Joint Workshop on Material Appearance Modeling and MANER Conference (2024)
H. Rushmeier and J. Y. Hardeberg (Editors)

A study of observer metamerism for reflectance-induced stimuli

L. Fascione[1] and J. Hanika[2]

[1]NVIDIA
[2]Karlsruhe Institute of Technology

Abstract
Cameras make images collecting per-pixel measurements of light reflected by the objects in the world. Commonly, these measurements undergo a transformation so that they become values in a standardized color space, such as the sRGB space. This makes it possible to send the values to a display device and produce a human a visual sensation as close as possible to what would have been caused by the original scene. In this work we aim to explore the difficulties and opportunities that arise in devising such non-bijective transformations, visualizing differences between device vision and human vision. In particular we are interested in the practical impact of observer metamerism: different camera devices and human observers can distinguish a different set of spectral stimuli presented to them. When characterizing a camera, this is usually ignored, missing potential to increase chromatic acuity where the camera can see more than the human observer. A question that arises is whether the metameric stimuli involved here do actually appear in practice in relevant cases. We run numeric experiments to investigate these questions.

1. Introduction

Cameras and humans perceive, or measure, color differently, as a consequence of the different *spectral response functions* (SRF) that characterize them. The light *stimulus* ϕ activates the n SRFs $A = \{\bar{r}_0(\lambda), \ldots, \bar{r}_{n-1}(\lambda)\}$ performing an integral over the wavelength domain $\Lambda = [\lambda_{\min}, \lambda_{\max}]$

$$a_i = \int_{\lambda \in \Lambda} \bar{r}_i(\lambda)\phi(\lambda)\, \mathrm{d}\,\lambda = \langle \bar{r}_i, \phi \rangle. \tag{1}$$

We call this the *action of observer A on the stimulus* ϕ, and we say that the result of the action is a *measurement vector* $(a_0, \ldots, a_{n-1})$.

The action of A is linear and in general, given two different observers A and B with the same number n of SRFs, there will be no linear mapping from the a_i resulting from A to the b_i from B. However, it is desirable to transform the data recorded from a digital camera into appropriate signals for a display device, so that the sensation observing the original scene in person is as close as possible to observing the reproduction on said display device. This is most accurately performed by *look-up tables* (LUTs), which for a given set of stimuli ϕ_i provide a mapping between measurements from the observers $A(\phi_i) \mapsto B(\phi_i)$. This gives rise to a few questions that we set out to address at least partially:

- While using such a LUT on camera RGB ensures correct gamut boundaries, it fails to resolve metamerism, i.e. one camera RGB coordinate will be due to one from a number of different spectra which might have induced different human measurements (and vice-versa). We know in theory this difference can be substantial, but how much does this matter in practice?
- For stimuli observed in photography, is there a significant ''chromatic fingerprint'' due to the SRFs of a specific camera model as compared to another?
- What is a relevant sample of stimuli (reflectance and illumination spectra, potentially including indirect and fluorescence) which would allow us to evaluate these questions in a practical setting?

2. Background

Let us recap our notation for our audience. We will not dive into post-vision phenomena or the limitations of display devices, these are important aspects of the discussion but we have not delved there for the present body of work.

Stimuli Our *stimulus functions* $\phi(\lambda) \geq 0$ (or *stimuli* for short), measure light arriving at an observer, as a function of wavelength over the so-called *visible range*. For this work

