https://doi.org/10.2352/EI.2024.36.9.IQSP-275

Objective color characterization of HDR videos captured by smartphones: laboratory setups and analysis framework

Anna Tigranyan, Paul Mathieu, Corentin Nannini, François-Xavier Thomas, Mauro Patti, Frédéric Guichard
DXOMARK, Boulogne-Billancourt, France

Abstract

This article provides elements to answer the question: how to judge general stylistic color rendering choices made by imaging devices capable of recording HDR formats in an objective manner? The goal of our work is to build a framework to analyze color rendering behaviors in targeted regions of any scene, supporting both HDR and SDR content. To this end, we discuss modeling of camera behavior and visualization methods based on the **IC_TC_P/ITP** *color spaces, alongside with example of lab as well as real scenes showcasing common issues and ambiguities in HDR rendering.*

Introduction

Recent technology improvements in both consumer and industrial cameras have enabled on-device capture and mastering of High Dynamic Range (HDR) content without user intervention. In our previous work [1] we highlighted the difficulty in handling the large differences of display luminance in HDR video content produced by such pipelines.

Evaluating color rendering in older SDR pipelines involved placing color charts in a few global conditions (e.g. a few illuminant types and illuminances), for which a limited set of comparisons and graphs in **CIELAB** a^*b^* planes at a narrow, constant lightness range were appropriate. In modern HDR pipelines, the user experience is linked to the ability of the camera to tweak their rendering of colors in any situation, including in localized areas of the image (e.g. [2]). The larger dimensionality of the data collected make it orders of magnitude more difficult to analyze, summarize and compare device behavior.

Some full-reference metrics compatible with HDR introduce quality correlates (e.g. [3]) that could be used as an alternative, but using images from different viewpoints makes their use difficult. Because controlled lab conditions are also much more reproducible, we will focus on color charts in this work. [4] provides a good overview on the literature on color spaces and colorimetry adapted to that purpose, and we note that concepts explored in our article could be adapted in any color space and color difference metric they describe. In all these cases, the difficulty of comparing a large number of measurements remains.

The goal of the following is to propose a framework for analyzing color rendering behaviors from colorimetry measurements on color charts placed in targeted regions of displayed images; the proposed framework should provide meaningful ways of comparing color rendering between a large number of regions, scenes, devices or formats, SDR or HDR, and in particular when they have different display luminance levels.

General methodology

In order to deal with varying display luminance levels, we use the null hypothesis H_0 throughout this paper that "free linear scaling (e.g. in **CIEXYZ**) results in the same perception of an image when adaptation reaches steady-state".

While that hypothesis is sometimes demonstrably wrong – in particular when considering HDR imagery – it is approximately valid in a lot of conditions, supported by [5] "between 20 – 2000 cd/m^2 we find approximate contrast constancy in most cases". This is particularly true for SDR content, where it is used implicitly for rendering images (such as in the default formulation of EOTFs such as ITU-R BT.1886[6]) or analyzing them (using the diffuse white luminance of **CIELAB**) ; both of which operate relative to the peak luminance of the SDR display, which is commonly found between 100 – 400 cd/m^2. Therefore, in this paper:

- We discuss several **visualizations of color rendering** through the use of *vectorscope* tools, designed to analyze linear scaling (following H_0) independently from color changes (*not* following H_0);
- We construct a simplified **model of color rendering** to relate the analysis of color to a function of a limited number of free parameters, and visualize its residuals in terms of linear scaling and color changes.

We then discuss their usefulness and validity as comparison and analysis tools, as well as the general applicability of H_0, with both SDR and HDR-encoded content.

Absolute colorimetry in HDR images

ΔE_{ITP} color difference metric and ITPJND space

We base the rest of our paper on the ITU-R BT.2124 [7] *color difference metric* called ΔE_{ITP}. This metric is implemented as the Euclidian norm embedded in a JND-scaled, perceptually approximately uniform opponent color space that we will call **ITPJND** hereafter, with values noted $(I^*T^*P^*)$. This space is based on **IC_TC_P** [8, 9] and absolute **CIEXYZ** colorimetry measurements. We also note that representations of colors in these spaces are equivalent through the following relations:

$$(I^*T^*P^*) = 720 \cdot (1, 0.5, 1) \cdot (IC_TC_P) = \mathbf{T}((XYZ))$$

ITPJND and **IC_TC_P** work similarly enough to **CIELAB**, and useful metrics can be implemented on top of them as shown figure 1 for them to feel familiar. However, they eschews decisions to include adaptation parameters (specifically *diffuse white*

Notation	Description
$h_{ITP} = \arctan(P^*/T^*)$	Hue angle [°]
$C_{ITP} = \sqrt{T^{*2} + P^{*2}}$	Absolute chroma [JND steps]
$\Delta C_{ITP} = C_{ITP1} - C_{ITP2}$	Chroma difference [JND steps]
$\Delta E_{ITP} = \|\|(I^*T^*P^*)\|\|_2$	Total color difference [JND]

Figure 1. *Related **ITPJND** color metrics and differences*

point scaling including the "wrong" Von Kries chromatic adaptation[10]) which are ill-defined in general for evaluating HDR content. These spaces have other differences, such as the fact that hue representations in the two spaces are oriented differently (as shown in figure 2).

