https://doi.org/10.2352/lim.2023.4.1.06

Quantifying visual differences between color visualizations on different displays

Eric Kirchner[1], Lan Njo[1], Esther Perales[2] and Aurora Larrosa Navarro[2]. 1. Department Surface Innovation & Analytics, AkzoNobel Paints and Coatings, Sassenheim, the Netherlands. 2. Department of Optics, University of Alicante, Alicante, Spain

Abstract

We introduce a quantitative metric to analyze the visually perceived differences between colors that are visualized on different displays, or that are visualized on the same display but using different visualization methods. This metric is validated by analyzing perceived visual differences as scored by observers using an iPad Air 2 display under different ambient light conditions. Our results show that the metric calculations are well aligned with the visual data from this experiment.

We use the new metric to investigate the reproducibility of spectroradiometer data from three different displays (iPad Air 2, and iPad models from 2017 and 2018). Our results show that color visualizations based on these datasets are virtually identical for the iPad Air 2 and iPad 2017. For the iPad 2018, in circa 10% of the colors a visually noticeable difference occurs between visualizations based on the older and on the new dataset. We use the same metric to also compare color visualizations on these three displays. Our results show that color visualizations between iPad 2017 and iPad Air 2 are often visually different, with color differences larger than CIEDE2000 = 4.0 for 50% of the colors. But comparing iPad 2017 with iPad 2018 the color visualizations are often visually identical. For 90% of the colors, color differences CIEDE2000 < 2.4.

Introduction

In the past decades there has been a strong growth in the demand for photorealistic rendering of objects. As an example, the rise of eCommerce and online retail for consumers has resulted in increasing demands on the accuracy of the digital representation of objects, in order to reduce the number of products that are returned to the supplier. The default technology to calculate digital color representations is by using the device-independent standard red, green, blue (sRGB) color space [1]. It is well-known that when using this method in cross-media color reproduction, a visual comparison of the digital representations of objects with the corresponding physical objects often shows these representations are not accurate.

The model parameters in sRGB color space were determined decades ago, when most displays were based on cathode-ray tube (CRT). Current displays are mostly based on either Organic LED (OLED) or Liquid Crystal Display (LCD) technologies. For displays in e.g. computer monitors and laptops color management systems including ICC profiles can be used to further optimize color representations on specific devices. For mobile displays such color management systems may be part of the operating system but they are not accessible to the user or to third party software. Modern mobile displays show widely varying color properties, and the color management system that is hidden to the user may not always result in sufficiently accurate color representations for certain tasks. In order to enable the user to create a device-dependent visualization method that results in more accurate color visualization than what is obtained by the standard visualization provided (sRGB color space and the manufacturer's color management system) we introduced the Mobile Display Characterization and Illumination Model (MDCIM) [2][3]. We showed that it results in better color accuracy in cross-media color reproduction both on a representative LCD [2] as well as on an OLED display [3]. The MDCIM model not only accounts for the technical specifications of displays, but also the influence of ambient lighting.

Here we present results for applying the MDCIM model on three different types of LCD display. We propose a quantitative method to analyze the differences in digital color representations for these three display models, and correlate the results of this analysis with the results from a visual assessment of the color accuracy of the cross-media color representations.

Parameters in the MDCIM model

Similar to the definitions of sRGB color space, in the MDCIM model the relation between tristimulus values X, Y, Z (which in turn are linked to colorimetric parameters such as CIELab) and the values of the luminance Y_R, Y_G and Y_B of the red, green and blue channel of the display follows the GOG-model from Berns [4][5]:

$$\begin{pmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{pmatrix} = M \begin{pmatrix} Y_R \\ Y_G \\ Y_B \end{pmatrix} \quad (1)$$

The Opto-Electronic Transfer Functions (OETF) that describe how the luminance of all three channels depend on the digital values d_R, d_G and d_B are assumed to be exponential:

$$\begin{pmatrix} Y_R \\ Y_G \\ Y_B \end{pmatrix} = \begin{pmatrix} (k_{1,R} d_R + k_{2,R})^{\gamma_R} \\ (k_{1,G} d_G + k_{2,G})^{\gamma_G} \\ (k_{1,B} d_B + k_{2,B})^{\gamma_B} \end{pmatrix} \quad (2)$$

although at low luminance values $d < d_0$ this dependence is assumed to scale linearly with a coefficient β. In our earlier work we showed that the values of parameters k_1, k_2 and β can be directly calculated from the numerical values of parameters d_0 and γ, in order to ensure continuity of the OETF functions and their derivatives [2][3]. For display systems with 3x8 bits color representation, d_R, d_G and d_B are found by dividing the common RGB values (ranging from 0 to 255) by 255 (i.e., $2^8 - 1$).

