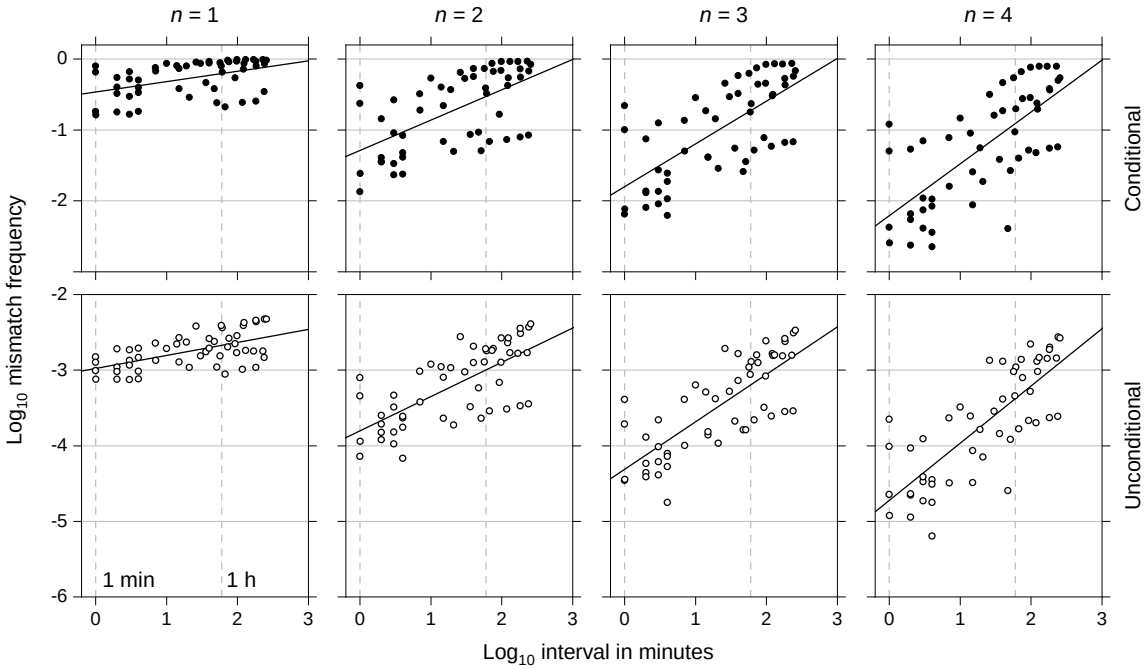


Fig. 6. Frequency of mismatching and image interval for the four scenes in the top row of Fig. 2. The logarithm of the conditional and unconditional relative frequency of mismatching is plotted against the logarithm of the image interval with criterion multiple n of threshold from 1 to 4. The slanting lines are linear regressions. The vertical dashed lines mark intervals of 1 min and 1 h. The CAM02-UCS [5] reference threshold ΔE^{thr} was 0.7.

Table 7. Conditional Relative Frequencies of Mismatching Pairs under Natural Illumination Changes for Selected Intervals with CAM02-UCS Threshold $\Delta E^{thr} = 0.7$[a]

	Criterion Multiple n of Threshold	
Time Interval	**1**	**4**
1 min	0.34 (0.26, 0.47)	0.01 (0.00, 0.02)
1 h	0.62 (0.53, 0.72)	0.12 (0.08, 0.17)
4 h	0.76 (0.57, 0.96)	0.34 (0.19, 0.55)

[a]Entries show regression estimates of conditional relative frequencies with 95% confidence intervals in parentheses. Other details as for Fig. 6.

Table 8. Conditional Relative Frequencies of Mismatching Pairs under Natural Illumination Changes for Selected Intervals with CAM02-UCS Threshold $\Delta E^{thr} = 1.5$[a]

	Criterion Multiple n of Threshold	
Time Interval	**1**	**4**
1 min	0.18 (0.14, 0.26)	–
1 h	0.49 (0.40, 0.57)	0.04 (0.02, 0.07)
4 h	0.69 (0.50, 0.86)	0.18 (0.07, 0.34)

[a]Details as for Table 7.

The parallelism of the regression fits for conditional and unconditional frequencies at each n in Fig. 6 is analyzed in Appendix B.

C. Simulated Changes in Illuminant Spectra

Replacing real illumination changes by simulated changes in daylight spectra provides a reference level of performance unaffected by changes in the distribution of light over the scene. The spectral changes, from a CCT of 25,000 K to one of 4000 K, were very much larger than natural changes in CCT over most of the day ([17], Fig. 5) ([18], Fig. 7). The panels on the left of Fig. 7 show the resulting column scatter plots of the logarithm of the relative frequency for the 18 scenes in Fig. 2 against the criterion multiple n of threshold. The CAM02-UCS threshold ΔE^{thr} was 0.7. Median values are tabulated in Table 10. For comparison, the two panels on the right of Fig. 7 show corresponding data, replotted from Fig. 5, for natural illumination

Table 9. Slopes of Linear Regression Fits of Log Relative Frequency of Mismatching Pairs on Log Interval with CAM02-UCS Threshold $\Delta E^{thr} = 0.7$[a]

		Criterion Multiple n of Threshold	
Estimate Type	**Time Intervals**	**1**	**4**
Conditional	4 extended	1.5×10^{-1} (5.1×10^{-2}, 2.3×10^{-1})	7.3×10^{-1} (5.0×10^{-1}, 9.1×10^{-1})
Unconditional	4 extended	1.7×10^{-1} (1.0×10^{-1}, 2.4×10^{-1})	7.5×10^{-1} (5.7×10^{-1}, 9.0×10^{-1})

[a]Entries show slopes and 95% confidence intervals in parentheses. Other details as for Fig. 6.

