This is a previous version of the article published in Journal of the Optical Society of America A. 2017, 34(1): 101-110. doi:10.1364/JOSAA.34.000101

Improving Color Reproduction Accuracy of LCD-based Mobile Displays

ERIC KIRCHNER,[1,*] IVO VAN DER LANS,[1] FRANCISCO M. MARTÍNEZ-VERDÚ,[2] ESTHER PERALES [2]

[1] *Color Research, BA Performance Coatings, AkzoNobel, Sassenheim, the Netherlands*
[2] *Department of Optics, Pharmacology and Anatomy, University of Alicante, Carretera de San Vicente del Raspeig s/n 03690, Alicante, Spain.*
**Corresponding author: eric.kirchner@akzonobel.com*

Received XX Month XXXX; revised XX Month, XXXX; accepted XX Month XXXX; posted XX Month XXXX (Doc. ID XXXXX); published XX Month XXXX

The default method for color representation on displays involves sRGB as device-independent encoding color space. For improving color reproduction accuracy, we develop a device-specific display characterization model for the Apple iPad Air 2. This is combined with an easy-to-implement new method to account for the influence of ambient ~~illuminance~~. The combination is called the Mobile Display Characterization and Illumination (MDCIM) Model for the iPad Air 2, representing modern LCD displays. Seven observers performed psychophysical tests at ambient illuminance levels from 600 to 3000 lx. They visually compared colors of calculated images with those of physical RAL samples. The MDCIM model achieves similar color reproduction accuracy as when using the default method involving sRGB encoding at 1000 lx, while considerably improving color accuracy at other illuminance levels. At 600 lx 98% of the observations prefer images directly generated with the MDCIM model over those created using the default method. The average color reproduction accuracy improves by two categories on a five-point scale. At 3000 lx, the percentage of colors that is represented at least reasonably well increased from 0% to 60%.

OCIS codes: *(330.1690) Color; (330.1715) Color, rendering and metamerism; (120.2040) Displays; (230.3720) Liquid-crystal devices.*

http://dx.doi.org/10.1364/AO.99.099999

1. INTRODUCTION

There are several applications that may benefit from higher color reproduction accuracy when displaying colors on a display. Examples of such applications are e-commerce of fashion and software applications for interior design. In such applications, it may not be sufficient to use a device-independent color space such as sRGB, which is the default encoding color space for non-high-end users of electronic displays. Instead, a device-specific display characterization model may be preferred.

In case the display is part of a mobile device, also ambient lighting conditions need to be considered for improving color reproduction accuracy. Mobile displays are designed to be used under a wide variety of ambient lighting conditions [1]. Even excluding outdoor conditions, illuminance levels for office applications already vary between typically 500 and 2000 lx [2]. It is well known that the ambient luminance level has a strong influence on the optimum color representation on a display.

For achieving good color reproduction accuracy, the mobile device should therefore use (i) a device-specific display characterization model and (ii) a method to account for ambient illuminance. In the current paper, we will develop a model for the Apple iPad Air 2 that accounts for both aspects, and that aims at an improved color reproduction accuracy as perceived by the user of the display.

The model developed here utilizes reflection data originating from spectrophotometers as device-independent colorimetric data. The model is not applicable for cases in which only color data is available from much more common image capturing devices, such as flatbed scanners or cameras. The new model also cannot be used for soft proofing tasks, where the target is to achieve a good color match between displayed images and printed images [3].

In this article we develop the Mobile Display Characterization and Illumination Model (MDCIM), specifically for use on mobile displays. The model is applied here for the case of the Apple iPad Air2 display, as a representative of liquid crystal display (LCD) technology. The MDCIM model exploits the fact that most mobile displays contain a light sensor which can be used to measure the ~~ambient luminance level~~. The new model also exploits the fact that users of a mobile display have hardly any possibility to change the display settings. This is in contrast to computer displays, which usually do allow users to change display ~~brightness,~~ contrast etc., often even without enabling software to determine those settings. Also, the number of different models of smartphones and tablets is much smaller than the number of combinations of computers and monitors, making the new model best suitable for mobile displays [4].

The display characterizing model for the Apple iPad Air 2 will be based on luminance measurements from the display. For quantifying the perceived color reproduction accuracy, we decided to use visual tests involving human observers in a range of different illumination conditions.

A number of relevant expressions for the display characterization model are briefly summarized in section 2A. In section 2B, we show how different ambient luminance levels are accounted for in the MDCIM model. The luminance measurements and the resulting display characterization model for the Apple iPad Air 2 are discussed in section 3A. The subsequent sections discuss the visual tests for assessing the perceived color reproduction accuracy. The results from the visual tests are discussed in section 4.

2. METHODS

2A. Display characterization model

Color measurements on physical samples are usually expressed in terms of colorimetric parameters such as CIELAB. From these, the tristimulus values X, Y, Z can be calculated. In order to be able to calculate parameters of an image representing the physical sample, the tristimulus values are usually converted into luminances Y_R, Y_G and Y_B of the red, green and blue channel of a display. The GOG-model from Berns et al. establishes this conversion, and was shown to be well applicable for CRT displays [5][6]. For LCD displays such as the Apple iPad Air 2, a modification is needed that will be discussed in section 3A.

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

In equation (1) the luminance of the red, green and blue channel is assumed to show the following functional dependence on the digital values d_R, d_G and d_B for each color channel (each scaled between 0 and 1):

$$\begin{pmatrix} Y_R \\ Y_G \\ Y_B \end{pmatrix} = \begin{pmatrix} (k_{1,R} d_R + k_{2,R})^{\gamma_R} \\ (k_{1,G} d_G + k_{2,G})^{\gamma_G} \\ (k_{1,B} d_B + k_{2,B})^{\gamma_B} \end{pmatrix} \quad if\ d_R > d_0 \quad (2a)$$

$$Y_R = \beta \cdot d_R \ if\ d_R \le d_0$$
$$Y_G = \beta \cdot d_G \ if\ d_G \le d_0$$
$$Y_B = \beta \cdot d_B \ if\ d_B \le d_0 \quad (2b)$$

The functions appearing in equation (2a) are the Tone Reproduction Curves (TRCs). In sRGB encoding space, it is assumed that the matrix M in eq. (1) is given by the following expression [7]:

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

and that the parameters gamma $\gamma = 2.4$, gain $k_1 = 1/1.055$ and offset $k_2 = 0.055/1.055$ are identical for all three channels. The threshold parameter d_0 is further discussed in the Appendix. The elements of matrix M as appearing in equation (3) are related to the chromaticity values (x, y) of the red, green and blue channel, and also to the maximum luminance Y_{max} of these channels [8][9]:

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

Therefore, the numerical values for matrix M correspond to a particular choice of values for the chroma and maximum luminance of the display channels. These display specifications underlying sRGB color space are summarized in Table 1.

