Analysis of a new HDR dataset of laboratory scenes images using the ICtCp color space

Elodie Souksava, François-Xavier Thomas, Hoang-Phi Nguyen, Laurent Chanas, Frédéric Guichard
DXOMARK, Boulogne-Billancourt, France

Abstract

This paper is the continuation of a previous work in [1], which aimed to develop a color rendering model using ICtCp color space, to evaluate SDR and HDR-encoded content. However, the model was only tested on an SDR image dataset. The focus of this paper is to provide an analysis of a new HDR dataset of laboratory scenes images using our model and additional color rendering visualization tools. The new HDR dataset, captured with different devices and formats in controlled laboratory setups, allows the estimation of HDR performances, encompassing several key aspects including color accuracy, contrast, and displayed brightness level, in a variety of lighting scenarios. The study provides valuable insights into the color reproduction capabilities of modern imaging devices, highlighting the advantages of HDR imaging compared to SDR and the impact of different HDR formats on visual quality.

Introduction

We noted previously that we didn't have a lot of public datasets of HDR formats with color charts. Here, we publicly provide image samples for developing measurements that work on both SDR and HDR images seamlessly.

In this accompanying paper, we aim to provide insights into how cameras with automatic capture pipelines render the proposed scenes in their captured images. Additionally, we want to explore what kind of rendering (using the photography pipeline and not the video pipeline) smartphone cameras (which heavily feature, for example, local tone- and gamut-mapping, multi-frame denoising and super-resolution algorithms) generate, including potential artifacts.

Because of these two reasons, unlike many datasets such as those mentioned in [2] [3], we're going to focus on lab setups where all elements of the scene are known in terms of luminance and chromaticity up to a manufacturing tolerance, in order to provide a quantitative view. Many images in these datasets are limited to full HD or sometimes 4K (3840×2160) resolution and rarely more, while our images are photographs with a minimum resolution of 12 megapixels (3024×4032).

Because we're interested in lab setups, full-reference approaches like [4] are not suitable, although they might be useful for later correlating image metrics measured on these setups. In the following, we discuss simple observations using both display luminances (in nits) and the IC_tC_p color space (and its associated ΔE_{ITP} color difference metric), as well as the color rendering model detailed in [1], in order to better understand the distribution of the renderings across devices and illumination conditions.

HDR formats

At the time of writing, two standards exist for still-image HDR formats in consumer electronics:

- ISO 22028-5 [5], so-called "dash five", is concerned with a simple encoding of HDR images. They provide pixel data directly to HDR displays using either the PQ or HLG transfer functions but do not support SDR displays without altering the image (through e.g. tone-mapping, clipping) in an unspecified and often device-specific manner.
- ISO 21496-1 [6] is concerned with standardization of so-called "gain map" images, which augment existing SDR file formats with some amount of upward compatibility for HDR displays in a well-specified manner. Previous formats used similar encoding schemes such as [7], which were often specific to each manufacturer.

While many smartphone manufacturers are switching to the second standard for storing HDR still images, support for reading and displaying them is still not widespread in all operating systems, and many devices still record their older, manufacturer-specific formats; additionally, "dash five" images are simpler to read: they contain pre-rendered display luminances and do not require an additional display adaptation process.

Among formats compatible with the "dash five" standard, AVIF [8] (AV1 [9] codec within a ISOBMFF [10] container) has the broadest support, notably across operating systems (at least the latest versions of the Chrome browser and the standard photo viewer on both OSX and Windows support it out of the box). The minimal metadata that is required is the CICP [11] transfer function and primaries. *Content Light Level* (CLL) metadata also provide information to display those images correctly according to the luminance range of their contents, where they would otherwise often assume a peak luminance of 10 000 cd/m^2 for the PQ transfer function in order to avoid recomputing statistics and perform unwanted tonemapping. All the images were therefore converted from their original formats (either Apple [12] or UltraHDR [7]) following specifications in figure 1.

Database contents

The database contains files captured by 11 smartphones, that originally captured the HDR content between 2020 and 2024. For each device, 3 setups were shot, with various illuminants, described below.