Table 10. Median Relative Frequencies of Mismatching Pairs under Simulated Changes in Illuminant Spectra with CAM02-UCS Threshold $\Delta E^{thr} = 0.7$[a]

		Criterion Multiple n of Threshold	
Estimate Type	**No. of Scenes**	**1**	**4**
Conditional	18	2.6×10^{-1} (1.9×10^{-1}, 4.2×10^{-1})	1.6×10^{-4} (4.5×10^{-5}, 3.6×10^{-3})
Unconditional	18	2.8×10^{-4} (2.1×10^{-4}, 4.5×10^{-4})	4.0×10^{-7} (7.3×10^{-8}, 1.5×10^{-6})

[a]Entries show median values of relative frequencies and 95% confidence intervals in parentheses. Other details as for Fig. 7.

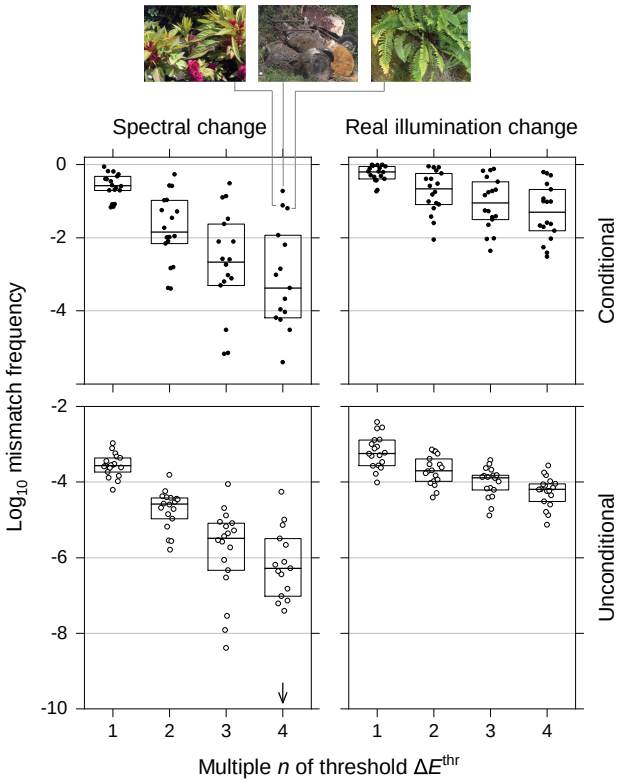


Fig. 7. Frequency of mismatching with simulated changes in illuminant spectra and real illumination changes replotted from Fig. 5. The spectral changes were from a daylight with CCT of 25,000 K to one of 4000 K, and the real illumination changes were over intervals of about 1–15 min. Relative frequencies were too low to record with spectral changes in three scenes for $n = 4$ (bottom left panel, arrowed). Other details as for Fig. 5. Notice the doubled vertical scale.

changes on the same 18 scenes with time intervals between the images of each scene of about 1–15 min.

Both unconditional and conditional relative frequencies with illuminant spectral changes were lower than with natural illumination changes, even with short intervals. Differences in $\log_{10}$ units ranged from about −0.33 (CI −0.46 to −0.15) for unconditional relative frequencies and $n = 1$ down to about −2.5 (CI −3.3 to −2.0) for conditional relative frequencies and $n = 4$. Differences were greater still with natural illumination changes over extended intervals.

D. Chromatic Matches

In the scene in Fig. 1, there were large differences in lightness between mismatched pairs, but only small differences in hue angle, not more than 3 deg either within or between image pairs. Across all scenes, lightness differences accounted on average for 47% (CI 33% to 72%) of the total variation at short intervals and 76% (CI 70% to 83%) with extended intervals.

To test the importance of lightness differences in determining the frequency of mismatching, the lightness term J' was omitted from the color difference formula, leaving only the chromatic attributes (a'_M, b'_M) (Section 2.E). In the event, conditional frequencies were not reliably different for any criterion multiple n of threshold. By contrast, unconditional frequencies were higher, and reliably so by factors ranging from about 10.8 (CI 6.5 to 20.0) to 18.5 (CI 8.5 to 25.5) across n.

Median relative frequencies for chromatic differences alone are summarized in Table 11 and may be compared with the estimates in Table 1 for total color differences.

E. Connected Surface Areas

Could random sampling by points overrepresent image areas that were susceptible to mismatching, thereby inflating frequency estimates? Partitioning images into spatially connected components with similar colors and choosing no more than one point from each component, as detailed in Section 2.G, should have countered this effect.