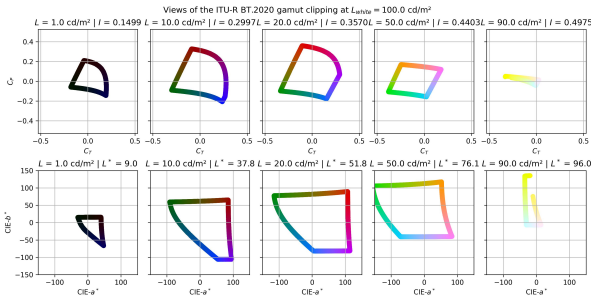


Figure 2. *Gamut boundary of an example ideal BT.2020 display peaking at 100 nits on each RGB channel, represented in IC_TC_P (top) and **CIELAB** $L^*a^*b^*$ (bottom) chroma planes for slices of different luminances (left to right).*

Color accuracy thresholds and ΔE_{ITP}

The **ITPJND** scale defines "potentially visible" color differences in the most sensitive adaptation state, unlike **CIELAB** which uses a fixed known adaptation state. Figure 3 shows a comparison to ΔE_{76} (the usual **CIELAB** color difference metric) and ΔE_{2000} (CIEDE2000 color difference metric) for a common SDR luminance, estimating between 2.5 (achromatic, neutral colors) to 5 (saturated colors for ΔE_{76}) or 10 (saturated colors for ΔE_{2000}) **ITPJND** steps per ΔE unit.

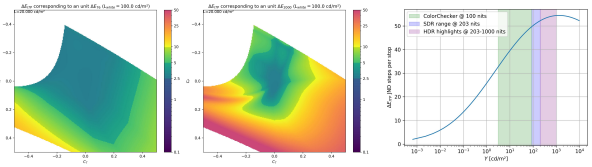


Figure 3. *Images of the constant luminance C_TC_P plane for a reflective object in SDR conditions (20% reflectance), with a colormap indicating a maximum ΔE_{ITP} on a sphere of radius $\Delta E_{76} = 1$ (left) and $\Delta E_{2000} = 1$ (middle) ; graph of the number of JND steps obtained when increasing the luminance by an EV, with common display luminance ranges overlaid (right).*

Common color accuracy thresholds in the literature can be used to get a sense of the acceptability thresholds for ΔE_{ITP}. For example, [11] links to two separate sets of standard color accuracy thresholds for technical photography in museums:

- Metamorfoze requires $\Delta E_{76} \leq 2.83$ for achromatic patches, corresponding to a $\Delta E_{ITP} \leq 7$;
- Metamorfoze also requires a mean $\Delta E_{76} \leq 4$ (resp. max $\Delta E_{76} \leq 10$) for all colors, corresponding to ΔE_{ITP} values between 10 and 50 ;
- FADGI requires average $\Delta E_{2000} \leq 2.0$, corresponding to ΔE_{ITP} between 5 and 20 in average.

ΔE_{ITP}	Description
≤ 5	Very good, strict color match
≤ 10	Good match for achromatic patches
≤ 20	Good match for color patches
$= 50$	Luminance increments/decrements of ≈ 1 to 1.5 EVs.

Figure 4. *Indicative thresholds used throughout this paper.*

Outside of accuracy thresholds, a decrement or increment of light (e.g. because of exposure changes on the camera) is often measured in EVs. How many **ITPJND** steps fit in that unit? The last graph in figure 3 gives us some elements towards that answer, setting an EV as approximately 30 to 50 **ITPJND** steps in the most common luminance ranges.

In the rest of this article, we shall assume the indicative thresholds in figure 4, while noting that they can vary depending on the scene and application being considered and should be the object of a further study.

Color rendering visualizations

Reference curves

We can define for each patch p of a ColorChecker:

- *Measured* color $(XYZ)_p$ or $(I^*T^*P^*)_p$, sampled as absolute display radiances from the image being evaluated ;
- *Reference* color $(XYZ)_{p,\text{ref}}$ or $(I^*T^*P^*)_{p,\text{ref}}$, result of an absolute colorimetric measurement with a colorimeter or a spectroradiometer.

