PERCEPTUAL COLOUR DIFFERENCE UNIFORMITY IN HIGH DYNAMIC RANGE AND WIDE COLOUR GAMUT

Thilan Costa[*1]*,Vincent Gaudet*[1]*, Edward R. Vrscay*[2] *and Zhou Wang*[1]

[1]Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
[2]Dept. of Applied Mathematics, University of Waterloo, Waterloo, ON, Canada

ABSTRACT

Perceptual uniformity is a highly desirable property of colour spaces or colour difference measures where equal level in colour value difference should result in equal perceptual difference. Designing colour spaces or colour difference measures of perceptual uniformity is a long standing problem in colour science. This has become increasingly important with the growing popularity of high dynamic range (HDR) and wide colour gamut (WCG) cameras, content and displays. We design an efficient testing framework to evaluate perceptual uniformity by subjective just noticeable difference (JND) measurement at a wide range of luminance levels followed by coefficient of variation (CV) computation. We carry out subjective testing on RGB, xyY, L*a*b*, YCbCr, CIECAM02-UCS and ICtCp colour spaces and ΔE2000 metric in ITU-R BT 2020 colour gamut across a wide range of luminance levels (from 0.01 to 500 nits) using a professional HDR/WCG display in a carefully controlled dark testing environment. Our results suggest that on average, the ICtCp space performs the best in the current test, but is still distant from achieving perceptual uniformity.

***Index Terms*—** colour difference, high dynamic range, wide color gamut, subjective testing, perceptual uniformity, colour spaces, ICtCp, L*a*b*, YCbCr, ΔE2000

1. INTRODUCTION

Since the foundations for the current system of colorimetry was laid in 1931 by the Commission Internationale de l'Eclairage [1], much research work has been performed to formulate a colour space that is perceptually uniform. The perceptually uniformity can be said to hold in a colour space if an approximately equal level of change in the colour values results in a perceptually equal difference across the entire colour space [2]. Perceptual uniformity of colour spaces and colour difference measures is highly desirable not only because of its scientific value, but also for the potential applications in a wide range of engineering problems. These include the design of novel cameras and displays, image/video quality assessment [3] [4], perceptual image/video coding [5], colour space conversion, and tone/gamut mapping [6]. The need has become increasingly strong with the growing popularity of high dynamic range (HDR) [7] and wide colour gamut [7] cameras, content and displays.

The ultimate judgment on colour differences should be given by human eyes, for which the most commonly used measure is the just noticeable difference (JND) [2], which denotes the smallest level of change that the human visual system (HVS) can perceive. MacAdams performed one of the earliest experiments for measuring the JND in 1942 [8]. The experiment was conducted with a constant luminance for the test area of approximately 48 candela per square meter (cd/m^2 or nit). Another similar experiment by Wyszecki and Fielder [9] was conducted at a constant luminance level of 12 nits. Recent colour difference tests performed by Dolby Laboratories on a newly introduced colour space ICtCp is presented in [10][11]. The results suggested that the ICtCp space exhibits better perceptual uniformity than RGB, YCbCr, L*a*b* and ΔE2000, especially at low luminance levels.

It is worth noting that a thorough test of JND color uniformity would require densely sampling the full color space, followed by gauging the JND along all directions in the color space at each sampling point. Unfortunately, the potential combinations prohibit such a test, resulting in very limited testing in reality [8][9][10][11]. In this work, we present a testing framework for evaluating the perceptual uniformity of colour spaces and colour difference measures. We evaluate the performance of five different colour spaces and one colour difference measure based on the JND measurements collected through subjective experiments. A comparison between the colour spaces are then performed using the coefficient of variation (CV). Compared with earlier works in the literature, our testing framework is more efficient, allowing us to test much more colours in a much larger luminance range. Specifically, while the experiments by MacAdams [8], Wyszecki and Fielder [9], and two experiments by Dolby Laboratories [10][11] covered 25, 30, 9 and 21 unique test colours, respectively, our test contains 245 unique test colours from within the ITU-R BT. 2020 [12] colour primaries, sampled at 7 luminance levels (35 unique test colours per constant luminance level). Moreover, the experiments prior to that of Dolby Laboratories was limited to $15 - 100\ cd/m^2$ [11], and the latest experiment was limited to 3 luminance levels (0.1, 25, and 1000 cd/m^2) [11]. As such, our test scope, methodology, and results provide a more comprehensive understanding of the perceptual uniformity performance of the colour spaces and colour difference measures.

