Colour Differences for Complex Images

Chikako Sano , Tao Song** and M Ronnier Luo***
**Sony Corporation Mobile Network. Co., Imag. Tech. Div. Japan*
*** Colour & Imaging Institute, University of Derby, United Kingdom*

Abstract

This paper describes the research work associated with the CIE TC8-02 *Colour Difference Evaluation in Images*, which aims to derive an industrial colour-difference evaluation method that is appropriate for complex colour images. The main aim of the TC is to recommend a colour-difference formula for complex image together with suitable perceptibility thresholds for industrial applications. With this in mind, a set of experimental results for assessing perceptibility thresholds were accumulated using printed images. The results from different data sets were found to be quite consistent. This set of images are currently being circulated and assessed at different imaging companies.

Introduction

Colour-difference assessments were typically carried out by experienced colourists. It is a time consuming and expensive process. Many colour difference formulae have been developed intended to replace human judgements. These have been widely used in the surface colour industries such as textiles, paint and packaging for evaluating the colour difference between two specimens. More recently, manufacturers of imaging products, such as printers and displays, are also interested in this subject for evaluating the performance of colour reproduction systems. This is normally evaluated by using the colour-difference values between colour patches in the original and its reproduction.

The accuracy of colour-difference formulae was typically evaluated using large uniform colour patches with grey background under specific viewing environments and formulae often modified to fit the perceived colour differences by observers. However, it is well known that perceived colour difference varies according to different viewing parameters such as sample size, separation, texture, and background colour.

The effects of viewing parameters, such as sample size, illumination, surround and background, on colour-difference evaluation in complex images, is more significant than that for colour patches because all these colour-difference equations do not take into account the spatial properties of human vision. There has been a large demand from imaging industry to standardise a method. With this in mind, CIE TC8-02 *Colour Difference Evaluation in Images* was established, which aims to derive an industrial colour-difference evaluation method that is appropriate for complex colour images together with suitable perceptibility thresholds for industrial applications.

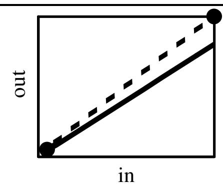
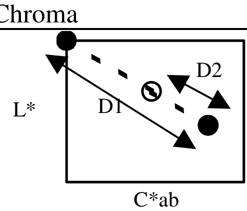
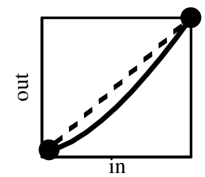
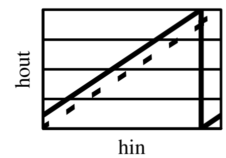
The current experimental results directly contribute to the work of TC8-02. A set of printed images were carefully produced and assessed by a panel of 11 observers. The results were used to test various colour-difference formulae and to establish the perceptibility thresholds for each formula. This set of images are currently being circulated and assessed at different sites including a number of industrial companies. This will lead to more data and more understanding of the variations of thresholds across different cultures and environments.

Experimental

An HP DesignJet20ps printer with 2400 by 1200 dpi resolution was used to reproduce test images and these were printed on A3 size glossy proofing papers with 6 inks (cyan, magenta, yellow, black plus a light cyan and a light magenta). A *GretagMacbeth Spectrolino* spectrophotometer with 45/0 geometry was used for all the colour measurements in this study. It was found that there was a large change in colour in the first three days after printing, due to the drying process of inks. Therefore, all experiments were carried out at least three days after when a print was made.

For evaluating printer's repeatability, a cube 6x6x6 test chart was used. Three tests were carried out according to three time periods: short-term, one day and one week. The former test was performed by printing the same chart three times in close succession. The other two charts were printed after one day and one week. Each colour in the chart was measured. The repeatability performances in terms of mean /maximum ΔE^*_{ab} values were 0.3/1.8, 0.5/2.7 and 1.7/5.0 for the short-term, one day and one week respectively. These results are considered to be quite satisfactory. Four test images, Picnic, Harbour, Fruit, and Wool were used as shown in Appendix. These images were selected because they include memory colours (sky, grass, skin and fruits), a variety of lightness, chroma and hue, different object size and indoor and outdoor scene, respectively.