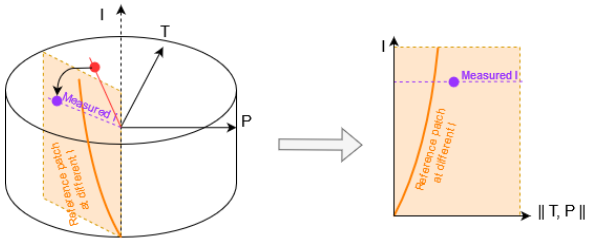


Figure 5. *Explanation showing a reference curve (orange) on a single-hue vertical plane in the **ITPJND** space, on which a single measured point (red) is projected (violet) at the same luminance.*

We define a *reference curve* by generating colors $\lambda \cdot (XYZ)_{\text{ref}}$ for any λ in an interval $]0; \lambda_{\text{max}}]$ where $\lambda_{max} \cdot Y_{\text{ref}} = 10\,000\ \text{cd}\,\text{m}^{-2}$ (peak luminance of PQ transfer function used in the **ITPJND** space). **CIEXYZ** values are then converted to form the reference curve $(I^*T^*P^*) = \rho(I^*)$ (in orange in fig. 5).

Scaled reference and associated ITP metrics

The traditional usage of chroma-only differences like ΔC^* or Δa^*b^* fails in presence of large luminance differences – for example, Metamorfoze uses ΔC^*_{76} but also places hard constraints on the ΔL^*_{76}). Used incorrectly on a brighter (resp. darker) color

than the reference, the total color difference will be dominated by the luminance difference, and they will underpredict (resp. overpredict) the amount of perceived chroma.

Following H_0, we propose instead to split the total color difference around the reference after performing luminance scaling. We define:

- The *scaled reference* $(XYZ)_{\text{s.ref}} = \lambda \cdot (XYZ)_{\text{ref}}$ such that $Y_{\text{s.ref}} = Y$ (the luminance of the scaled reference is equal to that of the measured color, with the chroma components linearly scaled accordingly). In figure 5, this corresponds to the intersection of the dotted measured line with the reference curve.
- $\widehat{NE}$, the number of ΔE_{ITP} JND steps from the *reference* to the *scaled reference* alongside this reference curve. This corresponds to "brightness" scaling according to our null hypothesis. This value is only dependant on the intensity, and corresponds to the curvilinear coordinate along the reference curve, expressed in JND steps:

$$\widehat{NE}(I_0, I_1) = \int_{I_0}^{I_1} \Delta E_{ITP}\left[\rho(I); \rho(I + dI)\right]$$

- $\widehat{\Delta E}$ and $\widehat{\Delta C}$ are regular color difference and chroma difference metrics from the *scaled reference* to the *measured* colors. They correspond to the chroma (colourfulness and hue) residuals that do not follow simple linear scaling.

$$\begin{aligned}\widehat{\Delta E}\left[(XYZ); (XYZ)_{\text{ref}}\right] &= \Delta E_{ITP}\left[(XYZ); (XYZ)_{\text{s.ref}}\right] \\ \widehat{\Delta C}\left[(XYZ); (XYZ)_{\text{ref}}\right] &= \Delta C_{ITP}\left[(XYZ); (XYZ)_{\text{s.ref}}\right]\end{aligned} \quad (1)$$

Vectorscopes

We shall discuss in this section a series of visualizations based on vectorscopes [12] a polar graph widely used in video applications to observe color information.

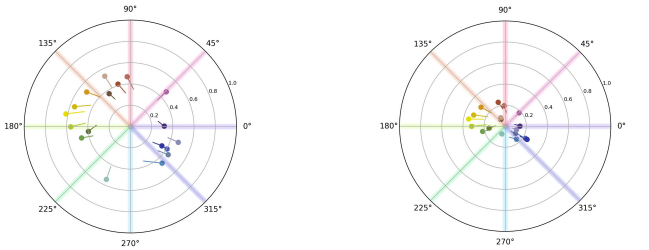


Figure 6. *Left: luma vectorscope, with radius I ; darker colors are situated close to the center of the vectorscope, while the brighter patches are further away. Right: chroma vectorscope, with radius $\sqrt{C_T^2 + C_P^2}$; the patches that are saturated (contain more color) are situated towards the edge of the graph, whereas the patches that lack color are close to the center, and grays in the middle.*

Vectorscopes whose radius is based on the encoded luma or chroma (Figure 6) are very common options. Neither, however, provides any complete visualization of color while taking linear scaling into account: while hue shifts are visible, it is impossible at first glance to tell whether the differences in the chroma vectorscope are over- or under-saturated with respect to the global luma differences.

Figure 7 exposes two proposed improvements ; the second (right) in particular shows both luma and chroma differences on the same graph and therefore provides a *consistent visualization* when color rendering deviates non-uniformly from H_0 (compare the dark green patch on the left to the other patches) ; it also provides a *sense of the perceptual scale* as all the units are in JND steps. On this graph, the more the red line is circular and residuals are small, the more color rendering follows H_0.