The dataset also contains DeviceDSelfie and DeviceFSelfie images, which correspond to the selfie camera of DeviceD and DeviceF, for comparison, with DeviceFSelfie having one additional image of a ColorChecker at 0 lux taken with the inbuilt flash.

Metadata	Description
Colorimetry (SDR)	Display P3 (i.e. sRGB [13] transfer function with P3-D65 colorimetry) [14])
Colorimetry (HDR)	ITU-R BT.2100 PQ [15]
Encoding (SDR)	Y'CbCr 4:4:4 8-bits encoded with AV1
Encoding (HDR)	Y'CbCr 4:4:4 10-bits encoded with AV1
MaxCLL, MaxFALL	Calculated according to CTA-861-G [16]

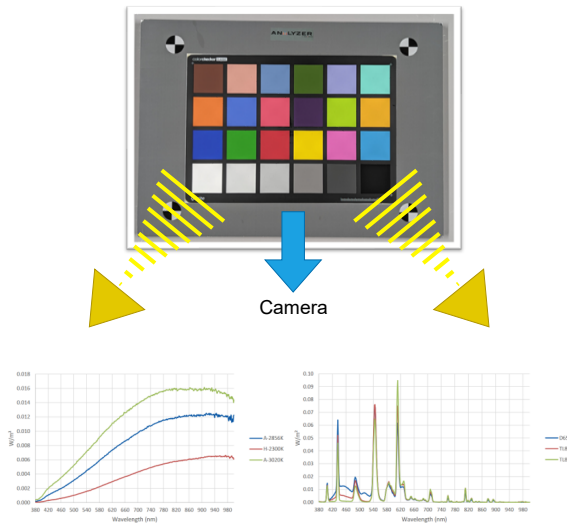**Figure 1.** *File format specifications for AVIF images in the database.*

Figure 2. *ColorChecker setup*

The database can be downloaded at https://dl.dxomark.com/analyzer/EI2025.

ColorChecker setup

The **ColorChecker** setup in figure 2 consists in a chart placed in the center of the image in front of a gray background, reflecting light (blue arrow) from 2 projectors (in yellow) creating a uniform (< 5% spatial variation) region of the specified illuminant on the chart. The spectra of the tungsten illuminants (A and H at respectively 2850K and 2300K) are close to that of a black body (bottom left), and the other illuminants are fluorescent tubes (bottom right) ; a summary of all illuminants in the database for this setup is provided in figure 3.

This setup notably allows measurements of luminance transfer function, noise, color rendering, white balance and chromatic adaptation properties.

Autofocus HDR setup

The **Autofocus HDR** setup in figure 4 consists in a Dead Leaves [17] chart placed in the middle-low region of the image reflecting light (middle blue arrow) from 2 projectors (in yellow) creating a uniform (< 5% spatial variation) region of the specified illuminant on the chart. The chart surrounded on the top and on the right by two transparent "Composite" slides [18] with sim-

Illuminant	6500K	4000K	3000K	2850K	2300K
1000 lux	✓	✓	✓		
300 lux	✓	✓	✓		
100 lux	✓	✓	✓	✓	
20 lux		✓	✓	✓	
5 lux			✓	✓	
1 lux					✓

Figure 3. *Illuminants included in the ColorChecker images*
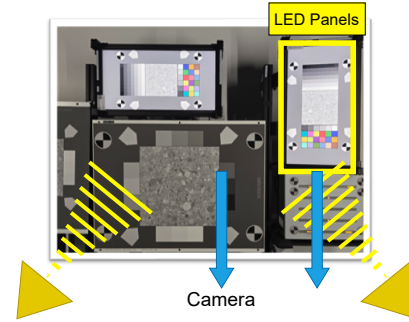

Figure 4. *Autofocus HDR setup*

ilar patches as the ColorChecker, each backlit by a LED panel (Kino Flo Celeb 250) and transmitting light towards the camera (right blue arrow). All objects are placed in front of a white background ; the spectra illuminating the reflective charts are the same as the ColorChecker setup in figure 2, and the different spectra emitted by the LED panels have similar correlated color temperature (CCT) as the reflected illuminant (< 5% tolerance). The luminance of the LED panels (measured without the transparent slide with a luminance meter) is compared to the luminance of a 100% reflective diffuse object in the center of the image (computed from measurements using an illuminance-meter facing the light sources), and reported as a nominal ΔEV value (in stops, with 0.3 stop tolerance). A summary of the illuminants in the database for this setup is provided in figure 5.