Figure 8 shows the logarithm of the median relative frequency for the 18 scenes of Fig. 2 plotted against the width of the CAM02-UCS color interval specifying the similarity of the colors in each connected component. As interval widths and therefore connected surface areas increased, the number of distinct connected components decreased, and eventually for widths greater than 10.0, one or more scenes failed to provide frequency estimates. Up to that limit, both unconditional and conditional relative frequencies varied little about the horizontal straight lines through the symbols, even though the number of

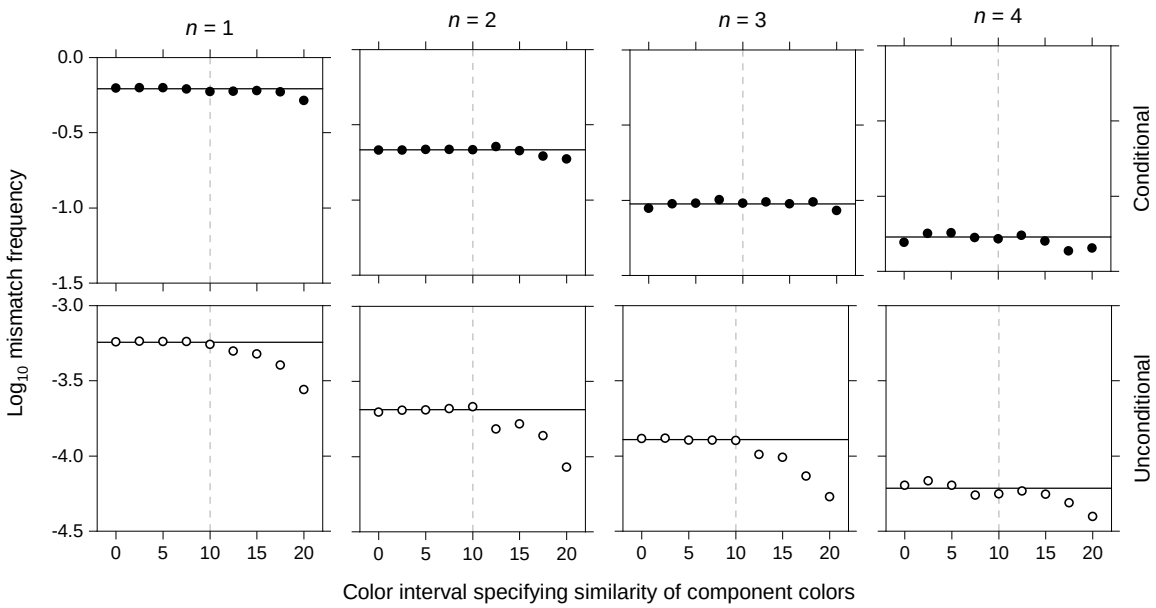


Fig. 8. Frequency of mismatching with increasing size of connected surface areas in partitioned images. The panels show the logarithm of the median conditional and unconditional relative frequency plotted against the width of the CAM02-UCS [5] color interval specifying the similarity of the colors in each connected component. Medians were taken over all 18 scenes in Fig. 2. The vertical line marks the limit beyond which the number of connected components decreased sufficiently to affect the number of sample pairs and therefore the frequency estimates. At that limit, the median number of sample points was 3% of the unpartitioned scenes, and the number of colors defining them was 67. Up to that limit, sample standard deviations of the logarithm of the relative frequency were less than 0.04. Other details as for Fig. 6.

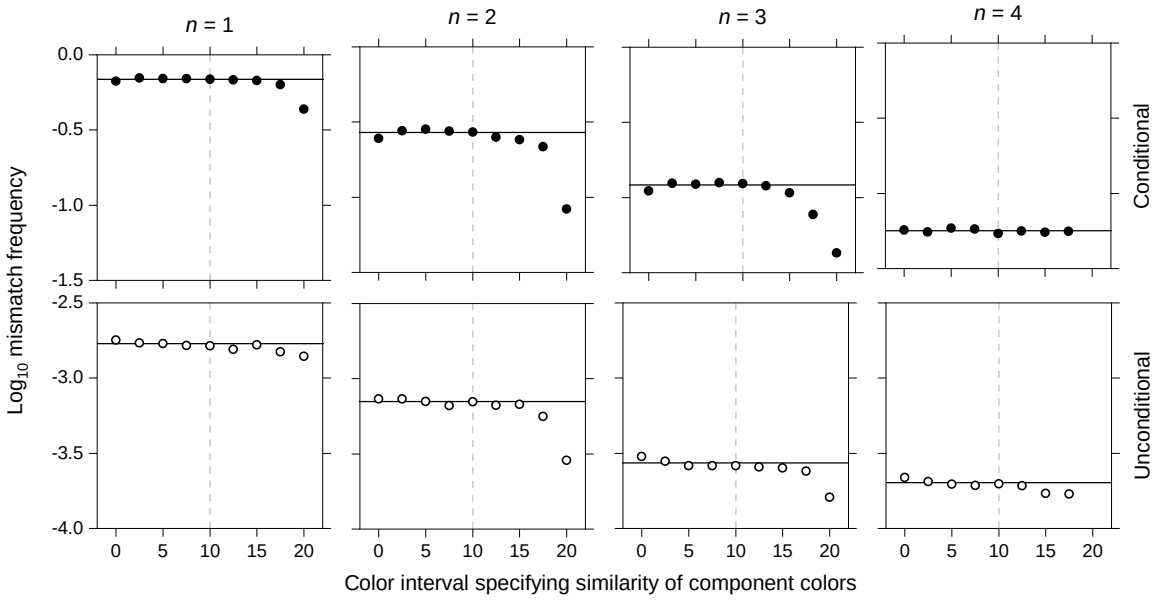


Fig. 9. Frequency of mismatching with increasing size of connected surface areas in partitioned images. Medians were taken over four scenes in the top row of Fig. 2 and multiple extended time intervals. Other details as for Fig. 8.

sample points fell to 3% of the total available. The pattern of performance was similar with the four scenes in the top row of Fig. 2 and extended time intervals (Fig. 9).

