OPTICAL REVIEW Vol. 8, No. 6 (2001) 487–494

Testing CIELAB-Based Color-Difference Formulae Using Large Color Differences

Haisong XU[1,2,*], Hirohisa YAGUCHI[1] and Satoshi SHIOIRI[1]

[1]*Department of Information and Image Sciences, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba, 263-8522 Japan,* [2]*State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China*

(Received April 9, 2001; Accepted August 3, 2001)

Three advanced CIELAB-based color-difference formulae, CMC, CIE94, and CIEDE2000, together with the basic CIELAB equation, were tested using large color-difference visual data (maximum average size was 12 CIELAB ΔE units) produced in this study. The color-difference comparison experiment was carried out at CIE Gray and Blue centers by a panel of 6 normal color-vision observers using CRT-generated stimuli based on the psychophysical method of constant stimuli. The experimental data, processed via probit analysis, were well fitted to chromaticity ellipses with a high reliance according to the observer accuracy in terms of $PF/3$ measure. A detailed comparison was performed to analyze the agreement between predicted color differences from all formulae and their corresponding visual scales in all measurement planes of CIELAB space. The results show that the CIEDE2000 marginally outperformed the others at all color centers while CIE94 was the worst in original formulae or with optimized k_L value, but the CIELAB performed worst when the parametric factors of k_L, k_C, and k_H were all optimized, with the CMC always lying between these extremes.

Key words: suprathreshold color difference, color-difference formula, color difference comparison, chromaticity ellipse, method of constant stimuli, probit analysis, CIELAB color space, CRT display

1. Introduction

Small color differences, including the color discrimination threshold, have been well studied since it is these differences that the color industries deal with primarily. However, in the field of color reproduction and industrial design concerning color image processing, the typical size of color difference could be more than 10 CIELAB ΔE units. Hence, large color-difference applications are also important and deserve to be researched. To date there have been few studies on large color differences,[1–4] and these with each other not agree very well. On the other hand, most of the existing color-difference formulae have been developed to evaluate small to medium color differences. Therefore, it is necessary to produce a new visual data set of large color differences and to estimate the performances of typical industrial color-difference formulae.

For the flexibility and ease of colorimetric characterization compared to object-color devices used in visual experiments,[5–8] CRT-generated stimuli were used in this study as in the authors' earlier study on color discrimination.[9] The aim of the present study was to test the CIELAB-based color-difference formulae, CMC,[10] CIE94,[11] and the newest CIEDE2000,[12] together with the basic CIELAB system,[13] using moderate to large color differences. The visual color-difference data were obtained from a specially designed psychophysical experiment carried out in the CIELAB color space based on the method of constant stimuli.

2. Methods

2.1 Apparatus and Stimuli

The experimental observations were performed in a booth which was totally dark. The color stimuli, displayed on the CRT monitor of a Sony Multiscan G500, were viewed by observers from a distance of 500 mm. The CRT display was driven by a visual stimulus generator system of Cambridge Research Systems VSG 2/4 with 15-bit resolution.

For comparison with the color discrimination study carried out previously,[9] the CIE Gray and Blue centers,[14,15] the most basic color and the most difficult color for perceptual characterization, respectively, were again selected as the test color centers. The CIELAB chromaticity parameters of these two centers are listed in Table 1.

The test stimulus pattern consisted of two color pairs, each of which included two 1° squares, in upper and lower positions, respectively, with a black frame of 0.1° at the center of a 6° background on the CRT display, as illustrated in Fig. 1. The two color pairs, designated as reference and test pair respectively, were presented, left and right, separated by a 0.5° visual angle. The total visual subtended angle size of the test stimulus was about 2.3°(height) × 2.7°(width), less than 4°, so the CIE1931 Standard Colorimetric Observer was used in calculations. The reference color was selected as gray with the chromaticity of D65 and a lightness (CIE Y) of 30, the same as the CIE Gray. The color differences of the reference pair were only the luminance variations or so-called gray scales along $+L^*$ direction in CIELAB space. For the test pair the color differences were the selected color distances, according to a predetermined step size, along 12 directions every 30° in (a^*, b^*)-plane and 8 directions every 45° in (a^*, L^*)- and (b^*, L^*)-plane from the two tested color centers. Hence the test pair was formed by the center colors and those stimuli evenly distributed around them. The background was set as Munsell N5 neutral gray with the chromaticity of D65, and was surrounded by a bright border of 8° visual angle with a luminance (L^*) of 100 cd/m^2 and also the chromaticity of D65. This border was presented to define the white point for the test pattern, and all the reference and test stimuli studied

---

*E-mail: chsxu@CISE.zju.edu.cn

Table 1. The CIELAB chromaticity parameters of the color centers studied. The CIE1931 Standard Colorimetric Observer was used in calculations.