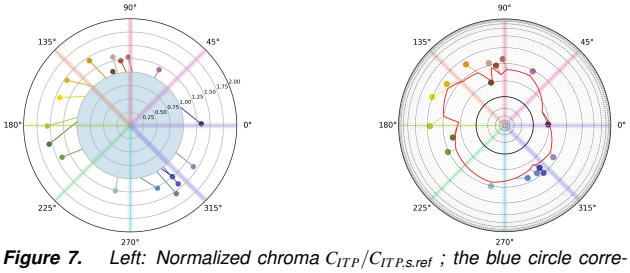


Figure 7. *Left: Normalized chroma $C_{ITP}/C_{ITP,s.ref}$; the blue circle corresponds to a ratio of 1, with anything inside being undersaturated and anything outside oversaturated. Right: Vectorscope showing simultaneously $\widehat{NE}$ (red line) to indicate linear scaling relative to the reference (darker inside the black circle, brighter outside), and $\widehat{\Delta C}$ (arrows) to indicate chroma changes relative to the red line (undersaturated inside the red line, oversaturated outside), with guidelines every 10 (light gray dashed lines) and 50 JND steps (dark gray dashed lines). The vectorscope radius is scaled with a weighting function $1/(1+e^{-(x-a)/b})$ so that it is possible to show large and small signed JND differences on the same graph.*

Global color rendering model

In this section, we describe our proposed global model of the color rendering as a parametric mathematical function $f_{\mathbf{x}} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ with parameters $\mathbf{x} = \lambda, n, K, \mathbf{b}, T, s$, that maps reference colors (source) into measured colors (target).

Given a set of patches with indices $p \in [0..N[$ (such as a ColorChecker, with $N = 24$), $f_{\mathbf{x}}$ is defined as a set of sequential transformations in figure 8, applied independantly for each patch p with initial reference colors $(XYZ) = (XYZ)_{\text{ref},p}$.

For a given chart in an image, we find $\mathbf{x}$ that minimizes the ΔE_{ITP} metric, solved numerically using a least squares solver such as `scipy.optimize.least_squares` from the SciPy library [13]. The optimization problem is formulated mathematically as:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_p \left\| \Delta E_{ITP}\left[(XYZ)_p, f_{\mathbf{x}}(XYZ)_{p,\text{ref}}\right] \right\|^2 \quad (2)$$

Naka-Rushton parametric contrast function

n, K are respectively the slope and highlight roll-off inflexion point controls of a *parametric contrast function* (Naka-Rushton from [14]), linked to properties of the camera such as its Opto-Optical Transfer Function (OOTF), as well as properties of the physical acquisition.

This function provides a way to describe common contrast curves with only two parameters. It typically presents S-curves for $K \in [0, 1]$ and $n > 1$, and good approximations of $\gamma = n$ power functions for sufficiently large $K > 1$, as shown in figure 9). Im-

	Description	Transform $(XYZ) \rightarrow \ldots$
n K	Contrast slope $(0-\infty)$ Rolloff knee $(0-\infty)$	$(1+K^n)\cdot\frac{(Y/Y_{\max})^n}{(Y/Y_{\max})^n+K^n}\cdot(XYZ)$
λ $\mathbf{b}$	Linear scaling factor XYZ glare offsets (nit)	$(\lambda Y_{\max}-\mathbf{b})\cdot\frac{(XYZ)}{Y_{\max}}+\mathbf{b}$
T	CCT (K)	$M_T\cdot(XYZ)$
s	Saturation factor $(0-\infty)$	$\mathbf{T}^{-1}\left[\mathbf{T}(XYZ)\cdot(1,s,s)\right]$

Figure 8. *Color model parameters and equations. To a given set of N reference values $(XYZ)_{ref,p}$ $(p \in [0..N[)$ corresponds a maximum luminance $Y_{max} = \max_p Y_{ref,p}$ (in the ColorChecker, the white patch).*

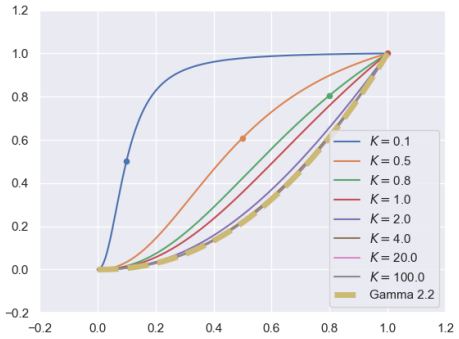


Figure 9. *Naka-Rushton curve with $n = 2.2$ and increasing values of K, compared to a $\gamma = 2.2$ power function.*

portantly, the function is continuous with relation to n, K, which makes it ideal for numerical optimization routines.

Linear scaling and glare offsets

The contrast function introduced above does not scale the maximum luminance, and neither does it offset the black point.

Linear scaling is achieved with λ, the *global exposure factor* discussed earlier, to account for differences between the actual display luminance and the reference.