4. DISCUSSION

The examples in Fig. 1 illustrate how surface colors in an outdoor scene can mismatch locally as solar elevation changes over time with little change in illumination spectrum. The analysis of the radiance images of all 18 scenes in Fig. 2 generalized this observation. Both conditional and unconditional relative frequencies of mismatching were much higher under natural illumination changes than under spectral changes alone, and both frequencies increased systematically as the intervals between images increased.

In the following sections, the results of this analysis are considered in more detail, together with their implications and

limitations. Issues addressed include differences in the frequency of mismatching; sampling by points, areas, and objects; lightness variations; relational color constancy; and surface color information.

A. Frequency of Mismatching

The difference between unconditional and conditional frequencies of mismatching is a consequence of initial conditions. The unconditional relative frequency of mismatching is necessarily low in outdoor scenes since two surfaces chosen at random are unlikely to match [6]. For the 18 scenes with intervals between images of 1–15 min, unconditional relative frequencies ranged from around 10^{-3} for a suprathreshold color difference to around 10^{-4} for a four-fold threshold color difference. But given two surfaces that already match, the conditional relative frequency is much higher. For the same scenes and intervals, conditional frequencies ranged from over 60% for a suprathreshold color difference to around 6% for a four-fold threshold color difference, the latter increasing markedly with longer intervals.

These estimates were for a CAM02-UCS discrimination threshold of 0.7. With a higher CAM02-UCS threshold of 1.5, appropriate for detecting changes in scenes [32,36], unconditional relative frequencies were higher, and conditional relative frequencies were lower, but still relevant. The pattern of performance was similar with the less uniform CIELAB color space and corresponding thresholds of 1.0 and 2.2. The roughly inverse relationship between conditional and unconditional frequencies is analyzed in Appendix A.

Interval duration had its expected effect on relative frequencies, but as a single explanatory variable, it can account only partially for the variation. A fuller description would require variables that were more directly related to the light reflected from scenes, for example, solar elevation, atmospheric transmission, the changes in each, and local scene structure.

B. Sampling by Points, Areas, and Objects

Many scenes contained extended, similarly colored physical surfaces that could have biased frequency estimates based on random sampling by points. But both conditional and unconditional relative frequencies remained almost the same when scene images were partitioned into connected components containing similar colors, and sampling was reduced to one point from each component. Random sampling by points seems not to have inflated frequency estimates (or potential detection performance).

An issue concerning point sampling itself is the mixing of reflected spectra. After downsampling, each pixel in a scene image represented an elementary area of approximately 0.6×0.6 arc min angular subtense. If images had been acquired at a greater distance, producing more spectral mixing, frequencies could differ. Yet, in previous simulations with illuminant spectral changes, image blurring had only a modest effect on the frequency of metamerism [6].

A more fundamental issue is whether sampling by points could be replaced by a higher-level sampling such as by objects. For some scenes, however, what constitutes an object may be difficult to decide, for example, in Fig. 2(a) whether it is an individual tree or clumps of trees, and in Fig. 2(c) whether it is part of or all of the rock face. In whatever way an object is defined, the problem of changes in appearance under changes in illumination still needs to be addressed, a point taken up later.

C. Contribution of Lightness

Most variation in color appearance in outdoor scenes is in lightness, not chromaticity [21,55,56]. This may be due to the way natural light is distributed [12,20,55,57,58], with the direct beam interrupted by nearby objects, including overhead foliage, to produce cast shadows [15,16]. Local surface geometry may also have a role [57] in that surface orientation is usually preserved over shorter distances than material identity, leading to more rapid changes in reflected intensity than in spectral composition [56].

Even so, as the present analysis showed, when lightness differences were excluded from the evaluation of color differences, conditional relative frequencies were little affected. Their high values, with or without lightness cues, seem not to be due to natural lightness variation, either within scenes or over time. In contrast, unconditional frequencies increased by an order of magnitude in the absence of lightness differences. The difference between unconditional and conditional frequencies is again due to the change in initial conditions. In any given scene, there is a much higher probability of finding pairs of surfaces indistinguishable in just chromaticity than in both chromaticity and lightness.

D. Relational Color Constancy and Mismatching

Failures of color matching under illumination changes are intimately related to the phenomenon of color constancy, the constant perceived or apparent color of a surface in the presence of changes in the intensity and spectral composition of the illumination [38,59–61]. In experimental studies, the illumination changes have, in general, been spatially uniform over the surfaces involved, as with the simulated spectral changes in this study and in laboratory experiments with real objects [62–64]. Levels of constancy have been found to vary, depending on the physical cues available and the experimental task [38,60].

As to explanatory mechanisms, chromatic adaptation, typified by von Kries scaling of cone responses [59,65–67], can help preserve color appearance. Where it cannot, observers can still attribute changes in appearance to changes either in illuminant or in the reflecting properties of the scene. This operational approach [68] may depend on relational color constancy, that is, the constancy of perceived spatial color relations, in turn quantified by spatial ratios of cone excitations or of their post-receptoral transformations [69,70]. Relational color constancy may also facilitate other kinds of surface-color comparisons [71–73]. In particular, mismatching under illumination changes may be seen as a failure of relational color constancy in which the perceived color relation between surfaces is one of identity.

