

Testing Color-Difference Formulae on Complex Images Using a CRT Monitor

Tao Song and Ronnier Luo
Colour & Imaging Institute, University of Derby
Derby, United Kingdom

Abstract

An experiment was carried out to evaluate the perceptibility and the acceptability of colour differences between pairs of CRT images. Four images were used. For each image, a series of images were systematically rendered following four directions: lightness, chroma, mixed lightness and chroma, and hue using different functions. The effects of rendering were assessed by a panel of observers. The results were used to test the performance of different colour difference equations. In addition, the perceptibility threshold and the acceptability tolerance for each formula were determined.

Introduction

CIE recommended the CIELUV and the CIELAB colour spaces and their corresponding colour-difference formulae to industries in 1976.[1] The CIELAB colour space and colour-difference formula are widely used for image applications.[2] But while the CIELAB colour space gained great success in industrial practice, the colour-difference formula was less.[3] Therefore, some new, more successful colour-difference formulae based on CIELAB were derived to fit colour discrimination data sets[4-6] based upon large sized surface samples. These include BFD,[7] CMC,[8] CIE94[3] and more recently, the new CIE colour difference equation, CIELAB2000.[9] The latter is a strong candidate for the new CIE colour difference equation.

However, the colour-difference formulae mentioned above were all developed based on the experimental data obtained using large size uniform surface patches such as textile and paint. These types of data are quite different from those used in pictorial images displayed on a CRT or a hardcopy print. There is a need from imaging industry to standardise a formula or a method for evaluating the quality of the original and reproduction images. Although some recent experiments for testing these formulae using large sized uniform colour patches on CRT were reported,[10,11] the colour-difference metrics used were only CIELAB and CIE94. Stokes[12] did experiment on 6 complex CRT images and concluded that CIELAB performed better than CMC. Recently, Uroz[13] did similar experiment on printed images and reported that CIELAB and CIE94 (2:1:1) describe the colour difference more accurately than CIE94(1:1:1). Image dependency is another important factor to affect the performance of colour difference equations. Some studies[12,14] show that the threshold of colour changes has nothing to do with image contents while some others give the opposite opinion.[13,15]

The aims of the current study are to: carry out experiments for determining perceptibility threshold and acceptability tolerance for a number of colour difference equations, test the performance of these equations, investigate the weighting factors for each individual colour difference component and to compare the present study with those from the earlier studies.

Experiment Setup

The four original images used by Uroz[13] were selected for the experiment: Fruit (F), Musicians (M), Harbor (H) and Gaudi (G). These are shown in Figure 1. Hence, the results obtained from the CRT can be directly compared with Uroz's results that are from prints.

Stokes[12] used four transfer functions in his experiment to simulate the changes in the practical colouring process of image devices such as contrast, gain, gamma controls and colour casts or shifts. These functions were applied to the dimensions of lightness, chroma and hue in CIELAB colour space. Four transform functions were used: additive offset, multiplicative, power and sigmoidal functions. The former three functions were used in the present study as shown in Table 1. (Stokes' study showed that similar results were obtained between the multiplicative and the sigmoidal functions. Hence, the latter function was discarded in this study.)

The reproduction images were processed on a HP Visualise C3000 Workstation. The output device is a BARCO CRT monitor with 72 dpi resolution, which is equipped with self-calibration software and a photo-sensor. During the experiment, the monitor was calibrated daily and the white point was set to D65. The GOG model[16] was used to characterise the CRT for transforming between CRT primaries and CIE specifications. Hence the colour differences between two images can be calculated pixel by pixel.

The reproductions for each image were made from the standard image by using the five transforms that are given in Table 1. For each transform, 11 levels of colour differences were decided from a pilot experiment. The RGB values of each pixel were converted to CIE tristimulus values, using the GOG model, and then to CIELAB values. All transforms were based upon CIELAB attributes and the transformed values were further converted back to the CRT's RGB values. During the transformation, when encountering out of gamut colours, they were simply clipped to the gamut boundary. (The numbers of out of colour gamut pixels vary according to the difference levels. The extreme situation is that: about 15% of the pixels in one image are out of gamut.)

Fruit (494x600)

Gaudi (800x542)

Harbor (800x600)

Musicians (590x600)

Figure 1. The Standard images

The experiment was conducted in a dark room. The images were displayed on the BARCO monitor. Each assessment included an original and a test image. But at any time only one image displayed on the screen, i.e. either the original or the reproduction. Observer used the TOGGLE button to switch these two images. This method was used due to insufficient screen size to accommodate two images simultaneously and to avoid problems due to screen non-uniformity. Each image was surrounded with a white border and a gray background. The border and background had the chromaticity values set to D65 with L* values of 100 and 50 respectively. Observers watched the images with a distance about 65 cm. The viewing angle for each image was about 26°. Before the experiment started, observers were asked to look at the gray background for one minute for adaptation. They then judged the colour differences between the original and the test images. Their results were recorded by pressing 1, 2 or 3 buttons on the screen for the answer of 'no difference', 'just perceptible difference' or 'just not acceptable difference', respectively.