Table 1. Transform functions for image rendering

Function	Multiplicative	Multiplicative	Power	Addtive offset
Formula	Chroma: $C^*_{out} = k_c C^*_{in}$ Lightness: $L^*_{out} = k_l(L^*_{in} - L_k) + L_k$ L_k : L^* of the black k_c ,$k_l < 1$: coefficient	$D_2 = kl_c D_1$ D_1:distance between reference white and original D_2 : distance between reference white and pixel after alternation $kl_c < 1$: coefficient	$L^*_{out} = L_k+(L_{\square} - L_k)*$ $((L^*_{in} - L_k)/(L_{\square} - L_k))^a$ L_k : L^* of the black L_w : L^* of the paper white $a > 1$: coefficient	$h_{out} = h_{in} + hoff$ $hoff > 0$
CIELAB directions	Lightness, Chroma	Mixture of lightness and Chroma	Lightness	Hue
Diagram				
Abb.	CM, LM	LCM	LP	HO

These four images are widely used for colour evaluation. Harbour and Wool were chosen from ISO/TC130 *Standard High Precision Pictures*, Picnic was selected from Sony sRGB standard images, and Fruit was selected from the earlier work by Uroz *et al* at Derby. [1] These images were adjusted to have the same resolution 1500 by 1250 pixel and their colour gamuts were compressed so that all colours in the original images were within that of the HP printer.

Each colour in the cube test chart including 9x9x9 colour patches was defined using printer RGB values. These colours were used to develop characterisation model of the printer. The procedure for generating test images is as follows: RGB digital input data of each pixels in an original image were initially transformed to CIELAB values using a tetrahedral interpolation characterisation model, colour transfer functions (see later) were then applied to render each image. Subsequently, the altered CIELAB values were converted to the printer RGB values using the inverse characterisation model. The rendered images were finally printed for visual assessments. Each original test image was altered using five transfer functions in four CIELAB dimensions: lightness, chroma, hue and a mixture of lightness and chroma.

Test images data were altered using five transfer functions at eight different colour-difference levels by pixel by pixel transformation from original image data. The transform functions are shown in Table 1. The eight levels for each function were carefully chosen by examining visually to ensure the threshold value of the colour difference located around the middle of its parameters. This resulted in different parameters used for each image.

A colour palette including 56 colours was extracted by sampling the most commonly used colours of each test image via the Photoshop software. These colour patches were rendered using the same transfer function and printed as the associated test image. These colour patches represent the present colours of the image and provide useful information the change of colours for a particular image over time. When carried out the psychophysical experiments, these patches were masked.

The psychophysical experiments were carried out in a dark room. The experimental situation is shown in Figure 1. A pair of test images was presented against a mid-grey wall under a CIE D65 simulator in a Macbeth Spectralight II viewing cabinet with a filtered tungsten halogen light source to simulate the CIE D65 illuminant. Each pair included an original and a test image having about A4 size with a white paper border. The size of each image was 25 by 19 cm^2. The test images were divided into 20 groups (5 transfer functions x 4 original images). The sequence of the test images presented in each group was randomised. In addition, four pairs which included two identical original images in a pair were included. In total, 164 judgements per observers were made. The whole assessment was divided into 4 sessions with each lasting approximately 25 minutes. Observers were asked whether they saw a colour difference between two images. Eleven observers participated in this experiment. They were students at the Colour & Imaging Institute and all had normal colour vision according to Ishihara test.

