

Parametric effects in color-difference evaluation

QIANG XU, KEYU SHI, AND MING RONNIER LUO*

State Key Laboratory of Modern Optical Instrument, Zhejiang University, Hangzhou, China

*m.r.luo@zju.edu.cn

Abstract: An experiment was conducted to investigate three parameters affecting color-difference evaluation on a display: 4 sample sizes (2° , 4° , 10° , and 20°), 2 color-difference magnitudes (4 and 8 CIELAB units), and 2 separations (inclusion or exclusion of the separation line between two colors in a pair). Sample pairs surrounding 5 CIE recommended color centers were prepared. In total, 1120 sample pairs of colors were assessed 20 times using the grey-scale method. The experimental results were used to reveal various parametric effects and to verify the performance of different color matching functions (CMFs) and four color difference formulae and uniform color spaces. It was found that there was little difference in terms of ΔE values calculated using different CMFs for all the color models tested. A parametric formula was proposed to predict three parametric effects for sample pairs having no-separation line: 1) differences in sample size, 2) media (surface and self-luminous colors), and 3) color-difference magnitudes.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Color-difference has been extensively investigated for over a century and, as a result, a number of color-difference formulae have been developed to evaluate color-difference. Significant progress was made after the recommendation of CIELAB [1] in 1976, leading to further advanced color-difference formulae [2–6] until the recommendation of CIEDE2000 [7]. The latter formula is the official recommendation of CIE [8] and it is widely used across many industries. In 2006, CAM02-UCS [9], based on the CIE 2002 color appearance model, CIECAM02 [10], was derived to give accurate prediction of color differences under different viewing conditions. More recently, CAM16-UCS [11], has been proposed to supersede CAM02-UCS [9] to overcome some mathematical drawbacks. Both these UCSs give very similar performance in predicting color differences. ZCAM-QMh [12] was also developed to include both one- and two-dimensional absolute color appearance attributes, and to fit the earlier published color-difference data, together with new data from high dynamic range conditions and from wide color gamut displays. The four formulae or spaces were used in this study.

According to the CIE [5], color-difference formulae should be restricted to apply to a set of reference viewing conditions, i.e., a pair of samples in direct edge contact, subtending a visual angle larger than 4° field of view (FoV) against a uniform neutral grey background ($L^*=50$) illuminated by a D65 simulator at 1000 lx, and the color-difference magnitude should be less than 5 CIELAB (ΔE_{ab}^*) units. Note that the pairs having edge contact means that observers can view a separation line between two samples in a pair. These conditions are usually achieved in industry by using lighting cabinets. However, the viewing conditions in real life could be very different. Hence, CIE also encouraged researchers [13] to investigate parametric effects, i.e., viewing conditions that differ from the above reference conditions. This resulted in the findings of some prominent parametric effects caused by, for example, lightness crispening [14], chromatic crispening [15,16], texture [17], physical size (FoV) [15,16], color-difference magnitude [18,19], separation [19], image [20] and illuminant [21]. The results from Mirjalili *et al.* [19] revealed that a separation effect occurred in the lightness difference affecting total color-difference. The present study extended their study to include the effects of the in/exclusion of a separation, and various FoVs and color-difference magnitudes.

Another issue that has been raised more recently is in regard to the application of different color matching functions (CMFs). The CIE recommended the use of the CIE 1931 2° and 1964 10° CMFs for FoVs less and greater than 4°, respectively. More recently, CIE 2006 CMFs [22,23] have been proposed, as cone fundamentals between 1° and 10° FoV, and observers from 20-80 years old, to take into consideration an individual's lens optical density, macular pigment, and visual pigment. The CIE 2006 2°, 10° and 4° CMFs [22,23] have been investigated to compare with the two conventional CMFs, Fig. 1. The present results include visual results from 4 FoVs that can be used to verify the CMFs. Figure 1 shows that differences occur in the peak wavelengths for the \bar{x} and \bar{z} functions, as well as differences between 450-500 nm for the \bar{y} function. The latter range could affect the calculation of colorimetric data, especially for lighting and display products that use blue LED emission in this region.

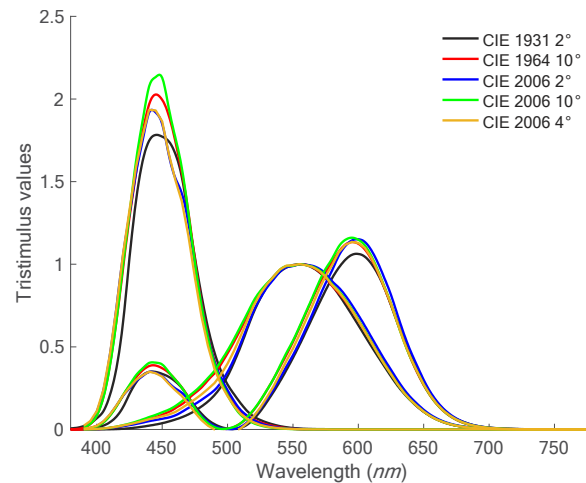


Fig. 1. Different CIE CMFs.

With the above in mind, the present experiment was designed to study three parametric color-difference effects: 1) inclusion or exclusion of a separation line between the two samples in a pair, 2) two color-difference magnitudes (4 and 8 CIELAB units), and 3) 4 fields of view (FoVs) (2°, 4°, 10° and 20°).