Color center	L^*	a^*	b^*	C^*	$h°$
Gray	61.65	0.11	0.04	0.12	20
Blue	35.60	4.83	−30.18	30.56	279

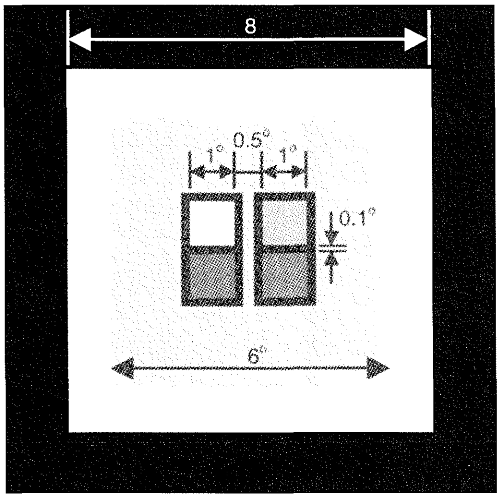


Fig. 1. The arrangement of test paradigm used in the present color-difference comparison experiment. Two color pairs, each formed by two 1° squares with a 0.1° black frame, represented the reference and test pair, respectively, presented on a 6° background of N5, with a separation of 0.5° visual angle between them. The test stimuli were surrounded by an 8° bright border set as the white point of the pattern.

here had a lower L^* than the surround, so they could be designed as simulated surface colors, as proposed by Indow *et al.*,[16] or related colors rather than aperture colors.

2.2 *Procedure*

A modified temporal gap condition, as shown in Fig. 2, was adopted in the present study. One cycle began with a 200 ms gap and ended when the response was received from the subject; this was no limitation on the subject's judgment time, so the period one cycle lasted differed for individual subjects. During gaps only the reference and test pairs were shut off with black, while the surrounding border and background remained so that the subject could maintain complete adaptation to the white point and background throughout the entire experiment.

Based on the principle of the psychophysical method of constant stimuli, the resultant data should be processed via the statistical method of probit analysis,[17–20] thus the measured stimuli were chosen by pilot experiments so that the color differences of the test pairs ranged from "always judged to be greater than the reference" to "never judged to be greater than the reference," with the majority lying between these two extremes.[14] The reference pair of samples had the required visual scales of color difference, including 4.0, 8.0, and 12.0 CIELAB ΔE units, in detail, in the present study.

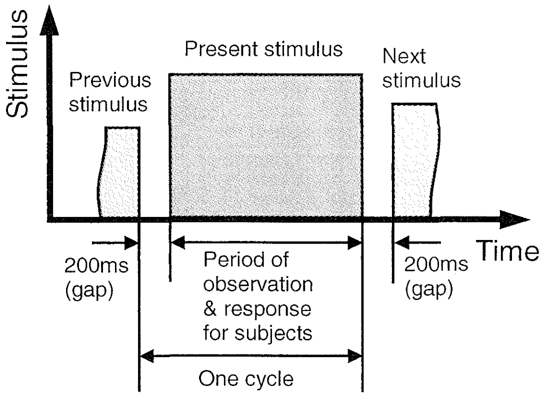


Fig. 2. The modified temporal gap condition adopted in the present experiment. One cycle began with a 200 ms gap and ended with receipt of the subject's response.

The task of the observer was to judge visually whether the color difference of the test pair was greater or less than that of the reference pair, and then to press one of the two keys on the keyboard which indicated his/her response; this stored the result and started the next trial. The judgments were repeated enough times to assign each test pair a probability of being judged to have a color difference greater than the reference pair. An iterative algorithm called probit analysis, a maximum-likelihood model that relates experimental response functions to probability-of-occurrence estimates, was used to find the most precise estimate at the tolerance of 50% rejection probability. This corresponded to the color-difference value visually equivalent to the reference color difference. Following the above procedure, the equal color-difference contours for each of the reference visual scales, 4.0, 8.0, and 12.0 CIELAB ΔE units, were determined to result in a new data set, which is analyzed below.

