Colour appearance modelling between physical samples and their representation on large liquid crystal display

Chrysiida Kitsara, M Ronnier Luo, Peter A Rhodes and Vien Cheung
School of Design, University of Leeds, Leeds, UK

Abstract

The use of large displays for purposes of colour communication is becoming increasingly popular and the need for high-fidelity reproduction of appearance is becoming even more demanding. In this work, the colour appearance of the ColorChecker chart was matched on a large liquid crystal display (LCD) and a comparison between the physical colours and the displayed image was made. Colour definition and colour perception spaces were used to derive appearance models that define the difference between the digital and physical stimuli. The procedure was repeated using a selection of coloured garments as stimuli. The results revealed a good agreement in the defining the appearance difference between digital and physical stimuli. In both cases the difference in lightness between the two media was found to be responsible for the variation in matching. This outcome was used to develop a colour-rendering chain for the display. The use of appearance modelling in the digital image reproduction chain enhanced the appearance of solid paint-coated surfaces and dyed-garment images.

Introduction

Both cross-media reproduction and appearance modelling have been active fields of research for more than 15 years [1-4]. The most recent related study involving comparing physical stimuli with those on a display was carried out by Oicherman *et al*. to examine the effect of observer metamerism in colour matching between physical stimuli and small displays [5].

This study concerns with the matching took place on a large rather than small display. Also, the physical samples in this study were viewed under studio lighting instead of inside a viewing cabinet. Digital images of coloured stimuli were shown on an LCD. Observers modified colour to achieve a match with physical stimuli. By combining physical measurements between both stimuli and the use of colour specification and perception spaces, appearance modelling was attempted.

The outcome of this study aims to define an appearance model that will aid achieving high-fidelity matches of images of different materials shown on a large LCD.

Measurement Device

All of the surface radiometric measurements were conducted using a Minolta CS-1000 tele-spectroradiometer (TSR).

Stimuli

Experimental planning began by selecting physical stimuli. For Experiment I, a set of uniform colours was sought. A number of standardised commercial calibration targets were considered and eventually the X-Rite ColorChecker was chosen.

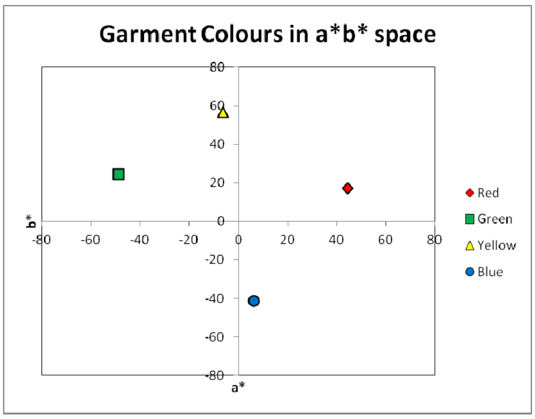
The GretagMacbeth ColorChecker chart consists of a 4×6 array of 50×50mm coloured patches, which are covered by a thick black frame. This size is sufficient to make patches discernible at a distance, which is critical for the colour-matching procedure as well as distant spectroradiometric measurement. The patches have a matt paint coating and the range of colours is sufficiently large as it includes primaries (often referred to as R, G, B, C, M, Y), a six-step greyscale and six colours that simulate objects found in nature [6].

Experiment II involved a number of garments that would cover the range of the four primary CIELAB hues (red, green, yellow and blue) without taking texture into account. Thus, four single-colour garments were chosen as depicted in (see Figure 1).

Figure 1. *Visual stimuli used in Experimental II*

The garments were measured using a Minolta CS-1000 tele-spectroradiometer (TSR) under the experiment lighting conditions. The a^*b^* and $L^*C_{ab}^*$ plots in Figure 2 indicate that stimuli were close to the CIELAB primaries and that they cover a medium to high C_{ab}^* range of 30-90 while holding a medium L^* range of 40-60.

(a)



(b)

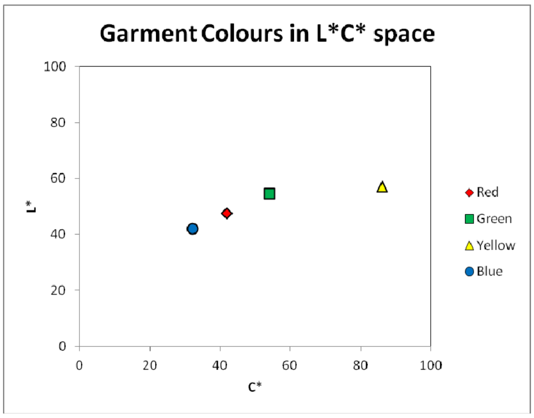


Figure 2. CIELAB specification of garments.