2. DESIGN OF TESTING FRAMEWORK

The colour matching subjective test consists of 245 reference colours sampled from within the ITU-R BT 2020 [12] defined colour primaries. The reference colours for each test are randomly sampled from within the xy$Y$ colour space [2]. The luminance value Y is fixed at 0.01, 0.1, 1, 10, 100, 300, and 500 nits. At each luminance level, 35 samples are chosen (for a total of $7 \times 35 = 245$ reference colours). The location of the 245 reference colour points are shown in the xy chromaticity diagram in Figure 2. For each reference, 1500 test colour points are then sampled in the xy$Y$ space along a straight line (of length 0.1 in the xy chromaticity diagram) from the reference, at a randomly chosen angle. The luminance Y for the test

* Corresponding authors. Email: tcosta@uwaterloo.ca

Table 1. Average JND in different colour spaces for each tested luminance level and the entire tested luminance range (denoted as "All")

Method	0.01 nits	0.1 nits	1 nits	10 nits	100 nits	300 nits	500 nits	All
ΔRGB	2.8930	4.6611	10.3475	15.6696	20.6305	19.8423	28.6965	14.6772
ΔICtCp	2.7888	3.9108	6.7452	10.4535	11.7651	11.8718	10.6578	8.3133
ΔYCbCr	0.0017	0.0027	0.0058	0.0091	0.0126	0.0107	0.0156	0.0083
ΔL*a*b*	11.0908	6.1569	11.3634	22.1332	42.1971	56.1250	66.8717	30.8483
ΔE2000	5.1055	2.5949	3.5564	3.4071	4.2322	5.0215	4.2041	4.0174
Δxy$Y$	0.0376	0.0206	0.0169	0.0148	0.0098	0.0090	0.0107	0.0171
CIECAM02-UCS	0.1629	0.0933	0.3203	0.2238	0.2090	0.1727	0.2125	1.6835

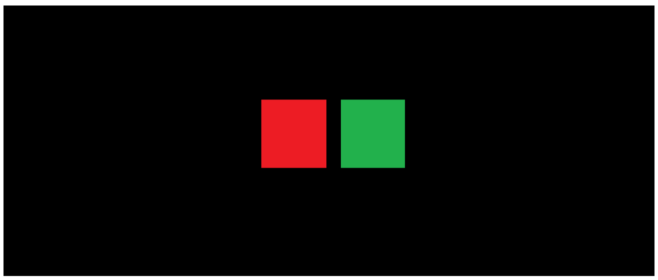

Fig. 1. Test screen layout.

colour points is held constant at the same value as the corresponding reference colour. The reference and test colour points are then converted to 10-bit RGB PQ encoded [13] [14] values that could be displayed on a reference monitor.

The test is carried out on a Canon DP-2420 reference display in a dark environment with no ambient lighting. Data transmission to the display is performed via four Serial Digital Interface (SDI) cables to prevent any possible loss of data that may occur in other standards such as the High Definition Multimedia Interface (HDMI). Each test subject compares two colour squares as shown in Fig. 1. Each test case consists of the reference square and the corresponding 1500 colour points sampled in a straight line as described above. One of the two squares, chosen at random for each test case, would contain the reference colour. At the start of each test case, the other square would contain the colour sample from within the 1500 sampled colour points that is furthest away from the reference colour. The colour of the test square would then be adjusted by the user to display the different colour points from the set of 1500 test colours corresponding to that particular reference test colour. The subject is given the ability to navigate through the 1500 test colours using a slider that is displayed on an adjacent display. This adjacent display is covered with blackout material to only reveal the slider. This reduces the impact from the backlight of the display. The DP-2420 reference display is also covered with the same blackout material to only reveal the colour squares. This covering is performed since the display itself is not capable of producing absolute blacks due to the LED technology on which it functioned.