Although illuminant changes can produce lapses in color constancy and relational color constancy with some reflecting surfaces, especially colorful ones [40,74], it is real-world

illumination differences that have the greatest impact on these constancies [13,20], in so far as they apply to appearance, as illustrated by the transient and progressive changes in the upper and lower pairs of surfaces in Fig. 1 with changes in the angle of the incident beam.

Why should natural illumination changes disrupt surface color appearance so much more than illuminant spectral changes? One explanation is that they may be optically indistinguishable, or nearly so, from genuine changes in surface reflectance, as noted elsewhere [40]. Spatially restricting von Kries scaling or spatial ratios of cone signals to compensate for local changes in illumination could lead to a failure to detect genuine differences in surface reflectance.

This is not to say that changes in appearance cannot be correctly judged if there is sufficient context, for example, allowing an appeal to continuity ([75], §67) [76]. Thus, observing individual surface areas continuously over time preserves identification despite changes in their reflected light. Analogously, observing the spatial continuity of surface areas preserves identification despite changes in their orientation. Neither is possible if observations are spatially or temporally discontinuous.

E. Mismatching and Color Information

The existence of mismatching demonstrates the limit on the information that color provides about scenes undergoing natural illumination changes. The amount of information preserved between one observation and another can be quantified with Shannon's notion of information [77], in turn interpreted as the logarithm of the number of surfaces retaining their identity across time [14], or, more precisely, as a least upper bound on the average number possible in the given conditions [78]. Consistent with such an interpretation, when this number was estimated with the four scenes in the top row of Fig. 2 [14], it declined rapidly with interval duration in a way similar to the decline in the relative frequency of mismatching.

These considerations suggest that the occurrence of mismatching could be predicted by informational quantities. In an analysis with 50 hyperspectral images of outdoor scenes, the variation of the relative frequency of mismatching with simulated changes in illuminant spectra was well modeled by combinations of informational quantities, if the criterion multiple of discrimination threshold was not too high [79]. The variation with natural illumination changes might be modeled similarly.

F. Conclusion

To return to the question posed at the beginning of this paper, if two surfaces in an outdoor scene look the same at one moment, then they may well look different later. The longer the interval, the greater the difference in appearance. Natural illumination changes outdoors seem more likely to impair surface identification by color than changes in illuminant spectra alone, and to much greater effect.

APPENDIX A: INTERDEPENDENCE OF UNCONDITIONAL AND CONDITIONAL FREQUENCIES

Recall from Section 2.H the sampling regime underpinning the frequency estimates. For each scene, a sample of N distinct pairs of points was drawn randomly from the scene. Suppose that N_0 of these pairs had color differences ΔE less than the reference threshold ΔE^{thr} at time t_1, where $N_0 > 0$, and that N_1 of these pairs had color differences ΔE greater than a certain multiple $n = 1, \ldots, 4$ of ΔE^{thr} at time t_2. The unconditional and conditional relative frequencies of mismatching pairs of points in the scene are given by N_1/N and N_1/N_0, respectively.

Since $N_1/N = (N_1/N_0)(N_0/N)$, the logarithm of N_1/N is given by

$$\log N_1/N = \log N_1/N_0 + \log N_0/N.$$

The variance of $\log N_1/N$, which is a linear combination of random variables, can be decomposed [80]; thus,

$$\text{var}(\log N_1/N) = \text{var}(\log N_1/N_0) + \text{var}(\log N_0/N) + 2\text{cov}(\log N_1/N_0, \log N_0/N). \quad \textbf{(A1)}$$

Table 11. Median Relative Frequencies of Chromatically Mismatching Pairs under Natural Illumination Changes with CAM02-UCS Threshold $\Delta E^{\text{thr}} = 0.7$[a]

		Criterion Multiple n of Threshold	
Estimate Type	No. of Scenes	1	4
Conditional	18	6.0×10^{-1} (4.2×10^{-1}, 7.5×10^{-1})	3.5×10^{-2} (1.4×10^{-2}, 1.4×10^{-1})
Unconditional	18	8.4×10^{-3} (5.8×10^{-3}, 1.6×10^{-2})	7.7×10^{-4} (3.1×10^{-4}, 1.3×10^{-3})

[a]Entries show median values of relative frequencies and 95% confidence intervals in parentheses. Other details as for Table 1.

Table 12. Variance–Covariance Decomposition of Unconditional Relative Frequencies[a]

n	var($\log N_1/N$)	var($\log N_1/N_0$)	var($\log N_0/N$)	2cov($\log N_1/N_0$, $\log N_0/N$)
1	0.19	0.05	0.32	−0.18
2	0.15	0.29	0.32	−0.46
3	0.15	0.43	0.32	−0.60
4	0.15	0.56	0.32	−0.73

[a]Entries show variances and covariances of data in Fig. 5, left panels, over 18 scenes for criterion multiple n of threshold from 1 to 4. Other details as for Fig. 5.

Table 12 gives values for each of the terms on the right-hand side of Eq. (A1) for the data in Fig. 5, left panels.