This setup allows measurements of color rendering in transmissive parts of the image, exposure time and other timing-related metrics, as well as luminance transfer functions, noise, white balance and chromatic adaptation properties, sharpness (MTF) and texture preservation in both reflective and transmissive parts of the image.

Portrait HDR setup

The **Portrait HDR** setup in figure 6 consists in a mannequin placed on the left-side of the image, reflecting light (middle blue arrow) from 2 projectors (in yellow) creating a uniform (< 5%

	Autofocus HDR			Portrait HDR	
↓ Illuminant / ΔEV →	7	4	2	7	4
1000 lux 6500K	✓	✓	✓	✓	✓
100 lux 4000K	✓	✓	✓	✓	✓
20 lux 2850K	✓	✓		✓	✓

Figure 5. *Illuminants included in the Autofocus HDR and PortraitHDR images*
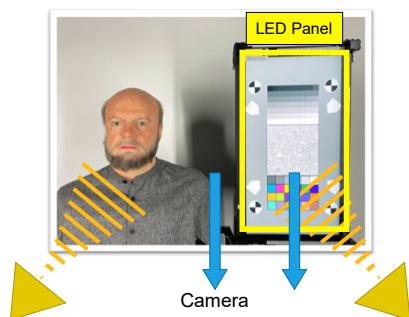
Figure 6. *Portrait HDR setup*

Figure 7. *Several ways of representing patch luminances. (a, top left) is a linear plot relating reference luminance to displayed luminance. (b, top right) is the same plot in a logarithmic scale. (c, bottom left) shows per-patch behavior as a JND delta, and (d, bottom right) shows per-patch behavior as a luminance delta (in stops).*

spatial variation) region of the specified illuminant ; and in a transparent slide on the right-side of the image with similar patches as the ColorChecker, backlit by a LED panel (Kino Flo Celeb 250) and transmitting light towards the camera (right blue arrow). The spectra illuminating the reflective charts are the same as the ColorChecker setup in figure 2, and the different spectra emitted by the LED panel (while not identical) have similar correlated color temperature (CCT) as the reflected illuminant ($<$ 5% tolerance). The luminance of the LED panels (measured without the transparent slide with a luminance meter) is compared to the luminance of a 100% reflective diffuse object in the center of the image (computed from measurements using an illuminance-meter facing the light sources), and reported as a nominal ΔEV value (in stops, with 1/3 stop tolerance). A summary of the illuminants in the database for this setup is provided in figure 5.

This setup allows measurements of properties of the rendering of the face in terms of brightness, color or details, but also of luminance transfer functions, noise, color rendering, white balance and chromatic adaptation, sharpness (MTF) and texture preservation in the transmissive parts of the image.

Observations and discussions

In the following we're going to use several data points to highlight rendering decisions made by devices in the dataset.

We reuse the "exposure factor compared to the reference" output (denoted f_{expo}) of our previous model [1], optimized to minimize the ΔE_{ITP} with an SDR reference. A f_{expo} value of 1.0 corresponds to the same as the reference, i.e. using a luminance of 100 nits to represent ideal diffuse white objects.

Each box plot shows the distribution of values, with the edges of the whiskers representing the minimum and maximum ; the box and its median line represents each quartile (25%, 75% and median) of the distribution.