Table 1. The mathematical models

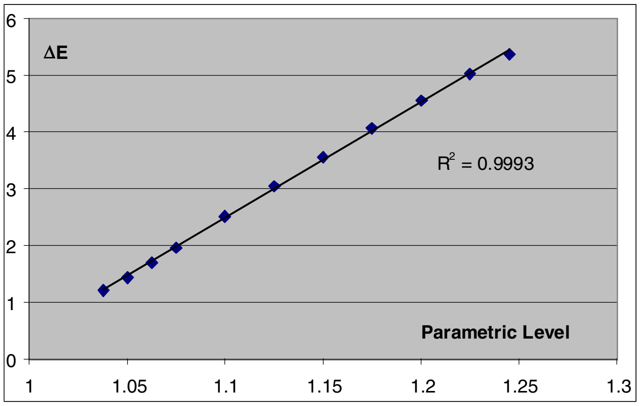
Function	Formula	Apply on	Abbr.
Multiplicative	Out = k * In Where $k < 1$	Lightness	LM
		Chroma	CM
		Lightness & Chroma	LC
Power	Out = 100 * (In / 100)^a Where $a > 1$.	Lightness	LP
Additive Offset	Out = In + *off* where *off* = 0 when C*<10.	Hue	HO

Ten observers with an average of 27 years old took part in this study and were the students at the Colour & Imaging Institute. They all had normal colour vision according to the Isihara test. Half of them had experience in attending psychophysical experiments.

The Method of Limits (or staircase method) was used to scale colour differences. The test images were divided into 20 groups (5 models and 4 images). In each group, the test images were displayed according to ascending or descending order of colour differences. Each observer was asked to perform both orders. Therefore, 440 pair images (20 groups x 2 orders x 11 colour difference levels) were assessed by each observer. The whole assessment was divided into 4 sessions so that one session lasted about 40 minutes to avoid observer fatigue.

Data Analysis

Probit Analysis[6,17] was employed in the data analysis. In our case the positive responses to stimulus intensities are 'just perceptible difference' and 'just not acceptable difference'. Therefore 50% positive responses correspond to the thresholds of perceptibility and acceptability in terms of parametric levels (the k, α and *off* factors in Table 1), which were used rather than average ΔE^*_{ab}. The reasons are that the parametric level has a strong linear relationship with the average ΔE^*_{ab} value (see Figure 2), and it is convenience to calculate colour differences between two images based upon parametric factors for each colour-difference formula.



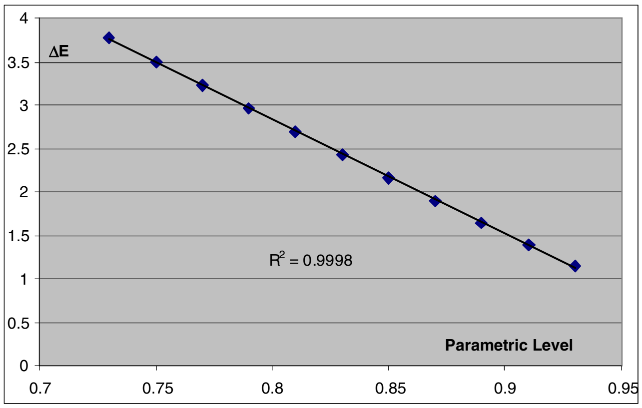


*Figure 2. The typical relationship between colour difference levels and ΔE*ab*

Results

According to the 50% perceptibility and acceptability levels for each transform function/image, 80 threshold images were reproduced (4 images x 5 functions x 2 orders x 2 types of data). The mean colour difference between the threshold and original images was calculated using CIELAB, BFD, CMC, CIE94, *CIELAB2000* colour difference formulae. At a later stage, the lightness and chroma weighting factors for each formula were optimised. Again, the mean colour difference between the threshold and original images was calculated between the threshold and original images for formulae with the optimised weights.

Comparing the Perceptibility and the Acceptability Thresholds

Table 2 lists the 50% level in terms of ΔE^*_{ab} and their differences between the ascending and descending orders. The ΔE^*_{ab} of ascending order is always greater than that of descending order for perceptibility. As for acceptability, this trend does not exist. The systematic difference between two orders in the perceptibility results is typical for staircase experimental data and is mainly caused by experimental setup. It was decided to average the results from the two orders to represent the overall perceptibility and acceptability results because the differences are small.