For $n = 1$, the largest contribution to the variance in the unconditional relative frequency N_1/N is from the variance in the relative frequency N_0/N of subthreshold differences. As n increases, this dependence changes, so that for $n = 4$, the contribution from the variance in N_0/N is the smallest. The negative sign for the covariance of N_1/N_0 and N_0/N implies that as N_0/N increases, N_1/N_0 tends to decrease. This inverse relationship is broadly consistent with the effects of changing ΔE^{thr}, noted in Section 4.A.

This analysis has parallels with an earlier analysis of the relationship between the relative frequency of mismatching and the entropy of colors in a scene under one illuminant and their conditional entropy under another illuminant [79]. When these two entropies were combined in a linear model, the coefficients weighting the entropies appeared with opposite signs.

APPENDIX B: CORRELATION OF UNCONDITIONAL AND CONDITIONAL FREQUENCIES WITH INTERVAL DURATION

Notation in the following is adapted from Appendix A. Based on the plots in Fig. 6, the variation of the logarithm of the unconditional relative frequency N_1/N across samples i and intervals Δt can be modeled for each criterion multiple n of threshold, as

$$\log N_1(i, \Delta t)/N = b_0 + b_1 \log \Delta t + e(i, \Delta t), \quad \textbf{(B1)}$$

where b_0 and b_1 are constants and $e(i, \Delta t)$ represent deviations about the linear fit. Analogously, the variation of the logarithm of the conditional relative frequency N_1/N_0 can be modeled as

$$\log N_1(i, \Delta t)/N_0(i) = b_0' + b_1' \log \Delta t + e'(i, \Delta t). \quad \textbf{(B2)}$$

The difference between the left-hand sides of Eqs. (B1) and (B2) is $\log N_0(i)/N$, which depends on i, not on Δt. Its variation can be modeled as

$$\log N_0(i)/N = b_0'' + e''(i).$$

It follows that the gradients b_1 and b_1' in Eqs. (B1) and (B2) should be equal and that the variance of the difference $\log N_0(i)/N$, which equals $\text{var}(e''(i))$, should be less than the variance of the residuals $\text{var}(e(i, \Delta t))$ and $\text{var}(e'(i, \Delta t))$, which depend on both i and Δt.

Both predictions are consistent with the data in Fig. 6 as the criterion multiple n of threshold increased and the independence of i and Δt increased. Thus, the differences between the gradients of the two fitted lines in each panel decreased from 12% to 3%, and the variance of the differences in unconditional and conditional relative frequencies as a proportion of the variance of the residuals decreased from 67% to 11%.

Funding. Engineering and Physical Sciences Research Council (EP/W033968/1); Leverhulme Trust (RPG-2022-266).

Acknowledgment. I am grateful to K. Amano and Y. Lu for critically reading the manuscript.

Disclosures. The author declares no conflicts of interest.

Data availability. The hyperspectral radiance data analyzed in this study are freely available at Refs. [54,81] and at Ref. [82]. Code for estimating mutual information for R, Python, and MATLAB computing environments is available at Ref. [83].

REFERENCES

1. M. R. Luo, "4-colour quality evaluation," in *Total Colour Management in Textiles*, J. H. Xin, ed. (Woodhead, 2006), pp. 57–75.
2. R. W. G. Hunt and M. R. Pointer, *Measuring Colour* (Wiley, 2011).
3. D. H. Foster and K. Amano, "Hyperspectral imaging in color vision research: tutorial," J. Opt. Soc. Am. A **36**, 606–627 (2019).
4. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae* (Wiley, 1982).
5. CIE, "Colorimetry," 4th ed. (CIE Central Bureau, 2018).
6. D. H. Foster, K. Amano, S. M. C. Nascimento, *et al.*, "Frequency of metamerism in natural scenes," J. Opt. Soc. Am. A **23**, 2359–2372 (2006).
7. A. Akbarinia and K. Gegenfurtner, "Metameric mismatching in natural and artificial reflectances," J. Vis. **17**(10):390 (2017).
8. G. D. Finlayson and P. Morovic, "Metamer sets," J. Opt. Soc. Am. A **22**, 810–819 (2005).
9. A. D. Logvinenko, B. Funt, and C. Godau, "Metamer mismatching," IEEE Trans. Image Process. **23**, 34–43 (2014).
10. P. Morovič and J. Morovič, "Real world metamer sets: or how we came to love noise," presented at Final Program and Proceedings-IS and T/SID Color Imaging Conference, USA, 2021.
11. X. Zhang, B. Funt, and H. Mirzaei, "Metamer mismatching in practice versus theory," J. Opt. Soc. Am. A **33**, A238–A247 (2016).
12. L. D. Griffin, "Reconciling the statistics of spectral reflectance and colour," PLoS One **14**, e0223069 (2019).
13. D. H. Foster, K. Amano, and S. M. C. Nascimento, "Time-lapse ratios of cone excitations in natural scenes," Vision Res. **120**, 45–60 (2016).
14. D. H. Foster, "Fluctuating environmental light limits number of surfaces visually recognizable by colour," Sci. Rep. **11**, 2102 (2021).
15. D. H. Miller, *Energy at the Surface of the Earth: An Introduction to the Energetics of Ecosystems* (Academic, 1981).
16. W. K. Smith and Z. C. Berry, "Sunflecks?" Tree Physiol. **33**, 233–237 (2013).
17. R. Pastilha and A. Hurlbert, "Chapter 14: seeing and sensing temporal variations in natural daylight," in *Progress in Brain Research*, N. Santhi and M. Spitschan, eds. (Elsevier, 2022), pp. 275–301.
18. C. Yu, M. Wijntjes, E. Eisemann, *et al.*, "Quantifying the spatial, temporal, angular and spectral structure of effective daylight in perceptually meaningful ways," Opt. Express **31**, 8953–8974 (2023).
19. P. M. Hubel, "The perception of color at dawn and dusk," J. Imaging Sci. Technol. **44**, 371–375 (2000).
20. S. M. C. Nascimento, K. Amano, and D. H. Foster, "Spatial distributions of local illumination color in natural scenes," Vision Res. **120**, 39–44 (2016).
21. J. M. M. Linhares, P. D. Pinto, and S. M. C. Nascimento, "The number of discernible colors in natural scenes," J. Opt. Soc. Am. A **25**, 2918–2924 (2008).
22. L. F. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," Prob. Inf. Transm. **23**, 95–101 (1987).
23. M. N. Goria, N. N. Leonenko, V. V. Mergel, *et al.*, "A new class of random vector entropy estimators and its applications in testing statistical hypotheses," J. Nonparametr. Stat. **17**, 277–297 (2005).
24. I. Marín-Franch, M. Sanz-Sabater, and D. H. Foster, "Application of offset estimator of differential entropy and mutual information with multivariate data," Exp. Results **3**, e16 (2022).
25. X. Olano, *Basic Measurements of Radiation at Station Cener (2010-06)* (Pangaea, 2011).
26. X. Olano, *Basic Measurements of Radiation at Station Cener (2010-10)* (Pangaea, 2011).
27. N. Ekpenyong, "Hyperspectral imaging: calibration and applications with natural scenes," Ph.D. thesis (University of Manchester, 2013).
28. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," Color Res. Appl. **26**, 340–350 (2001).