Global brightness and usage of headroom

Display brightness is chosen by the camera device as the combination of various heuristics (such as auto-exposure processes on the sensor, or brightness alterations in the final rendering). These try to render an impression of the scene brightness on the final photograph. Depending on the case, you (as a photographer) or the camera (with the default rendering chosen by the camera manufactruer) might want a night photograph to appear darker overall than one shot under bright outdoors skies at
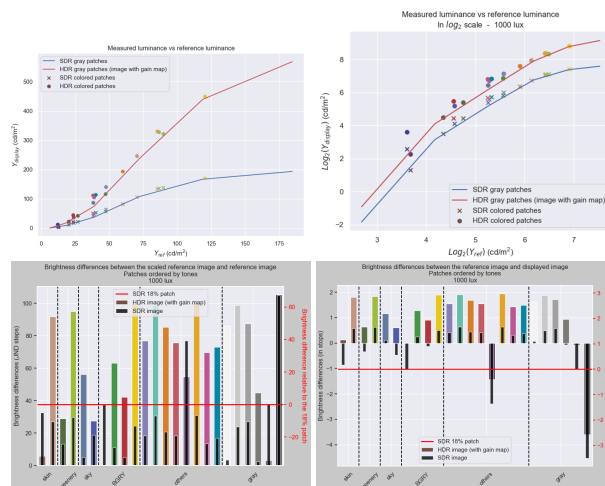
noon in order to convey the mood of the scene, even if the camera could very well place both photographs at the same overall display brightness.

The entire imaging system can be considered from the point of view of a global, glass-to-glass transfer function from reference rendering to displayed rendering. A simple measure of displayed luminance is shown on figure 7 (a) for one image, where the gain map globally increase the overall brightness of the image and not only highlights. This is even more obvious in the logarithmic version in (b).

Measures of luminance on color patches, as opposed to gray patches, also suggest a similar observation, albeit with a lot of variance between patches. A bar graph shows that in greater detail in terms of either ΔE_{ITP} JND in (c) or in stops in (d), where the differences with the reference are more apparent. One wonders how much the contrast adjustments of the gray patches influence a corresponding adjustment on colors, and further study of contrast adjustments and their effects on the perceived style of the image in HDR formats might be worthwhile.

When looking at these results through the lens of the color rendering model [1], we can see that, in SDR formats, the headroom is very limited. For a scene with limited dynamic such as the ColorChecker (figure 8, top row), a 10-stop difference in scene illuminant becomes a median display brightness difference of less than a stop in the final rendered image: the overall brightness of the rendered image never varies a lot. For objects in the highlights like the transparent panels in the Autofocus HDR and Portrait HDR scenes, there is even less variation as their renderings are limited by the maximum luminance of an SDR display. More notably, the interquartile difference for each illuminant is often less than 1/3 stop.

In HDR formats, the variation is much greater. For the ColorChecker (figure 8, bottom row), the median display brightness is generally brighter than an SDR rendering at 203 nits. It still varies by less than a stop for the same 10-stop scene luminance
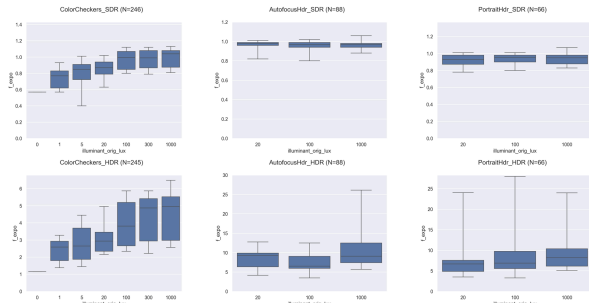
Figure 8. *Brightness decisions as the f_{expo} output of the color rendering model [1]. The top row (a), (b), (c) shows respectively the ColorChecker, Autofocus HDR and Portrait HDR setups for SDR formats displayed with an SDR reference white luminance of 100 nits ; the bottom row (a'), (b'), (c') shows the same setups for HDR formats displayed with an SDR reference Wwhite luminance of 203 nits.*

difference, but the per-illuminant interquartile difference is now more than 1 stop, highlighting a large variance in the rendering decisions for the same scene depending on the device! A similar trend is observed for the renderings of Autofocus HDR and Portrait HDR, including some very bright outliers at more than 2 stops compared to the median.

Relative brightness of highlights

We can explore these differences by looking at the display luminance of similarly reflecting or transmitting patches on both the dark and bright parts of the scene.