Output Media Device

To meet the aims of this work, a 46• LCD television having a resolution of 1920×1080 was chosen. Selection was based on the outcome of its evaluation in terms of temporal stability, additivity and spatial uniformity. The brightness and contrast levels were adjusted in order to provide the optimum tone reproduction. The data in Table 1 reveal the displays colorimetry, from which it can seen that the white luminance was setup at 264cd/m^2.

Table 1: Colorimetric data of the LCD primary colours.

	Lv(cd/m²)	x	y
White	264.20	0.32	0.33
Red	74.94	0.64	0.33
Green	168.60	0.24	0.67
Blue	35.10	0.15	0.08
Cyan	192.20	0.18	0.28
Magenta	113.40	0.31	0.16
Yellow	230.10	0.43	0.51
Black	0.65	0.30	0.31

The display primaries are plotted in Figure 3, revealing the device colour gamut as well that the device white was close to D65 white xy coordinates.

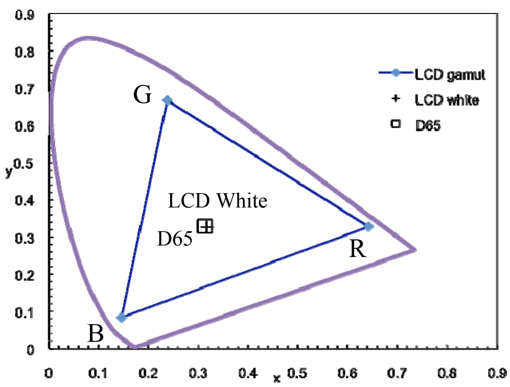


Figure 3. LCD gamut in CIE 1931 chromaticity diagram.

The display was characterised using a 13×13×13 3D-LUT. The model was tested using the 24 colours of the ColorChecker were used. The XYZ of the colours were used as input in the model which calculated their RGB values. The values were displayed on screen and measured using TSR. The input XYZ and measured XYZ values yielded a mean accuracy of 1.2 •E_{ab}^* units.

Input Media Device

Digital images were taken using a 6.1-megapixel Nikon D70S digital SLR. All were captured in raw 16-bit format to avoid the possibility of variations due to colour management from within the camera. The acquisition of 16-bit digital images provided sufficient information for high-fidelity colour rendering. To characterise the camera the GretagMacbeth ColorChecker DC (MCDC) consisting of 240 colour patches was acquired. The device was characterised using an 11-term polynomial model [7] and was tested using the 24 ColorChecker Chart colours. The performance of the model was found to exhibit a mean accuracy of 4.7 •E_{ab}^* units.

Light Source

The lighting in this experiment consisted of three fluorescent studio lights. The lights were positioned and luminance was adjusted to provide uniform illumination for the physical samples whilst limiting any flare light on the display. Measuring the calibrated diffuse white tile it was estimated that the light source had an approximate CCT of 5500K and luminance of 110cd/m^2. The same setup was used both to the take the digital images of the stimuli and to perform the colour-matching procedure.

Experimental Setup

The setup consisted of a medium grey background (L^* = 40), a Samsung 46• LCD TV display, an adjacent black plinth on which to place the physical samples and a set of three studio lights. The display was rotated into portrait orientation, as illustrated in Figure 4. The observer's seating position was chosen so that the LCD and stimulus could be viewed simultaneously from a distance of 2.5m from the LCD. This position was fixed to minimise variations due to the angular dependency of the device. The two media exhibited different luminance when it came to the white point. As it was reported above the display was setup in order to produce an optimum tone reproduction at 264cd/m^2 whereas the room lighting was adjusted to 110cd/m^2 in order to avoid the appearance of flare as much as possible from the surface of the display.

To minimise the effect of the luminance difference between the two media, observers were given about 10 minutes to adapt to the room lighting conditions prior to the initiation of the colour matching. In order to conduct the image matching, software was developed using Matlab software. Overall, observers were asked to match the colour of the digital stimulus shown on the LCD with that of the physical object next to it. The software manipulated colour in terms of $L^*a^*b^*C_{ab}^*$ and h°. During the course of the first experiment, the matched image and the RGB values of each chart patch were recorded. In the second experiment involving matching complete garments, a copy of each modified image was saved.