2. Experimental

2.1. Display

The experiment was conducted in a darkened room using a 31-inch NEC PA311D liquid crystal display, with a resolution of 4096×2160 pixels. The correlated color temperature (CCT) of the display peak white was set to 6500 K with a luminance of 300 cd/m^2 . The Gain-Offset-Gamma (GOG) model [24] was used to characterize the display. The predictive accuracy of the GOG model gave an average of 0.35 CIEDE2000 (ΔE_{00}) [25] units over 1120 sample colors in the present experiment, with a standard deviation of 0.20 ΔE_{00} units. In addition, sample pairs were measured 5 times during the experimental period, and their associated mean values were used in the analysis. The MCDM (mean color difference from the mean) of the 5 measurements was 0.12 ΔE_{00} units, with a standard deviation of 0.05 ΔE_{00} units. The above results indicate the display to be accurate and have high stability over time. All the measurements were made using a Konica Minolta CS2000A tele-spectroradiometer. The data were recorded in terms of the spectral power distribution of the samples on the display so that the XYZ tristimulus values could be computed using the different CMFs.

2.2. Stimuli

CIE has encouraged researchers to share the data in the areas of color difference [26]. The five CIE recommended color centers were typical examples to be extensively studied, i.e., grey, red, yellow, green, and blue [27]. Figure 2 shows the CIELAB color specification of each color center under illuminant D65 and the CIE 1964 10° CMFs, together with their L^* values. Figure 3 shows the distribution of the sample pairs for each center. For each color-difference magnitude (4 or 8 CIELAB units), 14 sample pairs were prepared surrounding each color center, including 1 pair with only a lightness difference, 6 pairs with mixed lightness and chromatic differences and 7 pairs with only chromatic differences, i.e., no lightness differences. Sample colors were distributed uniformly from 0° to 180° in the $\Delta a^* \Delta b^*$ plane, and from 0° to 90° in the $\Delta L^* \Delta a^*$ plane, or the $\Delta L^* \Delta b^*$ plane, respectively. Table 1 lists the viewing parameters for each of the 16 experimental phases (2 magnitudes, 2 separations, 4 FoVs). For each pair of samples, there were two separation conditions, corresponding to with-separation (S) or no-separation (NS). The former was produced by adding a black line one pixel wide across the edge between the two samples, to simulate the direct edge contact of the surface color pairs. Each sample pair had 4 FoVs, set at 2°, 4°, 10°, or 20°. In total, 1,120 sample pairs (14 samples \times 5 centers \times 2 magnitudes \times 2 (S or NS) \times 4 FoVs) were studied. In addition, 320 pairs (out of the 1120 pairs) were repeated (2 pairs in the $\Delta a^* \Delta b^*$ plane, 1 pair in the $\Delta L^* \Delta a^*$ plane and 1 pair in the $\Delta L^* \Delta b^*$ plane for each color center). The results from the repeated pairs were used to evaluate intra-observer variation. The visual result of each pair was obtained by averaging the results from 20 observers.

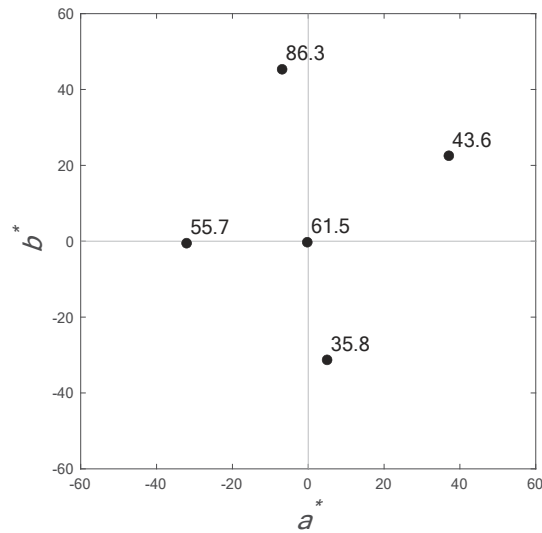


Fig. 2. The 5 CIE color centers under D65 and CIE 1964 10° CMF condition. The number denotes the L^* value of each center.

Table 1. The viewing parameters studied

Groups	ΔE_{ab}^*	Separation	FoVs
1	4	S	2°, 4°, 10°, 20°
2	4	NS	2°, 4°, 10°, 20°
3	8	S	2°, 4°, 10°, 20°
4	8	NS	2°, 4°, 10°, 20°

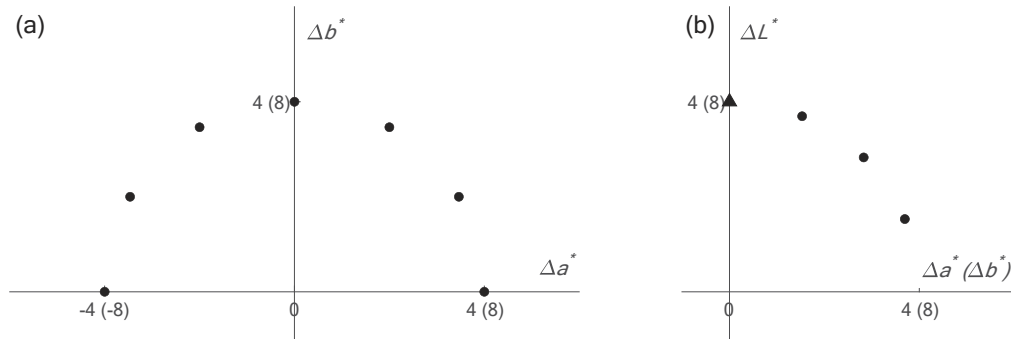


Fig. 3. The sample pair distribution in (a) $\Delta a^* \Delta b^*$ and (b) $\Delta L^* \Delta a^*$ (or Δb^*) planes for color difference magnitudes of 4 and 8 CIELAB units. (For (b), the triangle and dots represent the sample pairs along the ΔL^* axis and in the $\Delta L^* \Delta a^*$ or the $\Delta L^* \Delta b^*$ planes, respectively).