Each test subject would spend an hour on average to complete the test. A mandatory break was enforced at the halfway point of the test to reduce visual fatigue. Test subjects were also allowed additional breaks if required.

When taking the test, the subject is instructed to stop the movement on the slider at the furthest point at which the two squares would be perceived as having the same colour, and then press a button to move to the next test case. This action would record the current test colour sample values, and then display the next test case.

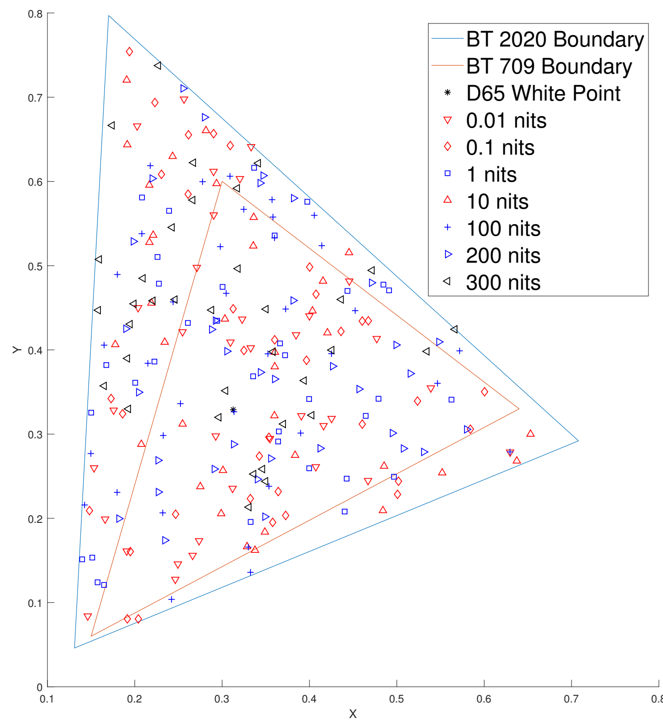

Fig. 2. Location of chosen reference colours on the xy chromaticity diagram.

The rationale behind this procedure is that the distance between the reference colour sample and the test colour sample at which the subject stops the movement, would then correspond to the JND for the reference colour in a particular direction.

Several key choices taken in the test framework are explained as follows.

- **Sampling in the xy$Y$ space:** We aim for providing a common ground for testing, and sampling the colour points from the same colour space being tested could result in a systematic bias (*eg. points could be sampled at sufficiently larger distance than the JND variance, thus giving the impression that it is perceptually uniform*). Therefore, we sample the colour points in the common xy$Y$ space where the standard colour gamuts such as ITU-R BT. 709 [15], BT. 2020 [12] and DCI P3 [16] are defined.

- **Conversion to RGB rather than YCbCr for display:** YCbCr space was designed mainly for industrial video distribution purposes with the intent of subsampling, which impacts the Cb and Cr components that are required for accurate colour representation.

- **Choice of luminance values:** The luminance range was chosen to cover a wide dynamic range with HDR applications in mind. It is also matched with the test carried out at the Dolby laboratories [10].

- **Difference in Experimental setup:** In the subjective experiment carried out at Dolby laboratories [10], each test case consists of 4 colour squares placed in 4 quadrants of the screen, with a fixed test colour in 3 of the squares and the reference colour displayed in one of them. The same test and reference colour combination will repeat 4 times, with the reference colour contained in a different quadrant each time. As a consequence, the pairs of reference and test colour combinations that can be tested according to this framework is fairly limited considering the fact that the subjects would need to finish the experiment within a reasonable period of time to avoid visual fatigue. With the framework described in this paper, each test case covers a single reference colour and 1500 unique test colour points (and could be simply scaled to even more points if necessary without affecting the length of the test), offering a far more efficient level of coverage.

