OPTICAL REVIEW Vol. 8, No. 2 (2001) 142–147

Estimation of Color-Difference Formulae at Color Discrimination Threshold Using CRT-Generated Stimuli

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 December 26, 2000; Accepted February 14, 2001)

The advanced color-difference formulae, CMC, CIE94, and the recently proposed CIEDE2000 formula, together with the basic CIELAB system, were estimated using the chromaticity discrimination threshold data at CIE Gray and Blue color centers. Gray is the most basic color and blue is, perceptually, in the most different region from other areas across the color space. The test stimuli, evenly distributed in the (a^*, b^*)-, (a^*, L^*)-, and (b^*, L^*)-plane of the CIELAB space, were generated on a CRT display, driven by a VSG system. Each direction from each color center was assessed 3 times by a panel of 9 normal color-vision observers with the psychophysical method of interleaved staircase. The experimental data were reliable and consistent with other studies according to the observer accuracy and fitted ellipse parameters. A comprehensive analysis shows the color discrimination tolerances could be well fitted by ellipses, and the CIEDE2000 and CIELAB formulae performed better than CIE94 while the CMC worst at the threshold level for the color centers studied.

Key words: color difference, color-difference formula, color discrimination threshold, CRT, method of staircase, chromaticity discrimination ellipse, CIELAB color space

1. Introduction

Since the CIELAB space[1)] was recommended by the International Commission on Illumination (CIE) in 1976, for promoting uniformity of color-difference practice, significant advances have been achieved in this field. Alongside the development in color vision research, new advanced color-difference formulae based on CIELAB have been proposed, aiming at a final goal[2)] of developing an universal standard of color-difference evaluation for most industrial applications. Among these, CMC,[3)] CIE94,[4)] and the newest CIEDE2000,[5)] proposed recently by CIE TC1-47 Hue and Lightness-Dependent Correction to Industrial Color-Difference Evaluation, are the representative models. All these color-difference formulae were developed to fit some experimental data sets. Therefore, their performances, in general, need to be tested with different data sets for practical application and further improvement.

An ideal color-difference formula is expected to perform well for the range from small, through moderate, to large color differences. In practice, it is very difficult for a single formula to give out the color-difference metrics equal to perceptual scales over the whole range. At least such a color-difference formula has not yet been achieved. However, different formulae can meet the needs of different applications, with respect to a specified range of color differences, with the level of color discrimination threshold being the common basis for all of these color-difference models.

On the other hand, the CRT is being widely used[6–8)] in vision research not only because of its efficiency, saving of time, labor, and cost needed to produce object-color samples, but also its flexibility; it makes easy the selection of test stimuli and makes it possible to study the parametric effects on color-difference judgments.

The main purpose of the present study was to estimate the performances of the advanced color-difference formulae, CMC, CIE94, and CIEDE2000, together with the basic CIELAB system, at the color discrimination threshold. The chromaticity discrimination data were obtained from a psychophysical experiment carried out in CIELAB color space using CRT-generated stimuli, described below.

2. Methods

2.1 *Apparatus and Stimuli*

A CRT monitor of Sony Multiscan G500 was used for presentation of color stimuli. The CRT was driven by a visual stimulus generator system, Cambridge Research Systems VSG 2/4, with 15-bit resolution. Observations were made in a booth in which the environment was black, and the viewing distance, from CRT to the eyes of observers, was 500 mm.

The stimuli to be measured were evenly distributed along 12 directions every 30° in (a^*, b^*)-plane and in 8 directions every 45° in (a^*, L^*)- and (b^*, L^*)-plane from the CIE Gray and Blue color centers,[2,9)] of which the CIELAB values are listed in Table 1. The CIE Gray and Blue centers were selected as the test color centers, because gray is considered the most basic color and blue the most difficult color to visually predict by color-difference formulae, due to the most different perceptual characteristics in the blue region with respect to the tolerance ellipses not pointing toward the neutral point.[5)]

