Predicting frequency of metamerism in natural scenes by entropy of colors

Gaoyang Feng and David H. Foster*

School of Electrical and Electronic Engineering, University of Manchester, Manchester M13 9PL, UK
*Corresponding author: d.h.foster@manchester.ac.uk

Estimating the frequency of metameric surfaces in natural scenes usually requires many comparisons of surface colors to determine which are visually indistinguishable under one light but distinguishable—by a certain criterion degree—under another. The aim here was to test the predictive power of a simpler approach to estimation based on the entropy of colors. In simulations with 50 hyperspectral images of natural scenes, the logarithm of the observed relative frequency of metamerism in each scene under two successive daylights was regressed on combinations of the estimated Shannon differential entropies of the colors of the scene under the same two daylights. The regression was strong, and it remained so when restricted to the estimated differential entropy under just the first daylight, providing that the criterion degree of metamerism was limited. When the criterion degree was made more extreme, however, the restricted regression failed. A possible explanation of the predictive power of differential entropy is briefly considered.   © 2012 Optical Society of America

OCIS codes: 330.1690, 330.1715, 330.1720, 330.1880, 330.4060, 110.3055.

1. INTRODUCTION

The phenomenon of metamerism may be revealed in several ways [1, Sec. 3.9], [2, Sec. 3.8], [3, Chap. 6], [4]. But in one form or another, it represents a loss of information in which a generally complex spectral distribution describing lights or reflecting surfaces is reduced typically to three numbers, corresponding to the excitations of the three classes of cone receptors of the normal eye (or sensor responses of an RGB color camera). The metamerism associated with reflecting surfaces is particularly important [1, Sec. 3.9], not least because it concerns the foundations of camouflage [5] and the relationship between visual and material identity in the natural world [6]. In this form of metamerism, two reflecting surfaces are defined as metameric if they have different spectral reflectances such that with some illumination spectrum they produce equal receptor responses (or equivalently equal tristimulus values) [2, Sec. 3.9]. Although suitable for counting metamers in the theoretical object-color solid [2, Sec. 3.10], this definition does not lend itself easily to making inferences about the frequency of metamerism in natural scenes.

In practice, a difference in the material identity of metameric surfaces usually becomes evident under a change in illuminant (illuminants are assumed here to be broadband, represented by daylights). Conventionally, metamerism in this context is measured by the CIE special metamerism index for illuminant changes [4], which quantifies the color mismatch under the new illuminant, but for the purpose of estimating the frequency of metamerism, it is more usefully expressed in relation to the limits of visual sensitivity [7]. That is, given some threshold for visual distinguishability, the frequency of metamerism in a scene is estimated by the number of pairs of surfaces whose color differences are less than this threshold under one phase of daylight and greater than this threshold by a certain multiple—the criterion degree of metamerism—under another phase of daylight [6]. The greater the criterion degree of metamerism, the more extreme the difference in colors.

Quantified in this way, metamerism in natural scenes is generally rare. Estimates of its frequency, expressed as a proportion of all possible pairs of surfaces in a scene, have been found to range from about 10^{-6} to 10^{-4} under successive daylights with correlated color temperatures (CCTs) of 25000 K and 4000 K [6]. Predicting this relative frequency by theoretical methods is, however, difficult, for it is a complicated function of the spectral reflecting properties of the surfaces in the scene, their relative frequencies of occurrence, the spectra of the illuminants, and the choice of threshold and criterion degree of metamerism, as well as the spectral sensitivities of the receptors of the eye. Thus, explicit numerical simulations are normally required in which many pairs of surface colors are tested to determine those that satisfy the definition, that is, visually indistinguishable under the one illuminant but distinguishable under the other.

One simpler alternative approach to estimation is to make use of the statistical properties of the colors of a scene. The aim of this work was to test the predictive power of one such property, namely the entropy of colors, which has been exploited in other applications related to color vision [8–11]. Informally, the entropy of colors in a scene may be thought of as a measure of the uncertainty of the color of a randomly chosen point in the scene. Formal definitions are given later.

In order to compare explicit estimates of the frequency of metamerism with entropy-based estimates, numerical simulations were undertaken in which reflected spectra were calculated at each point within each of 50 natural rural and urban scenes under daylights with different CCTs. Colorimetric quantities were evaluated within the CIE color space CIECAM02 [12], which was chosen not for its representation of particular appearance attributes, but for its approximate uniformity in defining color differences, an important factor in the evaluation of

differential entropy. As a control, colorimetric quantities were also evaluated within the color space CIELAB, which is rather less uniform than CIECAM02 [4], but which has been used previously [6]. The relative frequency of metamerism in each scene under two successive daylights was estimated, and its logarithm was regressed on combinations of the estimated Shannon differential entropies [13] of the colors of the scene under the two daylights. The goodness of fit was summarized by R^2, the proportion of variance accounted for by the entropy models, after due allowance for the degrees of freedom in the regression.

The regression was found to be strong, and it remained so when it was restricted to the estimated differential entropy of colors under just the first daylight, providing that the criterion degree of metamerism was limited. But if the criterion degree of metamerism was made more extreme, the restricted regression failed. A possible explanation of the predictive power of differential entropy and its failure is briefly considered.

2. METHODS

A. Reflected Spectra

Spectral-reflectance data were taken from 50 hyperspectral images of natural scenes [6,14]. Thumbnail illustrations of some of these scenes are available in [15], and four are shown in Fig. 1. None of the scenes used here contained visible sky or other self-luminous regions or objects. Each hyperspectral image, which had spatial dimensions $\leq 1344 \times 1024$ pixels and spectral range 400–720 nm sampled at 10 nm intervals, provided a discrete representation of an effective spectral reflectance $r(\lambda; x, y)$ at each wavelength λ and position (x, y). The effect of illuminating the scene with a particular illuminant with spectrum $e(\lambda)$ was simulated by multiplying $r(\lambda; x, y)$ at each point (x, y) by $e(\lambda)$. The assumptions and approximations involved in this approach have been discussed in [6, Appendix A]. As emphasized there, the intention is to model the effects of changes in natural illumination spectra, not the spatial changes in illumination resulting from changes in, e.g., the position of the sun and the distribution of cloud that give rise to those changes in illumination spectra. Because of the approximately 1.3 pixel line spread function of the camera system [6], only alternate pixels were sampled.