29. C. Li, M. R. Luo, B. Rigg, *et al.*, "CMC 2000 chromatic adaptation transform: CMCCAT2000," Color Res. Appl. **27**, 49–58 (2002).
30. M. D. Fairchild, *Color Appearance Models* (Wiley, 2013).
31. A. Akbarinia and K. R. Gegenfurtner, "Color metamerism and the structure of illuminant space," J. Opt. Soc. Am. A **35**, B231–B238 (2018).
32. M. A. Aldaba, J. M. M. Linhares, P. D. Pinto, *et al.*, "Visual sensitivity to color errors in images of natural scenes," Vis. Neurosci. **23**, 555–559 (2006).
33. B. Hill, T. Roger, and F. W. Vorhagen, "Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula," ACM Trans. Graph. **16**, 109–154 (1997).
34. W. S. Mokrzycki and M. Tatol, "Colour difference Delta E: a survey," Mach. Graph. Vis. **20**, 383–411 (2011).
35. S. Westland, C. Ripamonti, and V. Cheung, *Computational Colour Science Using MATLAB* (Wiley, 2012).
36. M. Stokes, M. D. Fairchild, and R. S. Berns, "Precision requirements for digital color reproduction," ACM Trans. Graph. **11**, 406–422 (1992).
37. H. Wang, M. R. Luo, G. Cui, *et al.*, "A comparison between perceptibility and acceptability methods," in *11th Congress of the International Colour Association (AIC)*, D. Smith, P. Green-Armytage, M. A. Pope, and N. Harkness, eds. (International Colour Association (AIC), 2009), pp. 1–7.
38. C. Witzel and K. R. Gegenfurtner, "Color perception: objects, constancy, and categories," Annu. Rev. Vis. Sci. **4**, 475–499 (2018).
39. P.-L. Sun and J. Morovic, "Inter-relating colour difference metrics," in *10th Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications* (Society for Imaging Science and Technology, 2002), pp. 55–60.
40. D. H. Foster and A. Reeves, "Colour constancy failures expected in colourful environments," Proc. R. Soc. B **289**, 20212483 (2022).
41. D. H. Foster, "The Verriest lecture: color vision in an uncertain world," J. Opt. Soc. Am. A **35**, B192–B201 (2018).
42. S. Plet and W. Gerbino, "Background articulation and relational colour constancy," Color Res. Appl. **26**, S201–S204 (2001).
43. K. J. Linnell and D. H. Foster, "Scene articulation: dependence of illuminant estimates on number of surfaces," Perception **31**, 151–159 (2002).
44. L. T. Maloney and J. A. Schirillo, "Color constancy, lightness constancy, and the articulation hypothesis," Perception **31**, 135–139 (2002).
45. B.-G. Khang and Q. Zaidi, "Illuminant color perception of spectrally filtered spotlights," J. Vis. **4**(9):2 (2004).
46. I. K. Zemach and M. E. Rudd, "Effects of surround articulation on lightness depend on the spatial arrangement of the articulated region," J. Opt. Soc. Am. A **24**, 1830–1841 (2007).
47. T. Vladusich, "Gamut relativity: a new computational approach to brightness and lightness perception," J. Vis. **13**(1):14 (2013).
48. J. L. Nieves, J. Ojeda, L. Gómez-Robledo, *et al.*, "Psychophysical determination of the relevant colours that describe the colour palette of paintings," J. Imaging **7**, 72 (2021).
49. Z. Tirandaz, D. H. Foster, J. Romero, *et al.*, "Efficient quantization of painting images by relevant colors," Sci. Rep. **13**, 3034 (2023).
50. C. Witzel, Z. Flack, E. Sanchez-Walker, *et al.*, "Colour category constancy and the development of colour naming," Vision Res. **187**, 41–54 (2021).
51. M. Olkkonen, C. Witzel, T. Hansen, *et al.*, "Categorical color constancy for real surfaces," J. Vis. **10**(9):16 (2010).
52. S. Peyvandi and S. H. Amirshahi, "Paramerism and reliable parameric correction," Color Res. Appl. **36**, 437–448 (2011).
53. B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, 1993).
54. D. H. Foster, "Hyperspectral radiance image pairs used in Foster, D.H. (2021). Fluctuating environmental light limits number of surfaces visually recognizable by colour. Scientific Reports, 11, 2102," figshare (2022), https://doi.org/10.6084/m9.figshare.c.5240420.
55. D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," J. Opt. Soc. Am. A **15**, 2036–2045 (1998).