Table 2 also shows that the hue threshold (or tolerance) is close to that of chroma, and the lightness threshold is mostly the largest. This implies that for image application, lightness differences are less noticeable than chroma or hue differences. This is similar to the textile applications with a ratio of 2:1 for lightness against chroma and hue when applying CIE94 colour difference equation. However, the acceptability tolerance for CRT images found here is about 250-400% larger than that typically used in the textile trade (about 1 ΔE^*_{ab} unit).

Table 2. List of Threshold ΔE^*_{ab}

	Perceptibility		Difference
	Ascend	Descend	(Des-Asc)
LP	2.16	1.93	-0.23
LM	3.18	2.84	-0.34
CM	1.56	1.34	-0.22
LC	3.02	2.52	-0.50
HO	1.62	1.65	0.03
	Acceptability		Difference
	Ascend	Descend	(Des-Asc)
LP	4.48	4.60	0.12
LM	6.05	5.43	-0.62
CM	2.75	2.71	-0.04
LC	5.87	6.12	0.26
HO	3.37	3.10	-0.28

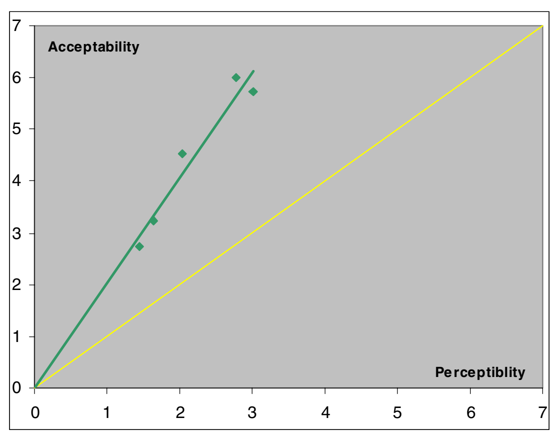


Figure 3. Average Perceptibility vs. Acceptability

Figure 3 shows the average perceptibility threshold results plotted against the average acceptability tolearance results. It clearly shows that there is a very good agreement between two sets of data. The present results indicate that the main difference between two types of judgement is the size, i.e. the acceptability tolerance is about twice of the perceptibility threshold in this study.

It is noted that the two lightness transform functions gave quite different results (see LP and LM in Table 2). This implies that for an image, a consistent change in lightness (overall darker or lighter) is less noticeable than a change of lightness contrast.

Colour-Difference Formulae Performances

The standard deviation (STD) and coefficient variation (CV) of the perceptibility thresholds and acceptability tolerances were calculated for each colour difference formula. CV is calculated by dividing STD by the mean and multiplying 100. These values indicate the performance for each formula. For perfect agreement between the formula predictions and visual results, CV and STD should equal to zero. A CV value of 30 means a 30% disagreement between two sets of data. The performance of each original formula is summarised in Table 3 (see 'Original', i.e. set lightness and chroma weights to one). The acceptability tolerances and perceptibility thresholds were further analysed to obtain the best lightness and chroma weighting factors (set hue factor to one) which gave the least STD and CV measures. It was found that the optimised weighting factors between the perceptibility and acceptability experiments were very similar. This is because the main difference between these two types of assessment is the size (as discussed in last section). Hence, it was decided to average them together to represent the overall results. The optimised best colour difference formulae are: CIELAB(1.8:0.85), BFD(1.4:0.65), CMC(1.8:0.6), CIE94(2.6:0.65) and CIELAB2000(1.85: 0.65). Their performances are also given in Table 3 (see 'Optimised').

Table 3. STD and CV values for the five colour-difference formulae performances

		Perceptibility				
		CIELAB	BFD	CMC	CIE94	CIELAB2000
Original	STD	0.71	**0.61**	0.74	0.91	0.62
	CV	32	**29**	37	48	37
Optimised	STD	0.31	0.29	0.22	**0.19**	0.20
	CV	19	15	**14**	17	16
		Acceptability				
		CIELAB	BFD	CMC	CIE94	CIELAB2000
Original	STD	1.51	**1.23**	1.49	1.95	1.31
	CV	34	**29**	38	50	39
Optimised	STD	0.65	0.49	**0.38**	0.39	0.39
	CV	20	**13**	13	18	16

The results show good agreement between the perceptibility and acceptability results. As described earlier, these two sets of results are very similar expect for size difference. As expected, the optimised equations perform significantly better than those original equations. This implies that the lightness, chroma and hue weighting factors should be differently weighted for assessing colour difference for pictorial images.

Comparing the performance between the different optimised colour difference formulae, the results show that the other colour difference formulae performed better than CIELAB. The BFD and the CMC predicted slightly more accurately than CIELAB2000 and CIE94 formulae. The CIELAB2000 is a strong candidate as the new CIE colour difference equation for surface colours. It performs the most accurate than the other formulae for the experimental data based upon large size surface patches. This indicates that there are some differences between the surface colours (colour patches) and pictorial images. However the performance of all the optimised formulae was excellent.