On either a synthetic scene like Autofocus HDR (figure 9) or a recognizable subject like in Portrait HDR (figure 10), we can see tone compression behavior in all the cases: HDR formats renders the 7EV case at a median between 3 and 4.1 EV, and the compression is even stronger for SDR formats. Tone compression on the face seems slightly stronger, but that would warrant further study.

On the Portrait HDR setup, the median face luminance stays at low levels for all these images, around 40 to 50 nits in SDR formats and around 60-70 nits for HDR formats ; all of which are higher than the luminances of the 40% reflective gray patch in Autofocus HDR. We also note that the luminances are higher in the HDR formats compared to SDR formats, even for reflective elements.

Color casts and chromatic adaptation

The "color cast" of an illuminant on a picture is similarly chosen by the camera as the combination of various heuristics (the auto-white balance processes on the sensor, color calibration matrices (CCM) or look-up tables (LUT), any alteration in the final rendering). Despite the common assertion that the perceived colors might be largely invariant with respect to the tested illuminants (and thus an photographed object would be rendered exactly the same under a 2850K and 6500K illuminant, for example), we are usually able to cognize both at the same time and say that the 2850K illuminant looks "warm" or yellow while recognizing the color of the skin on a face as such. Many photographers still choose to keep some amount of "color cast" to retain the mood of the scene.

Interestingly, when looking at the (T^*, P^*) chroma planes, the actual color cast in photographs, measured on gray patches in
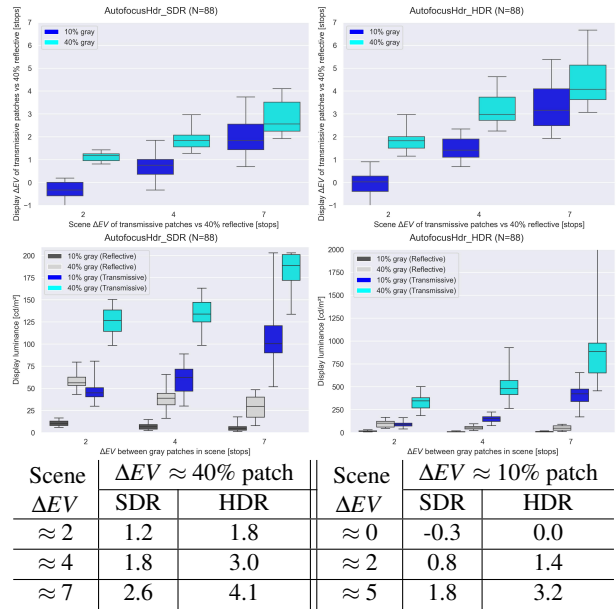

Scene ΔEV	$\Delta EV \approx 40\%$ patch		Scene ΔEV	$\Delta EV \approx 10\%$ patch	
	SDR	HDR		SDR	HDR
≈ 2	1.2	1.8	≈ 0	-0.3	0.0
≈ 4	1.8	3.0	≈ 2	0.8	1.4
≈ 7	2.6	4.1	≈ 5	1.8	3.2

Figure 9. *Distribution of differences in display luminances on the gray patches of the reflective chart and on patches with similar transmittance on the ColorChecker region of the Composite panel. (a, top left) and (b, top right) represent the ratios relative to a 40% reflective patch; (c, bottom left) and (d, bottom right) represent the luminances themselves, for respectively the SDR formats and HDR formats. The table (e) summarizes median ratios relative to the 40% reflective patch, in stops.*
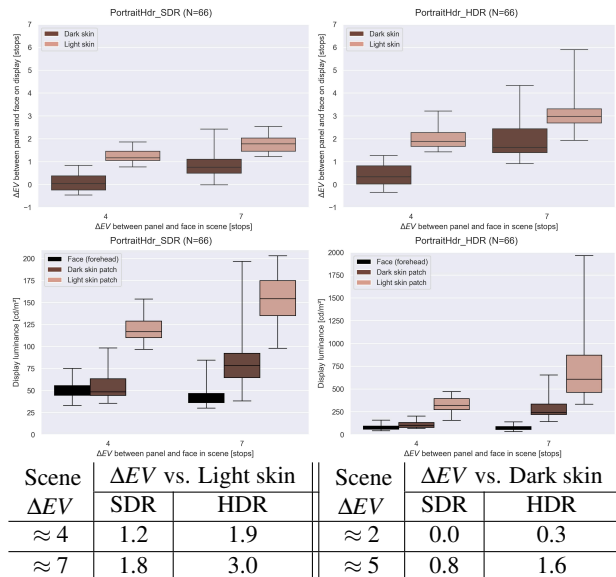