For sRGB space matrix M is given by:

$$M_{sRGB} = \begin{pmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{pmatrix} \quad (3)$$

In the MDCIM model, these matrix elements depend on display specific parameters, *i.e.* the chromaticity coordinates x, y and maximum luminance values Y_{max} of the red, green and blue channel. Comparing equation (3) and (4) it follows that the sRGB color space assumes particular values for the chromaticity coordinates and maximum luminance values of the red, green and blue channel. These values are mentioned in Table 1.

In equation (1), we have also included the tristimulus values X_0, Y_0, Z_0 of a completely black image. This black level correction was found to be necessary when applying the MDCIM model to LCD displays, since even when displaying pure black some backlighting is transmitted by the display [2]. For OLED displays the black level correction is not needed, since their representation of black is much darker than for LCD displays [3]. For more details on the MDCIM model and its derivation, and on the way it takes into account ambient lighting, we refer to our earlier publications [2][3].

$$M = \begin{pmatrix} \frac{x_R}{y_R} Y_{R,\max} & \frac{x_G}{y_G} Y_{G,\max} & \frac{x_B}{y_B} Y_{B,\max} \\ Y_{R,\max} & Y_{G,\max} & Y_{B,\max} \\ \frac{z_R}{y_R} Y_{R,\max} & \frac{z_G}{y_G} Y_{G,\max} & \frac{z_B}{y_B} Y_{B,\max} \end{pmatrix} \quad (4)$$

Experimental

Displays

In this investigation we included three types of LCD displays from manufacturer Apple. The first is an iPad Air 2, launched in 2014. The second is a 5th generation iPad, launched in 2017. The third display that we include is a 6th generation iPad, launched in 2018.

A CS-2000A spectroradiometer (Konica Minolta) was mounted inside a completely dark room. The display was placed against the aperture of the spectroradiometer, which was set at 0.1° viewing mode. Using the operating system of the tablet, we set the brightness of the display at maximum. In order to measure the parameters required for the MDCIM model, equations (1) - (4), we processed 56 well-chosen images that were shown on the display. The spectroradiometer measured the CIE 1931 chromaticity coordinates x, y as well as the luminance Y for the light emitted by the display for each separate image. For example, to determine the best fit parameters for the OETF of the red channel we included images with (R, G=0, B=0) and R=0; 5; 10; …; 50; 75; 100; 125; …; 200; 225; 255. This provided a good sampling of the complete tone rendering curve and enables us to determine values for the parameters x_R, y_R, z_R, $Y_{R,max}$, as well as X_0, Y_0 and Z_0. For more details on the experimental set-up, we refer to our previous publications [2][3]. The chromaticity coordinates and maximum luminance values that we measured for the three channels, and also for the black signal, are summarized in Table 1.

Optimizing MDCIM model parameters

By fitting the spectroradiometer data with the parametrization of equation (2) we determined the optimized values for parameters d_0, γ, as well as X_0, Y_0 and Z_0, which in turn (see Appendix of Ref.[2]) result in the corresponding values for model parameters k_1, k_2, β and d_0. These values are summarized in Table 2.

Previously we tested the accuracy of color reproduction when using the device-specific MDCIM model versus the case of using the default, device-independent sRGB color space. Since we described this test in detail before [2], we only present a summary here. Seven observers with normal color vision viewed 35 RAL colors on a neutral background. For each color, a physical low gloss color standard sample was compared to its digital representation on a display that was viewed from a perpendicular angle. Observers were asked to rate the quality of the color match between the physical sample and the digital image, using a scale that runs from 0 (no difference / hardly any difference) up to 5 (large difference; very bad match). For a range of ambient illuminance levels from 600-3000 lx these visual tests showed that the MDCIM model considerably improves color accuracy for the LCD display of the iPad Air 2 [2] and for the OLED display of the Samsung Galaxy S4 [3].