The test stimulus was a square array of four 1° squares

E-mail: hsxu@vision.tp.chiba-u.ac.jp

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

Color center	L^*	a^*	b^*
Gray	61.65	0.11	0.04
Blue	35.60	4.83	−30.18

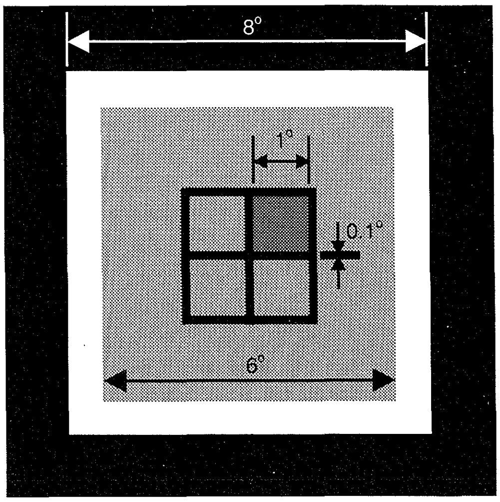

Fig. 1. The test stimulus paradigm used in the present color discrimination experiment. A four-square array stimulus, with 0.1° separation and frame, was presented on a 6° background of center color, surrounded by a bright border, at 8° visual angle, set as the white point of the image.

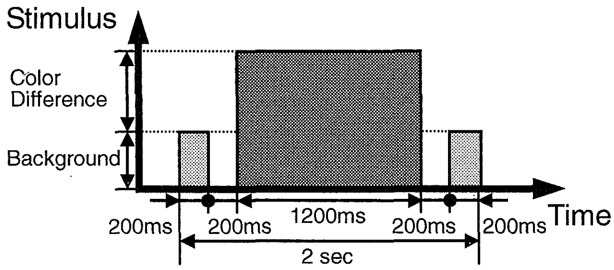

Fig. 2. The temporal gap condition used in the present experiment. One cycle of 2 seconds included two 200 ms gaps of background color and black at the beginning and the end of the 1200 ms test stimulus.

with a small black separation of 0.1° visual angle. The subtended visual angle of the test stimulus at the center of the CRT display was about 2°, less than 4°, so the CIE1931 Standard Colorimetric Observer was used in calculations. The square array, with a black frame of 0.1°, was presented on a 6° background set as the center color. The pattern, illustrated in Fig. 1, was surrounded by a bright border, with a visual subtended angle of 8°, having a luminance of 100 cd/m^2 and the chromaticity of D65. This "white" border was displayed in order to have the CRT stimuli appear as related colors and to define the white point for the image.

2.2 Procedure

A temporal gap condition was used in the experiment. As shown in Fig. 2, one cycle was 2 seconds, including two periods of 200 ms of background color and black gaps, respectively, at the beginning and end of the test stimulus which lasted for 1200 ms. A long but limited period of 2 seconds was selected to simulate the practical temporally unlimited viewing condition, and to control the whole duration of the experiment. In fact, the responding time generally was less than 2 seconds for all observers, so observer judgment was not influenced by the limited presenting time of test stimuli under such a temporal condition.

During gaps all areas, including the background and the 4-square array, were shut off with black except for the surrounding border, which remained to hold the complete adaptation of the observer to the white point. For each trial, the test color was presented on one of the four squares determined randomly by the software, with the other three squares remaining the same color as the background, according to the predicted color distance from the color center being studied by the psychophysical method of staircase (explained later). The visual task of the observer was to judge the position of the square where a color different from the background color was perceived, and then to press the corresponding key on the keyboard as his/her response; this stored the result and started the next trial.

The experiment for each color center was divided into three sessions, one for each measurement plane, i.e. (a^*, b^*)-, (a^*, L^*)-, and (b^*, L^*)-plane. Each session lasted no more than 25 min, including 3-min dark adaptation and 1-min background adaptation before the practical judgment. To avoid possible bias by the observers from the presenting sequence of the test stimuli, an interleaved staircase method was used. Each staircase involved a group of stimuli in four randomly selected directions with random presenting sequences. Each direction was assessed 3 times by individuals on a panel of 9 observers with normal color vision. Observers were students of Chiba University, and most of them had no experience in such a color discrimination experiment. In the initial trials of the staircase, an obviously discriminable step was presented. Then the step size decreased systematically until a criterion value, determined by pilot experiments to produce an efficient staircase, was reached to produce ten reversals. The averages of the ten reversals were calculated as thresholds.

3. Results

3.1 PF/3 Measure

To ease the comparison between two sets of data, a performance factor (PF), first devised by Luo and Rigg[10] and then modified to PF/3 by Guan and Luo,[11] and given in Eq. (1), was used in this study.