Scene ΔEV	ΔEV vs. Light skin		Scene ΔEV	ΔEV vs. Dark skin	
	SDR	HDR		SDR	HDR
≈ 4	1.2	1.9	≈ 2	0.0	0.3
≈ 7	1.8	3.0	≈ 5	0.8	1.6

Figure 10. *Distribution of differences in display luminances on the forehead and on skin patches of the ColorChecker region of the Composite panel. (a, top left) and (b, top right) represent the ratios relative to the forehead luminance ; (c, bottom left) and (d, bottom right) represent the luminances themselves, for respectively the SDR formats and HDR formats. The table (e) summarizes median ratios on the display, in stops; in the scene, the light skin patch has approximately the same average spectral transmittance as the face reflectance (around 40%), while the dark skin patch is approximately 2 stops darker than the light skin patch (around 10%).*
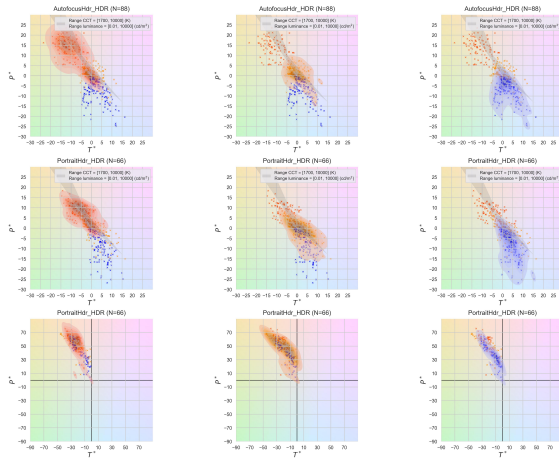
Figure 11. *Visualization of chroma values in IC_tC_p scaled for ΔE_{ITP} JND, i.e. the $(T^*, P^*) = (360 \times C_t, 720 \times C_p)$ plane, for HDR formats. Columns from left to right represent the CCT of the scene illuminant at respectively (1) 2850K, (2) 400K, and (3) 6500K, and rows from top to bottom correspond to respectively (a) all gray patches of the ColorChecker ROI in Composite panel in the Autofocus HDR setup, (b) the same patches in the Portrait HDR setup, (c) the forehead of the mannequin in the same Portrait HDR setup. In (a) and (b) rows, the gray area represents exact chroma values of the black body illuminants at various display luminances. In all graphs, the blue, orange and red shaded areas highlight an estimate of the distribution of (T^*, P^*) chroma values.*
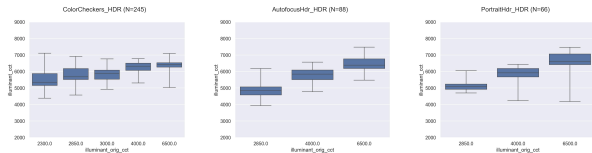

Figure 12. *Chromatic adaptation decisions on (a) ColorChecker, (b) Autofocus HDR and (c) Portrait HDR setups for HDR formats, represented as the CCT output of the color rendering model [1].*

figure 11 (a) and (b), has a per-illuminant variance in the same order of magnitude as the actual between-illuminant changes, whose order of magnitude $\Delta E_{ITP} \approx 20$. This is significant when testing the robustness of these complex image processing pipelines: assessing perceived color casts may become a challenging problem under a high amount of both measurement (due to manufacturing tolerances, lab variance, etc) and device noise (due to instability of the camera pipeline and heuristics). We observe a similar behavior with the same order of magnitude for skin tones in figure 11 row (c).