2.3. Visual assessment

The grey-scale method [28] was used to assess the color-differences of sample pairs. The grey-scale samples consisted of 9 ISO 105-A02 samples [28] (GS-1 to GS-5 with an interval of 0.5) and one additional sample (GS-0.5). The latter sample was added to extend the color difference range of the grey scale. This was done to avoid visual differences of some sample pairs exceeding that of the grey scale sample pairs. (It was found that pairs having the same ΔE_{ab}^* values, some having mixed ΔL^* and chromatic differences, appeared to have larger differences than those having only ΔL^* or chromatic differences.) The grey-scale pairs were constructed between the standard (GS-5) and each of GS-0.5 to GS-5 samples. The actual grey-scale pairs agreed well with those of the ISO standard, with a mean of 0.1 ΔE_{ab}^* units. Equation (1) was used to scale the visual judgments in terms of grey-scale values (GS) to visual color-difference values (ΔV), and was obtained by minimizing the ΔE_{ab}^* between the measured color-differences of the neighboring grey samples and the corresponding grey scale numbers (GS). Figure 4 shows the transformation between grey-scale (GS) and visual difference (ΔV), defined by Eq. (1). The predicted color-differences agreed with the actual grey-scale pairs specified by ISO within a mean of 0.2 ΔE_{ab}^* units.

$$\Delta V = 0.1172GS^4 - 1.7394GS^3 + 9.6987GS^2 - 26.0010GS + 31.8068. \quad (1)$$

The experiment was divided into 4 parts, according to the values of the FoVs of 2°, 4°, 10° and 20°. Each pair was assessed by 20 observers. Note that 7 observers participated all parts of the experiment; others participated in only parts of the experiment. In total, there were 46 normal color vision observers, including 21 males and 25 females, having an age range from 19 to 30 years with a mean of 24 years and a standard deviation of 2.9 years.

The experiment was carried out in a darkened room. Observers were seated 60 cm in front of the display. After a 1-minute adaptation period viewing only the mid-grey background ($L^*=50$), observers were asked to assess the color-difference of the sample pairs using the grey-scale method. They were encouraged to report their result to one decimal place, e.g., 3.3. Figure 5 shows the experimental situation for the 4 FoVs. Sample pairs were displayed in the center of the screen in a random order. The grey-scale pairs were displayed along the top of the screen from left to right or from right to left alternately. Observers clicked one of grey-scale pairs. The selected grey-scale pair was then displayed to the left or right of the sample pair. The distance between the two pairs was fixed at 2.5° for all FoVs. Observers were asked to find the grey-scale pair that appeared to have a similar color-difference to that of the sample pair, and then to report the score in terms of the GS value using the scroll bar at the bottom of the screen. The observers

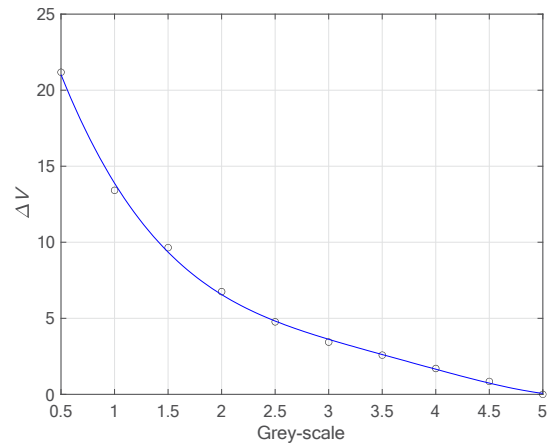


Fig. 4. The transformation between grey-scale (GS) and visual difference (ΔV).

could then press the ‘Previous’ or ‘Next’ button to redo the previous pair or to move on to the next pair, respectively. Each observer did two sessions, each session including 180 pairs, lasting approximately 60 minutes.

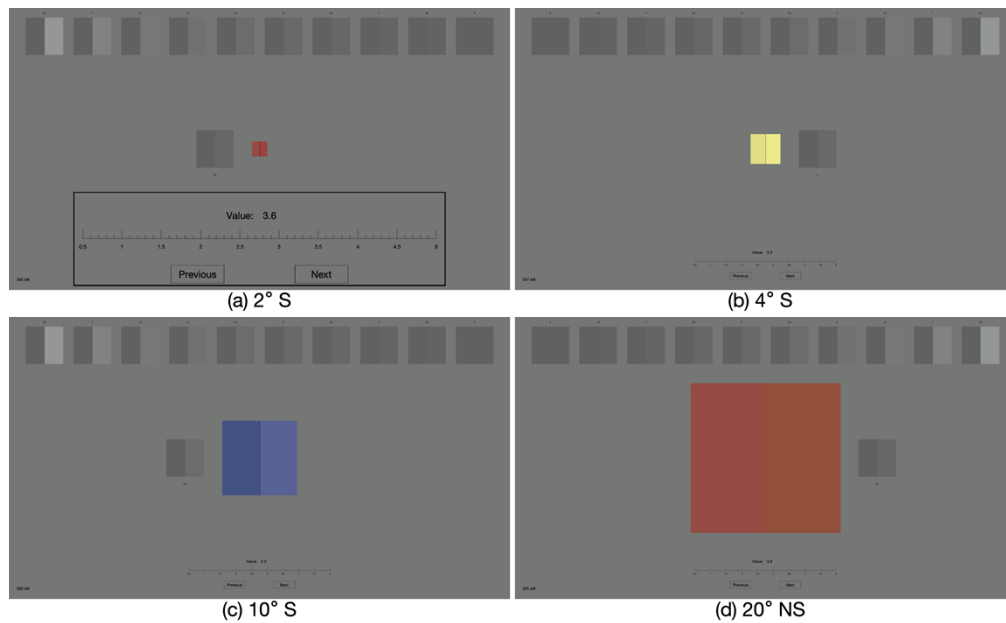


Fig. 5. The experimental interface for (a) 2°, (b) 4°, (c) 10°, and (d) 20° FoVs. The scrollbar in (a) is enlarged 250% for legibility. Note the scrollbar was the original size for all FoVs in the real experiment.