Differences between displays

For the three types of displays investigated here, Tables 1 and 2 show that the MDCIM model parameters are slightly different between these displays. However, it is unclear from these values how significant these differences are. For example, if we would use the MDCIM model parameters that we determined for the iPad Air 2 for visualizing colors on an iPad (model 2017), how good or bad would this visualization be as compared to the case where we would use the MDCIM model parameters that we determined specifically for the iPad (2017)?

Table 1: Spectroradiometer data relevant for MDCIM model parameters for three displays, compared to parameters implicitly assumed when using default sRGB color space.

	sRGB	iPad Air 2 v1	iPad (2017) v1	iPad (2018) v1
x Red	0.64	0.6421	0.6424	0.6395
y Red	0.33	0.3264	0.3293	0.3303
Y_{max} Red	21.26	80.89	101.6	93.82
x Green	0.30	0.3071	0.3031	0.3052
y Green	0.60	0.6079	0.6012	0.6029
Y_{max} Green	71.52	307.2	359.73	321.08
x Blue	0.15	0.1527	0.1567	0.1527
y Blue	0.06	0.0489	0.0607	0.0607
Y_{max} Blue	7.22	27.18	39.04	35.17
x Black	0.0	0.2458	0.2439	0.2647
y Black	0.0	0.2085	0.2078	0.2582
Y Black	0.0	0.3647	0.48	0.51

Table 2: Best-fit values for MDCIM model parameters for three displays, compared to default sRGB color space.

	sRGB	iPad Air 2 v1	iPad (2017) v1	iPad (2018) v1
Red channel				
γ	2.4	2.35	2.27	2.28
d_0	0.04045	0.007843	0.007843	0.003922
k_1	0.947867	0.989523	0.990137	0.995005
k_2	0.055	0.010477	0.009863	0.004995
β	0.077399	0.010443	0.013317	0.005380
Green channel				
γ	2.4	2.35	2.27	2.28
d_0	0.04045	0.007843	0.007843	0.003922
k_1	0.947867	0.989523	0.990137	0.995005
k_2	0.055	0.010477	0.009863	0.004995
β	0.077399	0.010443	0.013317	0.005380
Blue channel				
γ	2.4	2.35	2.47	2.33
d_0	0.04045	0.007843	0.007843	0.003922
k_1	0.947867	0.989523	0.988602	0.994811
k_2	0.055	0.010477	0.011398	0.005189
β	0.077399	0.010443	0.007287	0.004467

Table 1 shows that for example the chromaticity coordinates x, y of the red channel are (0.6421, 0.3264) for the iPad Air 2 and (0.6424, 0.3293) for the iPad (2017). Numerically these values are very close to each other, but would they result in visual differences when the parameters of the MDCIM model are exchanged for these two displays? Below we will define a new metric that quantifies the difference between displays, as far as the visual difference is concerned between colors visualized with the MDCIM model and using various display specific model parameters.

The need to quantify reproducibility

As with any other experimental method, spectroradiometer data such as those collected in Table 1 have a certain measurement error. This error may be due to repeatability errors: even if the same measurement equipment is used on the same display immediately after the first measurement, the numerical values of parameters that are measured will probably deviate from the first measurement. Another source of measurement errors are reproducibility errors: if the measurement is repeated after a long time, and/or executed on a different copy of the same display (i.e., having the same display model and type), and/or executed using a different copy of the spectroradiometer instrument (which may or may not be of the same brand and model as used for the earlier measurement), the new measurement data will probably deviate from the first measurement.

For example, when we repeated the spectroradiometer measurements for the iPad Air 2, we found that for the red channel the chromaticity coordinates changed from (x_R = 0.6421, y_R = 0.3264) to (x_R = 0.6428, y_R = 0.3317), which intuitively seem to be minor differences, whereas the maximum luminance changed from 80.89 to 83.84 cd/m^2. Intuitively it is almost impossible to estimate if the changes in measurement data can be expected to result in visually perceived differences for images calculated with methods based on either of the two datasets. It is therefore not clear what measurement variations in the measured chromaticity coordinates x,y or the maximum luminances Y_{max} would result in differences in displayed colors after these parameters are processed in the MDCIM model. The same is true for the model fit parameters from Table 2: when a new set of spectroradiometer data is used and slightly different values for the model fit parameters result, the visual impact on displayed colors is unclear.