Table 1. Display specifications assumed when sRGB color space was defined. These specifications include the chromaticity coordinates (*x*,*y*) and luminance maximum Y_{max} for the red, green and blue channel.

sRGB	*Red*	*Green*	*Blue*
x	**0.64**	**0.30**	**0.15**
y	**0.33**	**0.60**	**0.06**
Y_{max} (cd/m²)	**21.26**	**71.52**	**7.22**

The so-called display primaries defined in Table 1 were defined already in 1991 by the CCIR (Comité Consultatif International pour la Radio, now replaced by the International Telecommunication Union ITU) recommendation BT.709 for high definition television.

The values of the parameters k_1 (gain) and k_2 (offset) that are another part of the definition of sRGB color space are derived based on purely mathematical arguments, without considering the specifications of an actual display. This is shown in the Appendix.

Finally, the value of 2.4 for the gamma γ parameter in the definition of sRGB color space is partly a heritage from parametrizations dating back from before 1998 [7][10], and partly an optimization for a reference ambient illuminance level of 64 lx [7]. The international standard specifying sRGB color space remarks that the typical office or home viewing environment has an ambient illuminance level of 350 lx, i.e. exceeding 64 lx [7]. We note that the International Color Consortium recommends that for situations differing from the reference viewing environment, the value for the gamma parameter in the definition of sRGB color space should be modified [11].

Modern mobile displays are typically viewed under much brighter lighting conditions than 64 or 350 lx. Also, modern mobile displays have colorimetric specifications differing substantially from those listed in Table 1. For these reasons, although the use of sRGB color space is convenient as a device-independent encoding color space, improved color reproduction accuracy may be achieved by using a device-dependent display characterization model. Such improvements have been shown before for other, non-mobile LCD displays [12][13].

2B. The Mobile Display Characterization and Illumination Model

Inspired by a paradigm recently published to improve color reproduction accuracy for LCD displays in medical applications [14], but without aiming at specific display calibration methods for medical applications such as gray-scale standard display function (GSDF) [15], we derive the following method to achieve the same goal for mobile displays. We use equation (4) to account for displays with different colorimetric properties than those listed in Table 1. Note that the matrix elements defined in equation (4) result in tristimulus values in equation (1) that carry the dimension of luminance. For calculating colorimetric parameters such as CIELab values also the tristimulus values of reference white need to be expressed as luminance values, which requires their multiplication by the luminance L_{rw} of reference white:

$$L_{rw} = \frac{M_L}{\pi} = \frac{E\,\rho}{\pi} \tag{5}$$

where M_L is the luminous exitance of reference white, and where E is the ambient illumination level in lx [16]. Equation (5) also assumes that the reference white is a Lambertian scatterer, and in what follows we will also assume that it has ideal reflectivity $\rho = 1$. With equation (5) the illuminance of ambient lighting is accounted for in the MDCIM model.

We note that different approaches to account for ambient luminance have been proposed before. The most obvious choice would be to use the colour appearance model CIECAM02, as for example shown in Ref.[17], but CIECAM02 has been shown to be less useful for predicting the performance of mobile displays under a variety of illumination and surround conditions [18] (see also references cited in Ref.[19]). A refinement of CIECAM02 specific for mobile displays was recently proposed, but still leaves several areas needing improvement [19][18]. Here, we choose to develop a more easy-to-implement approach that also allows a more direct application of color difference equations than when using methods based on CIECAM02.

In preliminary tests of the MDCIM model we investigated the degree of adaptation of observers to ambient lighting. For observing images on mobile displays, we found that the best results were obtained by assuming observers to be completely adapted to the ambient lighting, and we will make the same assumption here as well. This adaptation state may be explained by noting that the physical dimensions of mobile displays of smartphones and tablets are relatively small, therefore having less impact on the adaptation state of the user than when watching television or monitor screens. Also, mobile displays are usually viewed for shorter time periods than displays of desktop computers and televisions. In the latter case, users are more often found to become visually immersed in the lighting environment of the display, and their adaptation state is more likely to shift.

3. EXPERIMENTAL

3A. Luminance measurements on Apple iPad Air 2

In this article, we investigate the proposed MDCIM model for a display representative of current LCD technology. For this, we use an Apple iPad Air2 device, which was commercially launched at the end of 2014.

For displays based on LCD technology, the luminance of the display remains visually noticeable even when displaying a purely black image. Therefore, for an accurate representation of the complete Tone Reproduction Curve, we needed luminance measurements on the display also for the low end of these curves. The measurement setup was chosen to be very similar to the common procedure of measuring colorimetric characteristics of electronic displays [20] [21].

Inside a completely dark room, display luminance and chroma measurements were obtained by mounting the display against the aperture of a CS-2000A spectroradiometer (Konica Minolta), aligned normal to the surface of the display. The display was set at maximum brightness, with automatic adjustment of display brightness switched off. A series of full screen images was shown on the display, and for each image the chromaticity coordinates x, y were measured, and also the luminance Y. As an example, the images to investigate the Tone Reproduction Curve for the red channel all have values G=0 and B=0, whereas R = 0, 5, 10, ..., 50, 75, 100, 125, ..., 200, 225, 255. In this way we obtain a good sampling of the complete Tone Reproduction Curve.