Affine offsets are added to each **CIEXYZ** channel using $\mathbf{b} = (XYZ)_{\text{offset}}$, to account for imperfect black levels and veiling glare.

The model is setup so that at this point, the maximum luminance becomes $\lambda Y_{\max}$, regardless of the offset used.

Bradford chromatic adaptation

T is the *target Correlated Color Temperature* (CCT) of a Bradford chromatic adaptation matrix M_T ([15]) that transforms the reference colors (under the display white point assumed to be D65) into the measured colors (under the target illuminant with a given correlated color temperature T), to account for the device white balance decisions.

Saturation control

Since **ITPJND** is approximately hue-linear and perceptually uniform, a simple multiplicative factor on both chroma channels T^*, P^* sounds like a simple saturation control, increasing or decreasing the perceived colourfulness for the same brightness.

Experimental results and discussion

In this section, we will put the previously defined color rendering model and visualization tools to use in specific conditions.

Consistency on existing SDR imagery

In this section, we discuss the behavior of the color rendering model in the following 3 datasets containing SDR images:

- **ColorCheckerLab**, an internal dataset of ColorChecker images in controlled lab conditions. Images come from 278 mobile camera devices from the last 5 years shot in 24 different lab conditions ranging from 1000 lux to 1 lux, with illuminant CCTs 6500K, 4000K, 3000K, 2850K and 2300K. Both the scene and the encoding are SDR-only.
- **ColorCheckerPub**, a public dataset [16] of ColorChecker images in various real life conditions. The encoding is SDR-only and only contains 7 devices, but the dataset contains a large number of real SDR and HDR scenes.
- **PortraitHdrLab**, an internal dataset of Portrait HDR [1] images in controlled lab conditions. Images come from 80 mobile camera devices from the last 5 years shot in 6 different lab conditions: 20 lux at 2850K, 100 lux at 4000K, 1000 lux at 6500K, each shot with the panel set to 4 and 7 EV brighter than the subject. The encoding is SDR-only, but the scene is HDR.

Figure 10 shows that the model is close but visibly different, even in SDR-encoded images. The achromatic patches are very well-supported, but color patches (in particular primary and complementary colors) have larger errors. The real dataset with few devices on a large number of real scenes appears to perform better than our dataset with many devices in a few controlled conditions, warranting further study.

Statistics for $\widehat{\Delta C}$ are shown in figure 11, showing that luminance scaling can account for a lot of the residuals: the median is within the set thresholds for a much larger portion of the patches, and outliers have much lower error. From this point, it is clear that any future color rendering modeling that includes a grayscale contrast curve must also include the possibility for it to be different outside of achromatic regions.

Both figures also presents an example of one of the worst-performing ColorCheckerLab images, and we note that most of the measured colors seem to be brighter than the estimates on Figure 10. Figure 11 on the other hand highlights the usage of the vectorscope for providing a more detailed analysis: the red line is often above the black reference (”brighter”, i.e. linearly scaled), but the chroma residuals are much smaller with a few large exceptions such as ”purple” patch 9 (not more saturated).

The model still showcases common recommendations of color rendering (figure 12): partial chromatic adaptation (the rendered CCT of a 2300K illuminant is closer to 5000K for most devices), and target exposure (scene brightness is similar to display brightness, resp. relative to illuminant and peak display luminance, with slightly darker target exposures for darker illuminants). It shows little bias in the saturation factor (the s parameter is in average close to 1): assertions that mobile renderings *oversaturate* images should most likely be reframed in terms of color contrast rather than a pure increase in chroma. We also noted previously the grayscale patches were particularly well-modeled by