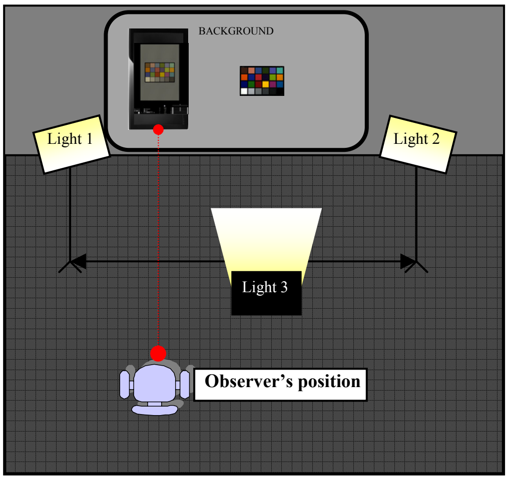


Figure 4. *Setup for the colour-matching experiment.*

Experimental I

The first experiment made use of ColorChecker chart colours as physical stimuli. Using the colour-matching software, observers were asked to match first the background colour, secondly the frame of the checker chart and finally each of the 24 colours in a random sequence. The data from this experiment were used to derive appearance characterisation models, as is described in the Data Acquisition and Analysis section.

Experimental II

The next experiment was concerned with matching images of 4 different garments on a mannequin, in a similar manner to Experiment I. The images were displayed on the LCD and observers were asked to match first the background, secondly the mannequin colour and finally the garment colour. The garments that were used covered the primary CIELAB hues red, green, yellow and blue. For each garment, 5 positions were identified, yielding a data set of 20 colours. This was later used to test the models derived from the data of Experiment I.

Observers

Observers numbering 10 and 15 (from China, Korea, Greece, Iran, Germany, Poland and the UK) participated in Experiments I and II respectively. Prior to each experiment, they were required to pass the Ishihara colour deficiency test. Their ages ranged from 24 to 36. Each observer had to attend 2 sessions in order to repeat the matching. This enabled the evaluation of their performance in terms of repeatability.

Data Acquisition and Analysis

Experiments I and II resulted in two digital images of the digital stimuli matched to the physical ones; for each observer and from each session. For each colour stimulus, the images were combined to produce a matched digital image that would represent that of the average observer. The matching of the average observer was later used to calculate observer variation.

Each observer's performance was evaluated by calculating the inter-observer variation and intra-observer variation measures [5]. The former indicated the agreement between each observer and the average observations and the latter the each observer's self-consistency. Inter-observer variation was determined by averaging the colour difference between an observer's matched colours and the overall average matched colours from all observers. On the other hand, intra-observer variation was determined for each observer by averaging the colour difference between colours matched in the first session and those of the second session.

In order to determine the relationship between the digital and physical stimuli, they were measured using a TSR. In the case of the ColorChecker image, the background and each of the 24 colours were measured. The TSR was placed at a fixed position from the stimuli corresponding to the observers' viewing location. The background was measured at 8 points surrounding the ColorChecker chart and each colour patch was measured once. In the case of the garment images, the background, the mannequin and the garment colours were measured. For each garment, 5 points were chosen to provide a data set of 20 colours.

The data acquired from the spectrophotometric measurements were converted into XYZ, CIELAB [7] and CIECAM02 appearance space [8-9]. The XYZ, $L^*C_{ab}^*$, JC and J'M' data from the ColorChecker chart in Experiment I were examined to reveal their relationship in each space.

Table 1: The models derived from the data acquired by matching the ColoChecker Chart in Experiment I.