The experiment for each color center was separated into three sessions, one for each measurement plane, i.e., (a^*, b^*)-, (a^*, L^*)-, and (b^*, L^*)-plane. Each session started with a 3-min dark adaptation and a 1-min background adaptation and lasted about 20 min for most subjects. Each test pair was assessed 20 times, carried out in two separate sessions with random orders of color stimulus presentation, by three subjects for Gray and Blue centers, respectively, on a panel of 6 observers with normal color vision. All observers were students of Chiba University, most of whom had no experience in such a color-difference comparison experiment. During the experiment the stimulus arrangements of left and right and upper and lower positions of the reference and test pair were determined randomly by the computer program to avoid any judgment bias by observers. In every direction around each color center for each reference scale, 9 test pairs, corresponding to 9 points of color stimuli distributed in this direction, were compared with the reference pair. Therefore, a total of 30,240 judgments were made by each subject: 2 color centers × 3 reference scales × [12 directions in (a^*, b^*)-plane + 8 × 2 directions in (a^*, L^*)- and (b^*, L^*)-plane] × 9 test points for each direction × 20 times assessment for every test pair.

Table 2. The parameters of chromaticity ellipses, fitted with ΔE metrics of CIELAB, CMC, CIE94, and CIEDE2000 color-difference formulae in their original forms ($k_L = k_C = k_H = 1$), for the visual scales of 4.0, 8.0, and 12.0 CIELAB ΔE units at CIE Gray and Blue color centers in (a^*, b^*)-plane of the CIELAB space.

Color center	ΔE formula	Visual scale	A	A/B	θ (deg)	$\sqrt{\pi AB}$
Gray	CIELAB	4.0	5.26	1.95	113	6.67
		8.0	9.04	1.98	115	11.38
		12.0	11.98	1.91	117	15.38
	CMC	4.0	6.52	1.76	114	8.71
		8.0	9.90	1.71	115	13.43
		12.0	12.05	1.60	116	16.87
	CIE94	4.0	5.08	1.93	113	6.48
		8.0	8.64	1.96	115	10.95
		12.0	11.37	1.88	117	14.71
	CIEDE2000	4.0	5.17	1.53	132	7.40
		8.0	8.32	1.56	132	11.78
		12.0	10.51	1.50	135	15.19
Blue	CIELAB	4.0	9.14	2.69	121	9.89
		8.0	12.92	2.58	122	14.26
		12.0	16.84	2.58	119	18.59
	CMC	4.0	4.99	1.91	134	6.39
		8.0	7.23	1.93	136	9.22
		12.0	9.16	1.85	131	11.92
	CIE94	4.0	4.15	1.97	132	5.24
		8.0	6.06	1.99	135	7.62
		12.0	7.77	1.92	130	9.94
	CIEDE2000	4.0	3.31	1.27	33	5.20
		8.0	4.74	1.19	23	7.71
		12.0	6.18	1.12	63	10.36

3. Results

3.1 Observer Variation

The same performance factor ($PF/3$)[21,22] as in the earlier study[9] was used to compare two sets of data [Eq. (1)]. A higher $PF/3$ value implies a worse agreement between data sets, and a $PF/3$ of 30 indicates a disagreement of about 30%.

$$PF/3 = 100[(\gamma - 1) + V_{AB} + CV/100]/3. \quad (1)$$

In the color-difference comparison experiment, the visual observations were made by a panel of 6 subjects. The observer accuracy was represented, in the $PF/3$ measure, by the average deviation between the individual and mean visual results. It was 14 $PF/3$ units ranging from 12 to 15 for the most and least accurate observers, respectively. This corresponds to a typical error of about 6% ($14/\sqrt{6}$).

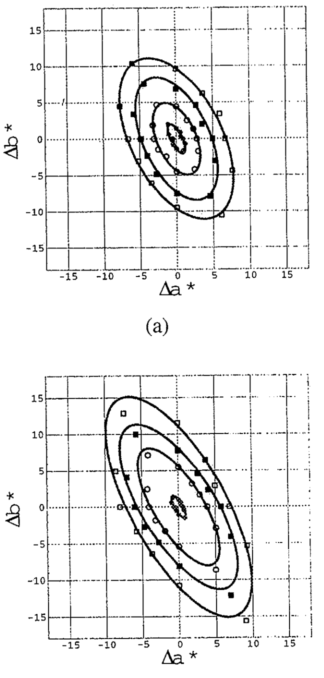
This observer variation was considered to be quite good compared with the earlier color discrimination study[9] with a $PF/3$ of 30 and other suprathreshold color-difference studies.[4,22–24] This is mainly due to the suprathreshold (large) rather than the threshold (very small) color differences being assessed, which are expected to result in higher accuracy especially when expressed as a percentage, and the pair comparison based on the method of constant stimuli was used in this study. In the present experiment, only greater or less color-difference judgments of the test pair compared with the reference pair were demanded from subjects which was much easier; thus more stable visual results could be obtained than by the gray scale method used by Guan and Luo.[4,22]