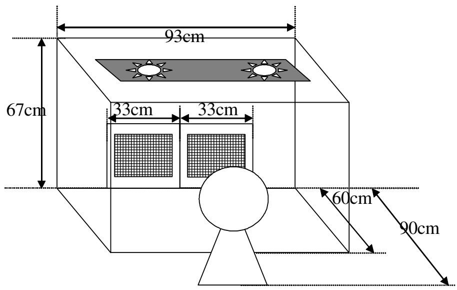


Figure 1. Viewing geometry

Results and Discussion

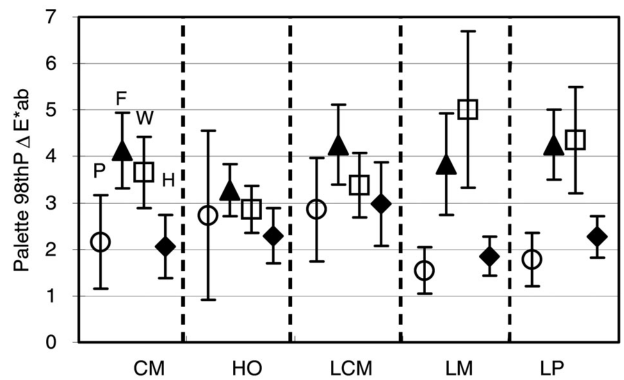
Two different techniques were used to calculate colour-differences between two images: pixel-by-pixel and colour palette measurement. The former calculates the mean or percentile of the colour-difference values between the original and a rendered image for all pixels, and between the measured palette of the original and that of the rendered image. In general, these two sets agreed well with each other except for few rendered images. It was found that these discrepancies were caused by the quantisation errors of the printer used, the sampling method of the Photoshop and the distribution of colour difference in an image. Finally, the palette data was used to represent the colour change because it represents the current state of the image to be assessed.

Perceptibility threshold expressed as the 98^{th} percentile ΔE^*_{ab} of all images for each transfer function and mean of that are shown in Figure 2a and 2b respectively. In each case, the colour differences were calculated as average values across entire image, using the CIELAB formula. The 98^{th} percentile is the ΔE^* for which 98% of the observers could detect a colour difference. The error bars show 95% confidence levels. An image having a lower threshold means its colour differences are more noticeable.

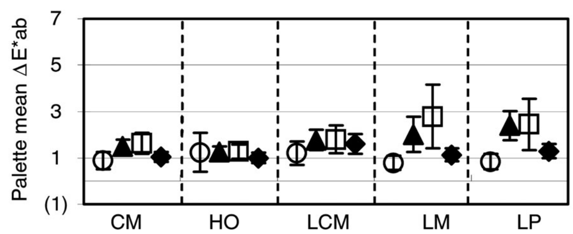
For the previous experiments, four colour difference metrics were used for data analysis: palette arithmetic mean, palette 98th percentile, pixel-by-pixel arithmetic mean, and pixel-by-pixel 99th percentile. Stokes[2] and Song[3] used mean colour difference and Uroz[1] used 99th percentile colour difference. If the colour difference of individual component, ΔL^*, Δa^* and Δb^*, have an approximately Gaussian distribution, one of them can be predicted from the other. In this study the results which used 98th percentile colour difference and mean colour difference showed good correlation. Only 98th percentile colour-difference threshold was used here in order to compare with Uroz's results because he also evaluated the colour-difference of prints.

Uroz[1] compared the goodness of fit to a normal distribution using Pearson's Chi-squared test between the colour difference of palette arithmetic mean and that of 99th percentile. The 99th percentile of colour differences between palette colours provided more approximate metric of perceptibility threshold of colour changes in images.

If the thresholds from all images studied are the same, there is no image dependence. The results show a weak dependence on image contents when chroma and hue attributes were altered, and a strong image dependence when lightness values were altered. For example, the perceptibility thresholds of Picnic and Harbour which were outdoor scenes, and had large sky areas, are lower than those of the other two images, especially when lightness was altered. This could be due to the fact that when lightness decreased, not only this area was darkened but also the contrast of the entire image was reduced.



(a) Palette 98th percentile



(b) Palette mean

○:Picnic(P) ▲:Fruit (F) □Wool (W) ♦:Harbor (H)

Figure 2. Perceptibility threshold of each test image with different transfer functions

Figure 2 also shows that the average thresholds corresponding to 98th percentile are about 3 CIELAB units for all attributes studied. For lightness, the thresholds vary according to image contents as mentioned earlier.