The chromaticity coordinates and maximum luminance values that we measured for the three channels, and also for black signal, are summarized in Table 2. These data show that the chromaticity coordinates for the iPad Air2 are generally well approximated by the values that are implicit in sRGB color space, as summarized in Table 1. The largest deviation is found for the blue primary, in agreement with the results from previous studies [4][22]. For the iPad Air2 we measure a white luminance of 407 cd/m², which is obviously much larger than the value of 80 cd/m² assumed in sRGB encoding color space.

Table 2. Values measured with the spectroradiometer on the Apple iPad Air 2, for chromaticity coordinates (x,y) and luminance maximum Y_{max} for the red, green and blue channel.

iPad Air 2	*Red*	*Green*	*Blue*	*Black*
x	0.6421	0.3071	0.1527	0.2458
y	0.3264	0.6079	0.0489	0.2085
Y_{max} (cd/m²)	80.89	307.2	27.18	0.3647

Table 2 also shows that a completely black image displayed on the iPad Air2 has a luminance of 0.3647 cd/m², due to the nonzero transmission of backlighting in the LCD display. To account for this black level we modify equation (1) as follows [12]:

$$\begin{pmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{pmatrix} = M \begin{pmatrix} Y_R \\ Y_G \\ Y_B \end{pmatrix} \tag{6}$$

where X_0, Y_0 and Z_0 are the tristimulus values of a completely black image. These values can be easily calculated from the measurement data in the last column in Table 2. So, $X_0 = (x_0/y_0)Y_0$ and $Z_0 = (1 - x_0 - y_0)(Y_0/y_0)$. With equation (6) one type of flare, due to light leaking through the LCD when displaying dark colors, is accounted for [19].

Figure 1 shows that the measurement data for the Tone Reproduction Curves of the iPad Air 2 are represented quite well by the GOG model, except for the values at low luminance values. The same figure also shows that the complete curves are represented very well if we use the black level correction from equation (6), combined with a different set of values for the parameters in the model. The best fit of the data was found by choosing d_0 = 2/255 = 0.007843 and gamma γ = 2.35. We note that in several recent studies on LCD displays the low luminance part of tone reproduction curves and the black level (equation 6) are not accounted for, thereby affecting the resulting values for gamma [23][24].

From these values, the values of all other parameters of the display characterization model can be calculated mathematically as explained in section 2. The resulting parameter values are listed in Table 3, and compared with the corresponding values when using sRGB color space.

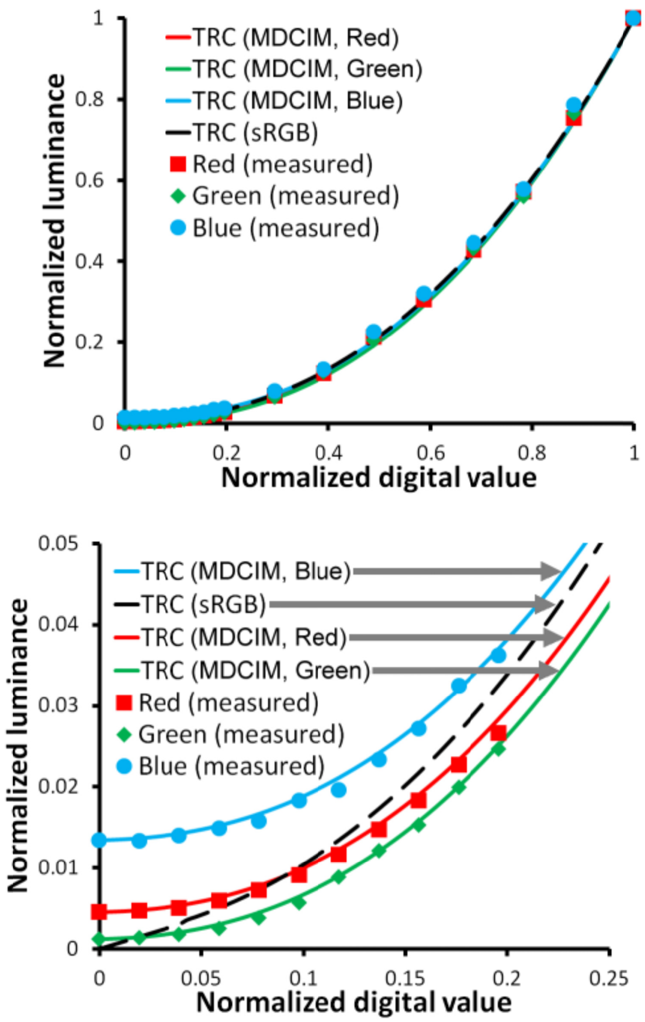

Fig. 1. Spectroradiometer measurement data for the Tone Reproduction Curves of the Apple iPad Air2, for the red, green and blue channel (markers). Solid lines represent optimized fits of the measurement data using the functions described in the text. The dashed line refers to the corresponding curve when assuming sRGB color space.

Table 3. Values for display characterization model for Apple iPad Air 2, derived from the measurement data shown in Figure 1.

	iPad Air 2	*sRGB*
γ	2.35	2.40
d_0	0.007843	0.04045
k_1	0.989523	0.947867
k_2	0.010477	0.055
β	0.010443	0.077399

The optimum value of γ= 2.35 for the parameter gamma that we found is close to the value of 2.4 that is assumed in the derivation of sRGB color space. However, the values of the other parameters have changed considerably.