3.2 Color-Difference Formula Performance

3.2.1 Ellipse fitting

The experimental visual results were first summarized and compared as chromatic ellipses, since it is appropriate and effective to represent the contours of equally perceived color differences around a given center as ellipses in a color space.[16,25–27] In order to compare the color-difference metrics predicted by the advanced color-difference formulae, CMC, CIE94, and CIEDE2000, equal color-difference ellipses were fitted not only in CIELAB ΔE units but also in the ΔE units of the three CIELAB-based formulae in their original forms, i.e., $k_L = k_C = k_H = 1$. The resultant ellipse parameters, corresponding to visual scales of 4.0, 8.0, and 12.0 CIELAB ΔE units around the CIE Gray and Blue centers in (a^*, b^*)-plane, are listed in Table 2, and involve the semi-major axis (A), ratio of semi-major and semi-minor axes (A/B), orientation angle (θ), and the square root of ellipse area.

The results of fitted ellipses are consistent with the study by Guan and Luo[4] using large color differences of surface colors, except that the present ellipses are more elongated. This discrepancy is due to the different chromaticity parame-



(a)

(b)

Fig. 3. Chromaticity ellipses fitted with CIELAB ΔE metrics in (a^*, b^*)-plane for (a) CIE Gray and (b) CIE Blue color centers. From center to outer are threshold, visually equal color-difference contours corresponding to reference visual scales of 4.0, 8.0, and 12.0 CIELAB ΔE units, respectively, with the raw data plotted in different symbols (filled circles for threshold, open circles for 4.0-, filled squares for 8.0-, and open squares for 12.0-visual scales in terms of CIELAB ΔE units).

Table 3. The ellipse fitting accuracy in $PF/3$ measure, together with correlation coefficient (r), for CIE Gray and Blue color centers in respect to CIELAB ΔE metrics.

Color center		Gray		Blue	
Plane	Visual scale	r	$PF/3$	r	$PF/3$
(a^*, b^*)	4.0	0.973	6	0.972	8
	8.0	0.983	5	0.977	7
	12.0	0.994	3	0.975	7
	All	0.995	5	0.983	7
(a^*, L^*)	4.0	0.780	14	0.936	5
	8.0	0.948	8	0.850	7
	12.0	0.965	7	0.696	10
	All	0.973	10	0.969	8
(b^*, L^*)	4.0	0.397	14	0.546	6
	8.0	0.721	11	0.226	4
	12.0	0.796	8	0.625	5
	All	0.948	11	0.985	5
All planes	4.0	0.873	12	0.971	7
	8.0	0.938	8	0.971	6
	12.0	0.962	6	0.955	7
	All	0.977	9	0.982	7

ters of color centers and the different psychophysical methods of color-difference comparison used in the two studies. The orientations of all ellipses for CIELAB, CMC, and CIE94 are almost the same, but rather different from that of CIEDE2000, especially in the blue region. Similar is the case of A/B values, in which the A/B values of CIEDE2000 are smaller by a considerable margin, especially at the Blue center, than those of other formulae. These differences indicate that the CIEDE2000 performed rather differently from other formulae due to its rotation item and its relative uniform performance for predicting visual color differences, which will be discussed further in the next section of this paper. In addition, it is shown for all color-difference formulae that the shapes (A/B) and orientations (θ) of ellipses for all visual scales are very close except for CIEDE2000 at the Blue center, in which the values of A/B are near one, i.e., the shapes of contours are near circles (representing the ideal uniformity), so the orientation angles (θ) are not stable. Hence, the magnitudes of color-difference metrics predicted by all color-difference formulae are, in general, related to the visual scales with some kind of proportion, for instance, on the basis of linear or sub-additive relation to some degree as proposed by Witt[28,29] for small color differences, though they are not the same for different directions. This relationship is obvious in Fig. 3, which involves the chromaticity ellipses fitted in (a^*, b^*)-plane with respect to the CIELAB ΔE metrics. The color discrimination threshold ellipses, measured by the method described previously,[9] are also plotted in the figure for comparison. From center to outer are threshold, equally perceived color-difference contours corresponding to reference visual scales of 4.0, 8.0, and 12.0 CIELAB ΔE units, respectively, with the raw data plotted in different symbols. The raw data were well fitted to the ellipses, shown in Fig. 3, which can be verified from the fitting errors in terms of $PF/3$ measure and correlation coefficients (r) listed in Table 3.