- **Spatial separation of test color squares:** While no spatial separation was used between the test and reference colours in the MacAdams experiment [8], studies on visual perception indicate that there are optical illusions (*Cornsweet illusion* [17] and *Mach band effect* [18]) that can cause the HVS to incorrectly perceive differences when two colours are placed immediately next to each other. Therefore, separating the two colours (reference and test) with a neutral background colour can alleviate such effects.

- **Spatial versus temporal separation of reference and test colour squares:** One possibility that was considered was the use of a single square that can be toggled to display between reference and test colour samples, i.e. temporal separation between test and reference colours. However, earlier studies have shown that due to the successive contrast effect, the HVS may be affected by the previously viewed colour [19] [20].

3. RESULTS & ANALYSIS

A total of 30 subjects, aged between 22 and 35, took part in the experiment. The JND value for each test case was then computed using the collected RGB data as follows.

$$\Delta\text{RGB} = \sqrt{(R_r - R_t)^2 + (G_r - G_t)^2 + (B_r - B_t)^2} \qquad (1)$$

where R_r,G_r,B_r denote the reference RGB values and R_t,G_t,B_t values denote the RGB values of the test colour sample chosen by the subject during the test. Outlier detection and removal were then performed on the collected data using these computed JND values as specified in [21].

The RGB data with the outliers removed were then converted to the YCbCr, ICtCp, Lab and xy$Y$ colour spaces. The JND for the RGB colour space was computed as given in (1) for the data with the outliers removed. The JND for YCbCr, L*a*b*, and ICtCp were calculated as follows [10].

$$\Delta\text{YCbCr} = \sqrt{(Y_r - Y_t)^2 + (Cb_r - Cb_t)^2 + (Cr_r - Cr_t)^2}$$

$$\Delta\text{L*a*b*} = \sqrt{(L_r^* - L_t^*)^2 + (a_r^* - a_t^*)^2 + (b_r^* - b_t^*)^2}$$

$$\Delta\text{ICtCp} = \sqrt{(I_r - I_t)^2 + 0.25(Ct_r - Ct_t)^2 + (Cp_r - Cp_t)^2}$$

Although not a colour space of its own and is based on the L*a*b* colour space, the JND based on ΔE2000 and CIECAM02-UCS [22] was also computed for each test case. ΔE2000 is recommended by CIE for measuring colour differences and the details of the computation are given in [23].

Given the JND J_k computed for each subject k for a particular test case i using the formulas given above, the final average JND value J_i for the test case i was then computed as (similar computation has been performed in [10][11])

$$J_i = \frac{1}{N}\sum_{k=1}^{N} J_k. \qquad (2)$$

Given the JND of J_i for each test case i, the JND for a particular luminance L (or all luminance values as denoted in the last column of Table 1), denoted J_L, was computed as

$$J_L = \frac{1}{n(L)}\sum_{i\in L} J_i, \qquad (3)$$

where $n(L)$ denotes the number of test cases corresponding to the luminance value. Table 1 shows the JND values obtained for different colour spaces. As can be seen from the table, none of the colour spaces produce a JND that is constant across luminance levels.

In order to have a better comparison between the variation of the computed JND J_i for each test case i in comparison to the computed JND for a particular luminance based subsets of the test cases J_L (or all luminance values), we computed the coefficient of variation (CV) as given by [24],

$$CV = \frac{\textit{standard deviation}}{\textit{mean}} = \frac{\sigma(\{J_{i\in L}\})}{J_L}. \qquad (4)$$

If the value of CV is closer to zero, it follows that J_i for each of the test cases within a particular subset L are closer to the computed global JND J_L. This is a strong indicator that the space is perceptually uniform within that luminance range. In the same manner, if the CV value is very high, it follows that the JND J_i of each test case within the particular subset L is largely varying. This would indicate

Table 2. Coefficient of Variation for each colour space at each tested luminance level and the entire tested luminance range (denoted as "All").