Based on the data from the spectroradiometer measurements on the display of the iPad Air 2, the values for the elements of matrix M that appear in equation (3) change considerably. Since often the inverse of this matrix is used in order to calculate R, G, B values, we will illustrate this by showing the values of the inverse matrix M^{-1}. For the iPad Air 2 display we find the following expression:

$$M^{-1}_{MDCIM} = \begin{pmatrix} 0.8415 & -0.4061 & -0.1361 \\ -0.2222 & 0.4356 & 0.0158 \\ 0.0067 & -0.0350 & 0.2262 \end{pmatrix} \quad \textbf{(7)}$$

The differences with the corresponding values from sRGB color space [7] are obvious:

$$M^{-1}_{sRGB} = \begin{pmatrix} 3.2406 & -1.5372 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{pmatrix} \quad \textbf{(8)}$$

3B. Visual set-up and samples

We tested the color reproduction accuracy when using the default method (which involves the device-independent encoding color space sRGB) and when using the device-dependent MDCIM model, by conducting a psychophysical experiment. We carried out visual tests in a room that allowed adjusting illumination levels at 600, 1000, 1500 and 3000 lx, in a randomized order that was different for each observer. This is illustrated in Figure 2. The light spectrum in the room represents day light (D65). Seven observers (5 males, 2 females, age between 35 and 55) participated in the visual experiments. All observers have normal color vision as confirmed by the Ishihara color vision test and the Farnsworth-Munsell 100 hue test.

For this test, we used 35 low-gloss RAL colors, as listed in Table 4. This selection includes samples from all main chromatic and achromatic color categories. We included four samples for each of the six chromatic color categories: yellow, orange, red, purple, blue and green. For each color category, both lighter and darker colors were included, and both large and small saturation. For the category of achromatic colors, we included 11 samples, since in previous studies we found that the accuracy of color reproduction is particularly challenging for this category [8]. We used samples from the RAL Classic collection, developed by the German Institute for Quality Assurance and Labeling (RAL) [25][26]. We chose samples from the RAL840HR collection because their low gloss levels make the color less strongly depending on observation angle, resulting in easier color evaluations.

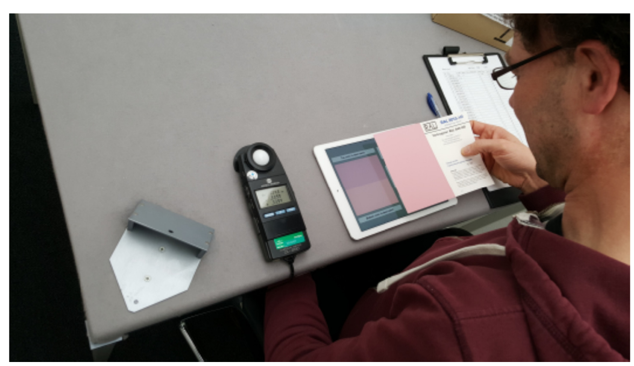

Fig. 2. Observer assessing accuracy of color representation on the iPad Air 2.

Spectral reflectance data were measured for all samples, using a multi-angle spectrophotometer (BYK-Mac from BYK-Gardner) at the 110° aspecular angle. This angle was chosen to minimize the influence from direct surface reflection (gloss), and for this sample set checked to be very similar to d/0 spectral reflectance data.

For four samples, we found that due to fluorescence the reflection data were less accurate. For these samples, we substituted spectral reflectance data obtained with a Datacolor SF600 instrument, including a UV-filter to suppress fluorescence. Using Specular Component Excluded data, we ensured the best compatibility with the spectral reflectance data of the other samples.

Table 4. Selection of RAL colors used in the visual experiment.

Category	RAL number	Category	RAL number
Yellow	1013	Green	6019
	1016		6021
	1020		6026
	1024		6029
Orange	2000	Achromatic	7006
	2001		7013
	2004		7021
	2008		7022
Red	3003		7023
	3012		7035
	3015		7040
	3017		7044
Purple	4004		7047
	4005		9010
	4008		9011
	4009		
Blue	5002		
	5010		
	5012		
	5024		

3C. Images and test procedure

The observer is given an observer-specific list with the RAL color names in a randomized order specific for each observer. The ambient illuminance level is adjusted until the required setting is reached, as controlled with a chromameter positioned next to the tablet. Apart from the four sessions at different illuminance levels, also a fifth session was held that repeated a randomly chosen earlier session. For one observer, a sixth session was held, also being a repeat session. In this way, we collected data for seven observers viewing 35 samples under four different illuminance levels, i.e. 980 assessments. Apart from that, we collected repeatability data from 280 assessments.

The session-specific list with RAL sample names also contained a code that varies for each RAL color name, and that specified which of two pre-loaded PNG images needed to be displayed on the iPad Air 2. As illustrated in Figure 3, each image shows two color patches, one calculated using the default method (involving sRGB color space), the other with the MDCIM model. Since the code differs for each sample, observers were not able to determine which patch was generated by which method. The color patches were limited in size and did not cover the full display, because the display has a relatively large dependence of displayed color on viewing angle.

In the images, the background color was calculated as to imitate the color of the cloth on the table where the visual experiment was performed. This was done to reduce the effect of simultaneous contrast.

According to the instructions the iPad Air 2 was put flat on the table, and viewed from a straight angle. The physical RAL sample was held next to the display, making it possible to directly compare the color on the tablet display with the color of the physical sample (cf. Figure 2). Observers were advised to slightly rotate the physical samples in order to better determine their colors. The quality of the color match between physical sample and displayed image was assessed, as if the displayed image would be used to indicate the color of a wall paint to a consumer.

The observers were instructed to note down whether the top image or the bottom image was perceived to be most similar in color to the physical sample. The observer also needed to note down a visual score for both images, representing the perceived quality of the color match between image and physical sample. The visual scores are given on a scale between 0 and 5, with steps of 0.25. The descriptions for the scores are shown to the observer, and reproduced in Table 5. A score of 0 refers to the situation where no or hardly any color difference can be seen between the physical sample and the displayed image. The perceived quality of the color match gradually becomes worse for larger values of the visual score. A score of 3 indicates that the perceived color difference becomes doubtful, and at a value of 4 the match is judged to be not correct. This procedure was repeated for all 35 RAL samples.