In (a^*, b^*)-plane, the fitting accuracy for the Gray center was better than for the Blue center, but not in (a^*, L^*)- or (b^*, L^*)-plane. This would be due to the luminance component making the ellipse fitting less accurate in (a^*, L^*)- and (b^*, L^*)-plane for the Gray center, but this influence was not as strong for the Blue center, and thus resulted in a little better total accuracy of ellipse fitting. For any single measurement plane of the Gray center, the fitting error for a large visual color-difference scale was smaller than for a small one, but was not as obvious for the Blue center. This is mainly because the $PF/3$ unit is a relative measure expressed as a percentage, and the same absolute error for large scales corresponds to a smaller $PF/3$ value. That means the absolute accuracy for the large visual scale was worse than for the small one in the case of the Blue center, which is consistent with the correlation coefficients of ellipse fitting. Thus, in the blue region the determination of visual color differences is rather difficult. However, the total fitting errors of 9 and 7 $PF/3$ units for all three planes at the Gray and Blue centers, respectively, are good in comparison with the observer accuracy of 14 units, so the experimental results are reliable.

3.2.2 Visual Comparison

All the advanced color-difference formulae, as represented by CMC, CIE94, and the newest CIEDE2000, were derived

Table 4. Summary of the color-difference formula performance evaluated in $PF/3$ measure at CIE Gray and Blue color centers.

Color center		CIELAB	CMC	CIE94	CIEDE2000
For different visual scales with $k_L = k_C = k_H = 1$					
	4.0	26	19	27	**17**
Gray	8.0	25	17	26	**16**
	12.0	24	**16**	25	17
	4.0	29	27	30	**16**
Blue	8.0	26	27	30	**15**
	12.0	25	30	31	**18**
	4.0	28	28	32	**25**
Gray & Blue	8.0	26	28	34	**26**
	12.0	25	27	33	**25**
For different measurement planes with $k_L = k_C = k_H = 1$					
	(a^*, b^*)	27	26	27	**21**
Gray	(a^*, L^*)	31	**22**	32	**22**
	(b^*, L^*)	**21**	**21**	22	22
	(a^*, b^*)	41	31	33	**21**
Blue	(a^*, L^*)	**17**	27	29	**17**
	(b^*, L^*)	**18**	34	36	22
	(a^*, b^*)	37	35	36	**30**
Gray & Blue	(a^*, L^*)	**25**	26	33	28
	(b^*, L^*)	**20**	32	36	28
For all visual scales and measurement planes with $k_L = k_C = k_H = 1$					
Gray		29	24	30	**22**
Blue		32	33	35	**22**
Gray & Blue	31	32	37	**29**	
For all visual scales and measurement planes with optimized k_L					
Gray	k_L	1.02	1.19	1.02	1.21
	$PF/3$	29	25	29	**21**
Blue	k_L	0.79	1.15	0.75	0.78
	$PF/3$	32	**29**	46	**29**
Gray & Blue	k_L	0.93	1.17	0.91	1.06
	$PF/3$	31	31	39	**29**
For all visual scales and measurement planes with optimized k_L, k_C, and k_H					
Gray	k_L	1.07	1.16	1.07	1.23
	k_C	0.86	1.09	0.82	0.94
	k_H	0.31	1.24	0.31	0.89
	$PF/3$	27	24	26	**21**
Blue	k_L	0.80	1.19	0.80	0.82
	k_C	1.38	0.84	0.57	0.65
	k_H	0.68	0.73	0.48	0.69
	$PF/3$	26	26	26	**18**
Gray & Blue	k_L	0.94	1.18	0.96	1.09
	k_C	1.03	1.11	0.77	0.95
	k_H	0.72	0.70	0.45	0.64
	$PF/3$	30	28	29	**25**

by modifying the CIELAB equation, so these formulae are called CIELAB-based color-difference formulae and have the following common structure:

$$\Delta E = \sqrt{\left(\frac{\Delta L}{k_L S_L}\right)^2 + \left(\frac{\Delta C}{k_C S_C}\right)^2 + \left(\frac{\Delta H}{k_H S_H}\right)^2 + \Delta R}, \quad (2)$$