Daylight spectra were simulated from those described by the CIE [4]. They included average daylight with CCT 6500 K and the extremes with CCTs of 4000 K and 25000 K, characteristic of the sun and sky at different times of the day. On a reciprocal color-temperature scale, 4000 K and 25000 K are approximately equidistant from 6500 K.

B. Color Signals

The color signal at each point in a scene was represented by its 1931 CIE XYZ tristimulus values, calculated in the usual way; that is, at each point (x, y) with spectral reflectance $r(\lambda; x, y)$ and illuminant $e(\lambda)$, the values of X, Y, Z are given by

$$X = \int \bar{x}(\lambda)e(\lambda)r(\lambda; x, y)\mathrm{d}\lambda,$$

$$Y = \int \bar{y}(\lambda)e(\lambda)r(\lambda; x, y)\mathrm{d}\lambda,$$

$$Z = \int \bar{z}(\lambda)e(\lambda)r(\lambda; x, y)\mathrm{d}\lambda, \qquad (1)$$

where $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$ are the CIE 1931 color-matching functions [4] and the evaluation of the integral is over the visible wavelength range.

These values of X, Y, Z were then transformed into the corresponding tristimulus values under a reference daylight illuminant with CCT 6500 K according to the CIECAM02 specification with default values [12], including those for chromatic adaptation, for compatibility with [11], and for CIELAB [4] with the chromatic adaptation transform CMCCAT2000 [16], assumed to be complete for compatibility with [6]. These corresponding tristimulus values were then transformed, respectively, into CIECAM02 coordinates J, a_C, b_C [12] and CIELAB coordinates L^*, a^*, b^* [4] with respect to D65. Color differences between points in CIECAM02, which, as noted earlier, is approximately uniform, were calculated from the corresponding Euclidean distance [17,18], but in CIELAB, because of its more limited uniformity, from the color-difference formula CIEDE2000 [4,19].

C. Frequency of Metamerism

The relative frequency of metamerism was estimated much as in [6]. Each scene was assumed to be illuminated in turn by one of two selected daylights. With spatially uniform random sampling, 50000 points (about 15% of the total available) were chosen from each scene, yielding $N = 50000 \times 49999/2 \approx 1.25 \times 10^9$ pairs. These large samples were taken to allow accurate estimates of relative frequency for individual scenes; smaller samples of 3000 points were also taken for comparison with previous estimates [6]. Uniform random sampling was used, rather than one determined by an analysis of scene structure, to ensure the neutrality of the estimates with respect to scene contents. For each color space CIECAM02 and CIELAB, a nominal color-difference threshold value ΔE^{thr} was chosen, and, from the set of N pairs of points, the N_0 pairs with color differences less than ΔE^{thr} were determined for the first daylight. From this subset of N_0 pairs, the N_1 pairs with color differences greater than a certain multiple $n = 1, 2, 3, 4$ of ΔE^{thr} were determined for the second daylight. These N_1 pairs were the metameric pairs of the sample for the criterion degree of metamerism defined by the multiplier n. The relative frequency of metamerism in the scene as a whole was estimated by N_1/N.

In principle, the nominal threshold value ΔE^{thr} could have been replaced by a psychometric function [11], but for the present purpose this would have unnecessarily increased the complexity of the analysis. The nominal threshold ΔE^{thr} was set, in turn, to 0.5 and 1.0 for both color spaces. These values encompass typical thresholds for the two spaces [20–22].