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

where VS and γ, proposed by Coates *et al.*,[12] and V_{AB}, derived by Schultz,[13] are given in Eqs. (2)–(4), respectively.

$$CV=\frac{\sqrt{\frac{1}{N}\sum (X_i-fY_i)^2}}{\bar{X}}\times 100, \quad (2)$$

and

$$f=\frac{\sum X_iY_i}{\sum Y_i^2},$$

$$\log_{10}(\gamma)=\sqrt{\frac{1}{N}\sum\left[\log_{10}\left(\frac{X_i}{Y_i}\right)-\overline{\log_{10}\left(\frac{X_i}{Y_i}\right)}\right]^2}, \quad (3)$$

$$V_{AB}=\sqrt{\frac{1}{N}\sum\frac{(X_i-FY_i)^2}{X_iFY_i}}, \quad (4)$$

and

$$F=\sqrt{\sum\frac{X_i}{Y_i}\Big/\sum\frac{Y_i}{X_i}},$$

where N is the number of compared pairs, and X_i and Y_i are values of pair i. When evaluating the goodness of fit of a color-difference formula using the above measures, X and Y sets are the color-difference metrics calculated from the tested formula and visual scales, respectively. A PF/3 of 30 means a 30% prediction error by that formula.

3.2 *Observer Accuracy*

Each session of the color discrimination experiment was carried out 3 times by a panel of 9 observers, and the mean values were used in the next analysis. Thus the PF/3 measure between the results of an individual observer and the mean visual results can be considered as the inter-observer accuracy. It was 30 units ranging from 16 to 41 for the most and least accurate observers, respectively. This corresponds to an error in the mean of the measurements for each direction of about 10% ($30/\sqrt{9}$).

The intra-observer accuracy was also calculated between each measurement and the mean value of the 3 ones made by each observer. It was found the typical observer repeatability was also 30 PF/3 units ranging from 19 to 40, corresponding to a standard deviation of 10%.

Such observer accuracy was considered to be good compared with other studies.[11,14–16] The observer repeatability performance was not very satisfactory, which was expected to be better than the inter-observer accuracy. However, the observer accuracy confirms that the experimental results in the present study are believable and repeatable.

3.3 *Color Discrimination Ellipses*

Since the pioneering work carried out by MacAdam in 1942,[17] contours of equally perceived color differences around a given center are usually represented as ellipses or ellipsoids in a color space.[18–21] The present experimental results were also fitted into color discrimination ellipses, defined by the semi-major axis (A), ratio of the semi-major and semi-minor axes (A/B), and orientation angle (θ). These parameters (together with the square root of ellipse area) of the fitted ellipses corresponding to the CIE Gray and Blue centers are summarized in Table 2.

Table 2. Color discrimination ellipse parameters for CIE Gray and Blue color centers in CIELAB space.

Color center	Plane	A	A/B	θ(deg)	$\sqrt{\pi AB}$
Gray	(a^*, b^*)	1.86	3.18	109	1.85
	(a^*, L^*)	1.50	2.30	86	1.75
	(b^*, L^*)	1.46	1.19	55	2.37
Blue	(a^*, b^*)	1.85	5.13	125	1.45
	(a^*, L^*)	1.41	3.46	114	1.35
	(b^*, L^*)	0.86	1.83	129	1.13

Table 3. The fitting accuracy of chromaticity discrimination ellipses in PF/3 measure for CIE Gray and Blue color centers.

Color center	Gray	Blue
(a^*, b^*)-plane	7	16
(a^*, L^*)-plane	9	22
(b^*, L^*)-plane	12	16
All planes	10	18

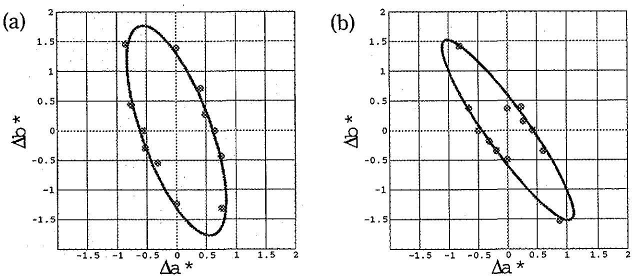

Fig. 3. Chromaticity discrimination ellipses in (a^*, b^*)-plane for (a) CIE Gray and (b) CIE Blue color centers with the raw data plotted by solid dots.