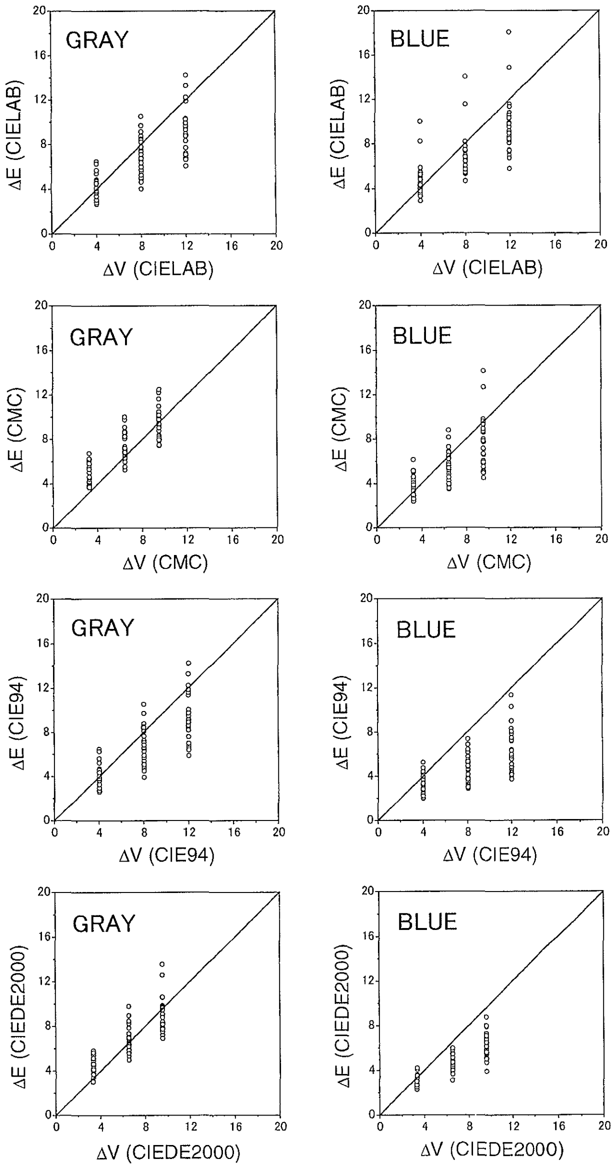
and

$$\Delta R = R_T \cdot f(\Delta C, \Delta H),$$

where ΔL, ΔC, and ΔH are the CIELAB metrics lightness, chroma, and hue differences ΔL^*, ΔC^*, and ΔH^*, respectively, calculated between the standard and sample in a color pair for CMC and CIE94 formulae; for the CIEDE2000 formula, however, only ΔL is the same as ΔL^*, and ΔC and ΔH are calculated differently as $\Delta C'_{ab}$ and $\Delta H'_{ab}$. Also, the interactive term ΔR between chroma and hue differences exists only in the CIEDE2000 system. The S_L, S_C, and S_H are the weighting functions for the lightness, chroma, and hue components, respectively, to improve the perceptual uniformity of CIELAB; and the k_L, k_C, and k_H are the parametric factors to be adjusted according to different viewing parameters such as textures, backgrounds, and separations for the lightness, chroma, and hue components, respectively. For CIE94, S_L equals one, and $k_L = k_C = k_H = 1$ under reference condition when the color-difference formula is designated as the original form.

In the present study, the three color-difference formulae, CMC, CIE94, and CIEDE2000, together with the basic CIELAB system, were tested with respect to their performance in predicting moderate to large visual color differences at CIE Gray and Blue color centers. The comparisons between color differences (ΔE) predicted by different formulae and the corresponding visual scales (ΔV) were carried out and resulted in the $PF/3$ units representing the visual prediction performances of individual color-difference formulae. The evaluation was first made using the original forms of all formulae, i.e., $k_L = k_C = k_H = 1$. In this case, the agreement accuracies between ΔE and ΔV were calculated separately for different visual scales (4.0, 8.0, and 12.0 CIELAB ΔE units) and measurement planes. Then, for each formula, the parametric factors, k_L, k_C, and k_H, were optimized for the data sets of Gray, Blue, and their combination, respectively, to account for the influence of the present experimental conditions on perceived color differences. To compare with other studies[4,12,22,24] and to analyze the relationship between lightness and chromatic differences of the four formulae studied, the optimizations were carried out by two methods: each formula's k_L value only was optimized with $k_C = k_H = 1$ to give the best fit to the visual scales; and k_L, k_C, and k_H of each formula were optimized at the same time to fit the visual color-difference scales. The test results are summarized in Table 4 with the best performance in each data set printed in bold font for ease of comparison.