56. D. H. Foster and S. M. C. Nascimento, "Little information loss with red-green color deficient vision in natural environments," iScience **26**, 107421 (2023).
57. L. Arend, "Environmental challenges to color constancy," presented at Photonics West, San Jose, California, 2001.
58. J. J. Koenderink, "The prior statistics of object colors," J. Opt. Soc. Am. A **27**, 206–217 (2010).
59. H. E. Smithson, "Sensory, computational and cognitive components of human colour constancy," Philos. Trans. R. Soc. Lond. B **360**, 1329–1346 (2005).
60. D. H. Foster, "Color constancy," Vision Res. **51**, 674–700 (2011).
61. D. H. Brainard and A. Radonjić, "Colorconstancy," in *The New Visual Neurosciences*, J. S. Werner and L. M. Chalupa, eds. (MIT, 2014), pp. 545–556.
62. J. M. Kraft and D. H. Brainard, "Mechanisms of color constancy under nearly natural viewing," Proc. Natl. Acad. Sci. USA **96**, 307–312 (1999).
63. V. M. N. de Almeida and S. M. C. Nascimento, "Perception of illuminant colour changes across real scenes," Perception **38**, 1109–1117 (2009).
64. K. R. Gegenfurtner, D. Weiss, and M. Bloj, "Color constancy in real-world settings," J. Vis. **24**(2):12 (2024).
65. M. A. Webster and J. D. Mollon, "Adaptation and the color statistics of natural images," Vision Res. **37**, 3283–3298 (1997).
66. O. Rinner and K. R. Gegenfurtner, "Time course of chromatic adaptation for color appearance and discrimination," Vision Res. **40**, 1813–1826 (2000).
67. A. Werner, "Spatial and temporal aspects of chromatic adaptation and their functional significance for colour constancy," Vision Res. **104**, 80–89 (2014).
68. B. J. Craven and D. H. Foster, "An operational approach to colour constancy," Vision Res. **32**, 1359–1366 (1992).
69. D. H. Foster and S. M. C. Nascimento, "Relational colour constancy from invariant cone-excitation ratios," Proc. R. Soc. B **257**, 115–121 (1994).
70. S. M. C. Nascimento and D. H. Foster, "Relational color constancy in achromatic and isoluminant images," J. Opt. Soc. Am. A **17**, 225–231 (2000).
71. D. H. Foster, S. M. C. Nascimento, K. Amano, *et al.*, "Parallel detection of violations of color constancy," Proc. Natl. Acad. Sci. USA **98**, 8151–8156 (2001).
72. J. J. van Es, T. Vladusich, and F. W. Cornelissen, "Local and relational judgements of surface colour: constancy indices and discrimination performance," Spat. Vis. **20**, 139–154 (2007).
73. H. Karimipour and C. Witzel, "Colour expectations across illumination changes," Vision Res. **222**, 108451 (2024).
74. S. M. C. Nascimento and D. H. Foster, "Misidentifying illuminant changes in natural scenes due to failures in relational colour constancy," Proc. R. Soc. B **290**, 20231676 (2023).
75. D. Katz, *The World of Colour* (Routledge, Trench, Trubner & Co., Ltd., 1935).
76. M. Manassi and D. Whitney, "Continuity fields enhance visual perception through positive serial dependence," Nat. Rev. Psychol. **3**, 352–366 (2024).
77. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, 2006).
78. I. Marín-Franch and D. H. Foster, "Estimating information from image colors: an application to digital cameras and natural scenes," IEEE Trans. Pattern Anal. Mach. Intell. **35**, 78–91 (2013).
79. G. Feng and D. H. Foster, "Predicting frequency of metamerism in natural scenes by entropy of colors," J. Opt. Soc. Am. A **29**, A200–A208 (2012).
80. J. A. Rice, *Mathematical Statistics and Data Analysis* (Cengage Learning, 2007).
81. University of Manchester, https://personalpages.manchester.ac.uk/staff/d.h.foster/.
82. S. Nascimento, "Scientific data," https://sites.google.com/view/sergionascimento/home/scientific-data.
83. I. Marín-Franch, M. Sanz-Sabater, and D. H. Foster, "KLo: application of offset estimator of differential entropy and mutual information with multivariate data. Experimental results," GitHub (2022), https://github.com/imarinfr/klo.