Four commonly used colour-difference formulae were tested: CIELAB,[4] CMC,[5] CIE94[6] and CIEDE2000.[7] The Coefficient of Variation (CV), defined as the standard deviation divided by the mean and multiplying by 100, was used to compare the perceptibility thresholds in terms of 98th percentile with four colour-difference formulae. For a perfect agreement, CV should be zero. This means that a formula has the same thresholds for all the different colour directions investigated. A CV value of 30 means a 30% disagreement between the formula and visual results. The results are summarised in Table 2. The mean from five different transfer functions together with the CV value calculated from all 20 thresholds are also given in Table 2 (the results under 'Mean' and 'Overall' respectively). For each colour-difference formula, two forms were tested: the original formula with K_L =1 and modified formula as given in Equation (1).

$$\Delta E(K_L) = [(\Delta L/K_L)^2 + (\Delta C)^2 + (\Delta H)^2]^{1/2} \tag{1}$$

where ΔL, ΔC, ΔH and ΔE are the lightness, chroma, hue and total difference of a colour-difference formula, and K_L is the lightness parameter for adjusting the balance between the lightness and chromatic difference.

The three colour difference equations except for CIELAB have a weighing parameter for the colour difference of lightness. K_L of one is used for most application and for specific application such as textile, K_L of two is recommended. The results in CV units for both K_L values are given in Table 2. The CIELAB formulae does not have the parametric factor which can be varied according to the experimental conditions, in order to compare other equations, CIELAB(K_L=2) was introduced:

$$\Delta E^*_{ab}(2:1:1)=[(\Delta L^*/K_L)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2} \tag{2}$$

where K_L = 2

When set K_L to 1, CIELAB and CIEDE2000 showed relatively good performance. The perceptibility threshold of lightness was larger than that of chroma and hue. This means that lightness difference is less noticeable when colour-difference equations except CIELAB are used.

When set K_L to 2, the performance of the colour difference formulae except CIELAB improved, significantly, especially CIEDE2000 was the best of all. The change of K_L from 1 to 2 made the difference in performance amongst different transfer functions decreased. These results agree with previous experiments well (Uroz, Stokes, Song). The increase in weighting coefficient improved the performance of CIE94 and CIEDE2000. However, ΔE_{CMC} (K_L=1.5) performed better than both ΔE_{CMC}(K_L=1) and ΔE_{CMC}(K_L=2). In addition, the difference between, ΔE^*_{ab} (K_L =2) and ΔE^*_{ab} (K_L=1) were small. The increase in the weighting coefficient of lightness not always improve the performance of equations, the optimisation of the weighting coefficient is necessary for further improvement.

In order to compare the independence of colour-difference formulae on images and colour dimension, the Coefficient of Variation, CV, was again used. The smaller values show small dependence, *i.e.* higher performance. The CV values for the image change of different colour dimension are summarised in Table 2. "Mean" means the mean of five different transfer functions and "Overall" means the CV of 20 thresholds, 4 images 5 transfer functions. Again the performance of ΔE_{00} (K_L=2) was the best of all, followed by ΔE_{00}(K_L=1.5).

Table 2. Testing different colour difference formulae using various data sets in CV unit.

	CIELAB		CIEDE2000			CMC			CIE94		
K_L	1	2	1	1.5	2	1	1.5	2	1	1.5	2
CM	35	38	56	46	**19**	44	26	55	37	38	48
HO	14	18	12	16	9	22	14	18	21	15	**8**
LCM	19	24	14	15	**4**	26	16	20	29	23	19
LM	54	30	47	30	**27**	63	42	54	60	48	47
LP	42	48	41	29	**27**	52	34	33	47	31	34
Mean	33	32	33	26	**17**	41	26	34	40	31	30
Overall	33	31	39	27	**17**	48	27	36	46	34	32

Overall perceptibility threshold in terms of 98th percentile and mean ΔE^*_{ab} in the present study were compared with those founded by Uroz, Stokes and Song in Table 3. The results of this study gave good agreement with the other studies, i.e. a larger lightness weight comparing with chroma and hue weights. The present results have slightly lower threshold values than the others.

Table 3. Perceptibility threshold of complex images

	Present	Uroz	Stokes	Song
98th Percentile	3.0	4.4	n/a	n/a
mean	1.5	2.3	2.0	2.2

The perceptibility threshold in terms of 98th percentile ΔE^*_{ab} was approximately twice of mean. This aslo agrees with Uroz's results, which means that the distribution of colour difference in the test image was similar.