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
ColorCheckerLab	Min	0.1	0.3	0.4	0.6	0.7	0.3	1.1	1.0	0.4	1.3	0.2	0.9	1.7	0.7	2.6	1.1	0.6	0.5	0.0	0.2	0.1	0.1	0.2	0.2
	Q1	14.5	7.4	5.9	11.6	8.0	5.8	14.2	17.0	13.6	14.8	10.7	16.4	25.4	11.6	22.0	16.7	13.4	18.6	3.7	2.9	2.9	3.4	3.5	2.6
	Med	20.2	10.1	8.4	18.5	11.9	8.3	20.1	22.8	21.3	23.3	14.7	22.4	33.3	16.0	31.3	20.7	19.6	24.6	5.9	4.0	4.1	4.9	5.6	4.7
	Q3	26.5	13.7	11.7	24.5	17.2	11.1	26.0	30.5	30.7	36.2	19.2	28.7	43.0	20.9	43.1	26.0	27.9	30.5	9.0	5.4	5.7	7.1	8.8	8.5
	Max	64.6	65.3	65.3	57.6	51.4	60.0	77.8	65.6	86.4	129.6	52.2	56.5	112.2	66.8	107.7	93.4	93.7	66.5	44.3	21.6	86.9	41.9	47.1	66.2
ColorCheckerPub	Min	0.3	0.3	0.4	0.4	0.2	0.5	0.4	0.5	0.3	0.2	0.1	0.4	0.2	0.2	0.1	0.5	0.3	0.1	0.1	0.1	0.0	0.1	0.1	0.1
	Q1	8.7	7.1	6.8	8.8	6.7	9.6	14.3	14.1	8.3	7.0	6.4	13.9	17.0	9.4	9.9	10.2	10.9	16.9	3.8	2.7	2.3	2.7	2.9	3.3
	Med	14.6	11.1	10.6	13.5	10.8	13.0	20.3	18.8	14.5	11.5	10.5	19.6	23.6	12.7	16.9	15.1	15.7	23.0	7.4	5.2	4.2	5.3	5.8	6.9
	Q3	23.3	17.7	14.8	19.7	15.8	17.6	28.1	24.6	23.8	20.3	17.0	27.4	32.2	17.9	27.5	23.3	23.4	28.7	12.8	9.3	8.3	11.0	12.3	13.8
	Max	118.2	90.3	123.9	79.7	65.6	94.3	108.8	107.7	122.1	112.5	71.5	90.3	137.6	98.7	116.1	71.9	116.2	94.6	65.5	55.7	55.6	66.0	80.7	76.9
PortraitHdrLab	Min	1.7	0.9	0.8	1.0	1.1	1.2	1.5	0.9	1.0	0.8	1.1	0.6	3.2	1.4	2.0	1.4	2.1	1.1	0.4	0.5	0.4	0.4	0.7	0.7
	Q1	9.4	5.3	5.7	5.8	9.0	7.2	14.3	12.6	11.2	10.9	8.4	11.3	20.1	9.5	20.0	9.4	11.1	9.2	2.9	2.7	2.2	2.4	3.6	5.2
	Med	14.5	8.1	8.6	9.0	11.9	10.1	20.4	18.1	15.8	16.7	11.8	15.8	30.1	14.0	30.1	14.8	15.5	12.7	4.2	3.9	3.2	3.6	5.3	8.2
	Q3	19.8	11.3	12.3	12.8	14.8	13.4	26.2	24.4	22.1	25.1	16.3	21.2	40.6	18.2	41.0	20.7	21.0	18.0	6.5	5.1	4.4	5.1	7.8	11.3
	Max	50.0	48.4	47.0	70.1	43.4	58.3	53.9	68.3	49.6	54.9	60.8	47.0	86.8	49.7	115.4	49.5	44.7	45.3	19.4	13.0	12.9	28.2	17.2	28.6

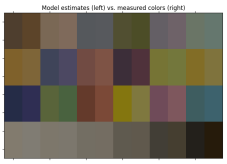


Figure 10. *ΔE_{ITP} total color difference residuals between modeled colors and measured colors on SDR-encoded datasets. The colors corresponds to differences of 5 (green), 10 for achromatic / 20 for chromatic (black) and above 50 JND steps (red). On the left is an example of the 99th percentile of the median $\widehat{\Delta C}$ in the ColorCheckerLab dataset, showing the limits of the model.*

the Naka-Rushton curve: we see usual values of n from 1 to 3, and note that usually these curves most often converge to S-curves and not gamma curves or linear functions (the knee-point K is in the majority of cases below 1).

Behavior on HDR-encoded formats

A setup that is suitable for assessing HDR formats must also provide data for assessing color in conditions that are particularly challenging. In particular, this means assessing both dark regions (which are potentially subject to veiling glare and/or quantization, noise) and bright regions, bright regions (which are potentially subject to saturation issues), including with recognizable objects (e.g. a diverse selection of human faces, which are potentially subject to object-specific color rendering adjustments) and other items such as motion in the scene (preventing the device from optimizing a static scene).

We will use the installation described in figure 13, using a controllable backlit portrait ranging from 0 to 7 EV, to present example of results in figure 14.

Conclusion

The first conclusion here is that the model provides a good overview of the manufacturer color rendering of most camera devices, showing that a surprisingly little number of parameters can account for a lot of the rendering behaviors, particularly when close to neutral colors ; linear scaling also appears a good proxy for a lot of the residual differences. Using the color model in combination with residual vectorscopes and grayscale curves is an interesting option for providing a quick decomposition of color rendering into understandable components.

The model fails in some scene conditions: high chroma illuminants, non-uniform veiling glare are cases that are still out of scope for such a simple illuminant. There is also an ambiguity between some of the parameters: the offset chromaticity and chromatic adaptation parameter both generate ”color casts”, and the saturation factor does not seem to model chroma behavior that seem better modeled by varying contrast curves.

An obvious next step would be to apply this model to a reference image (for example a RAW or a DSLR with little processing) to generate image samples for use with full-reference metrics, which can then use to predict visibility and quality. Would that correlate with the user experience of camera devices?