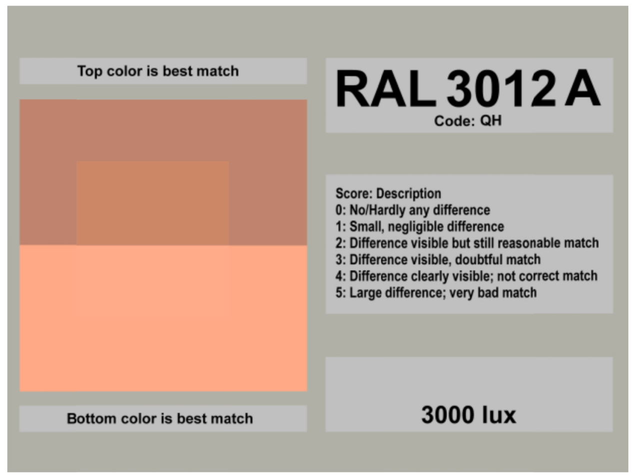

Fig. 3. One of the images used for the visual test.

Table 5. Visual scores and descriptions used in the visual test.

Visual score	Description
0	No / hardly any difference
1	Small, negligible difference
2	Difference visible but still reasonable
3	Difference visible, doubtful match
4	Difference clearly visible; not correct match
5	Large difference; very bad match

3D. Gamut mapping

Just like any other method that aims at representing colors on a display, there are situations in which the calculated values of the R, G and B parameters fall outside the numerical range that the display can handle (0-255 for 8-bit representation). For these out-of-gamut cases, we used two different gamut mapping methods. Many other methods have been proposed in the past [27][28], but since the aim of this study was not to investigate various gamut mapping techniques, we limited ourselves to two relatively simple techniques.

The first technique that we used is PHL-mapping, i.e. preserving both hue angle and lightness. With this technique, the hue angle and lightness of a color are kept constant whereas the chroma is reduced until all calculated RGB-values reach the required numerical range. Good results have been reported for this technique [29]. Similarly, in PHC-mapping the hue angle and chromaticity values are both kept constant, whereas the lightness value is increased or decreased until the color is no longer out-of-gamut.

For all colors in the test that are out-of-gamut, we tested both the PHL and the PHC mapping method. The method for which the color needed the smallest adjustment in terms of calculated color difference was selected.

When using the default method involving sRGB encoding color space, 5 from the 35 RAL colors were predicted to be out-of-gamut: RAL 2000, 5002, 5010, 6026 and 6029. In all 5 cases, the procedure just described resulted in PHL-mapping providing the best solution. We were interested if these 5 colors would also be out-of-gamut when using the MDCIM model, so we kept these colors in the investigation. When using the MDCIM model in the case of 600 and 1000 lx illumination, the same 5 colors were predicted to be out-of-gamut, and the same was also the case for RAL 5012. For all six cases the problem was best solved using PHL-mapping.

For the MDCIM model at 1500 lx illumination, one more color became out-of-gamut: RAL2008. In this case, for two colors the procedure outlined above resulted in a preference for PHC-mapping. This happened for RAL2008 and RAL 5002. Both colors are characterized by not only high or low lightness values ($L^*>60$ or $L^*<25$), but also by large chromaticity values. Apparently this combination makes that the aim to reduce chromaticity until the color is no longer out-of-gamut is only reached by largely distorting the color, making a change in lightness for achieving the same preferable.

For the MDCIM model at 3000 lx illumination, a total of 20 colors becomes out-of-gamut. This large number is obviously caused by the display being not bright enough to match many surface colors under such bright lighting conditions. Indeed, in 15 from the 20 cases, the procedure outlined above identified the PHC-method as preferable, showing that it is mainly the lightness dimension which made the display matching problematic. The remaining five colors that are out-of-gamut at 3000 lx illumination are exactly the same five colors that were also out-of-gamut at 600 lx. For those five colors, PHL mapping again provided the best solution.

4. RESULTS

4A. Results on repeatability

As described in section 3, 280 from the 980 assessments (28%) were repeated in separate experimental sessions. In the repeated sessions, the order of samples and also the relative position of the image representations generated using the default method (involving sRGB encoding color space) and the MDCIM model were independently randomized. In this way, we made sure that observers were not aware which image was calculated by which method, and also they were unlikely to remember their previous assessments.

The first part of each assessment refers to the preference about the color of which of two images best matches the color of the physical RAL sample. After ending all visual tests, the identification of all images with either default method (involving sRGB encoding color space) or with the MDCIM model was made. In this way we found that in the repeated session the preference for images created by the default method or by the MDCIM model was the same as in the original session in 91.4% of the cases. This shows that the preference data are highly repeatable.

The second part of each assessment consists of visual scores for the color match with the image created using the default method that involves sRGB encoding color space (which we will call default images), and with the image created by the MDCIM model (MDCIM images). For the default images, the average absolute difference between the visual score during repeat session and original session is 0.44. For the MDCIM images, the corresponding difference is 0.64. Since the visual scores are given on a five-point scale, as shown in Table 5, these values for the average absolute difference can be considered to represent a good repeatability.

The values for the repeatability reported here are important as a reference. When further analyzing the results from the visual tests, the focus should be on effects that are larger than this repeatability error.

4B. Results on preference

We analyzed the preference for MDCIM images over default images. The results are shown in Table 6. The percentage of assessments favoring MDCIM images over default images is found to be lowest at an illuminance level of 1000 lx. This correlates with the result that the color reproduction accuracy obtained when using sRGB color space is largest at 1000 lx, as shown below. According to Table 6, at 1000 lx ambient lighting there is no preference for MDCIM images over default images, because 50% of the assessments prefer either method.

Table 6. Preference of observers for MDCIM images over default images, when using the iPad Air 2 under ambient illuminance levels of 600, 1000, 1500 and 3000 lx.

Ambient illuminance (lx):	600	1000	1500	3000
Percentage preferring MDCIM	98%	50%	66%	87%
Percentage where 6 from 7 observers prefer MDCIM	100%	43%	26%	77%
Percentage where the majority of observers prefers MDCIM	100%	46%	80%	94%

At other illuminance levels of ambient lighting, the situation is very different. At 600 lx, the preference for MDCIM images has increased to 98%, which is almost unanimously. At 1500 lx, the preference of 66% shows that the number of samples where the MDCIM model is preferred is two times larger than the number for which default images are preferred. At 3000 lx, the preference increases to 87%, showing that for the vast majority of samples the MDCIM model produces images with an improved color reproduction accuracy.