The results in Table 2 are summarised below:

- Comparing the original formulae (K_L=1), CIELAB performed the best followed by CIEDE2000, CMC and CIE94 which performed the worst.
- Comparing the K_L formulae, CIEDE2000(K_L=2) outperformed the others by a large margin with a CV value of 17, followed by CMC(K_L=1.5) and CIEDE2000(K_L =1.5).
- All formulae improved their performances by varying the K_L value. The improvements are large for all formulae except CIELAB (only 2 CV units). For imaging applications, the present results show that K_L values should be set at 1.5 and 2 for CMC and the rest of the formulae respectively.

- The CV values for lightness variations are larger than those for chroma and hue variations. When K_L is equal to 2, the performance of all the formulae except CIELAB improved significantly, especially CIEDE2000. The increase of K_L would reduce the effect of lightness differences, which means that lightness change is less noticeable compared to chroma and hue changes. These results agree with previous experimental results conducted by Uroz,[1] Stokes[2] and Song.[3]

The results from sCIELAB were not reported using the present data because the results are very similar to CIELAB. This is because that although sCIELAB takes into account spatial properties of human vision and uses low pass filters before colour-difference calculation, the colour difference calculation was based upon prominent colour patches which depend on its appearance on the image has similar effect to low pass filter.

Conclusions

A set of images including four original images were prepared by rendering each original image via five transfer functions: lightness multiplicative, lightness power, chroma multiplicative, chroma and lightness multiplicative, and hue additive offset. The experiment was conducted to establish the perceptibility thresholds at different colour directions. In addition, the performance of four different colour difference formulae, CIELAB, CMC, CIEDE2000, CIE94 were compared.

The perceptibility thresholds in terms of the 98th percentile of ΔE^*_{ab} and mean were 3.0 and 1.5 respectively. The results also showed image dependence, especially in the lightness direction. This could be due to the fact that the two outdoor test images studied having a large blue sky area gave smaller perceptibility threshold (or more noticeable perceived difference) than the other images. The results also showed that lightness differences were less noticeable than chroma and hue differences, which agreed with the previous experiments.

Concerning the performance of the four colour difference formulae, CIEDE2000(K_L =2) outperformed the others by a large margin. This is encouraging because this formula has been recommended by the CIE to be used for estimating small magnitude colour-differences with uniform colour patch samples.

Currently, the same set of samples are being circulated and assessed at different sites including some industrial companies. The results will be finally used to recommend a formula together with suitable thresholds for industrial applications.

References

1. J. Uroz, M. R. Luo and J. Morovic, Perception of colour difference between printed images, L. W. MacDonald and M. R. Luo (Eds), Colour Science: Exploiting digital media, John Wiley & Sons Ltd, (2002) 49-73.
2. M. Stokes, Colorimetric tolerances of digital images, MSc Thesis, College of Graphic Arts and Photography, Rochester Institute of Technology, Rochester, New York, U.S. 1991.
3. T. Song and M. R. Luo, Testing colour difference formulae on complex images using a CRT monitor, The eighth Color Imaging Conference, IS&T and SID, Scottsdale, Arizona, 7-10 November, (2000) 44-48.
4. CIE Publication No. 15.2, *Colorimetry*, CIE Central Bureau, Vienna, Austria, 1986.
5. F. J. J. Clarke, R. McDonald, and B. Rigg, Modification to the JPC79 colour-difference formula, *J. Soc. Dyers Col.*, 100, (1984) pp128-132.
6. CIE Publication No. 116, *Industrial Colour-Difference Evaluation*,, CIE Central Bureau , Vienna, Austria, 1995.
7. M. R. Luo, G. Cui and B. Rigg, The development of the CIE 2000 colour difference formula, *Color Res. Appl.* 26 (2001) 340-350.

Biography

Chikako Sano received her B. Eng. degree and M.Eng.degree in electrical engineering in 1988 and 1990 from Shizuoka University, Japan. In 1997 she joined Sony Corporation where she engaged the development of signal processing LSIs for digital still cameras. She received an MS degree in imaging science in 2003 from the University of Derby, UK. She is currently focusing on the color image processing and the image quality evaluation of digital photography.

Appendix A: Test Images

(1) Picnic

(2) Fruit

(3) Harbour

(4) Wool