The quotient N_1/N can be decomposed [6] into two factors N_0/N and N_1/N_0. The first of these factors N_0/N estimates the relative frequency of pairs of points in the scene with color differences less than ΔE^{thr} under the first daylight. The second factor N_1/N_0 estimates a conditional relative frequency, namely, given pairs of points with color differences less than ΔE^{thr} under the first daylight, the relative frequency of their color differences being greater than $n\Delta E^{\mathrm{thr}}$ under the second daylight. To anticipate the later analysis, all three quotients N_1/N, N_0/N, and N_1/N_0 proved relevant, and values of all three were recorded for each scene. These quotients were expressed as logarithms to the base 10, where defined.
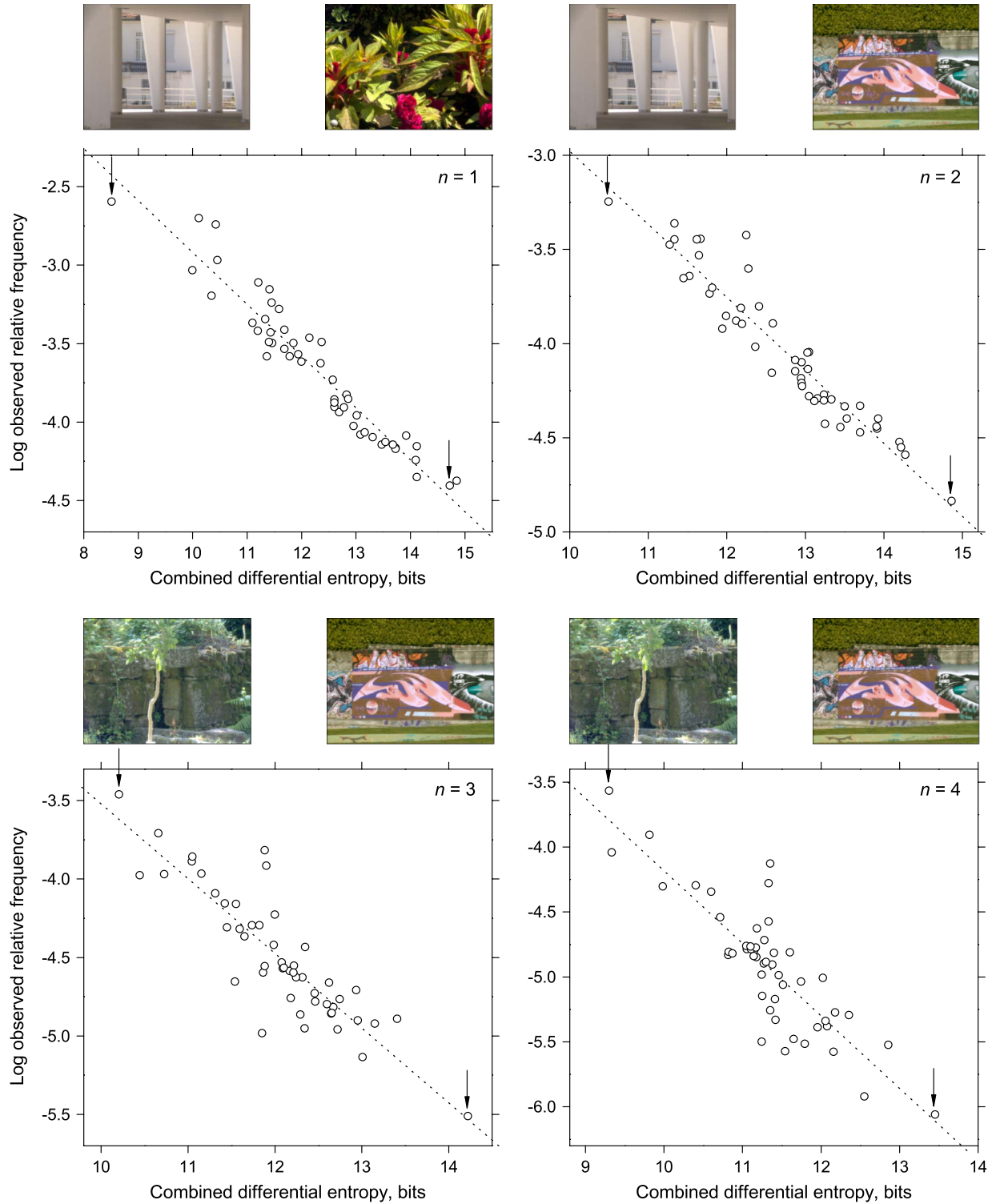
Fig. 1.   (Color online) Prediction of relative frequency of metamerism by combinations of differential entropy. In each panel, the logarithm of the observed relative frequency of metamerism for daylights with CCTs of 4000 K and 25000 K and nominal color-difference threshold ΔE^{thr} of 0.5 is plotted against the combination of estimated entropies defined by Eq. (6′). The dotted lines represent perfect fits. Each panel is based on data for 50 natural scenes with a different criterion degree of metamerism n of 1, 2, 3, and 4, as indicated. Notice the different abscissa and ordinate ranges. The images above each panel show the scenes with the lowest and highest estimated relative frequencies of metamerism.

D. Differential Entropy

Since points were sampled randomly from each scene under a selected daylight, the triplets of values (J, a_C, b_C) in CIECAM02 or of values (L^*, a^*, b^*) in CIELAB may be treated as instances of a trivariate continuous random variable, U say,

governed by a probability density function (pdf), f say, which in general varies with the scene and daylight. As in [11], the degree of uncertainty of the random variable U was quantified by the Shannon differential entropy $h(U)$, defined [13] as

$$h(U) = -\int f(u) \log f(u) \mathrm{d}u, \qquad (2)$$

where the integration is taken over the support set $f(u) > 0$. The conventional symbol h for differential entropy should not be confused with the symbol h for hue in CIECAM02. If the logarithm is to the base 2, then the differential entropy is measured in bits.

The relationship between the notions of differential entropy for continuous random variables and of entropy for discrete random variables is discussed in [13]. In particular, differential entropy is not the limit of the corresponding summation defining the entropy of the discretized version of U (assuming that the pdf f is Riemann integrable).

With two continuous random variables, U_1 and U_2 say, the Shannon conditional differential entropy $h(U_2|U_1)$ of U_2 given U_1 may also be defined as in Eq. (2), but, more usefully here, it can be obtained in the following way. Let $h(U_1)$ and $h(U_2)$ be the differential entropies of the colors of the scene under the first and second daylights, respectively, and let $h(U_2, U_1)$ be the joint differential entropy of the pair (U_2, U_1). Then [13]

$$h(U_2|U_1) = h(U_2, U_1) - h(U_1). \qquad (3)$$

In obtaining estimates of the differential entropy, however, two technical issues need to be addressed. First, the estimate varies with the coordinate system in which U takes its values [13]. Fortunately, this variation does not affect the strength of the regression of the frequency of metamerism on differential entropy, since the same coordinate system was used for all the entropy estimates (either CIECAM02 or CIELAB). Second, naive estimators of differential entropy based on, e.g., binning frequency counts to estimate the pdf f in Eq. (2) are known to be susceptible to bias [23]. An asymptotically bias-free, k-nearest-neighbor estimator due to Kozachenko and Leonenko [24,25] was therefore used. It was applied in a computationally efficient offset form due to Marín-Franch and Foster [9,26].

E. Modeling Frequency of Metamerism

The relationship between differential entropy and the relative frequency of metamerism depends on the decomposition $N_1/N = (N_0/N)(N_1/N_0)$ (Subsection 2.C). Consider, first, the factor N_0/N. Since it estimates the relative frequency of pairs of points in a scene with color differences less than ΔE^{thr} under the first daylight, it necessarily depends on the degree of uncertainty of U_1: the more unpredictable U_1, the more likely that color differences will be greater than ΔE^{thr} and, therefore, the lower the value of N_0/N. Accordingly, the observed value of $\log N_0/N$ for each scene and pair of daylights might be modeled, to first order, by the estimate of $h(U_1)$ for that scene and the first daylight, but with a negative coefficient of proportionality; that is,

$$\mathrm{E}[\log(N_0/N)] = -\beta_1 h(U_1) + \alpha_1, \qquad (4)$$

where E is the expectation, β_1 and α_1 are scalars, and $\beta_1 > 0$.