Model		
1	$X' = 0.73X$ $Y' = 0.74Y$ $Z' = 0.71Z$	***Equation 1***
2	$X' = 0.73X + 1.9041$ $Y' = 0.74Y + 1.7762$ $Z' = 0.71Z + 1.795$	***Equation 2***
3	$L^{*\prime} = 0.89L^*$ $C_{ab}^{*\prime} = 0.84C_{ab}^*$	***Equation 3***
4	$L^{*\prime} = 0.83L^* + 4.3335$ $C^*_{ab}{}' = 0.88C^* - 2.9449$	***Equation 4***
5	$J'' = 0.91J$ $C'' = 0.88C$	***Equation 5***
6	$J'' = 0.79J + 0.8781$ $C'' = 0.88C - 1.9755$	***Equation 6***
7	$J''' = 0.90J'$ $M''' = 0.91M'$	***Equation 7***
8	$J''' = 0.90J' - 3.271$ $M''' = 0.91M' - 1.8969$	***Equation 8***

Plots describing the relationship between the digital and physical stimuli were used to determine linear trend lines a) with and b) without intercepts. In total, 8 models were established using 4 colour specification and colour-appearance spaces, and are given in Table 1.

The models derived from the ColorChecker in Experiment I were applied to scale the data of the digital image stimuli and were examined to see how well they fitted the physical object stimuli data for both the ColorChecker and garments. Prior to modelling, the flare from the surface of the LCD TV was measured and removed from the data of the digital stimuli. Model performance was tested by calculating the colour difference between: a) the scaled data of the digital stimuli and b) those of the physical stimuli using the CAM02-UCS [9]. The reason for choosing CAM02-UCS was that it is a uniform colour-appearance space and therefore able to describe colour-appearance differences more accurately. Due to fact that observers adapted to the room lighting conditions prior to the colour matching, the data from the experimental room were used to compute the CAM02-UCS attributes.

Results

Average observer performances in Experiments I and II are shown in Table 2 in terms of mean CIELAB $^{*}_{ab}$ units. The inter-observer variation values reveal that an observer's matching has similar performance compared to the whole group for both experiments. Noticing the intra-observer variation values it was determined that the ability of the observers to provide repeatable results was similar in both experiments.

Table 2: Inter-observer and intra-observer variation in both experiments in CIELAB colour-different units.

	Variability (CIELAB ΔE)	
	inter-observer	**intra-observer**
Experiment I	5.7	5.9
Experiment II	5.4	5.2

Comparing the values in Table 2 with those from Oicherman's work [5], the magnitude of inter-observer and intra-observer variation is higher. Also, in Oicherman's work the intra-observer variability was less than that of inter-observer; whereas in this study the two measures for each experiment have a similar range. This is mainly caused by the size effect of the display and the significant flare that is introduced from the light source used to illuminate the garments. This phenomenon is of particular interest as it is similar to viewing conditions used in real world situations.

The models shown in Table 1 describe the relationship of the components in various colour spaces, between the colours of the physical chart and their match on the LCD. Examining the relationship of the tristimulus values in Model 1 and Model 2, those of the digital stimuli have to be scaled down by approximately 30% to match those of the physical sample. Models 3 and 4 reveal that the digital stimuli appear approximately 15% lighter. Considering the CIECAM02 appearance space, Models 5 and 6 indicated that the lightness *J* and chroma *C* of the colours shown on screen were about 10% higher. To conclude, the relationships in the CAM02-UCS appearance space also show that the LCD colours were 10% lighter and 10% more colourful than those in the physical chart.