Comparing with the Previous Studies

There is excellent agreement between the present results with those found by Uroz[13] and Stokes.[12] The former experiment was conducted using large size prints and the latter using a CRT. Comparing the perceptibility thresholds in terms of ΔE^*_{ab} units, these are 2.0, 2.3 and 2.2 for Stokes, Uroz and present studies respectively. All results indicate that there is a need to have larger lightness weight than chroma and hue weights. Comparing the acceptability tolerance between the present and Stokes's studies, these are 4.4 and 6.6 respectively. This indicates that the observers in present study are more rigorous than his observers. It is well known that the acceptability criteria could have a large variation depending upon the observers used. However, the earlier results showed that there is a linear correlation between thresholds and tolerances. In other words, if the threshold of each colour difference components are known, we can simply multiply a factor to approximate the tolerance for acceptability applications.

Conclusions

An experiment was carried out to evaluate the perceptibility and acceptability of colour differences between pairs of CRT images. Four images were used. For each image, a series of images were systematically generated following four directions: lightness, chroma, mixed lightness and chroma, and hue using different functions. The results are summarised below:

- The difference between the perceptibility threshold and acceptability tolerance is only in magnitude, not the ratio between lightness, chroma and hue differences. The ΔE^*_{ab} for perceptibility is about 2.2 and 4.5 for acceptability. Lightness difference is less noticeable than chroma and hue differences.
- There is not much difference between different colour difference formulae using optimised weights. In general, all formulae performed slightly better than CIELAB, and all performed very well.
- There is a very good agreement between the present and earlier studies in terms of perceptibility thresholds

and weighting factors for each colour difference components.

References

1. Colourimetry. CIE Publication 15.2. Vienna: CIE Central Bureau, (1986).
2. H. Komatsubara, and G. Kawakami, Research on the application of colour-difference formulas to industry, *Study Colour*, 35, pg. 6-20, (1988)
3. *Industrial colour-difference evaluation*. CIE Publication 116. Vienna: CIE Central Bureau, (1995).
4. M. R. Luo, and B. Rigg, Chromaticity-discrimination ellipses for surface colours, *Colour Res. Appl.*, 11, pg. 25-42, (1986).
5. K. Witt, Three-dimensional threshold of colour-difference perceptibility in painted samples: variability of observers in four CIE colour regions, *Colour Res. Appl.,* 12, pg. 128-134, (1987).
6. R. S. Berns, D. H. Alman, L. Reniff, G. D. Snyder, and M. Balonon-Rosen, Visual determination of suprathreshold colour-difference tolerances using probit analysis, *Colour Res.Appl.*, 16, pg. 297-316, (1991).
7. M. R. Luo, and B. Rigg, BFD(l:c) colour difference formula, Part I - Development of formula, *J. Soc. Dyers Col.*, 103, pg. 86-94, (1987).
8. F. J. J. Clarke, R. McDonald, and B. Rigg, Modification to the JPC79 colour difference formula, *J. Soc. Dyers Col*., 100, pg.128-132, (1984).
9. M. R. Luo, G. Cui, and B. Rigg, Further Improvements to the CIELAB based Colour Difference Formulae, Submitted to *Colour Res. Appl.*
10. E. D. Montag, and R. S. Berns, Visual Determination of Hue Suprathreshold Colour-Difference Tolernaces Using CRT-Generated Stimuli, *Colour Res. Appl*., 24, pg. 164-176, (1999).
11. M Melgosa, M. M. Perez, A. E. I. Moraghi, and E. Hita, Colour Discrimination Results from a CRT Device: Influence of Luminance, *Colour Res. Appl.*, 24, pg. 38-44, (1999).
12. M. Stokes, *Colourimeteric tolerances of digital images,* MSc Thesis, RIT, University of Rochester, New York, USA. (1991).
13. J. Uroz, Colour difference perceptibilityfor large-size printed images, MSc thesis, Colour & Imaging Institute. The University of Derby, Derby, UK. (1999).
14. E. Fedorovskaya, Huib de Ridder, and F. J. J. Blommaert, Chroma variations and perceived quality of colour images of natural scenes, *Colour Res. Appl.*, 22, 2, pg. 96-110, (1997).
15. S. P. Farnand, The effect of image contents on colour difference perceptibility, *Proc. 4th IS&T/SID Colour Imaging Conference,* pg. 101-104, (1996).
16. R. S. Berns, Methods for characterizing CRT displays, *Displays*, 16, 4, pg. 173-182, (1996).
17. D. J. Finney, *Probit Analysis,3rd Edition,* Cambridge University Press, Cambridge, UK. (1971).