The trends just discussed are not limited to average score values. Similar trends are found when we examine the assessments in more detail. As an example, Table 6 also shows the percentage of samples for which the majority of the seven observers prefers the representations calculated with the MDCIM model over those produced by using sRGB color space. This percentage is 46% for an illuminance level of 1000 lx, again signifying that at 1000 lx both types of images produce similar color reproduction accuracy. However, with ambient lighting dropping to 600 lx this percentage reaches the full 100%. At increased luminance levels of 1500 and 3000 lx, the corresponding percentage values are 80% and 94%. This confirms that for ambient lighting levels other than 1000 lx, the MDCIM model represents a substantial improvement in color reproduction accuracy.

4C. Color reproduction accuracy of default images

Apart from preference, which tests the relative color reproduction accuracy of MDCIM images versus default images, we also tested the absolute color reproduction accuracy of both types of images, as visually perceived by the observers. Observers gave visual scores to quantify the color accuracy of displayed images when visually compared to the corresponding physical RAL sample (cf. Table 5). If we average the visual scores over all samples when using default images, we find the average scores reproduced in Table 7.

For ambient lighting at an illuminance level of 1000 or 1500 lx, Table 7 shows that the average visual score is 3.6. According to the descriptions of the visual scores in Table 5, the average color reproduction accuracy of the images is therefore assessed to lie between "difference visible, doubtful match" and "difference clearly visible, not correct match". This shows that on average, default images do not provide good color reproduction accuracy at 1000 and 1500 lx.

The visual scores become worse at illuminance levels of 600 and 3000 lx. The corresponding average visual scores are 4.6 and 4.4, respectively. This qualifies the color reproduction accuracy as "not correct", and frequently even becoming "very bad".

A different way to analyze the visual data is to calculate the percentage of samples for which the average visual score is smaller than 3, i.e. for which the color reproduction accuracy was assessed to be reasonable or better. Table 7 shows that this percentage is only 34% for default images for the sessions held under 1000 lx lighting. At 1500 lx, the percentage becomes even worse at 17%. At 3000 lx, and also at 600 lx, the corresponding percentage is 0%.

Similar trends are found by determining the percentage of samples with an average visual score of 4 ("Difference clearly visible, not correct match") or worse. This percentage is 46% at 1000 lx ambient lighting, and 23% at 1500 lx. At 3000 lx it is 83%, and at 600 lx it is even 94%.

Our results confirm that the color reproduction accuracy of default images is not good, especially at 600 (typical for indoor conditions) and 3000 lx (typical for outdoor conditions) ambient lighting.

Table 7. Visual scores for default images, observed under ambient illuminance levels of 600, 1000, 1500 and 3000 lx.

Ambient illuminance (lx):	600	1000	1500	3000
Average visual score	4.6	3.6	3.6	4.4
Average score > 4	94%	46%	23%	83%
Average score < 3	0%	34%	17%	0%

4D. Accuracy of MDCIM images vs default images

Table 8 shows that at 1000 lx ambient lighting, the average visual score for MDCIM images is 3.6. This is equal to the value that we found for default images at 1000 lx lighting (cf. Table 7). Below, we will see that at other illuminance levels MDCIM images do represent an improvement in color reproduction accuracy, but this is apparently not the case at an illuminance level of 1000 lx. This may be understood by noting that according to the results obtained in section 4C, default images seem to produce the most accurate color reproduction at 1000 lx ambient lighting. This is possibly due to a fortunate cancellation of errors at that illuminance level. An ambient illuminance of 1000 lx is much brighter than what was assumed in the derivation of sRGB color space, but this is roughly compensated by the display being much brighter than what was assumed in the same derivation.

When reducing the illuminance level to 600 lx, the average visual score for MDCIM images becomes 2.6. This makes the color reproduction accuracy a full 2 units better than the value of 4.6 that we found for default images (cf. Table 7). Also for illuminance levels of 1500 and 3000 lx we find that the average visual score is greatly improved when using MDCIM images, with average improvements of 0.51 and 1.42 units on the five-point scale, respectively.

A different way to express these improvements is to calculate again the percentage of samples for which the average visual score is smaller than 3, i.e. for which the color reproduction accuracy was assessed to be reasonable or better. This percentage is 26% for MDCIM images at 1000 lx, which is similar (or even slightly worse) than when using default images at 1000 lx. But at 600, 1500 and 3000 lx this percentage is 83%, 49% and 60% for MDCIM images, respectively. The corresponding numbers for default images are 0%, 17% and 0%. This confirms that for a large majority of samples MDCIM images represent a substantial improvement in color reproduction accuracy at 600, 1500 and 3000 lx ambient lighting.

Table 8. Visual scores for MDCIM images, observed under ambient illuminance levels of 600, 1000, 1500 and 3000 lx.

Ambient illuminance (lx):	600	1000	1500	3000
Average visual score	2.6	3.6	3.0	2.9
Average score > 4	0%	34%	3%	9%
Average score < 3	83%	26%	49%	60%

4E. Auto-brightness

Most mobile displays offer the user an option to make the ~~brightness~~ of the display automatically depend on the ~~illuminance~~ its sensors detect. In the iPad Air 2, this option is called Auto-brightness. It may be a good idea if manufacturers like Apple would couple the MDCIM model to the Auto-brightness function on these devices, using input from the light sensors.

The software algorithms currently used for auto-~~brightness~~ have not been published, so it is not possible to compare it directly with the MDCIM model. However, it is straightforward to prove that the MDCIM model does correct for color reproduction accuracy errors on the iPad Air 2, also when Auto-brightness is switched on. As an example, several reviewers of the iPad Air 2 have noted that when showing pure white its display looks "slightly too blue" [30][22]. This affects the color reproduction accuracy of all low saturated colors [22]. A blue cast has also been noticed for dark images and for bright ambient light conditions [22]. This shift to blue is clearly visible also when auto-brightness is switched on.

The MDCIM model compensates for the blue cast. For example, according to the MDCIM model an almost white sample with colorimetric parameter values L^*=90, $a^*=b^*$=0 observed under 600 lx is represented by R=164, G=163, B=154. This shows that the blue channel is much stronger suppressed than when using only sRGB color space, which uses $R=G$=226 and B=225 to show the same color.