3. Results and discussions

The standard residual sum of squares (STRESS) [29] (see Eq. (2)) metric was used to evaluate the disagreement between the two datasets considered.

$$STRESS = 100 \sqrt{\frac{\sum (F \Delta E_i - \Delta V_i)^2}{\sum \Delta V_i^2}}, \quad (2)$$

where $F = \frac{\sum \Delta E_i \Delta V_i}{\sum \Delta E_i^2}$, a scaling factor to adjust the visual color-difference, ΔV and measured color-difference, ΔE to be on the same scale. A STRESS value of zero indicates perfect agreement.

Two types of statistical test were used to test the significant difference between two datasets. The first was ANOVA test (alpha level 0.05), used to reveal parametric effect such as separation, FoV. The second test was F -test (alpha level 0.05) based on STRESS values, to test statistical difference, e.g., between two color models to predict a visual dataset.

3.1. Observer variations

Intra-observer and inter-observer variations were first investigated. Figure 6 shows the inter-observer variations of each FoV group, in STRESS units according to five groups (the S, NS, $\Delta E_{ab}^* = 8$ and $\Delta E_{ab}^* = 4$ groups and total) together with the intra-observer variation of the total data. In addition, the 95% confidence interval was plotted for each bar. For the inter-observer variation, the STRESS values were 43 (2°), 41 (4°), 42 (10°) and 39 (20°), with an average of 41. Note that typical inter-observer variation for assessing color differences has been found to be in the range about 32-38. For assessing parametric color differences on displays using the grey scale method this can increase to 40-45 as found here [15,30,31].

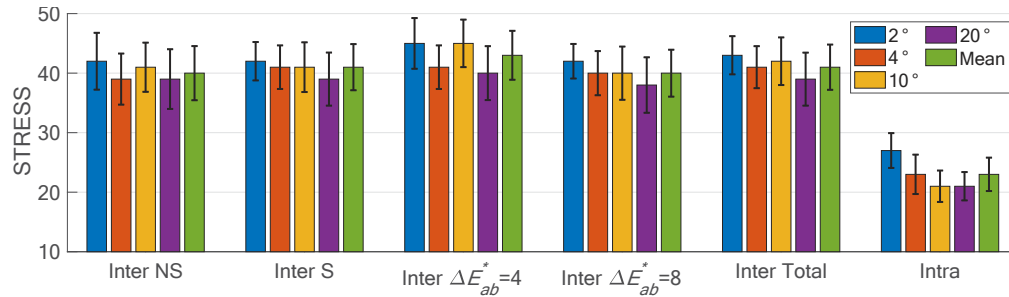


Fig. 6. Inter- and intra-observer variation in STRESS units. Error bar denotes the 95% confidence interval.

Comparing the observer variations between the S and NS groups, it was found that the former had more variation than the latter by approximately 5%. Comparing the large and small ΔE_{ab}^* magnitude groups, the visual difference of the 8 ΔE_{ab}^* group was consistently smaller than that of the 4 ΔE_{ab}^* group by approximately 7%. However, there was no significant difference in the above comparisons according to the F -test ($F = 1.00 \sim 1.27$, $F_C = 0.72$, $1/F_C = 1.40$).

Comparing the four different FoVs, the 2° and 20° fields had the worst and best consistency, respectively. This leads to the suggestion that observers performed more consistently as the FoV increases.

The intra-observer variations (320 out of 1120 pairs were repeated by each observer) were 27 (2°), 23 (4°), 21 (10°) and 21 (20°) in STRESS units, with an average of 23. The intra-observer variations were thus approximately half the STRESS values for inter-observer variation. According to the F -test, a significant difference was found between 2° vs 10° ($F = 1.65$, $F_C = 0.64$, $1/F_C = 1.56$) and between 2° vs 20° ($F = 1.65$).

3.2. Visual effect

3.2.1. Size effect

The size effect was investigated by calculating between ΔV values between two individual FoVs, and between each FoV and overall mean. Figure 7 shows the scatter plots for the above comparisons.