Consider, next, the factor N_1/N_0. Since it estimates a conditional relative frequency, namely, given pairs of points with color differences less than ΔE^{thr} under the first daylight, the relative frequency of their color differences being greater than $n\Delta E^{\mathrm{thr}}$ under the second daylight (Subsection 2.C), it necessarily depends on the degree of uncertainty of U_2 given U_1

(although, as explained later, this dependency is complicated by the criterion degree of metamerism n). That is, the more unpredictable U_2 is given U_1, the more likely that color differences will be greater than $n\Delta E^{\mathrm{thr}}$, and, therefore, the higher the value of N_1/N_0. Accordingly, the observed value of $\log N_1/N_0$ for each scene and pair of daylights might be modeled, to first order, by the estimate of $h(U_2|U_1)$ for that scene and daylights, with a positive coefficient of proportionality; that is,

$$\mathrm{E}[\log(N_1/N_0)] = \beta_2 h(U_2|U_1) + \alpha_2, \qquad (5)$$

where, in general, β_2 and α_2 are different from β_1 and α_1 in Eq. (4), but again $\beta_2 > 0$.

Since $\log N_1/N = \log N_0/N + \log N_1/N_0$, Eqs. (4) and (5) may be combined. The observed value of $\log N_1/N$ for each scene and pair of daylights might thus be modeled, to first order, by

$$\mathrm{E}[\log(N_1/N)] = -\beta_1 h(U_1) + \beta_2 h(U_2|U_1) + \alpha_0, \qquad (6)$$

where β_1 and β_2 are as in Eqs. (4) and (5), and $\alpha_0 = \alpha_1 + \alpha_2$, also as in Eqs. (4) and (5). These two entropies are not, however, independent in practice. If their combination in Eq. (6) is rewritten as $\beta_1' h(U_1) - \beta_2' h(U_2|U_1)$, where $\beta_1' = \beta_1/(\beta_1^2 + \beta_2^2)^{1/2}$ and $\beta_2' = \beta_2/(\beta_1^2 + \beta_2^2)^{1/2}$, then Eq. (6) becomes

$$\mathrm{E}[\log(N_1/N)] = -\beta_0[\beta_1' h(U_1) - \beta_2' h(U_2|U_1)] + \alpha_0, \qquad (6')$$

with $\beta_0 = (\beta_1^2 + \beta_2^2)^{1/2}$. When $\beta_2' = 0$, the combination reduces to $h(U_1)$, and when $\beta_1' = 0$, to $h(U_2|U_1)$. The formulation of Eq. (6') is convenient for plotting the results of the regression.

The observed values of $\log N_1/N$ (Subsection 2.C) were regressed on estimates of $h(U_1)$ and $h(U_2|U_1)$ over the 50 scenes, with estimates of $h(U_2|U_1)$ obtained by applying Eq. (3) to the separate estimates of $h(U_2, U_1)$ and $h(U_1)$ (Subsection 2.D). Goodness of fit was summarized by R^2, the proportion of variance accounted for by the model. Except where the regression failed, all values of R^2 were adjusted for the degrees of freedom associated with different numbers of regressor variables [27].

For comparison, goodness of fit was calculated for the restricted models obtained by setting $\beta_2' = 0$ in Eq. (6'), that is,

$$\mathrm{E}[\log(N_1/N)] = -\beta_1 h(U_1) + \alpha_1, \qquad (7)$$

and by setting $\beta_1' = 0$ in Eq. (6'), that is,

$$\mathrm{E}[\log(N_1/N)] = \beta_2 h(U_2|U_1) + \alpha_2. \qquad (8)$$

Estimates of the standard error (SE) for R^2 were obtained by a bootstrap with at least 1000 iterations [28].

3. RESULTS AND COMMENT

Figure 1 illustrates the dependence of the frequency of metamerism on the entropy of colors for the extreme pair of daylights with CCTs of 4000 K and 25000 K. In each panel, the logarithm of the observed relative frequency of metamerism is plotted against the combination of estimated entropies defined by Eq. (6'). Each point represents data from one of the 50 natural scenes. The color space was CIECAM02, and the nominal color-difference threshold ΔE^{thr} was 0.5 (results

were closely similar to those with ΔE^{thr} of 1.0). Each of the panels is for a different criterion degree of metamerism n of 1, 2, 3, and 4, as indicated. The dotted lines represent perfect fits. The coefficients β_1' and β_2' of the full model defined by Eq. (6′) were all positive.

The images above each panel show the scenes with the lowest and highest estimated relative frequencies of metamerism. Their identity varies with the criterion degree of metamerism because the frequency distributions of the color differences in each scene have maxima in different positions and tails of differing length, as illustrated in [6].

Table 1 gives numerical details of the regressions for these and other conditions, all with CIECAM02. For each combination of nominal threshold ΔE^{thr}, criterion degree of metamerism n, and CCTs of the first and second daylights, the entries show the ranges of the logarithm of the observed relative frequencies $\log N_0/N$, $\log N_1/N_0$, and $\log N_1/N$ over the 50 scenes (Subsection 2.C) and the proportions R^2 of the variance accounted for by the corresponding entropy models (Subsection 2.E). The first row of entries in the table summarizes the data in the top left panel of Fig. 1; the fourth, seventh, and tenth rows summarize the data in the top right, bottom left, and bottom right panels of Fig. 1.

The goodness of fit of the full model defined by Eq. (6) is shown in last column of Table 1. For $n = 1$, the regression was strong for all three pairs of daylights and both thresholds, with

mean $R^2 = 0.90$ (mean SE = 0.02). The regression was less strong, however, for $n \geq 2$, with mean $R^2 = 0.66$ (mean SE = 0.09). From plots of the kind shown in Fig. 1, it can be seen that this weakening of the regression with increased n is due not to a loss in linearity of the dependence (i.e., departure from degree 1) but to a loss in the dependence itself. Even so, the size of n is not the only predictor of the loss in dependence. The difference in the ranges of $\log N_0/N$ and $\log N_1/N_0$ also acted as a predictor. Figure 2 shows R^2 for the full model of Eq. (6) (last column Table 1) plotted against the difference in ranges of $\log N_0/N$ (fifth column Table 1) and $\log N_1/N_0$ (sixth column Table 1). The dotted line shows a linear regression, for which $R^2 = 0.74$ (SE = 0.08), as elsewhere adjusted for degrees of freedom [27].