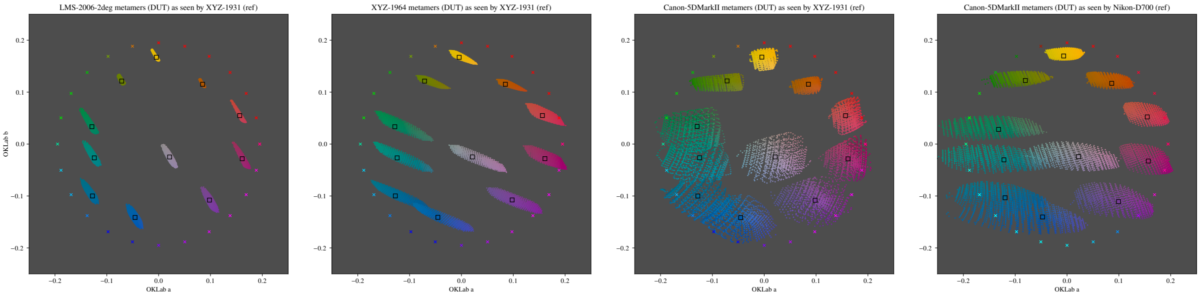

Figure 1: *Color mismatch volumes for 2deg-SCO. From left: CIE LMS 2006 data, CIE 1964* $10°$ *Supplementary Observer and Canon 5D MkII camera. Fourth plot shows Canon metamers seen by Nikon D700. Details in section 3.1*

we used $\Lambda = [400\text{nm}, 700\text{nm}]$. Light comes into scenes from *illuminants* and turns into stimuli by multiplication with *reflectance distributions* $\rho(\lambda)$. For the present work all stimuli discussed are products of this kind.

Spectral response functions Our observers are ordered collections of SRFs: $A = \{\bar{r}_0(\lambda), \ldots, \bar{r}_{n-1}(\lambda)\}$, as used in eq. (1) to convert stimuli ϕ to measurement vectors, these are called *tristimulus coordinates* when $n = 3$. The over bar indicates that the functions have been scaled so that their collective maximum is 1. This aligns with the names $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ used by the CIE to define the $2°$, 1931, Standard Colorimetric Observer (2deg-SCO in short) [WS82].

Ensembles A pair of different stimuli $\phi_1, \phi_2 \in S$ can map to the same values under the action of A, in other words $A(\phi_1) = A(\phi_2)$. In this case we will say that ϕ_1 is an *A-metamer* of ϕ_2, and that ϕ_1 and ϕ_2 form an *A-metameric pair*. Through this we obtain the *metameric ensemble over* ϕ: $M_A(\phi) = \{\psi \in S | A(\phi) = A(\psi)\}$ with respect to A. Sometimes a pair of stimuli ϕ_1 and ϕ_2 will instead map to the values that are ''close'' under the action of A, in other words $A(\phi_1) \simeq A(\phi_2)$. Such pairs are called *A-parameric*, and their ensemble is $P_A(\phi) = \{\psi \in S | A(\phi) \simeq A(\psi)\}$.

3. Experiments

We compare two observers A and B: A will be the *device under test* (DUT), while B will be our *reference* (ref). Importantly in all our experiments we vary reflectance functions $\rho(\lambda)$ under well-defined illuminants, chosen per-experiment. This constrains the space of stimuli under analysis in important ways, and is a different approach from other work in computational vision.

All the computation in this work was executed in *Python*. This paper focuses on color matters, so we have obtained a PDF file with colors encoded in *sRGB* color space passing the `natural` option to the `xcolor` LaTeX package.

The OKLab color space To quantify and visualize color differences and the location of color coordinates, we use the OKLab color space [Ott20].

$$\begin{bmatrix} L_{\text{ok}} \\ a_{\text{ok}} \\ b_{\text{ok}} \end{bmatrix} = M_2 \begin{bmatrix} \sqrt[3]{l_{\text{ok}}} \\ \sqrt[3]{m_{\text{ok}}} \\ \sqrt[3]{s_{\text{ok}}} \end{bmatrix} \qquad \begin{bmatrix} l_{\text{ok}} \\ m_{\text{ok}} \\ s_{\text{ok}} \end{bmatrix} = M_1 \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2}$$

The defining matrices M_1 and M_2 were numerically optimized to place the colors obtained from the Munsell Book of Colors (MBoC) under Standard Illuminant C (StdIllC) as seen by 2deg-SCO on shapes as close to concentric circles as the degrees of freedom in the transform will allow.