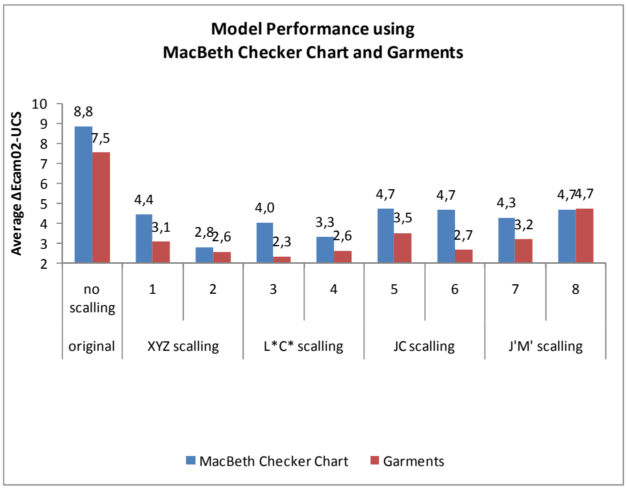


***Figure 5.** Model performances using the ColorChecker Chart and garments as test set.*

The bars in Figure 5 illustrate the performance of the models in Table 1 when it came to fitting the digital colour data to the physical ones. The original data for the ColorChecker Chart had an average of 8.8 CAM02-UCS E units, which is 1.3 units different from those of the garments. This means that the difference between digital and physical colour from both datasets was similar; however, the small difference between the two datasets is most likely to have occurred because of the disparity in stimuli size. The data presented in Figure 3 suggest that the models fit the datasets very well, reducing the colour difference in most case by more than 50%. In all cases, the colour difference of the garments was smaller than for the ColorChecker colours. Considering the fact that the ColorChecker data set had a higher colour difference than that of the garments, the models actually had a similar performance in both datasets. The scaling in tristimulus values and CIELAB space resulted in a better fitting of the digital stimuli data to those of the physical stimuli for the garments. On the other hand, the ColorChecker test set had a performance of 4.0 to 4.7 units on all models except Models 2 and 4. The model with the best fit for both datasets was Model 2, with a performance of 2.8 and 2.6 units for the ColorChecker chart and garment dataset respectively. This means that when scaling the tristimulus values using a first-order linear equation when showing colours on the LCD, the difference from physical samples was about 2.8 units.

The values in Table 3 are the components that determine the colour difference between the digital and physical stimuli. The term that contributes the most in forming the colour difference between the stimuli was that of the J' lightness difference. Taking into account that the physical stimulus was used as standard, the negative sign of the J' clearly shows that the digital stimuli had higher lightness than the physical ones. The fact that the digital stimuli were lighter was also predicted by the majority of models in Table 1. Comparing this finding with the previous study between small displays and physical samples [5], there is an agreement since it was found that the lightness difference was responsible for the variation in matching.

Table 2: Analytical average CAM02-UCS colour-difference measures between digital and physical values from the ColorChecker Chart and garments datasets.

	ΔJ'	Δa'	Δb'	ΔE CAM02-UCS
ColorChecker	-6.89	-0.42	0.34	8.83
garments	-5.99	-0.11	-0.21	7.53

Conclusions

Two colour datasets of physical stimuli, varying in material, were photographed and reproduced on a 46• LCD TV under fixed lighting conditions. Their digital representations were modified to match the physical samples. Measuring the two types of stimuli and comparing them using the CAM02-UCS colour difference revealed an absence of material dependency. It was also demonstrated that although both stimuli appeared the same, the digital stimuli were significantly lighter that the physical ones. A simple linear equation to fit the tristimulus between the datasets is able to compensate for their physical differences. The main difference between the physical and displayed images was found to be due to lightness. This was caused by the flare light that was introduced to illuminate the adjacent physical samples. The surface of the output media should be measured in order to retrieve the amount of surface flare. This should be removed in advanced before colour appearance modelling takes place.

References

[1] P.A. Rhodes, S.A.R. Scrivener and M.R. Luo, ColorTalk: a system for colour communication, *Displays*, **13** (2), 89-96 (1992).

[2] P.A. Rhodes and M.R. Luo, A system for WYSIWYG colour communication, *Displays*, **16** (4), 213-221 (1996).

[3] M. Olkkonen, T. Hansen, and K.R. Gegenfurtner, Color appearance of familiar objects: Effects of object shape, texture, and illumination changes, *Journal of Vision*, **8** (5):13, 1-16 (2008).

[4] M.R. Pointer, Measuring visual appearance: a framework for the future, *NPL Report: COAM19* (2003).

[5] B. Oicherman, M.R. Luo, B. Rigg and A.R. Robertson, Effect of observer metamerism on color matching of display and surface colours, *Color Research and Application,* **33** (5), 346-359 (2008).

[6] C.S. McCamy, H. Marcus and J.G. Davidson. A color-rendition chart. *Journal of Applied Photographic Engineering,* **2** (3), 95-99 (1976).

[7] G. Hong, M.R. Luo and P.A. Rhodes, A Study of Digital Camera Colorimetric Characterisation Based on Polynomial Modeling, *Color Research and Application,* **26** (1), 76-84 (2000).

[8] CIE Technical Report, CIE15:2004 *Colorimetry* (3rd edition).

[9] CIE Technical Report CIE159:2004 *A color appearance model for color management systems: CIECAM02*.

[10] M.R. Luo, G. Cui and C. Li, Uniform color spaces based on CIECAM02 appearance model, *Color Research and Application,* **31** (4), 320-330 (2006).

Author Biography

Chrysiida Kitsara is a PhD Student from the University of Leeds. Her current research interest is cross-media colour reproduction.