The color rendering model [1] provides a similar observation in terms of its CCT output in figure 12, where a lot of the distribution is gathered outside of the interquartile. It also reiterates the same observation as our last work: the median device does a very partial chromatic adaptation: they represent very warm illuminants (2300K-2850K) using similar chromaticities as a 5000K black body.

Color saturation

The last observation about our dataset is related to perceived saturation of colors. Following our observations about luminance
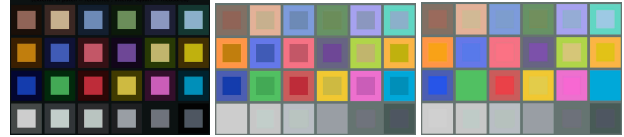

Figure 13. *Proposed visualization of patch saturation between measured colors of one HDR rendering of the ColorChecker ROI of a Composite transmissive slide (center color) and original SDR references (border color). All images are scaled down for preview on an SDR display for this paper, but were originally meant to be displayed on an HDR display. (a, left) shows original values; (b, middle) shows the output of the color rendering model [1]; (c, right) same as (b), with colors output by the model additionally scaled to match the luminance of the measured HDR colors.*
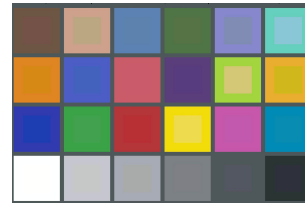

Figure 14. *Proposed visualization of patch saturation on SDR displays, where the measured center color alone, initially supposed to be displayed on an HDR display, is scaled down in luminance to match the luminance of the reference. All colors displayed on this image are therefore displayable on an SDR display.*

contrast and relative brightness in earlier sections, similar alterations can affect the perceived saturation of colors: a similar S-curve as those in figure 7 can affect chroma values, causing low-chroma colors to swerve towards gray and high-chroma colors to become even more saturated.

Interestingly, it can be fairly difficult to showcase such differences in a paper. A typical representation is done via tone-mapping, or even a simple luminance scaling in CIE-XYZ, but the luminance differences (or even color cast differences) between references make it impossible to compare colors between different representations, such as in 13 (a). Alternative representations can involve comparing measured colors with the outputs of color models such as our own work [1] in (b); they effectively eliminate most of the color cast issues and provide a good representation for gray patches... but luminance scaling of the model outputs is often still necessary in order to isolate color differences in (c).

It is also possible to luminance-scale the measured colors, and not the reference (figure 14) and have a visualization that's compatible with all SDR displays. This is a very useful trait for analysis, especially since we don't always have HDR displays when comparing color measurements!

Lastly, we note that the observation we made in our previous work [1] about the fact that a simple modeling of saturation as a simple chroma factor does not really hold - this is very visible in the fact that some patches in figure 13 (b) are more saturated, and some are less ; and indeed, the distribution of chroma factors in the proposed dataset shown in figure 15 does not deviate strongly from unity in average.
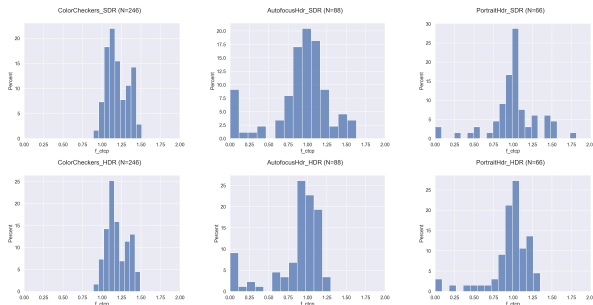
Figure 15. *Global saturation decisions as the $f_{C_tC_p}$ output of the color rendering model [1]. The top row (a), (b), (c) shows respectively the ColorChecker, Autofocus HDR and Portrait HDR setups for SDR formats ; the bottom row (a'), (b'), (c') shows the same setups for HDR formats.*

Conclusion and further work

The images in this dataset contain measurement areas that we haven't explored, choosing to focus on the same topic as our previous work (color and brightness): notably, measurement of noise, sharpness and texture/detail preservations which may be the subject of future papers.