We introduce the *camera OKLab transform*, which maps camera RGB coordinates into OKLab coordinates directly. We optimize the matrices M_1^{cam} and M_2^{cam} which define the OKLab space so that they minimize the mapping error between the stimuli generated by MBoC under StdIllC with respect to the prediction of their OKLab coordinates obtained through the 2deg-SCO observer. Because the match is necessarily imperfect, we will talk in the rest of the paper of a *camera OKLab space* to signify the result of mapping camera RGB data through its matching camera OKLab transform.

3.1. Color mismatch volumes

We have analyzed the difference in chromatic acuity between pairs of observers. The metameric ensembles for DUT observer A, as seen through the action of ref observer B are called *color mismatch volumes* (CMV) in [Sch76].

We plotted *maximal* CMVs in camera OKLab space for ref observer B as follows: we selected several patches from MBoC, lit them with StdIllC, obtaining stimuli $\{\phi_1, \ldots, \phi_k\}$. For each ϕ_i we found the range of values Y_B (y being ref observer B's second coordinate) among all reflectances in $M_A(\phi_i)$. We picked several values $y_B \in Y_B$ and for each solved again to find an X_B range (being the first coordinate as produced by ref observer B), constrained to be in $M_A(\phi)$ as well as having the chosen y_B. We repeated this a third

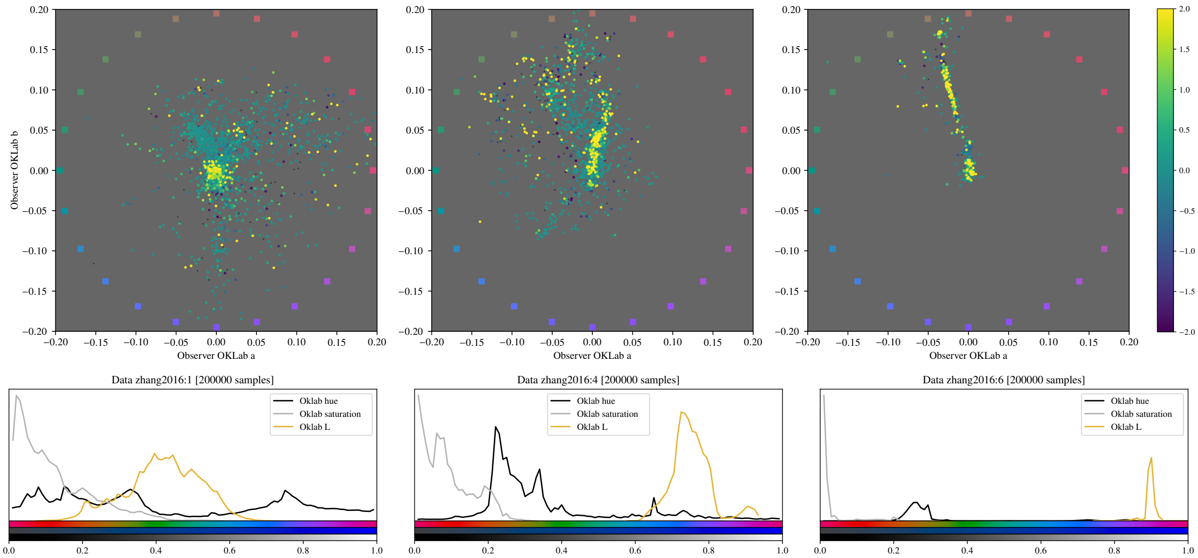

Figure 2: ***Top row:*** *change in volumes for empirical paramerc ensembles for a Canon 5D Mk II with respect to the 2deg-SCO. We plotted* $\log_2(\mathrm{vol}(P_{\mathrm{DUT}}(\phi))/\mathrm{vol}(P_{\mathrm{ref}}(\phi)))$ *encoded as the false color scale on the right, volumes from the PCA-aligned OBB of each ensemble, the location of the dots is the center of the OBB, see section 3.1.1.* ***Bottom row:*** *The OKLab Hue, Saturation and Luminance distributions of the data in the plots in the top row, plotted as a PDF. The gradients under the plot illustrate the hue scale, saturation scale and luminance scale of OKLab*

time picking $x_B \in X_B$, to find a Z_B range among spectra in $M_A(\phi_i)$ also meeting both our y_B and x_B targets. We then plotted the resulting point clouds and obtained fig. 1. A direct approach to describing CMVs was presented in [LFG14].