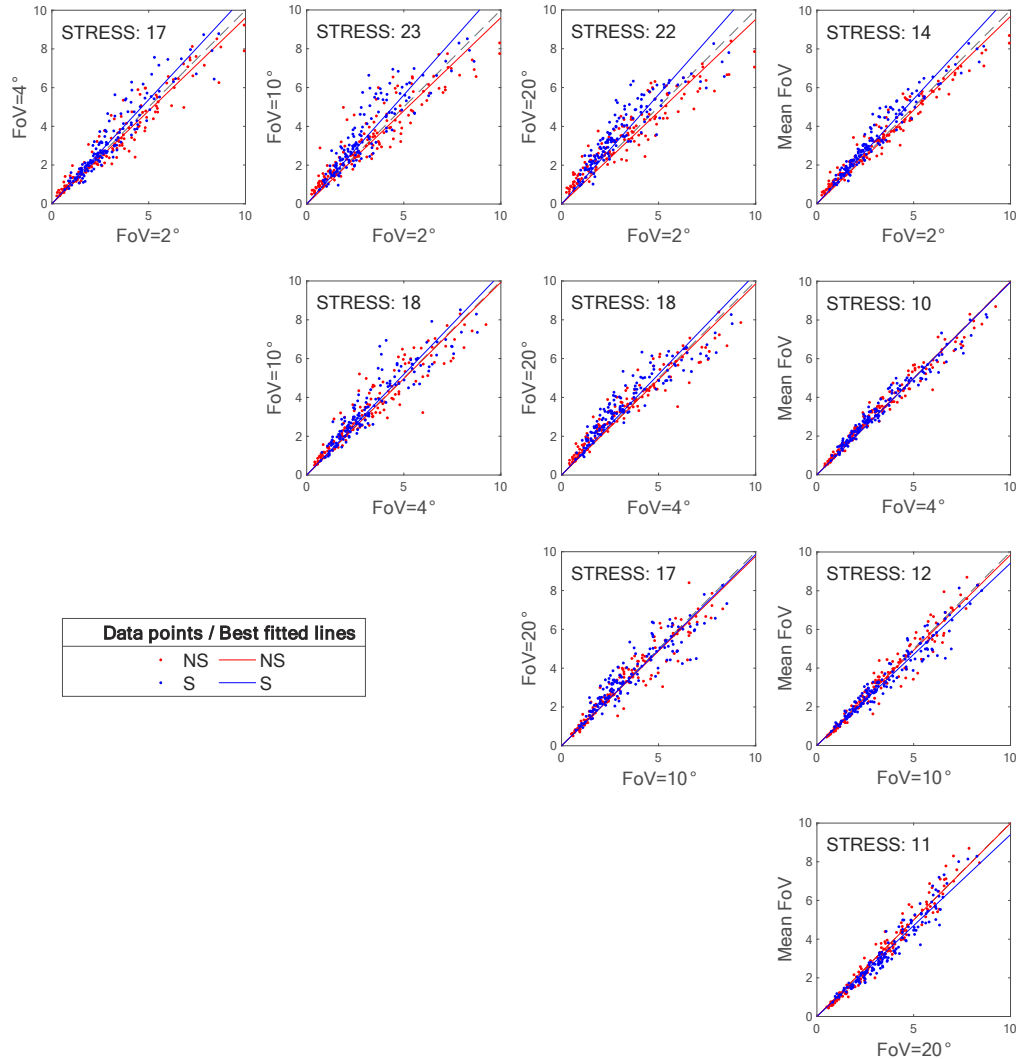


Fig. 7. The scatter diagrams and STRESS values calculated between different FoV phases, and between each phase and the overall mean (diagrams in the last column).

Comparing to the overall mean results, the STRESS values were very similar, 14 (2°), 10 (4°), 12 (10°) and 11 (20°). Comparing each FoV against the others, the mean STRESS values were 21 (2°), 18 (4°), 19 (10°) and 19 (20°), where the 2° and 4° FoVs, and the 10° and 20° FoVs were close to each other (17 STRESS units). The ΔV values between the 2° and 10° FoVs had the largest difference (23 STRESS units). According to the ANOVA test, there was no significant difference between all pairs of the FoVs ($F = 0.37$, $p = 0.773$).

3.2.2. Separation effect

To reveal the separation effect, the visual differences (ΔV) of the sample pairs between the NS and S groups were evaluated in terms of STRESS units. Figure 8 shows the ΔV values between the S and NS groups, where the black and red dots represent sample pairs without or with a lightness difference, respectively. A systematic trend can be seen in that sample pairs with and without lightness differences have larger or smaller visual differences, respectively. This reveals the parametric effect due to separation. When a pair of samples include a lightness difference, with no separation line between the two samples, the visual difference will appear much greater than the equivalent pair with a separation line. Note that all four datasets that include sample pairs with a separation line (BFD [3,4], RIT-DuPont [32], Leeds [6] and Witt [33]) were used to develop advanced color models (CIEDE2000, CAM16-UCS, ZCAM-QMh). More recently, Mirjalili *et al.* [19], Zhao *et al.* [30] and Xu *et al.* [31] investigated color-differences using sample pairs with no separation line. According to the CIE reference conditions, sample pair should have a separation line [5].

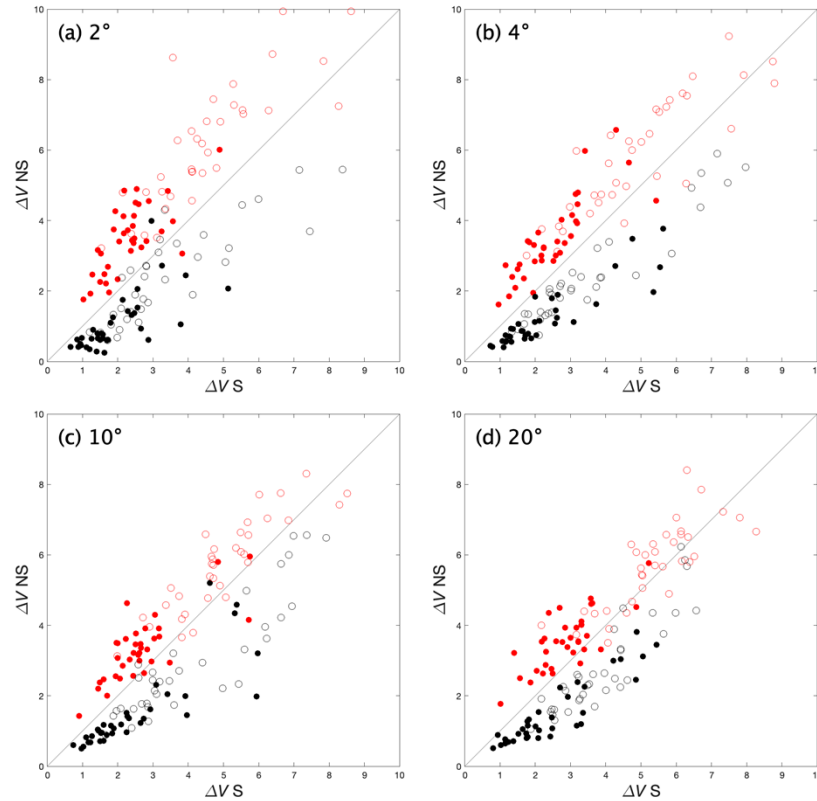


Fig. 8. Scatter plots of ΔV values between the NS and S groups: (a) 2° , (b) 4° , (c) 10° , (d) 20° . Red solid, red open, black solid, black open dots denote $\Delta E_{ab}^* = 4$ including ΔL^* , $\Delta E_{ab}^* = 8$ including ΔL^* , $\Delta E_{ab}^* = 4$ excluding ΔL^* , $\Delta E_{ab}^* = 8$ excluding ΔL^* pairs, respectively.