Method	0.01 nits	0.1 nits	1 nits	10 nits	100 nits	300 nits	500 nits	All
ΔRGB	0.4387	0.5285	0.8584	0.7705	1.0061	0.9764	1.3197	1.3943
ΔICtCp	**0.3536**	**0.3474**	**0.4111**	**0.4765**	0.5347	**0.3345**	**0.2861**	**0.6182**
ΔYCbCr	0.4307	0.5760	0.8034	0.7755	1.0506	0.8671	1.2565	1.3339
ΔL*a*b*	1.5192	0.6029	0.5277	0.4959	0.4463	0.4070	0.3925	0.9091
ΔE2000	1.4293	0.5206	0.4891	0.5054	0.6369	0.8487	0.5929	0.9186
Δxy$Y$	0.4264	0.6450	0.7956	0.8653	0.8180	1.1841	1.3462	0.9230
CIECAM02-UCS	1.4427	1.3239	0.6096	1.2456	**0.4418**	0.6302	0.6524	0.7801

that the perceptual uniformity of the space is lacking. Thus, it can be used as a measure of the perceptual uniformity of the colour spaces for a particular luminance range for which the test data is available. Since CV is normalized, it also allows us to compare the performance (the perceptual uniformity) across colour spaces and colour difference measures. Table 2 contains the CV for each of the luminance values and for the entire range of luminance values (rightmost column of the table).

From the CV for the entire range of luminance values, ICtCp colour space appears to be the best performer with the least variation of the JND. The CIECAM02-UCS is also seen to perform better than ΔE2000 and the L*a*b* colour space, indicating that it is more perceptually uniform as indicated in [25] Both the JNDs computed in the L*a*b* space and using ΔE2000, which also operates on the L*a*b* space, perform worse in the lowest luminance range of 0.01 nits. These results are consistent with the findings in [10]. Removing the test samples from the 0.01 nits luminance range improves the CV to 0.8262 for the L*a*b* space and 0.6900 for ΔE2000. Moreover, CIECAM02-UCS performance increases to 0.5825 CV when the luminance under 100 nits range is excluded. For the higher end of the HDR range of 300 and 500 nits, while ICtCp performs the best, the L*a*b* space has a comparatively lower CV as seen from the table. In fact, computing the JND in the L*a*b* space by excluding all luminance ranges below 300 nits resulted in a CV of 0.4503, much better than its CV for the entire range of luminance of 0.9091. Somewhat surprisingly, RGB, YCbCr, and xy$Y$ colour spaces also show good performance at the lowest luminance range of 0.01 nits.

4. CONCLUSION

We have presented an efficient test framework to measure the perceptual uniformity of existing colour spaces and the ΔE2000 colour difference measure. We also propose to use CV for perceptual uniformity assessment of a colour space. Analysis of our subjective test data suggests that on average, the ICtCp colour space appears to be more perceptually uniform in comparison to the other colour spaces. Meanwhile, the performance of ΔE2000 is comparable if the lowest luminance level of 0.01 nits is excluded, and CIECAM02-UCS performance is better than ICtCp for luminances greater than 100 nits. The analysis also indicates that the RGB and YCbCr colour spaces have a reasonable degree of perceptual uniformity at the lowest luminance range of 0.01 nits. This is especially interesting since these colour spaces are widely used today for distribution of image and video content. Given that ICtCp, the most perceptually uniform colour space under test, still has only a significantly high overall CV of 0.6182, our work suggests that there is still a large room for improvement in the development of a perceptually uniform colour space and colour difference measures.

Compared with existing works, our current work focuses more on global uniformity (i.e. JND uniformity across the color space) by sampling significantly larger color samples in the color space. In the future, to achieve a thorough evaluation of color uniformity, it is necessary to perform local uniformity testing, for which the JND at every test color sample is desired to have a sphere shape in the 3D color space.