Corresponding to different visual scales, the $PF/3$ units of each formula are very close, which means the absolute visual prediction errors of all formulae increased with the magnitude of visual color difference. In (a^*, b^*)-plane, the CIEDE2000 outperformed all other formulae, but this was not the case in (a^*, L^*)- or (b^*, L^*)-plane. This indicates that the relationship between the lightness and chromatic differences of the four formulae studied are quite different. For all visual scales and measurement planes, with the original forms of $k_L = k_C = k_H = 1$, the CIELAB, CMC, and CIE94 performed better at the Gray center than at the Blue center, while the performance of CIEDE2000 was the same for the two centers and the best (29 $PF/3$ units) among all formulae, followed by CIELAB (31) and CMC (32) with CIE94 the worst (37). This implies that the inclusion of a rotation function, R_T, in the CIEDE2000 is effective for improving the prediction in the blue region, and other modifications of CIEDE2000 for rescaling a^* axis in the near neutral area, correcting lightness and hue weighting functions *etc.* make it outperform all other formulae in the whole gamut. The CIELAB was a little better than CMC as a whole but, in fact, the CMC performed better than CIELAB at the Gray center. This means the CIELAB is still usable in practice; since the CMC was originally developed for the textile industry it is suitable for large color-difference prediction. The CIE94 has the simplest structure excluding CIELAB, and was defined as being for use in its original form only under its reference condition. The poor performance of CIE94 may be due to the present experimental parameters being different from its reference viewing condition. Figures 4(a)–(b) show the predicted color differences (ΔE) by four color-difference formulae with $k_L = k_C = k_H = 1$ against the visual cololor-difference scales (ΔV) for Gray and Blue data sets, respectively. It can be seen that the scatter from the CIEDE2000 is narrowest among these four formulae. Although the scatter of CMC at the Gray center is a little narrower than those of CIELAB and CIE94, the reference visual color-difference values from CMC were smaller, so a larger percentage error



(a) (b)

Fig. 4. Predicted color difference (ΔE) by CIELAB, CMC, CIE94, and CIEDE2000 color-difference formulae, with their original forms ($k_L = k_C = k_H = 1$), plotted against the visual color difference scale (ΔV) for (a) Gray and (b) Blue color centers.

in $PF/3$ units occurred for the CMC formula. The predicted color differences (ΔE) from CIE94 did not scatter wider than those of CIELAB and CMC, but their correlation with ΔV was worse (correlation coefficient $r = 0.669$) than those of CIELAB ($r = 0.732$) and CMC ($r = 0.719$). These analyses agree with the $PF/3$ values listed in Table 4.

With the optimized k_L and $k_C = k_H = 1$, none of the formulae improved their performance except for the CMC, of which the $PF/3$ value slightly decreased, but the CIEDE2000 still outperformed others. In fact, the optimized k_L values are near one, i.e. 0.93, 1.17, 0.91, and 1.06 for CIELAB, CMC, CIE94, and CIEDE2000, respectively, for the combined data set of the Gray and Blue centers. This indicates that the parametric effects of the present viewing condition on lightness difference were weak for all four formulae studied here. When the three parametric factors k_L, k_C, and k_H were optimized at the same time, all formulae improved their performance. The CIE94 improved the most and CIELAB the least with CIEDE2000 and CMC lying between them. At the Blue center, the $PF/3$ measure of CIEDE2000 was reduced to 18 units, much better than those of other formulae, which all performed the same. This shows the excellent visual prediction performance of the CIEDE2000 formula in the blue region as well as the gray region, due to its uniformity of color-difference metrics across the whole color space of this model. The CIE94 was sensitive to k_C and k_H, implying that the balance of chroma and hue weighting functions is not good; the $PF/3$ value of CIELAB remained almost unchanged even with the optimized k_L, k_C, and k_H, indicating this formula does not have an appropriate structure to improve its performance on the basis of viewing parametric modification for the condition of this study. The CMC could be improved by adjusting the parametric factors, but its balance of lightness, chroma, and hue weighting functions was worse than that of CIEDE2000, of which the optimized k_L and k_C were very near one, i.e. 1.09 and 0.95, respectively, for the combined data set of Gray and Blue centers, although its optimized k_H (0.64) was somewhat smaller due to the influence of blue colors. With regard to the behavior of the interaction among lightness, chroma, and hue components in the color-difference perception for the four formulae with optimized k_L, k_C, and k_H values, their performance ranks (from best to worst) changed from CIEDE2000, CIELAB, CMC, and CIE94 to CIEDE2000, CMC, CIE94, and CIELAB. These ranks are also different from that at the color discrimination threshold in the earlier study,[9)] where the CIEDE2000 and CIELAB performed better than CIE94 while the CMC was worst for the same Gray and Blue centers. This shows that the perception of large color differences is very different from that of small (including threshold) color differences.

From the above comparison, a robust tendency is found that the CIEDE2000 outperformed all other formulae in all situations, with original forms or optimized parametric factors, across the whole gamut. The CMC performed better than CIE94 due to its ability to predict large color differences, while the CIELAB was not bad in its original form but could not be improved by optimizing k_L, k_C, and k_H factors due to its poor formula structure.