Table 2 lists the STRESS values calculated between ΔV values from the NS and S groups, with pairs including ΔL^* (Fig. 3(b)), excluding ΔL^* (Fig. 3(a)) and combined respectively. The smallest STRESS values of each set are underlined. Comparing the S and NS groups, the pairs including lightness differences had better agreement than those with excluding lightness differences, and the larger FoV subsets had better agreement than the smaller FoVs subsets. As

can be seen, when combining all the pairs with different FoVs (see Overall set), the disagreement between the S and NS groups further increased.

Table 2. The STRESS values of ΔV values calculated between the S and NS groups^a

S vs NS subsets	2°	4°	10°	20°	Mean	Overall
Including ΔL^* set (70)	21	19	<u>16</u>	<u>16</u>	18	20
Excluding ΔL^* set (70)	31	<u>21</u>	25	23	25	25
Combined (140)	36	32	29	<u>26</u>	31	32

^a(The numbers in the brackets represent the number of sample pairs.)

According to ANOVA test, for the combined set, there was significant difference between all pairs for the 2° and 20° FoVs ($F = 4.04$ and 5.09 , $p = 0.048$ and 0.027 , respectively). However, for the inclusion and exclusion of ΔL^* groups, there was a significant difference between the NS and S groups for all FoVs ($F = 7.58$, $p = 0.007$ for the including ΔL^* group 20° FoV. $F = 13.98 \sim 56.52$, $p < 0.001$ for the others).

3.2.3. Color-difference magnitude effect

The visual differences (ΔV) of sample pairs having color-differences of $4 \Delta E_{ab}^*$ group were compared with those with $8 \Delta E_{ab}^*$ group. Figure 9 shows a very similar trend for all FoVs, i.e., the visual results of the $8 \Delta E_{ab}^*$ group to be about twice larger than those of the $4 \Delta E_{ab}^*$ group. Sample pairs in the S and NS groups had a similar color-difference magnitude effect. Table 3 shows the differences expressed in STRESS units. These values were quite similar (about 19) for all FoVs and overall results.

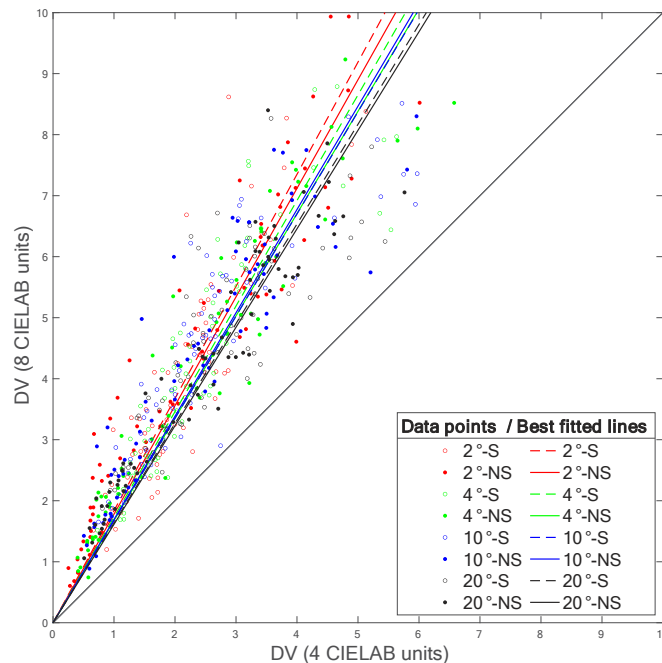


Fig. 9. A scatter plot between ΔV values of sample pairs that have $4 \Delta E_{ab}^*$ units and those that have $8 \Delta E_{ab}^*$ units.

Table 3. The STRESS values of ΔV values calculated between 4 and 8 ΔE_{ab}^* groups^a

ΔE_{ab}^* 4 vs 8 subsets	2°	4°	10°	20°	Mean	Overall
S set (70)	19	19	21	21	20	20
NS set (70)	19	17	21	16	18	17
S + NS (140)	19	18	21	19	19	18

^a(The numbers in the brackets represent the number of pairs.)

3.2.4. Comparing chromatic difference ellipses

Chromatic difference ellipses for the S and NS groups were fitted in CIELAB a^*b^* and CAM16-UCS $a'b'$ planes, respectively. A color-difference ellipse is given in Eq. (3).

$$\Delta E = \sqrt{g_{11}\Delta a^2 + g_{12}\Delta a\Delta b + g_{22}\Delta b^2}, \quad (3)$$

where coefficients g_{11} to g_{22} were optimized to give the least STRESS value between color-differences predicted by the ellipse equation (ΔE) and the visual differences (ΔV). Only 7 excluding ΔL^* (chromatic) pairs for each color center were used to fit a chromatic ellipse. The ΔE values from Eq. (3) were fixed to have a visual difference ΔV of 4 or 8 to plot ellipses for the 4 and 8 ΔE_{ab}^* groups, respectively. Note that ΔV values of 4 and 8 correspond to 9.8 (19.3) and 22.3 (34.8) ΔE_{ab}^* units for S (NS) parameters, respectively.