A second next step is that there is not, at the moment, an annotated dataset of HDR images with color checkers with enough devices and formats: the relative recency of their introduction means that there is much less variety, in particular for low-end and middle-range devices.

Bibliography

[1] Cyril Lajarge et al. “Objective image quality evaluation of HDR videos captured by smartphones”. In: *Electronic Imaging* 34.9 (Jan. 2022), pp. 312-1-312–6. ISSN: 2470-1173. DOI: 10.2352/ei.2022.34.9.iqsp-312.

[2] Longwen Zhang et al. *Neural Video Portrait Relighting in Real-time via Consistency Modeling*. 2021. DOI: 10.48550/ARXIV.2104.00484.

[3] Rafal K. Mantiuk, Dounia Hammou, and Param Hanji. “HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content”. In: (Apr. 2023). DOI: 10.48550/ARXIV.2304.13625. arXiv: 2304.13625 [eess.IV].

[4] Anustup Choudhury et al. “Image quality evaluation for high dynamic range and wide color gamut applications using visual spatial processing of color differences”. In: *Color Research & Application* 46.1 (Nov. 2020), pp. 46–64. ISSN: 1520-6378. DOI: 10.1002/col.22588.

[5] Maliha Ashraf et al. “Suprathreshold Contrast Matching between Different Luminance Levels”. In: *Color and Imaging Conference* 30.1 (Nov. 2022), pp. 219–224. ISSN: 2166-9635. DOI: 10.2352/cic.2022.30.1.38.

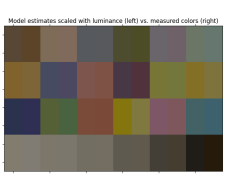
[6] International Telecommunication Union. *Recommendation ITU-R BT.1886 - Reference electro-optical transfer function for flat panel displays used in HDTV studio production*. 2011.

[7] International Telecommunication Union. *Recommendation ITU-R BT.2124-0 - Objective metric for the assessment of the potential visibility of colour differences in television*. 2019.

[8] Dolby Laboratories. *ICtCp White Paper*. URL: https://professional.dolby.com/siteassets/pdfs/ictcp_dolbywhitepaper_v071.pdf.

[9] International Telecommunication Union. *Recommendation ITU-R BT.2100-2 - Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange*. 2018.

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
 ColorCheckerLab	Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Q1	7.1	2.7	1.0	6.6	1.7	1.5	3.7	2.4	2.9	6.3	3.1	5.6	3.5	4.3	3.4	7.0	3.9	10.4	1.6	0.8	0.7	1.0	1.2	1.1
	Med	12.3	5.7	2.2	12.9	3.5	3.5	7.8	5.3	6.9	12.1	7.3	10.1	7.6	9.3	7.7	12.2	8.6	16.7	3.5	1.6	1.4	2.2	2.7	2.7
	Q3	17.7	9.4	3.7	19.6	6.2	6.1	12.8	9.3	12.9	21.0	12.7	15.5	12.9	16.5	14.4	18.2	14.2	22.9	6.3	2.7	2.4	4.4	5.3	6.2
	Max	61.8	42.5	64.8	45.5	46.1	23.7	40.3	39.1	79.2	113.9	51.5	56.3	49.9	56.5	80.3	84.0	64.6	48.4	34.4	17.3	80.8	37.0	32.6	58.2
 ColorCheckerPub	Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Q1	2.3	2.2	1.4	4.2	1.5	2.9	3.2	2.1	2.1	2.1	2.0	3.6	2.5	2.2	3.0	3.0	3.0	8.3	2.2	1.2	0.9	1.2	1.2	1.8
	Med	4.6	4.8	2.9	7.8	3.4	5.9	6.8	4.8	5.6	5.3	4.5	7.2	5.7	5.0	7.5	6.9	6.7	13.4	5.5	3.0	2.0	3.2	3.3	5.1
	Q3	9.8	8.9	5.6	13.5	6.9	9.2	12.1	9.2	12.4	13.0	8.6	12.4	10.3	9.2	17.2	13.5	11.3	17.8	11.0	6.6	4.9	8.9	9.8	12.0
	Max	55.0	50.2	52.7	57.2	50.4	32.4	78.7	51.3	50.5	66.2	47.3	57.1	64.3	44.5	89.9	49.9	50.8	46.5	63.2	52.7	56.1	66.7	78.8	75.6
 PortraitHdrLab	Min	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Q1	4.3	2.1	1.6	2.3	5.5	1.9	2.3	2.7	2.3	3.3	2.2	2.6	6.0	2.3	9.1	3.6	2.0	3.1	1.2	0.7	0.6	0.6	0.8	1.5
	Med	9.8	4.8	2.9	4.6	8.4	4.3	4.7	6.0	5.1	8.7	5.7	6.6	12.3	5.2	21.7	8.5	4.3	7.3	2.5	1.7	1.3	1.2	2.1	3.4
	Q3	16.2	7.6	5.2	8.3	11.2	8.0	8.8	10.2	9.5	16.7	11.0	11.5	24.1	9.2	34.1	15.5	8.6	12.3	4.6	2.7	2.1	2.5	3.6	6.5
	Max	43.3	21.3	23.8	37.2	23.6	21.7	44.1	38.3	31.3	48.4	55.1	42.7	86.1	40.0	98.2	49.4	43.3	36.9	15.6	12.5	9.0	14.5	13.1	25.7