We observed that the volumes are sizable. While we expected larger sizes for neutral colors, we were somewhat surprised to see how the blue-dominant region seemed rather large even at higher saturation levels. This might correspond to the larger responsivity gap between the short and medium wavelength cone responses with respect to the conceptually equivalent blue versus green response, except that we expected the region to be advantageous for the DUT, as consequence of the even count of red versus green versus blue photosites (often in a 1:2:1 proportion) as opposed to much less even distribution in the retina (l:m:s are roughly 8:5:1 proportion in the foveal region). As the OKLab volume of these projected metameric ensembles exhibited a strong dependency on the color coordinate, we came to the question of what should be important stimuli to map with low error.

3.1.1. Empirical parameric ensembles

For this purpose, starting from a database of roughly 42 million spectral reflectances [ZFM16], we ran the following experiment: All reflectance spectra in our database were turned into stimuli by multiplication with a reference illuminant. For a given stimulus we have selected all the spectra A-parameric to it, based on a $\Delta E_{\mathrm{OKLab}}$ threshold of $0.01 \simeq 0.5$JND.

We then plotted $\log_2(\mathrm{vol}(P_A(\phi))/\mathrm{vol}(P_B(\phi)))$ in OKLab space, using the color of the plotted mark to represent the change in volume from A to B obtaining fig. 2. This is effectively an empirical (as opposed to maximal) form of fig. 1 built from spectra actually available in a given dataset. The intuition for this visualization is that where a parameric ensemble gives a larger volume, there the corresponding observer has a better ability to separate the stimuli in it, the tempering from available data being a way to explore where this difference would actually make a difference in practice. The logarithmic plotted scale gives us positive values indicating that A separates the set better than B, and vice-versa for negative values.

Repeating the same plotting for different parts of the dataset, as visible in fig. 2, reveals an influence from the starting database on the results obtained. One part of the explanation might well be that more saturated reflection spectra are scarcer than neutral colored ones, or medium brightness ones would be more common than very dark or light ones. So we asked ourselves how to visualize the distribution of large databases of spectral reflectances.

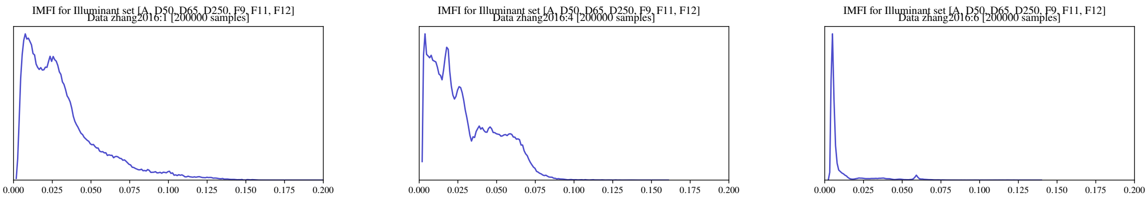

Figure 3: *The distribution of IMFI values for the data used to generate fig. 2 Each spectral reflectance sample was lit by CIE illuminants A, D50, D65, D250, F9, F11, F12. The resulting 2deg-SCO coordinates were adapted back to D65 using CIE CAT16 method. The IMFI is the RMS average of the* ΔE_{OKLab} *resulting from this process*

3.2. Analysis of large bodies of spectral reflectance data

Our basis for this analysis was again the data provided with [ZFM16]. It is distributed in six files, so we ran some analyses on these separately to explore structural differences.