The blue shift is also clear from the spectroradiometric measurements we have done on the iPad Air 2 display. For a purely white image, we find chromaticity coordinates values x=0.3045 and y=0.3168. These values are clearly shifted towards the blue direction, with respect to the D65 and D50 whitepoints. Also for dark images a clear shift to blue is noticeable on the iPad Air 2 display. The spectroradiometer data show that for dark images there is a considerable leakage of blue light, even when for example a dark red image with R=20, $G=B$=0 is displayed. Also visually such an image looks purple rather than red. This observation is confirmed by Figure 1b, showing that this display transmits a considerable amount of blue light also for dark images. In the MDCIM model, these effects are accounted for by the way in which the spectroradiometer data are used to derive the model parameters.

5. CONCLUSIONS

In this article we developed a new method for color encoding and representation, aiming specifically at mobile displays. In the derivation of this MDCIM model we account for the variety of ambient illuminance levels under which mobile displays are typically used. We also take into account the colorimetric properties of the display, exploiting the fact that variations in colorimetric properties within models of smartphones and tablets are relatively small [4].

From spectroradiometer measurement data to characterize the key colorimetric properties of a display, we constructed a display-specific matrix for converting tristimulus values into RGB-values (equation 4). We also used spectroradiometric data to find the display-specific tone reproduction curves (equation 2). From optimized values for gamma γ and threshold value d_0, values for all other parameters are produced by utilizing the mathematical relations between the underlying parameters (appendix). The MDCIM model takes into account the illuminance level of ambient light (equation 5).

Independent of the model training data, we used visual tests to investigate if the MDCIM model produces color representations on a mobile device that are more accurate than when using the default method that involves device-independent sRGB color space, when the colors of the resulting images are visually compared to the colors of physical RAL samples. This was tested for an iPad Air 2 display, representing modern LCD-based mobile displays. The experiments were performed under a range of ambient illuminance levels varying from 600 to 3000 lx.

We found that at an illuminance level of 1000 lx, the MDCIM model results in a color reproduction accuracy that is very similar to what is obtained when using the default method (involving sRGB color space). The color reproduction accuracy obtained by using images created by the default method (involving sRGB encoding color space) was assessed to be best at 1000 lx. At 1000 lx the display luminance being much brighter than what is assumed in the derivation of sRGB color space is roughly compensated for by the ambient illuminance level being much brighter than what is assumed in the derivation of sRGB color space. However, our results show that at other illuminance levels a much better visualization of colors is possible than what is achieved when using default images.

At 600, 1500 and 3000 lx we did find a clear preference for the MDCIM model. Especially at 600 and 3000 lx, the preferences reach values well above 90%. We have seen that for a large majority of samples, MDCIM images represent a substantial improvement in color reproduction accuracy over default images at 600, 1500 and 3000 lx ambient lighting. For example, the percentage of samples showing at least reasonable color reproduction accuracy improves from 0% to 83% at 600 lx, and from 0% to 60% at 3000 lx.

For future work, we will investigate the effect of light reflection directly from the display face, which represent an external type of flare, different from the internal flare accounted for by equation (6) [19].

6. APPENDIX

In this appendix we will show how to derive the numerical values for the gain and offset parameters that are used in the derivation of sRGB color space.

When *R*, *G* and *B* values in an image ~~tend to~~ zero, the corresponding values of d_R, d_G and d_B approximate zero as well. According to equation (2) the resulting luminance of the channels ~~tends to~~ zero as well only if the offset parameters k_2 are equal to zero. However, the model would then not show sufficient flexibility anymore to describe measurements of the Tone Reproduction Curves of a particular display. We therefore need an alternative solution: the exponential form of the tone reproduction curves as expressed in equation (2) is only assumed to be valid for values of d_R, d_G and d_B that are larger than a threshold value d_0. For smaller values of d_R, d_G and d_B a linear approximation is assumed. For example, for the red channel:

$$Y_R = \beta \cdot d_R \text{ if } d_R \le d_0 \qquad \textbf{(A1)}$$

The numerical value of the linear coefficient β will be determined below. The linear interpolation of the luminance values *Y* between $d=0$ and $d=d_0$ makes the tone reproduction curve becomes a function of *d* that has an inverse for integer *R*, *G* and *B* values. This is important for applications where the inverse of equation (1) is needed, i.e. calculating *R*, *G* and *B* values from tristimulus or CIELab values.

For a white pixel with $R=G=B=255$ (i.e., for $d_R = d_G = d_B = 1$) one should find a total luminance $Y = Y_{R,max} + Y_{G,max} + Y_{B,max}$. From this it follows that:

$$k_2 = 1 - k_1 \qquad \textbf{(A2)}$$

The parts of the tone reproduction curve below and above the threshold need to be smoothly connected. This is only the case if the corresponding function values and their derivatives of these functions are continuous. Continuity of the functions defined in equations (2) and (A1) is ensured if

$$\left(k_1 d_0 + k_2\right)^{\gamma} = \beta\, d_0 \qquad \textbf{(A3)}$$

whereas demanding continuity of their derivatives results in

$$k_1\, \gamma \left(k_1 d_0 + k_2\right)^{\gamma-1} = \beta_0 \qquad \textbf{(A4)}$$

Substituting equation (A2) into (A3) and (A4) we find that

$$k_1 = \frac{1}{1+(\gamma-1)\, d_{0\ 0}} \qquad \textbf{(A5)}$$

$$k_2 = \frac{(\gamma-1)\, d_0}{1+(\gamma-1)\, d_0}\ {}_0 \qquad \textbf{(A6)}$$

$$\beta = \frac{1}{d_0} \left(\frac{\gamma\, d_0}{1+(\gamma-1)\, d_0}\right)_0 \qquad \textbf{(A7)}$$

For 3x8 bits color depth, the threshold d_0 can be chosen to correspond to *R*, *G*, *B* values of 10. Then $d_0 = 10/255 = 0.0392$. For the given value of gamma $\gamma = 2.4$, equations (A5), (A6) and (A7) produce numerical values for the gain parameter $k_1 = 1/1.055$, the offset parameter $k_2 = 0.055/1.055$, and the linear coefficient $\beta = 1/12.92$.