Figure 10 shows the chromatic difference ellipses plotted in CIELAB and CAM16-UCS, respectively. Circles corresponding to ΔV values of 4 and 8 were also plotted. The parameters for CAM16-UCS (adapting luminance L_a , luminance factor of the neutral background Y_b and the surround) were 59.0 (using CIE 1964 CMFs), 18.4 and 'dim', respectively. Note that the stimuli sampled surrounding a circle in CIELAB space could be slightly distorted in CAM16-UCS space. However, the ellipses were fitted to the visual results in CAM16-UCS space, and not fitted using the data transformed from those in CIELAB ellipses. Thus, the ellipses should be a true representation in CAM16-UCS.

Some patterns in the ellipses can be summarized as below:

- The ellipses can be divided into 2 distinct groups according to the ellipse size: those that represent small magnitudes ($\Delta E_{ab}^*=4$) and those that represent large magnitudes ($\Delta E_{ab}^*=8$). For the NS group, the effect becomes less distinct.
- The NS group ellipses have larger sizes than the S group ellipses due to the excluding ΔL^* (chromatic) pairs having smaller visual differences for the NS pairs than those for the S pairs (as mentioned in section 3.2.2, also see Fig. 8).
- For the S group, different FoVs ellipses had similar shapes and orientations. For the NS ellipses, 2° FoV ellipse was in general more elongated (longer and thinner) than the others in the Grey, Red and Yellow color centers.

From Fig. 10, all the NS ellipses are larger than the S ellipses by a factor of approximately 1.90. Also, all CAM16-UCS ellipses are closer to the constant-sized circle than the CIELAB ellipses. This indicates that the CIELAB space is less uniform than the CAM16-UCS space. This was further verified by using measures of the global and local uniformity as calculated by Eqs. (4) and (5) [31] respectively, where A and B are semi-major and semi-minor axes, and S is the area (πAB) respectively. A smaller *Local* and *Global* values indicate a more uniform color space. The results showed that CAM16-UCS performed better than CIELAB, by 1.9 and 1.3

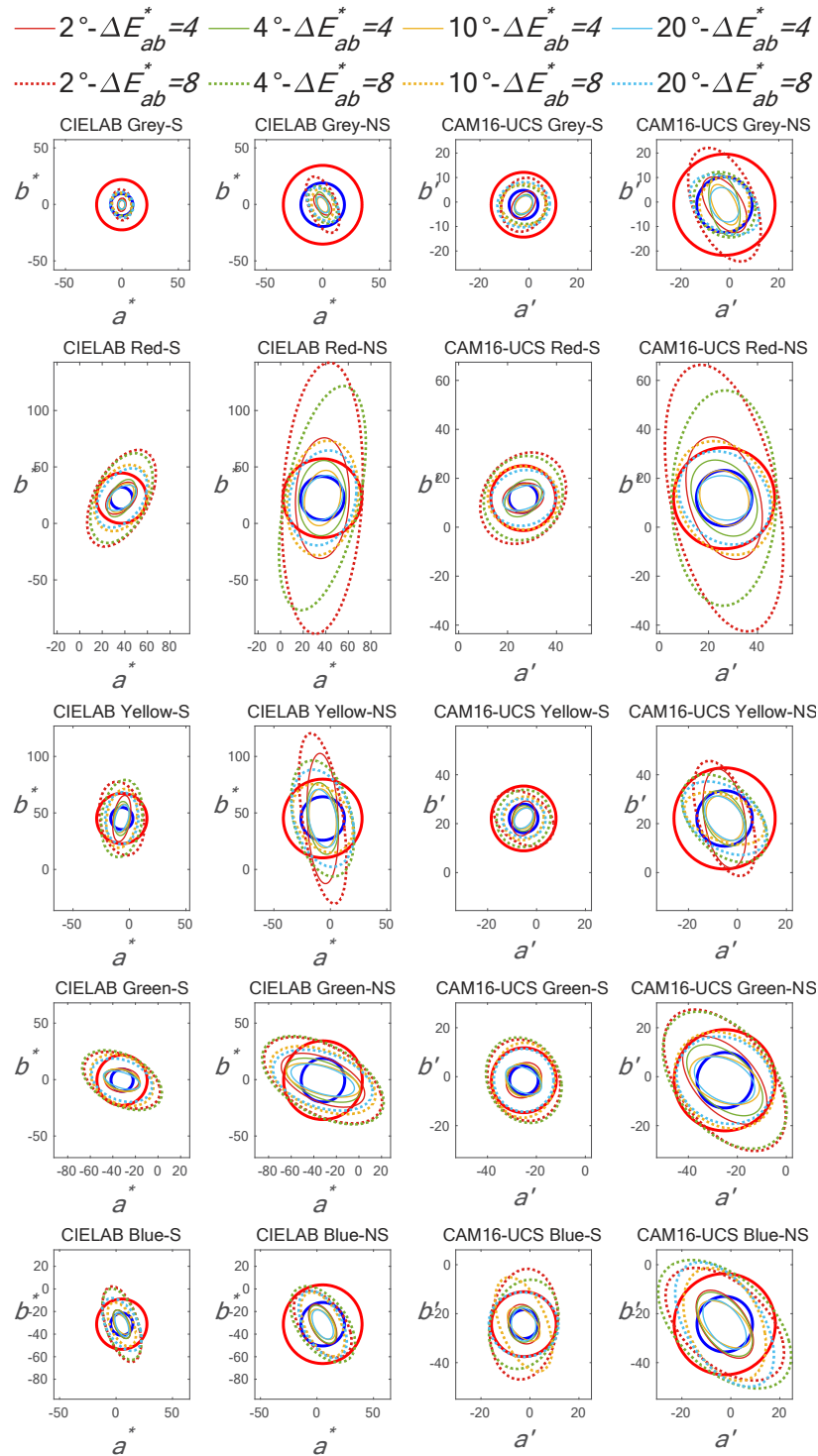


Fig. 10. Chromatic difference ellipses from the left to right: CIELAB S, CIELAB NS, CAM16-UCS S, and CAM16-UCS NS, and from top to bottom: the CIE color centers of grey, red, yellow, green, and blue, respectively. For each center, there are 4 FoV ellipses for ΔV values of 4 and 8, respectively. The thick red and blue circles in each space represent the constant sizes corresponding to ΔV values of 4 and 8, respectively.