The bottom row of fig. 2 shows statistics on the OKLab HSV coordinates of the datase, plotted as probability distribution functions. We find that this visualization makes it clear how the selection of spectra has a big impact on the frequency of certain colour coordinates. The top row of fig. 2 shows large variations in parameric volume change even for very similar tristimulus coordinates, manifesting as dark blue dots near yellow dots, for example.

We suspect that the shape of the input reflectance spectra might have an impact here as well. So we came up with a metric for what's sometimes called *color inconstancy*, which we called *Illuminant induced Metamerism Failure Index* (IMFI). IMFI is similar to the notion of a Color Inconstancy Index such as CMCCON02 [LRS03]. In particular, the IMFI is an estimate of how much a given reflectance $\rho(\lambda)$ makes a specific adaptation method fail, averaged over several illuminants. In this case we have lit $\rho(\lambda)$ with Standard Illuminants A, D50, D250, F9, F11, F12 and then used CIECAT16 to adapt the resulting 2deg-SCO tristimulus back to D65. The IMFI is the RMS average of the ΔE_{OKLab} between the adapted predictions and the D65 ground truth. We plotted the distribution of these values in fig. 3 for 3 datasets from fig. 2.

4. Discussion, Limitations and Future Work

When comparing how two different observers A and B act on the space of metamers, there are two competing perspectives. From one point of view the question arises as to how well observer A is able to simulate the vision behavior of B. Typically when B is the 2deg-SCO this question is expressed as how far the device under test A is from meeting the Luther-Ives condition. The concern with a failure to do so lies in the fact that the reference observer B would be able to separate stimuli that the device under test A might not. This seems like would limit or impair the possibility of accurate color or tone reproduction from data recorded through A.

The opposite point of view instead embraces the complementary condition: when observers A and B have substantially different SRFs, A becomes able to separate stimuli that B cannot, opening opportunities in terms of going beyond what the reference observer can achieve unaided. It seems that for the common case where B is a model of human vision, such as 2deg-SCO, this could come in service of artistic purposes as well as scientific ones. At the same time, constructing an accurate LUT to characterize A will collapse the metameric ensembles into a single point as perceived by B which may be undesirable.

Another important question is whether human perception resolves colors well enough to be able to distinguish the ''chromatic fingerprint'' of certain camera devices from others. To quantify this it would be interesting to render a set of relevant hyperspectral images through several DUT observers and then compare the images rendered through different input device transforms. From our analyses we expect that there will be significant visible differences, depending a lot on the spectral shape of input stimuli. It remains to answer what are *relevant* stimuli. Will this be dominated by natural objects, which have shaped the sensitivity of the eye over the evolutionary time-scale, or by the things that people like to photograph today?

References

[LFG14] Logvinenko A. D., Funt B., Godau C.: Metamer mismatching. *IEEE Transactions on Image Processing 23*, 1 (2014), 34–43. doi:10.1109/TIP.2013.2283148. 3

[LRS03] Luo M., Rigg B., Smith K.: CMC 2002 colour inconstancy index CMCCON02. *Coloration Technology 119* (06 2003), 280 – 285. doi:10.1111/j.1478-4408.2003.tb00184.x. 4

[Ott20] Ottosson B.: A perceptual color space for image processing, Dec. 2020. https://bottosson.github.io/posts/oklab/. 2

[Sch76] Schmitt F. J. M.: A method for the treatment of metamerism in colorimetry. *J. Opt. Soc. Am. 66*, 6 (Jun 1976), 601–608. doi:10.1364/JOSA.66.000601. 2

[WS82] Wyszecki G., Stiles W. S.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*, second edition ed. Wiley Classics Library. John Wiley & Sons, Aug. 1982. 2

[ZFM16] Zhang X., Funt B., Mirzaei H.: Metamer mismatching in practice versus theory. *J. Opt. Soc. Am. A 33*, 3 (Mar 2016), A238–A247. doi:10.1364/JOSAA.33.00A238. 3, 4