The goodness of fit of the restricted model defined by Eq. (7) is shown in the third column from the last of Table 1. For $n = 1$, the regression remained strong for all three pairs of daylights and both thresholds, with mean $R^2 = 0.88$ (mean SE = 0.03). But, by contrast with the full model, the regression generally failed for $n \geq 2$, with mean $R^2 = 0.12$ (mean SE = 0.06), little different from chance.

As with the full model, the difference in the ranges of $\log N_0/N$ and $\log N_1/N_0$ acted as a predictor of the loss in dependence. Figure 3 shows R^2 for the restricted model of Eq. (7) plotted against the difference in ranges of $\log N_0/N$ (fifth column Table 1) and $\log N_1/N_0$ (sixth column Table 1). The

Table 1. Regression over 50 Scenes of the Logarithm of Observed Relative Frequency of Metamerism on Estimated Differential Entropy Calculated within the Color Space CIECAM02 [12] for Various Nominal Color-Difference Thresholds, Criterion Degrees of Metamerism, and Daylights with Different Correlated Color Temperatures

Threshold[a] ΔE^{thr}	Criterion degree[b] n	First daylight CCT, K	Second daylight CCT, K	Range[c] $\log N_0/N$	Range[d] $\log N_1/N_0$	Range[e] $\log N_1/N$	R^2 Eq. (7)[f]	R^2 Eq. (8)[g]	R^2 Eq. (6)[h]
0.5	1	4000	25000	1.94	0.24	1.81	0.92	0.43	0.92
		4000	6500	1.95	0.56	1.90	0.86	0.31	0.88
		6500	25000	1.90	0.68	1.71	0.89	0.31	0.90
	2	4000	25000	1.94	1.15	1.59	0.74	0.10	0.90
		4000	6500	1.94	3.17	2.40	0.12	0.02	0.62
		6500	25000	1.88	2.29	1.91	0.17	0.03	0.64
	3	4000	25000	1.94	2.35	2.05	0.23	0.00	0.76
		4000	6500	1.94	4.15	3.52	0.00[i]	0.29	0.63
		6500	25000	1.90	3.45	2.88	0.00[i]	0.30	0.61
	4[j]	4000	25000	1.95	2.87	2.49	0.03	0.16	0.72
		4000	6500	1.95	4.88	4.66	0.04	0.45	0.60
		6500	25000	1.90	4.42	3.76	0.02[i]	0.44	0.61
1.0	1	4000	25000	1.76	0.33	1.59	0.93	0.35	0.94
		4000	6500	1.77	0.68	1.80	0.82	0.25	0.87
		6500	25000	1.74	0.83	1.57	0.84	0.22	0.86
	2	4000	25000	1.76	1.51	1.77	0.46	0.00[i]	0.84
		4000	6500	1.75	3.85	3.40	0.01	0.07	0.42
		6500	25000	1.72	2.95	2.85	0.01	0.18	0.57
	3[j]	4000	25000	1.76	3.14	2.84	0.01[i]	0.26	0.73
		6500	25000	1.74	5.56	5.38	0.01	0.44	0.59
	4[j]	4000	25000	1.76	4.03	3.70	0.01[i]	0.42	0.70

[a]Nominal threshold for color differences under first daylight, giving N_0 pairs out of $N \approx 1.25 \times 10^9$ pairs chosen randomly from each of the 50 scenes
[b]Multiple of ΔE^{thr} for color differences under second daylight, giving N_1 pairs out of N_0.
[c]Difference between maximum and minimum of $\log N_0/N$.
[d]Difference between maximum and minimum of $\log N_1/N_0$.
[e]Difference between maximum and minimum of $\log N_1/N$.
[f]Proportion of variance accounted for by restricted linear model (7), adjusted for degrees of freedom.
[g]Proportion of variance accounted for by restricted linear model (8), adjusted for degrees of freedom.
[h]Proportion of variance accounted for by full linear model (6), adjusted for degrees of freedom.
[i]Adjustment for degrees of freedom omitted, as invalid.
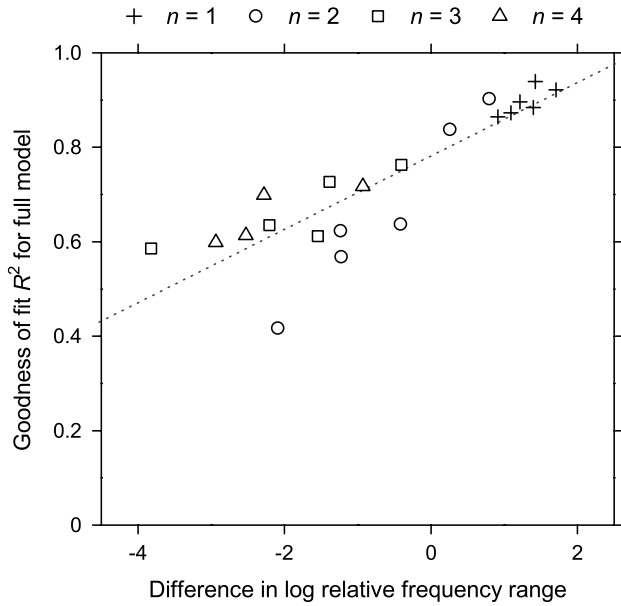[j]Results for other daylight pair or pairs omitted as $N_1 = 0$ for one or more scenes.