New metric to estimate visual differences between displays and reproducibility

We developed a new metric to quantify differences between displays (as far as colors visualized on them are concerned) as well as the reproducibility of MDCIM model parameters. Since statements on deviations between visualized colors may depend strongly on which colors are included in the visualization, we need to include a set of colors that occupies a representative part of color space. Therefore we chose to include all 197 RAL colors [6]. Secondly when assessing the difference between using two different methods to visualize colors we cannot assume that one of these methods is accurate and the other is not. Instead, we need to use a different visualization method that is independent of the two methods that are assessed with respect to each other. For this independent visualization method we use default sRGB color space [1].

In this approach we assume sRGB color space is sufficiently accurate for the particular display to show good correlation with visual assessments of color differences. Based on these arguments for this metric we calculate the RGB representation for each of the 197 RAL colors. We do this for both visualization methods that we need to compare with each other, resulting in two sets of RGB values for each RAL color. In order to quantify the visually perceived color difference between an image with values (R_1, G_1, B_1) with an image with values (R_2, G_2, B_2) we use default sRGB color space to convert RGB to CIE-Lab values (L_1*, a_1*, b_1*) and (L_2*, a_2*, b_2*). Next, we calculate color differences CIEDE2000 and ΔE_{ab} between the CIE-Lab values for each RAL color. This way we estimate the distribution of visual color differences that result when replacing one color visualization method by another, and hence derive statistical parameters related to that distribution: average, median, and percentiles of the distribution.

Results

Validate metric with visual assessments

For one visualization method to be visually significantly better than another method, the visualizations by both methods need to be visually distinguishable from each other. This is exactly what is quantified by our metric. Although our metric contains different options such as median values versus 90% percentile values, and CIEDE2000 versus ΔE_{ab} values, here we want to express the metric in one single number. Since earlier thresholds for cross-media color reproduction on displays proposed to use mean values of ΔE_{ab} = 3 (for professionals) and 6 (for consumers) [7], we test our criterium by calculating the percentage of RAL colors for which the sRGB and the MDCIM visualization for iPad Air 2 at 1500 lux differ by more than ΔE_{ab} = 4.5, exactly in between the thresholds just mentioned. For the 197 RAL colors, this percentage is 40%. We assume that (i) MDCIM visualizations are always more accurate than sRGB visualizations, and (ii) that for the remaining 60% of the cases where both visualizations are expected to be visually indistinguishable, a (random) half of the observations would also choose MDCIM to be more accurate. This results in an expected 70% of observers preferring MDCIM over sRGB for the iPad Air 2 at 1500 lux ambient lighting. Earlier we published the results of psychophysical tests where we tested (among others) the preference of observers on using color visualization on an iPad Air 2 display by the default sRGB color space or by the newly developed MDCIM method [2]. At an ambient illuminance level of 1500 lux, 66% of the visual observations the MDCIM method was preferred.

Table 3: Preference for visualization by display-specific MDCIM method versus device-independent sRGB color space for the iPad Air 2 display. Rows indicate level of ambient lighting.

	Preference for MDCIM model over sRGB color space	
	Predicted by new metric	Found experimentally [2]
600 lux	95%	98%
1000 lux	68%	50%
1500 lux	70%	66%
3000 lux	100%	87%

If we repeat this analysis also for the psychophysical data on three more levels of ambient lighting, we obtain the results shown in Table 3. A good alignment is found between the preferences predicted by the proposed metric and the experimental results from a visual test with observers. From these results we conclude that the proposed metric gives a good indication on differences between colors visualizations between different methods.

Reproducibility on three displays

We repeated the spectroradiometer measurements for all three displays. In all cases, the spectroradiometer measurements were executed with a different CS-2000A spectroradiometer (Konica Minolta) instrument than what was used in the original measurements. The second set of measurements was taken between two and seven years after taking the first measurements. For the iPad Air 2 and the iPad (2017), the displays that we used for the repeat measurements were the same as for the original measurements. For the iPad (2018), a different display was used, although obviously it was still an iPad (2018).