These numerical values are identical to the values defined in sRGB color space [7]. This shows that these parameter values are not based on an analysis of technical properties of displays, but on mathematical and pragmatic arguments only.

References

1. W. Cummings and T.G. Fiske, "Simplified Ambient Performance Assessment for Mobile Displays Using Easy Measurements," Soc. Inf. Display Symposium Digest **45**, 528-531 (2014).
2. CIE International Commission on Illumination, Specification of colour appearance for reflective media and self-luminous display comparisons (CIE 195:2011).
3. ISO 14861:2015, Graphic technology -- Requirements for colour soft proofing systems.
4. S.J. Dain, B. Kwan, and L. Wong, "Consistency of color representation in smart phones," J. Opt. Soc. Am. A **33**, 300-305 (2016).
5. R.S. Berns, R.J. Motta, and M.E. Gorzynski, "CRT colorimetry, Part I: Theory and practice," Col. Res. Appl. **18**, 299–314 (1993).
6. R.S. Berns, "Methods for characterizing CRT displays," Displays **16**, 173-182 (1996).
7. International Electrotechnical Commission (IEC) 61966-2-1, International Standard, Multimedia systems and equipment – Colour measurement and management - Part 2.1: Default RGB colour space – sRGB (IEC, Geneva, 1999). Amendment 1 (2003).
8. E.J.J. Kirchner, I.B.N. van der Lans, and S.L. Njo, "Perception-based Accurate Colour Calibration of Electronic Displays," in *Proceedings 12th Congress of the Int'l Colour Assn.* (AIC, Newcastle, 2013), pp. 1077-1080.
9. R.S. Berns, *Billmeyer and Saltzman's Principles of color technology* (John Wiley & Sons, 2000). Third edition.
10. C. Poynton, *Digital video and HDTV - Algorithms and Interfaces*, The Morgan Kaufmann Series in Computer Graphics (Morgan Kaufmann, 2003) p. 267.
11. International Color Consortium, A standard default color space for the internet: sRGB, http://www.color.org/srgb.xalter, accessed: March 2016.

12. M.D. Fairchild, D.R. Wyble, "Colorimetric characterization of the Apple Studio Display (Flat Panel LCD)". Munsell Color Science Laboratory Technical Report 1998.
13. E.A. Day, L. Taplin and R.S. Berns, "Colorimetric characterization of a computer-controlled liquid crystal display", Col. Res. Appl. **29**, 365-373 (2004).
14. L.D. Silverstein, S.F. Hashmi, K. Lang, and E.A. Krupinski, "Paradigm for achieving color-reproduction accuracy in LCDs for medical imaging," Journal of the Soc. Inf. Display **20**, 53-62 (2012).
15. G.C. Kagadis, A. Walz-Flannigan, E.A. Krupinski, P.G. Nagy, K. Katsanos, A. Diamantopoulos and S.G. Langer, "Medical imaging and their use in image interpretation", Radiographics **33**, 275-290 (2013).
16. A. Ryer, *Light measurement handbook* (International Light., 1998).
17 X. Gao, E. Khodamoradi, L. Guo, X. Yang, S. Tang, W. Guo and Y. Wang, "Evaluation of colour appearances displaying on smartphones," in *Proceedings of the Midterm meeting of the Int'l Colour Assn.* (AIC, Tokyo, 2015), pp. 539-544.
18. Y.K. Park, C.J. Li , M.R. Luo, Y. Kwak, D.S. Park and C. Kim, "Applying CIECAM02 for mobile display viewing conditions", *Proceedings IS&T/SID 15th Color Imaging Conference* (IS&T, Springfield, VA, 2007) 169-173 .
19. Y.K. Park, M.R. Luo, C.J. Li and Y. Kwak, "Refined CIECAM02 for bright surround conditions", Col. Res. Appl. **40**, 114-124 (2015).
20. Display Mate, Testing and Evaluating Displays and Projectors, http://www.displaymate.com/testing.html, accessed: March 2016.
21. Cnet, How we test tablets, http://www.cnet.com/news/cnet-tablet-battery-life-results/, accessed: March 2016.
22. R..M. Soneira, iPad 2014 Display Technology shoot-out, http://www.displaymate.com/iPad6_ShootOut.htm, accessed March 2016.
23. T.M. Aslam, I.J. Murray, M.Y.T. Lai, E. Linton, H.J. Tahir, and N.R.A. Parry, "An assessment of a modern touch-screen tablet computer with reference to core physical characteristics necessary for clinical vision testing," J. R. Soc. Interface **10**, 20130239.1-4 (2013).
24. L. Bodduluri, M.Y. Boon, and S.J. Dain, "Evaluation of tablet computers for visual function assessment," Behav. Res. Methods (2016, in press).
25. Gall L, "Das RAL Design System-ein neuer Farbenatlas nach CIELab," Farbe Lack. **11**, 863–869 (1992).
26. 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, accessed: March 2016.
27. B.H. Kang, M.S. Cho, J. Morovic, and M.R. Luo, "Gamut compression algorithm development on the basis of observer experimental data," in *Proceedings IS&T/SID Eighth Color Imaging Conf.* (IS&T, Springfield, VA, 2000), pp. 268-272.
28. J. Morovic and M.R. Luo, "The fundamentals of gamut mapping," J. Imaging Sci. Technol. **45**, 283-290 (2001).
29. E.A. Fedorovskaya, H. De Ridder, and F.J.J. Blommaert, "Chroma variations and perceived quality of color images of natural scenes," Color Res. Appl. **22**, 96-110 (1997).
30. D. Schmidt, Apple iPad Air 2 (A1567 / 128 GB / LTE) Tablet Review, http://www.notebookcheck.net/Apple-iPad-Air-2-A1567-128-GB-LTE-Tablet-Review.129396.0.html, accessed March 2016.