Fig. 2. Dependence of goodness of fit of the full entropy model on the difference in relative frequency ranges. Proportion R^2 of variance accounted for by Eq. (6) is plotted against the difference in ranges of $\log N_0/N$ and $\log N_1/N_0$, values taken from Table 1. Each point represents data from 50 scenes under different pairs of daylights, with different nominal color-difference thresholds, and different criterion degrees of metamerism n. The dotted line is a linear regression.
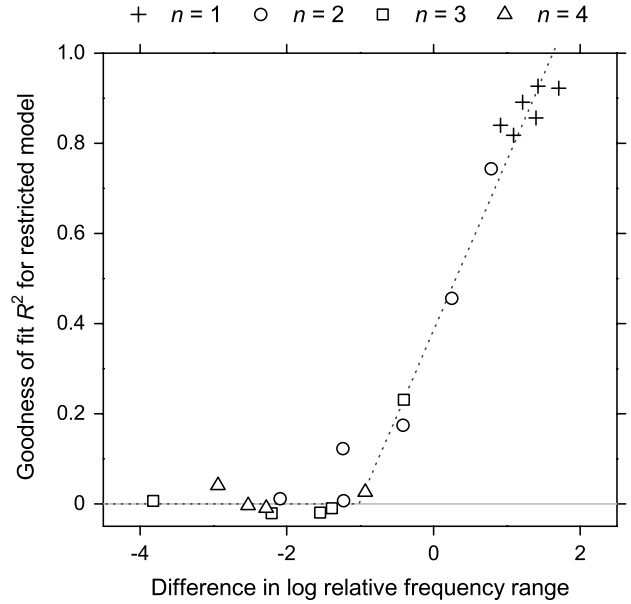

Fig. 3. Dependence of goodness of fit of the restricted entropy model on the difference in relative frequency ranges. Proportion R^2 of variance accounted for by Eq. (7) is plotted against the difference in ranges of $\log N_0/N$ and $\log N_1/N_0$, values taken from Table 1. Each point represents data from 50 scenes under different pairs of daylights, with different nominal color-difference thresholds and different criterion degrees of metamerism n. The dotted line is a piecewise linear regression.

dotted line shows a piecewise linear regression with the first segment of the model forced to zero. For this regression, $R^2 = 0.98$ (SE = 0.01), again adjusted for degrees of freedom [27].

The goodness of fit of the other restricted model defined by Eq. (8) is shown in the second column from the last of Table 1. For all n, all three pairs of daylights, and both thresholds, the regression was weak, with mean $R^2 = 0.24$ (mean SE = 0.11).

A physical interpretation of the regressions shown in Figs. 1–3 is considered in Section 4.

With the color space CIELAB, there was a similar pattern of performance. Table 2 summarizes the numerical details of the regressions (there were fewer entries, as samples from each scene were based on just 3000 points). The regressions were, however, less strong. For example, with $n = 1$, for the full model of Eq. (6), mean $R^2 = 0.73$ (mean SE = 0.11); for the restricted model of Eq. (7), mean $R^2 = 0.62$ (mean SE = 0.11); and for the restricted model of Eq. (8), mean $R^2 = 0.12$ (mean SE = 0.10).

Table 2. Regression over 50 Scenes of the Logarithm of Observed Relative Frequency of Metamerism on Estimated Differential Entropy Calculated within the Color Space CIELAB with Color-Difference Formula CIEDE2000 [4] for Various Nominal Color-Difference Thresholds, Criterion Degrees of Metamerism, and Daylights with Different Correlated Color Temperatures

Threshold[a] ΔE^{thr}	Criterion degree[b] n	First daylight CCT, K	Second daylight CCT, K	Range[c] $\log N_0/N$	Range[d] $\log N_1/N_0$	Range[e] $\log N_1/N$	R^2 Eq. (7)[f]	R^2 Eq. (8)[g]	R^2 Eq. (6)[h]
0.5	1	4000	25000	2.43	0.94	1.77	0.75	0.24	0.78
		4000	6500	2.50	1.31	1.67	0.67	0.13	0.74
		6500	25000	2.48	0.95	2.00	0.77	0.22	0.82
	2[i]	4000	25000	2.48	2.37	2.73	0.34	0.00[j]	0.67
		4000	6500	2.51	4.48	3.20	0.03	0.08	0.51
	3[i]	4000	25000	2.59	4.08	2.86	0.01	0.10	0.47
1.0	1	4000	25000	1.69	1.49	1.49	0.54	0.05	0.68
		4000	6500	1.72	1.61	1.66	0.49	0.01	0.67
		6500	25000	1.72	1.65	1.56	0.53	0.04	0.71
	2[i]	4000	25000	1.72	3.43	3.13	0.01	0.12	0.49

[a]Nominal threshold for color differences under first daylight, giving N_0 pairs out of $N \approx 1.25 \times 10^9$ pairs chosen randomly from each of the 50 scenes.
[b]Multiple of ΔE^{thr} for color differences under second daylight, giving N_1 pairs out of N_0.
[c]Difference between maximum and minimum of $\log N_0/N$.
[d]Difference between maximum and minimum of $\log N_1/N_0$.
[e]Difference between maximum and minimum of $\log N_1/N$.
[f]Proportion of variance accounted for by restricted linear model (7), adjusted for degrees of freedom.
[g]Proportion of variance accounted for by restricted linear model (8), adjusted for degrees of freedom.
[h]Proportion of variance accounted for by full linear model (6), adjusted for degrees of freedom.
[i]Results for other daylight pair or pairs omitted as $N_1 = 0$ for one or more scenes.
[j]Adjustment for degrees of freedom omitted, as invalid.

The poorer performance with CIELAB was unsurprising, with possible explanatory factors including differences in the extent of the assumed adaptation and the use with CIE-LAB of the color-difference formula CIEDE2000 for estimates of the relative frequency of metamerism (Subsection 2.B), but not for entropy.

With CIECAM02, reversing the direction of the change in daylights had, on average, a small effect. For $n = 1$, mean R^2 fell by 0.08, and for $n \geq 2$, mean R^2 fell by 0.06. With CIE-LAB and CIEDE2000, reversing the direction of the change in daylights also had a small effect (cf. [6]). For $n = 1$, mean R^2 rose by 0.07, and for $n \geq 2$, mean R^2 rose by 0.09.

The subsequent discussion will concentrate on results obtained with CIECAM02.

4. DISCUSSION

The relationship between the frequency of metamerism in natural scenes and the entropy of colors is not a direct one. As proposed here, two entropies need to be considered: the differential entropy of the colors in a scene under one daylight and the conditional differential entropy of those colors under another daylight. Importantly, when these two entropies are combined in a linear model to predict the relative frequency of metamerism, the coefficients weighting the entropies appear with opposite signs.