The parameters of fitted ellipses are consistent with other studies[11,16] for small color difference with surface colors, except that the present ellipses are more elongated. The present ellipse size for the Blue center is smaller, while for the Gray center it is a little larger. This discrepancy is due to the use of the four-square stimuli on CRT and the staircase method for the color discrimination measurement used in the present study, rather than the surface color samples and gray scale method used by Guan and Luo.[11,16]

Figure 3 shows the obtained chromaticity ellipses in (a^*, b^*)-plane with the raw data plotted in solid dots. The raw data were well fitted to the ellipses, which can be clearly seen from the fitting errors in PF/3 measure listed in Table 3. The fitting accuracy for the Gray center was better than for the Blue center, and that in the (a^*, b^*)-plane was the best among the three measurement planes, since in the (a^*, L^*)- and (b^*, L^*)-planes the luminance component made the ellipse fitting less accurate. However, for the combined data sets of all three planes, the total error of 10 and 18 PF/3 units at Gray and Blue centers, respectively, is quite good in comparison with the observer accuracy of 30 units, so the resultant data are reliable. In addition, it is well confirmed that the color discrimination data can be fitted accurately by ellipses as the mathematical model. Furthermore, the chromatic and luminance channels function differently in the color vision system, and determination of the characteristics at the Blue center is rather difficult.

3.4 Color-Difference Formula Evaluation

As mentioned above, the advanced color-difference formulae CMC[3] and CIE94[4] are modifications from the basic CIELAB,[1] and so is the recently proposed CIEDE2000 formula.[5] Compared to CIE94, the CIEDE2000 formula includes the following modifications: (1) rescale the a^* axis in the near neutral area, (2) modify the lightness weighting function, (3) take hue-angle into account in the hue weighting function, and (4) an ellipse rotation part similar to BFD[10,14] and LCD[22] formulae is included to reflect the phenomenon that in the blue area the tilt of the major axis of the tolerance ellipse is in the counterclockwise direction and away from the direction of constant hue angle. These modifications make the calculation of the CIEDE2000, given in Eq. (5), much more complicated than that of the CIE94, but the CIEDE2000 is expected to give more accurate results,

$$\Delta E_{00}=\sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2+\left(\frac{\Delta C'_{ab}}{k_C S_C}\right)^2+\left(\frac{\Delta H'_{ab}}{k_H S_H}\right)^2+R_T\left(\frac{\Delta C'_{ab}}{k_C S_C}\right)\left(\frac{\Delta H'_{ab}}{k_H S_H}\right)}, \quad (5)$$

where

$S_L=1+\frac{0.015(\overline{L'}-50)^2}{\sqrt{20+(\overline{L'}-50)^2}}$, $S_c=1+0.045\overline{C'_{ab}}$, and

$S_H=1+0.015\overline{C'_{ab}}T$, where

$$T=1-0.17Cos(\overline{h'_{ab}}-30°)+0.24Cos(2\overline{h'_{ab}})$$
$$+0.32Cos(3\overline{h'_{ab}}+6°)-0.20Cos(4\overline{h'_{ab}}-63°),$$

and

$\Delta H'_{ab}=2\sqrt{C'_{ab,b}C'_{ab,s}}Sin\left(\frac{\Delta h'_{ab}}{2}\right)$, $R_T=-Sin(2\Delta\theta)R_C$, where

$\Delta\theta=30\exp\{-[(\overline{h'_{ab}}-275°)/25]^2\}$, and $R_C=2\sqrt{\frac{\overline{C'_{ab}}^7}{\overline{C'_{ab}}^7+25^7}}$,

where

$L'=L^*$, $a'=(1+G)a^*$, $b'=b^*$, $C'_{ab}=\sqrt{a'^2+b'^2}$, and

$h'_{ab}=\tan^{-1}(b'/a')$, where

$$G=0.5\left(1-\sqrt{\frac{\overline{C^*_{ab}}^7}{\overline{C^*_{ab}}^7+25^7}}\right),$$

where $\overline{L'}$, $\overline{C'_{ab}}$, $\overline{h'_{ab}}$, and $\overline{C^*_{ab}}$ are the arithmetic means of the L', C'_{ab}, h'_{ab} (in degree), and C^*_{ab} values for a pair of samples.