The resulting measurement data and derived model fit parameters are shown in Tables 4 and 5, which are to be compared with the results from the original measurements in Table 1 and 2.

For the iPad Air 2, the measurement values and optimized MDCIM model parameters from the first measurement (Table 1 and

2) and those from the repeat measurement (Table 3 and 4) show small differences. The relatively largest differences are for the maximum luminance of the blue channel (27.18 and 37.76 cd/m²,[2] respectively), but also for the other two channels the differences are several percent. For the MDCIM model parameters the optimized values for the original and the repeated dataset are slightly different as well. For example, γ = 2.35 (original data) and 2.34 (new data). Applying our metric to this case, we find that the old and the new data for the iPad Air 2 result in visually almost identical results. For 89% of the RAL colors the color difference is below the ΔE_{ab} = 4.5 threshold. The median color difference is CIEDE2000 = 1.0, the 90% percentile is CIEDE2000 = 2.2. We conclude that both datasets result in virtually identical color visualizations for the iPad Air 2. Similarly, also for the iPad (2017) we find visually almost identical color visualizations when using the older measurement data or the newer data. The color difference is below the ΔE_{ab} = 4.5 threshold for 95% of the RAL colors. The color difference is CIEDE2000 = 0.7 (median) and CIEDE2000 = 1.0 (90% percentile).

Table 4: Repeated measurements compared to Table 1.

	sRGB	iPad Air 2 v2	iPad (2017) v2	iPad (2018) v2
x Red	0.64	0.6428	0.6405	0.6403
y Red	0.33	0.3317	0.3295	0.3317
Y_{max} Red	21.26	83.84	102.62	115.33
x Green	0.30	0.2995	0.3009	0.2992
y Green	0.60	0.5987	0.6013	0.6029
Y_{max} Green	71.52	297.35	361.18	394.31
x Blue	0.15	0.1528	0.1569	0.1519
y Blue	0.06	0.0615	0.0606	0.0579
Y_{max} Blue	7.22	37.76	39.67	41.02
x Black	0.0	0.2549	0.2501	0.2571
y Black	0.0	0.2268	0.2085	0.2447
Y Black	0.0	0.45	0.48	0.57

Table 5: Best-fit values for MDCIM model parameters based on repeated measurements compared to Table 2.

	sRGB	iPad Air 2 v2	iPad (2017) v2	iPad (2018) v2
Red channel				
γ	2.4	2.34	2.26	2.31
d_0	0.04045	0.007843	0.007843	0.003922
k_1	0.947867	0.989572	0.990196	0.994907
k_2	0.055	0.010428	0.009804	0.005093
β	0.077399	0.010648	0.013629	0.004896
Green channel				
γ	2.4	2.34	2.31	2.32
d_0	0.04045	0.007843	0.007843	0.003922
k_1	0.947867	0.989572	0.989841	0.994831
k_2	0.055	0.010428	0.010159	0.005169
β	0.077399	0.010648	0.011839	0.004553
Blue channel				
γ	2.4	2.34	2.49	2.38
d_0	0.04045	0.007843	0.007843	0.003922
k_1	0.947867	0.989572	0.988451	0.994598
k_2	0.055	0.010428	0.011549	0.005402
β	0.077399	0.010648	0.006872	0.003646

Also for the iPad (2018) similar results are obtained. With 87% of the RAL colors resulting in a color difference below the ΔE_{ab} = 4.5 threshold the color visualizations when using the new versus the older set of measurement data are visually almost identical. The median color difference between both color visualizations for the 197 RAL colors is CIEDE2000 = 1.8, the 90% percentile at CIEDE2000 = 3.4. This shows that from the three displays, the results for the iPad (2018) are the least reproducible. For this display, for a minority of circa 10% of the colors the two visualizations result in visually well recognizable differences. A comparison of the data in Tables 1 and 4 reveals that the older dataset for the iPad (2018) was probably less accurate, since the maximum luminance levels of all three channels were measured to be lower than those for the iPad (2017), which is unexpected. Generally, newer versions of displays show higher display brightness. Table 4 shows that the new set of measurement data for the iPad (2018) results in higher display brightness for the iPad (2018), so this newer dataset is expected to be most accurate.