5. REFERENCES

[1] H. S. Fairman, M. H. Brill, and H. Hemmendinger, "How the cie 1931 color-matching functions were derived from wright-guild data," *Color Research & Application*, vol. 22, no. 1, pp. 11–23, 1997.

[2] C. Poynton, *Digital video and HD: Algorithms and Interfaces*, Elsevier, 2012.

[3] K. Ma, T. Zhao, K. Zeng, and Z. Wang, "Objective quality assessment for color-to-gray image conversion," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4673–4685, 2015.

[4] U. Rajashekar, Z. Wang, and E. P. Simoncelli, "Perceptual quality assessment of color images using adaptive signal representation," in *Human Vision and Electronic Imaging XV*. International Society for Optics and Photonics, 2010, vol. 7527.

[5] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1215–1226, 1998.

[6] C. Chinnock, "Dolby Vision and HDR10," *White Paper of Insight Media*, 2016.

[7] P. H. Putman, "SMPTE Progress Report–Displays 2017," *SMPTE Motion Imaging Journal*, vol. 126, no. 7, pp. 83–87, 2017.

[8] D. L. MacAdam, "Visual sensitivities to color differences in daylight," *JOSA*, vol. 32, no. 5, pp. 247–274, 1942.

[9] G. Wyszecki and G. H. Fielder, "New color-matching ellipses," *JOSA*, vol. 61, no. 9, pp. 1135–1152, 1971.

[10] J. Pytlarz, E. Pieri, and R. Atkins, "Objectively evaluating high dynamic range and wide color gamut color accuracy," *SMPTE Motion Imaging Journal*, vol. 126, no. 2, pp. 27–32, 2017.

[11] E. Pieri and J. Pytlarz, "Hitting the mark-A new color difference metric for HDR and WCG imagery," *SMPTE Motion Imaging Journal*, vol. 127, no. 3, pp. 18–25, 2018.

[12] ITU-R Recommendation BT.2020-2, "Parameter values for ultra-high definition television systems for production and international programme exchange," October 2015.

[13] ST SMPTE, "2084: 2014," *High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays*, pp. 1–14, 2014.

[14] "BT.2100: Image parameter values for high dynamic range television for use in production and international programme exchange," Standard, Radiocommunication Sector of International Telecommunication Union, 2017.

[15] ITU-R Recommendation BT.709-6, "Parameter values for the HDTV standards for production and international programme exchange," June 2015.

[16] EG SMPTE, "432-1:2010," *SMPTE Engineering Guideline - Digital Source Processing Color Processing for D-Cinema*, 2010.

[17] R. P. Dooley and M. I. Greenfield, "Measurements of edge-induced visual contrast and a spatial-frequency interaction of the Cornsweet illusion," *JOSA*, vol. 67, no. 6, pp. 761–765, 1977.

[18] A. Fiorentini, "Mach band phenomena," in *Visual psychophysics*, pp. 188–201. Springer, 1972.

[19] G. Wyszecki and W. S. Stiles, *Color science*, vol. 8, Wiley New York, 1982.

[20] G. Wyszecki, K. R. Boff, L. Kaufman, and J. P. Thomas, *Handbook of perception and human performance: Sensory processes and perception*, 1986.

[21] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," January 2012.

[22] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M R. Luo, and T. Newman, "The CIECAM02 color appearance model," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2002, vol. 2002, pp. 23–27.

[23] P. Goldstein, "Non-Macadam color discrimination ellipses," in *Novel optical systems design and optimization XV*. International Society for Optics and Photonics, 2012, vol. 8487, p. 84870A.

[24] H. Abdi, "Coefficient of variation," *Encyclopedia of Research Design*, vol. 1, pp. 169–171, 2010.

[25] N. Moroney and Z. Huan, "Field trials of the CIECAM02 color appearance," *CIE 25th Quadrennium*, 2003.