In this study, the above four color-difference formulae, CIELAB, CMC, CIE94, and CIEDE2000, were tested in terms of their performance in predicting the color discrimination threshold at CIE Gray and Blue color centers. In general, it would be desirable for computed color differences ΔE between the color center and points on the contour of color discrimination threshold to have a constant value ΔV, and the ideal value of ΔV should be 1 ΔE unit. Therefore, the comparison between ΔE predicted by the four color-difference formulae and the objective visual scale $\Delta V=1$ was carried out to obtain the performance results, in terms of PF/3 measure, summarized in Table 4. All color-difference formulae were used in their original forms, i.e., $k_L=k_C=k_H=1$.

Table 4. Evaluating color-difference formulae in PF/3 units at threshold for CIE Gray and Blue color centers.

Color center	ΔE formula	(a^*, b^*)	(a^*, L^*)	(b^*, L^*)	All planes
Gray	CIELAB	42	34	12	37
	CMC	42	12	24	35
	CIE94	42	34	12	37
	CIEDE2000	45	30	11	36
Blue	CIELAB	62	42	28	50
	CMC	56	53	43	52
	CIE94	55	53	44	55
	CIEDE2000	55	42	38	48
Gray & Blue	CIELAB	57	46	47	52
	CMC	89	52	74	76
	CIE94	74	56	69	69
	CIEDE2000	71	51	69	65

It was found from these results that, for the combined data set of Gray and Blue, CIELAB performed best, followed by CIEDE2000 and CIE94, with CMC worst, while CIEDE2000 performed best for the Blue center. This finding is basically meaningful, but the PF/3 units listed in Table 4 are not completely consistent with their scatter diagrams as shown in Fig. 4, in which the ΔE values calculated from the four formulae are plotted against the distribution (directions in degree) of the test stimuli in all three measurement planes, i.e., (a^*, b^*)-, (a^*, L^*)-, and (b^*, L^*)-plane. The dotted lines represent the ideal visual scale $\Delta V=1$, and the solid lines are the fitted lines. Although the correlation coefficients of linear fitting are rather low, the slopes and intercepts are useful for evaluating the uniformity of the color differences predicted by the individual color-difference formulae. The nearer to zero is the slope of a fitted line, the better is the local uniformity of the corresponding ΔE predicted by a given color-difference formula in different directions

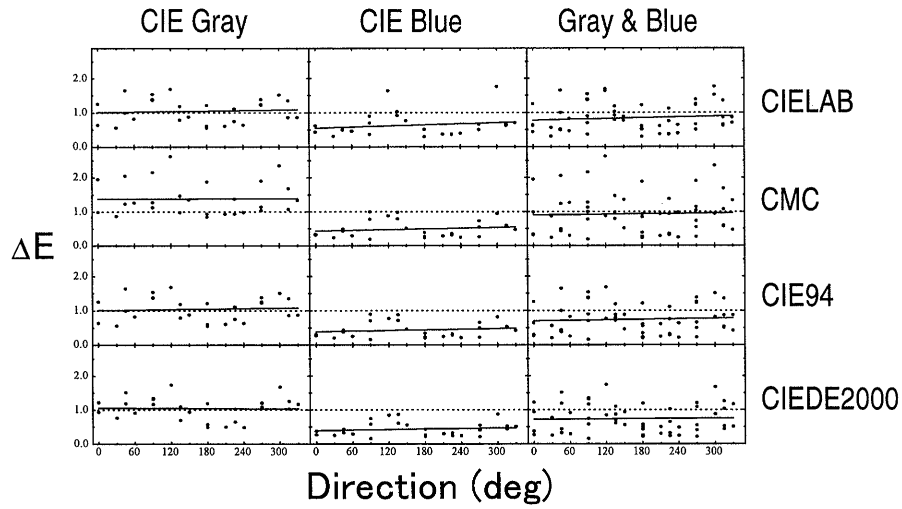

Fig. 4. Predicted ΔE values calculated from four color-difference formulae plotted against the distribution (directions in degree) of test stimuli in CIELAB space. The solid lines are linear fitted results and dotted lines represent the ideal objective visual scale $\Delta V=1$ for color discrimination.