Quantifying differences between displays

Using the same metric, we can also quantify the visual difference when using the same RGB image on different displays. For example, if we compare the iPad Air 2 display with the iPad (2017) display, the color difference is below the ΔE_{ab} = 4.5 threshold for only 15% of the RAL colors. The median color difference is CIEDE2000 = 4.0, the 90% percentile is 5.8. This means that for a vast majority of the colors, these two displays result in visually different color visualizations. If we compare the iPad Air 2 display with the iPad (2018) display, the color difference is below the ΔE_{ab} = 4.5 threshold for 31% of the RAL colors, more than what we found for the iPad (2017) display. In this case the median color difference is CIEDE2000 = 3.6 and the 90% percentile 6.9. For most colors, both the iPad (2017) and the iPad (2018) show color visualizations that are visually different from colors displayed on the iPad Air 2.

But when comparing the iPad (2017) with the iPad (2018) display, we find different results. In this case, for 89% of the RAL colors visualized on these displays the color difference between the visualizations is below the ΔE_{ab} = 4.5 threshold. The median color difference is only CIEDE2000 = 1.3, and the 90% percentile is CIEDE2000 = 2.4. We can conclude that the color visualizations between iPad (2017) and iPad (2018) are only slightly different, and visually for circa 90% of the colors no color difference will be perceived between colors visualized on these displays.

Conclusions

In this conference proceeding we apply a display characterization model, the MDCIM model, on three different types of LCD display. We introduce a quantitative metric to analyze the visually perceived differences between colors visualized specifically on these three displays. We validate this metric by analyzing the perceived visual differences as scored by observers using an iPad Air 2 display under different ambient light conditions. Our results show that the metric calculations are well aligned with the visual data from this experiment. We use the new metric to investigate if a new set of spectroradiometer data from three different displays (iPad Air 2, iPad 2017 and iPad 2018) result in visually distinct color visualizations as compared to using an older set of spectroradiometer data from the same three different displays. Our results show that color visualizations are virtually identical for both datasets for the iPad Air 2 and iPad 2017. For the iPad 2018, in circa 10% of the colors a visually noticeable difference occurs. We showed that this was probably due to a measurement error in the older dataset for this display. Finally, we used the same metric to compare color visualizations on all three displays. This showed that color visualizations between iPad 2017 and iPad Air 2 are often visually different, with color differences larger than CIEDE2000 = 4.0 for 50% of the colors. In contrast, color visualizations between iPad 2017 and iPad 2018 are often visually identical, with color differences smaller than CIEDE2000 = 2.4 for 90% of the colors.

References

[1] International Electrotechnical Commission (IEC), 100/PT61966(PL) 34, International Standard, “Multimedia systems and equipment–colour measurement and management, part 2.1: default RGB colour space–sRGB” (1999), Amendment 1 (2003).

[2] E. Kirchner, I. van der Lans, F. Martínez-Verdú, E. Perales. "Improving color reproduction accuracy of a mobile liquid crystal display". J Opt Soc Am. A, 34, 101-110 (2017).

[3] E. Kirchner, I. van der Lans, E. Perales, F. Martínez-Verdú. "Improving color reproduction accuracy of an OLED-based mobile display". Color Res Appl., 43, 34–46 (2018). https://doi.org/10.1002/col.22148.

[4] R.S. Berns, R.J. Motta, M.E. Gorzynski, “CRT colorimetry, Part I: Theory and practice.” Color Res. Appl., 18, 299–314 (1993).

[5] R.S. Berns, “Methods for characterizing CRT displays.” Displays, 16, 173-182 (1996).

[6] German Institute for Quality Assurance and Labeling, RAL Classic, http://www.ral-farben.de/en/PRODUCTS-SHOP/RAL-CLASSIC/RAL-840-HR-primary-standards.html.

[7] J.B. Thomas, J.Y. Hardeberg, A. Trémeau. “Cross-media color reproduction and display characterization”. In: C. Fernandez-Maloigne (eds) Advanced Color Image Processing and Analysis (Springer, 2013). https://doi.org/10.1007/978-1-4419-6190-7_4.