Although this model has only three degrees of freedom, it produced strong regressions across the 50 natural scenes and three pairs of daylight illuminants used in this analysis. With the approximately uniform color space CIECAM02, and a criterion degree of metamerism of $n = 1$, the mean proportion R^2 of variance accounted for by the model was 0.90; with $n \geq 2$ it was 0.66. Lower values of R^2 were obtained with the color space CIELAB and the color-difference formula CIEDE2000.

Some loss in the strength of the regression with larger criterion degrees n was expected. As n increased, the tails of the distribution of colors under the second daylight would have become more and more influential. Consequently, their representation by the conditional differential entropy would have been less and less accurate, given that it measured the degree of uncertainty of all colors under the second daylight, rather than of only those with color differences exceeding some multiple of a nominal color-difference threshold.

What was unexpected, however, was that a restricted version of the model, based on the differential entropy of the colors of the scene under just the first daylight, would have such predictive power. For a criterion degree of $n = 1$, the mean proportion R^2 of variance accounted for by this restricted model was 0.88, almost identical to that for the full model. Yet, for $n \geq 2$, this restricted regression failed, with mean R^2 falling to 0.12.

Why should the differential entropy of colors under just the first daylight be such a powerful predictor when the criterion degree of metamerism is limited? Some possible contributory factors are considered in the following subsections.

A. Variation of Metamerism over Scenes

As Fig. 1 and Table 1 showed, the relative frequency of metamerism varied markedly from scene to scene. For a criterion degree of metamerism of $n = 1$, the range defined by the logarithm of the ratio of the highest to the lowest observed relative frequency over the 50 natural scenes was, on average, 1.73, a

ratio greater than 50. For $n \geq 2$, the range was even greater (Table 1, seventh column).

This variation in observed relative frequency N_1/N depends on the variation in its factors N_0/N and N_1/N_0 in the decomposition $N_1/N = (N_0/N)(N_1/N_0)$ (Subsection 2.E). Crucially, the ranges of N_0/N and N_1/N_0 were very different. For $n = 1$, the mean range of $\log N_0/N$ was 1.84, and the mean range of $\log N_1/N_0$ was 0.55 (Table 1, fifth and sixth columns). For $n \geq 2$, the ranges were more nearly equal or reversed in size.

What this difference in ranges implies is that for $n = 1$, most of the variation in the relative frequency of metamerism is determined by the initial probability of finding a pair of surfaces that are indistinguishable under the first daylight, rather than by the conditional probability of their becoming distinguishable under the second daylight. Therefore, in a regression of $\log N_1/N$ on differential entropy, it is the factor N_0/N that accounts for most of the variation in N_1/N, and it is this factor that is related to the differential entropy of colors under the first daylight.

B. Differential Entropy and Distinguishable Pairs

In Appendix A, two extreme model distributions are defined to demonstrate more formally the relationship between differential entropy and the relative frequency of indistinguishable pairs of points in a scene. The distributions are assumed to be of one color attribute, say hue. One distribution has pdf f_1, which is constant over the range of possible hue values (i.e., all hue values are equally likely in the scene), and the other has pdf f_2, which is concentrated at a particular hue value (i.e., there is only one hue value in the scene). Of such distributions, f_1 has the lowest frequency of pairs of points with color differences less than ΔE^{thr}, and f_2 has the highest. On the other hand, f_1 has the highest differential entropy, and f_2 has the lowest. The extrema of the relative frequencies of indistinguishable pairs and of differential entropies are thus inversely related to each other. The plot shown in Fig. 1, for a criterion degree of metamerism of $n = 1$, illustrates this inverse relationship, as the ordinate is almost identical to the logarithm of the observed relative frequency N_0/N. The gradient of the dependence of $\log N_0/N$ on the differential entropy of colors is –0.30, which, when entropy is expressed with respect to the same base as the logarithm of relative frequency, transforms to –1.00.

C. Uniqueness and Limitations of Differential Entropy

The distribution of colors in a scene under a given daylight can of course be summarized in more than one way. Are there summary measures, therefore, other than differential entropy, which have similar predictive power? One obvious candidate is the standard deviation or variance of the distribution.

To test its relevance, the variances of the two extreme model distributions f_1 and f_2 of Subsection 4.B are compared in Appendix B. Although the distribution f_1 has a high variance and f_2 has a low variance (in the limit, the lowest), it does not follow generally that the extrema of the relative frequencies of indistinguishable pairs and the extrema of the variances of the distributions are inversely related to each other. A model distribution f_3 is also defined in Appendix B whose variance violates this inverse dependence, and whose differential entropy does not.

Instead of the variance, other measures of the distribution of surface colors might be considered, for example, the estimated color gamut boundaries of the scene. Yet, as illustrated with the model distribution f_3, simply specifying the bounds on colors, such as their convex hull [29,30], is not sufficient. Any measure of gamut needs to take into account regions of color space both where colors are absent or sparse and where they are dense.

Nevertheless, measures such as variance and gamut boundary, although not optimal, are easy to compute, whereas differential entropy may seem less so. But with the availability of asymptotically unbiased estimators due to Kozachenko and Leonenko [24,25] and with the offset modification proposed by Marín-Franch and Foster [9,26], it is possible to make estimates of differential entropy both accurately and efficiently.

The predictive power of simple entropy models is necessarily imperfect. The present analysis has shown that for a limited criterion degree of metamerism, a combination of differential entropies under two successive daylights and differential entropy under just the first daylight were both able to account for almost all of the variation in the frequency of metamerism across 50 natural scenes. With more extreme criterion degrees of metamerism, however, only a combination of entropies was able to provide adequate predictive power. Even then, not all the variation in the frequency of metamerism could be accounted for. To achieve that goal, statistical models more comprehensive than those considered here are needed.

APPENDIX A: RELATIVE FREQUENCY AND DIFFERENTIAL ENTROPY

This example uses two extreme model distributions to illustrate the inverse relationship between the relative frequency of finding points close to each other and differential entropy.

First, consider the model pdf f_1, which is constant over a continuous range, $[-a, a]$ say, with $a > 0$, and zero elsewhere. The relative frequency of finding two points at random within a certain distance, Δa say, of each other, where $\Delta a \ll 2a$, is approximately $\Delta a / a$. In fact, this is the lowest relative frequency for all such distributions over the interval $[-a, a]$. Second, consider the model pdf f_2, which is constant over a region of size Δa centered at the origin, and zero elsewhere. The relative frequency of finding two points at random within Δa of each other is exactly 1. Therefore, f_1 has the lowest relative frequency, and f_2 has the highest relative frequency.