4. Conclusions

Based on the psychophysical method of constant stimuli, a color-difference visual comparison experiment was carried out using large color-difference (maximum average size was 12 CIELAB ΔE units) stimuli on a CRT display. The resultant equally perceived color-difference contours corresponding to different visual scales were well fitted to chromaticity ellipses, which were found to extend in a regular proportion with the magnitude of increasing visual scale.

The new data sets obtained in the present study were used to test the three advanced CIELAB-based color-difference formulae, CMC, CIE94, and CIEDE2000, together with the basic CIELAB equation. In terms of $PF/3$ measure, the CIEDE2000 formula performed best no matter its original form or with k_L alone or all of k_L, k_C, and k_H being optimized. With optimized k_L and $k_C = k_H = 1$, the performance ranks of all formulae were the same as those in their original forms, and there was no obvious difference between CIELAB and CMC, both of which outperformed CIE94. Due to the high sensitivity of CIE94 to the parametric factors, its performance was greatly improved to be better than CIELAB with optimized k_L, k_C, and k_H. The CMC formula was only inferior to CIEDE2000 and still better than CIE94, while the CIELAB performed worst since its poor formation was hard to improve by optimizing parametric factors. It is confirmed that the CIEDE2000 outperformed, at the cost of complex calculation, other formulae much better in predicting suprathreshold color differences than color discrimination thresholds.

Acknowledgments

The authors would like to thank Professor M. Ronnier Luo for providing technical guidance and all the observers for their patience and time.

References

1) G. Wyszecki and H. Wright: J. Opt. Soc. Am. **55** (1965) 1166.
2) D. L. MacAdam: J. Opt. Soc. Am. **64** (1974) 1691.
3) T. Fuchida, W. B. Cowan and G. Wyszecki: Color Res. Appl. **10** (1985) 92.
4) S. S. Guan and M. R. Luo: Color Res. Appl. **24** (1999) 356.
5) R. S. Berns, R. J. Motta and M. E. Gorzynski: Color Res. Appl. **18** (1993) 299.
6) E. D. Montag and R. S. Berns: Color Res. Appl. **24** (1999) 164.
7) M. Melgosa, M. M. Perez, A. El Moraghi and E. Hita: Color Res. Appl. **24** (1999) 38.
8) V. C. Smith, J. Pokorny and H. Sun: Color Res. Appl. **25** (2000) 105.
9) H. Xu, H. Yaguchi and S. Shioiri: Opt. Rev. **8** (2001) 142.
10) F. J. J. Clarke, R. McDonald and B. Rigg: J. Soc. Dyers Colour **100** (1984) 128.
11) CIE: *Industrial colour difference evaluation*, CIE Publication No. 116 (CIE Central Bureau, Vienna, 1995).
12) M. R. Luo, G. Cui and B. Rigg: Color Res. Appl. **26** (2001) 340.
13) CIE: *Colorimetry*, CIE Publication 15.2 (CIE Central Bureau, Vienna, 1986).
14) A. R. Robertson: Color Res. Appl. **3** (1978) 149.
15) K. Witt: Color Res. Appl. **20** (1995) 399.
16) T. Indow, A. R. Robertson, M. von Grunau and G. H. Fielder: Color Res. Appl. **17** (1992) 6.
17) D. Strocka, A. Brockes and W. Paffhausen: Color Res. Appl. **8** (1983) 169.
18) D. H. Alman, R. S. Berns, G. D. Snyder and W. A. Larsen: Color Res. Appl. **14** (1989) 139.
19) T. Indow and M. L. Morrison: Color Res. Appl. **16** (1991) 42.

20) R. S. Berns, D. H. Alman, L. Reniff, G. D. Snyder and M. R. Balonon-Rosen: Color Res. Appl. **16** (1991) 297.
21) M. R. Luo and B. Rigg: J. Soc. Dyers Colour **103** (1987) 126.
22) S. S. Guan and M. R. Luo: Color Res. Appl. **24** (1999) 331.
23) M. R. Luo and B. Rigg: J. Soc. Dyers Colour **103** (1987) 86.
24) S. S. Guan and M. R. Luo: Color Res. Appl. **24** (1999) 344.
25) M. Cheung and B. Rigg: Color Res. Appl. **11** (1986) 185.
26) K. Witt: Color Res. Appl. **15** (1990) 189.
27) M. Melgosa, E. Hita, A. J. Poza, D. H. Alman and R. S. Berns: Color Res. Appl. **22** (1997) 148.
28) K. Witt: Color Res. Appl. **20** (1995) 36.
29) K. Witt: Color Res. Appl. **24** (1999) 78.