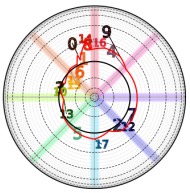


Figure 11. *$\widehat{\Delta C}$ chroma residuals between modeled colors and measured colors on SDR-encoded datasets. The colors corresponds to differences of 5 (green), 10 for achromatic / 20 for chromatic (black) and above 50 JND steps (red). On the right is the same 99th percentils of the median $\widehat{\Delta C}$ as fig. 10 showing comparison between scaled model estimates and measured colors, representing the comparatively much smaller remaining $\widehat{\Delta C}$. A corresponding vectorscope showing both $\widehat{NE}$ and $\widehat{\Delta C}$ between model estimates and measured colors is also shown.*

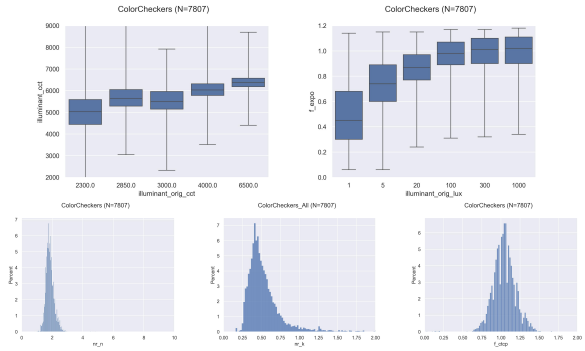


Figure 12. *Distribution of model parameters in the ColorCheckerLab dataset, showing (top left) real illuminant CCT against modeled CCT (T), (top right) real illuminance vs exposure factor (λ), (bottom left) contrast slope (n), (bottom middle) knee point (K), and (bottom right) saturation factor (s).*

Figure 13. *An illustration of setups with (a) a ColorChecker, allowing to track color rendering using the methods described in this paper and (b) a human face, allowing to track alterations of color rendering on select ROIs compared to the ColorChecker, (c) controllable backlit elements, with either diffusers or additional charts and (d) motors under either faces or the camera for switching between setups.*

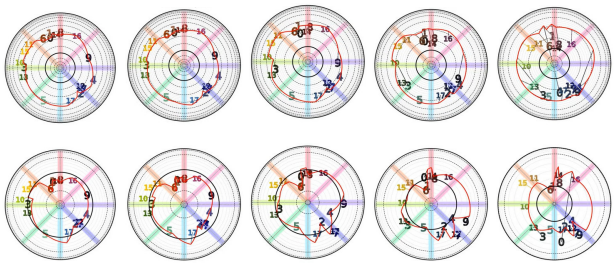


Figure 14. *Example vectorscopes for 5 conditions under the same illuminant type but increasing backlit glare, in the first row of fig. 13. The top row show the difference between reference and measured colors, and the second row shows residuals between modeled and measured colors.*

[10] Kevin A. G. Smet et al. “Study of chromatic adaptation using memory color matches, Part II: colored illuminants”. In: *Optics Express* 25.7 (Mar. 2017), p. 8350. ISSN: 1094-4087. DOI: 10.1364/oe.25.008350.

[11] Eric Kirchner et al. “Exploring the limits of color accuracy in technical photography”. In: *Heritage Science* 9.1 (May 2021). ISSN: 2050-7445. DOI: 10.1186/s40494-021-00536-x.

[12] Dave Rodriguez. “Introduction to Audiovisual Transcoding, Editing, and Color Analysis with FFmpeg”. In: *Programming Historian* 7 (Dec. 2018). Ed. by Brandon Walsh. ISSN: 2397-2068. DOI: 10.46430/phen0077.

[13] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[14] K. I. Naka and W. A. H. Rushton. “S-potentials from colour units in the retina of fish (Cyprinidae)”. In: *The Journal of Physiology* 185.3 (Aug. 1966), pp. 536–555. ISSN: 1469-7793. DOI: 10.1113/jphysiol.1966.sp008001.

[15] K. M. Lam and University of Bradford. “Metamerism and colour constancy”. eng. PhD thesis. University of Bradford, 1985.

[16] Mahmoud Afifi et al. “When Color Constancy Goes Wrong: Correcting Improperly White-Balanced Images”. In: *CVPR*. 2019.