around a color center; the nearer to 1.0 is the intercept of a fitted line or, more accurately, the mean of the ΔE values, the better is the consistency of the predicted color differences between different regions of the color space. These two evaluations were called isotropy and homogeneity, respectively, by Melgosa.[23] For example, Table 4 shows the CMC performed best at the Gray center, while the left-hand plots of Fig. 4 show the CIEDE2000 performed best and the CMC worst from both of slopes and intercepts of fitted lines. At the Blue center, the predicted ΔE values are all below the ideal line of $\Delta V=1$, so the area of the color discrimination contour is relatively small as mentioned above. However, the isotropy performance, the more important indicator of the goodness of the color-difference formula tested,[23] of CIEDE2000 was a little better than the others, which is consistent with the PF/3 measure in Table 4. The CIEDE2000 did not perform as well as expected due to the value of added rotation item R_T being very small, in fact almost zero, for the Blue center studied with the chroma not high enough, so its improvement in the blue region was not obvious in the present study. In addition, the present experimental conditions were different from the reference conditions, under which the original form of the CIEDE2000 formula was developed with the parametric factors, $k_L=k_C=k_H=1$. Again, for the combined data set of Gray and Blue, the uniformity of CIEDE2000 was the best, while the predicted ΔE by CIELAB was nearer to $\Delta V=1$. For the CMC, the fitted line seemed nearest to $\Delta V=1$, but this was the result of the counteraction of the plus errors for Gray and the minus errors for Blue, rather than actually a good performance.

As a whole, for the range studied in the present experiment, the performances of CIELAB and CIEDE2000 were better than CIE94, while CMC was worst. Hence CIELAB is still useful at the level of color discrimination threshold, and CIEDE2000 is indeed improved in comparison with CIE94, while the CMC formula was originally proposed for textile application, whose viewing condition and based data sets are very different from the present experiment.

4. Conclusions

A chromaticity discrimination experiment was carried out using CRT-generated stimuli with the psychophysical method of staircase. The resultant threshold data were well represented by ellipses as the fitting model. The present data are quite reliable and consistent with other studies according to the PF/3 measure and ellipse parameters.

Three advanced color-difference formulae, CMC, CIE94, and CIEDE2000, together with the basic CIELAB system, were evaluated using the color discrimination data in the present study. According to the overall consistency (homogeneity) and local uniformity (isotropy) and based on the PF/3 assessment and the scatter diagram plots, the CIEDE2000 and CIELAB performed better than CIE94, while the CMC was worst, at the level of color discrimination threshold for the CIE Gray and Blue color centers studied.

Hence, at the color discrimination threshold level CIELAB is still meaningful and CIEDE2000 outperforms CIE94 despite its complexity. It is expected that the CIEDE2000 will perform much better for suprathreshold color differences, while the predicting performance for visual scales at threshold is at least one of the basic aspects.

References

1) CIE: *Colorimetry*, CIE Publication 15.2 (CIE Central Bureau, Vienna, 1986).
2) K. Witt: Color Res. Appl. **20** (1995) 399.

3) F. J. J. Clarke, R. McDonald and B. Rigg: J. Soc. Dyers Col. **100** (1984) 128.
4) CIE: *Industrial colour difference evaluation*, CIE Publication No.116 (CIE Central Bureau, Vienna, 1995).
5) M. R. Luo, G. Cui and B. Rigg: Color Res. Appl. **26** (2001) 000.
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) A. R. Robertson: Color Res. Appl. **3** (1978) 149.
10) M. R. Luo and B. Rigg: J. Soc. Dyers Col. **103** (1987) 126.
11) S. S. Guan and M. R. Luo: Color Res. Appl. **24** (1999) 331.
12) E. Coates, K. Y. Fong and B. Rigg: J. Soc. Dyers Col. **97** (1981) 179.
13) W. Schultz: *Color Metrics* (AIC, Soesterberg, 1972) p. 245.
14) M. R. Luo and B. Rigg: J. Soc. Dyers Col. **103** (1987) 86.
15) S. S. Guan and M. R. Luo: Color Res. Appl. **24** (1999) 344.
16) S. S. Guan and M. R. Luo: Color Res. Appl. **24** (1999) 356.
17) D. L. MacAdam: J. Opt. Soc. Am. **32** (1942) 247.
18) M. Cheung and B. Rigg: Color Res. Appl. **11** (1986) 185.
19) K. Witt: Color Res. Appl. **15** (1990) 189.
20) T. Indow: Color Res. Appl. **17** (1992) 6.
21) M. Melgosa, E. Hita, A. J. Poza, D. H. Alman and R. S. Berns: Color Res. Appl. **22** (1997) 148.
22) D.-H. Kim and J. H. Nobbs: *Proc. Colour 97, Kyoto, Japan*, (1997) Vol. 1, p. 446.
23) M. Melgosa: Color Res. Appl. **25** (2000) 49.