However, they only contain lab data and we still have very little datasets of charts (e.g. ColorChecker) inside high-resolution images of natural scenes.

This work reiterates the importance of working on a cohesive assessment of contrast in high-fidelity HDR images, including both luminance and color contrast.

Finally, a long-term end-goal of this and previous works is to eventually link well-studied full-reference HDR image quality metrics with lab measurements, and provide conditions in which they are valid for quantifying image quailty of automatic camera devices, including for use in automated Q&A testing.

Bibliography

[1] Anna Tigranyan et al. "Objective Color Characterization of HDR Videos Captured by Smartphones: Laboratory Setups and Analysis Framework". In: *Electronic Imaging* 36.9 (2024), pp. 275-1–275-1. DOI: 10.2352/EI.2024.36.9.IQSP-275. URL: https://library.imaging.org/ei/articles/36/9/IQSP-275.

[2] Emin Zerman, Giuseppe Valenzise, and Frederic Dufaux. "An extensive performance evaluation of full-reference HDR image quality metrics". In: *Quality and User Experience* 2.1 (Apr. 2017). ISSN: 2366-0147. DOI: 10.1007/s41233-017-0007-4.

[3] 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.

[4] Anustup Choudhury and Scott Daly. "Comparing common still image quality metrics in recent High Dynamic Range (HDR) and Wide Color Gamut (WCG) representations". In: *Electronic Imaging* 32.9 (Jan. 2020), pp. 214-1-214–9. ISSN: 2470-1173. DOI: 10.2352/issn.2470-1173.2020.9.iqsp-214.

[5] *Photography and graphic technology — Extended colour encodings for digital image storage, manipulation and interchange — Part 5: High dynamic range and wide colour gamut encoding for still images (HDR/WCG)*. 2023. URL: https://www.iso.org/standard/81863.html.

[6] *Digital Photography: Gain map metadata for image conversion, Part 1: Dynamic Range Conversion*. Draft International Standard. URL: https://www.iso.org/standard/86775.html.

[7] Google. *Ultra HDR Image Format v1.0*. URL: https://developer.android.com/guide/topics/media/platform/hdr-image-format.

[8] *AV1 Codec ISO Media File Format Binding*. The Alliance for Open Media. URL: https://aomediacodec.github.io/av1-isobmff/.

[9] *AV1 Bitstream & Decoding Process Specification*. The Alliance for Open Media, 2019. URL: https://aomediacodec.github.io/av1-spec/av1-spec.pdf.

[10] *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*. 2022. URL: https://www.iso.org/standard/83102.html.

[11] International Telecommunication Union. *Recommendation ITU-T H.273 - Coding-independent code points for video signal type identification*. 2024.

[12] Apple. *Applying Apple HDR effect to your photos*. URL: https://developer.apple.com/documentation/appkit/applying-apple-hdr-effect-to-your-photos.

[13] *IEC 61966-2-1:1999 Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB*. International Electrotechnical Commission, 1999.

[14] *SMPTE EG 432-1:2010 Digital Source Processing —Color Processing for D-Cinema*. Society for Motion Pictures Engineers, 2010.

[15] 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.

[16] *A DTV Profile for Uncompressed High Speed Digital Interfaces (CTA-861-G)*. Consumer Technology Association, 2016.

[17] Frédéric Cao, Frédéric Guichard, and Hervé Hornung. "Dead leaves model for measuring texture quality on a digital camera". In: *Digital Photography VI*. Ed. by Francisco Imai, Nitin Sampat, and Feng Xiao. Vol. 7537. SPIE, Jan. 2010, 75370E. DOI: 10.1117/12.838902.

[18] Gabriele Facciolo et al. "Quantitative measurement of contrast, texture, color, and noise for digital photography of high dynamic range scenes". In: *Electronic Imaging* 30.12 (Jan. 2018), pp. 170-1-170–10. ISSN: 2470-1173. DOI: 10.2352/issn.2470-1173.2018.12.iqsp-170.