By contrast, the differential entropy of f_1 is log $2a$, which is the highest for all pdfs over the interval $[-a, a]$ (see, e.g., [13]), and the differential entropy of f_2 is log Δa, which, as Δa tends to zero, tends to minus infinity. Therefore, f_1, which has the lowest relative frequency, has the highest differential entropy, and f_2, which has the highest relative frequency, has the lowest differential entropy.

APPENDIX B: VARIANCE AND RELATIVE FREQUENCY

This example illustrates the fact that variance is not inversely related to relative frequency. For the model function f_1 of Appendix A, constant over the interval $[-a, a]$ and zero elsewhere, its variance is $a^2/3$. For the model function f_2 of Appendix A, constant over a region of size Δa centered at the origin and zero elsewhere, its variance is $\Delta a^2/12$, which tends

to zero as Δa tends to zero. Therefore the distribution f_1, which has the lowest relative frequency of finding two points at random within Δa of each other, has a high variance, and the distribution f_2, which has the highest relative frequency of finding two points at random within Δa of each other, has, in the limit, the lowest variance.

But it is possible to construct a model distribution, f_3 say, that has the same bounded support as f_1 and that violates this inverse dependence. Suppose f_3 is constant over two regions, each of size Δa centered at $-a$ and a, and zero elsewhere. Its variance is approximately a^2, which is larger than the variance of both f_1 and f_2, whereas the corresponding relative frequency of finding two points at random within Δa of each other is 0.5. This is between the values for f_1 and f_2, rather than being lower than the value for f_1.

For comparison, the differential entropy of f_3 is log $2\Delta a$, which, as required, falls between the values for f_1 and f_2, thereby preserving the inverse relationship.

REFERENCES

1. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas* (Wiley, 1967).
2. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (Wiley, 1982).
3. R. W. G. Hunt, *Measuring Colour*, 3rd ed. (Fountain Press, 1998).
4. CIE, "Colorimetry," 3rd ed., CIE Publication 15:2004 (CIE Central Bureau, 2004).
5. T. Troscianko, C. P. Benton, P. G. Lovell, D. J. Tolhurst, and Z. Pizlo, "Camouflage and visual perception," Phil. Trans. R. Soc. B **364**, 449–461 (2009).
6. D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," J. Opt. Soc. Am. A **23**, 2359–2372 (2006).
7. R. G. Kuehni, "Metamerism, exact and approximate," Color Res. Appl. **8**, 192 (1983).
8. T. Wachtler, T.-W. Lee, and T. J. Sejnowski, "Chromatic structure of natural scenes," J. Opt. Soc. Am. A **18**, 65–77 (2001).
9. D. H. Foster, I. Marín-Franch, K. Amano, and S. M. C. Nascimento, "Approaching ideal observer efficiency in using color to retrieve information from natural scenes," J. Opt. Soc. Am. A **26**, B14–B24 (2009).
10. J. J. Clark and S. Skaff, "A spectral theory of color perception," J. Opt. Soc. Am. A **26**, 2488–2502 (2009).
11. I. Marín-Franch and D. H. Foster, "Number of perceptually distinct surface colors in natural scenes," J. Vision **10**(9), 9 (2010).
12. CIE, "A colour appearance model for colour management systems: CIECAM02," CIE Publication 159:2004 (CIE Central Bureau, 2004).
13. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, 2006).
14. S. M. C. Nascimento, F. P. Ferreira, and D. H. Foster, "Statistics of spatial cone-excitation ratios in natural scenes," J. Opt. Soc. Am. A **19**, 1484–1490 (2002).
15. 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).

16. C.-J. Li, M. R. Luo, B. Rigg, and R. W. G. Hunt, "CMC 2000 chromatic adaptation transform: CMCCAT2000," Color Res. Appl. **27**, 49–58 (2002).

17. M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," Color Res. Appl. **31**, 320–330 (2006).

18. M. Melgosa, R. Huertas, and R. S. Berns, "Performance of recent advanced color-difference formulas using the standardized residual sum of squares index," J. Opt. Soc. Am. A **25**, 1828–1834 (2008).

19. 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).

20. G. Cui, M. R. Luo, B. Rigg, G. Roesler, and K. Witt, "Uniform colour spaces based on the DIN99 colour-difference formula," Color Res. Appl. **27**, 282–290 (2002).

21. P.-L. Sun and J. Morovic, "Inter-relating colour difference metrics," in *Proceedings of Tenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications* (Society for Imaging Science and Technology, 2002) pp. 55–60.

22. S. M. C. Nascimento, D. H. Foster, and K. Amano, "Psychophysical estimates of the number of spectral-reflectance basis functions needed to reproduce natural scenes," J. Opt. Soc. Am. A **22**, 1017–1022 (2005).

23. A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," Phys. Rev. E **69**, 066138 (2004).

24. L. F. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," Prob. Peredachi Inf. **23**, 9–16 (1987).

25. M. N. Goria, N. N. Leonenko, V. V. Mergel, and P. L. Novi Inverardi, "A new class of random vector entropy estimators and its applications in testing statistical hypotheses," J. Nonparametric Stat. **17**, 277–297 (2005).

26. I. Marín-Franch and D. H. Foster, "Estimating information from color signals: an application to digital cameras and natural scenes," IEEE Trans. Pattern Anal. Mach. Intell. (submitted).

27. D. B. Montgomery and D. G. Morrison, "A note on adjusting R^2," J. Finance **28**, 1009–1013 (1973).

28. B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, 1993).

29. J. Morovic and M. R. Luo, "The fundamentals of gamut mapping: a survey," J. Imaging Sci. Technol. **45**, 283–290 (2001).

30. A. M. Bakke, I. Farup, and J. Y. Hardeberg, "Evaluation of algorithms for the determination of color gamut boundaries," J. Imaging Sci. Technol. **54**, 050502 (2010).