times for local uniformity and global uniformity respectively.

$$Local = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{A_i}{B_i} - 1 \right)^2} \times 100\%, \quad (4)$$

$$Global = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - \bar{S})^2}}{\bar{S}} \times 100\%. \quad (5)$$

The F test was also conducted, i.e., the F value is considered to be significant when outside the range of $F_C = 0.53$, $1/F_C = 1.89$. The results showed that for local uniformity, the S ellipses are significantly closer to circle than NS ellipses ($F = 2.10$ and 3.61 for CIELAB and CAM16-UCS respectively). For global uniformity, the S ellipses were closer to the constant size than NS ellipses ($F = 2.08$ and 2.00 for CIELAB and CAM16-UCS respectively). The differences of local and global uniformity between 4 and 8 ΔE_{ab}^* groups are statistically insignificant ($F = 1.16$, 1.17 , 0.99 and 0.89 for CIELAB local, CAM16-UCS local, CIELAB global and CAM16-UCS global uniformity respectively).

Overall, CAM16-UCS performed significantly better in both local uniformity ($F = 3.46$, $F_C = 0.64$, $1/F_C = 1.56$) and global uniformity ($F = 1.79$) than CIELAB.

Figure 11 is taken from Merjialili *et al.*'s paper [19]. It shows a great similarity for each CIE center between 5 datasets: their own, Witt [34], Cheung and Rigg [35], RIT-DuPont [32] and Cui and Luo's LMG0N [15]. It can be seen in Fig. 11, the ellipses from all datasets had good agreement in terms of ellipse shape and orientation, but not in size. Very similar pattern as Fig. 11 can be found for all ellipses in Fig. 10 in CIELAB space.

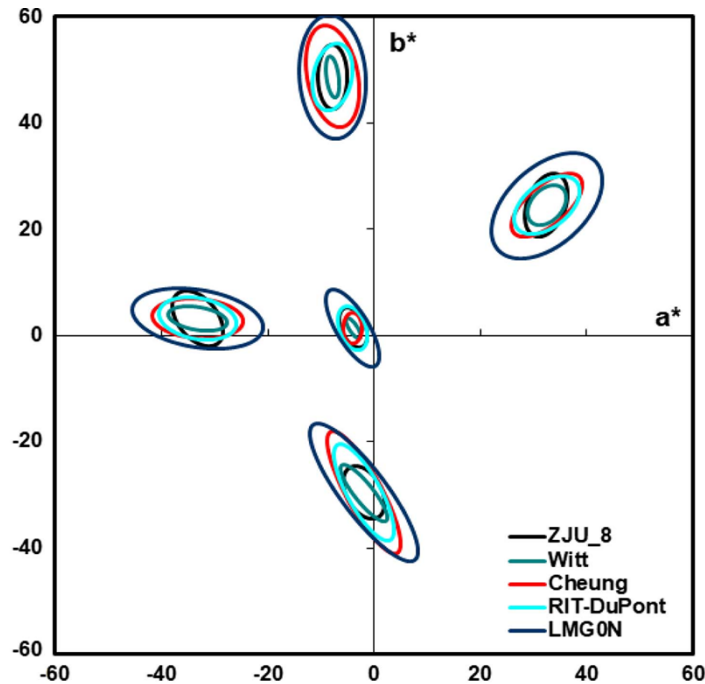


Fig. 11. Comparing the ellipses obtained from various studies. (The figure comes from [19].)

4. Comparing the performances of metrics formed by CMFs and color models

As noted earlier, 5 CMFs (CIE 1931 2°, CIE 1964 10°, CIE 2006 2°, CIE 2006 10°, CIE 2006 4°) and 4 color models (CIELAB, CIEDE2000, CAM16-UCS and ZCAM-QMh) were investigated in this study. Note that a mean age of 24 years was used as a parameter to obtain the CIE 2006 4° CMFs. The name 'metric' is used here to represent a combination of CMFs and color model. The data were used to test the performance of each metric in terms of STRESS values. The results should reveal which CMFs and which model gave the best performance.

The testing follows the CIE specification to use 2° and 10° CMFs for the FoVs larger and smaller than 4°, respectively. For 4° FoV data, all 5 CMFs were tested. Table 4 lists the testing results of each metric. For each FoV, the best color models and the best CMFs are in bold and underlined, respectively. The results are summarized below:

- All metrics consistently performed better for the larger FoV (10° and 20°) than for the smaller FoV (2° and 4°) data.
- For each color model, the CMFs compared gave almost exactly the same performance regardless of color model. Also, there was no significant difference between CMFs according to the F -test ($F = 0.86 \sim 1.17$, $F_C = 0.79$, $1/F_C = 1.27$). Comparing the STRESS values, ZCAM-QMh outperformed CAM16-UCS and CIEDE2000, and CIELAB the worst. Only CIELAB performed significantly worse than the other models according to the F -test ($F = 1.81 \sim 3.31$, $F_C = 0.79$, $1/F_C = 1.27$).

Table 4. Comparison of the performance of different color models using different CMFs in STRESS units.^a

CMFs (FoV data)	CIELAB	CIEDE2000	CAM16-UCS	ZCAM-QMh
CIE 1931 (2°)	<u>48</u>	<u>34</u>	<u>35</u>	<u>29</u>
CIE 2006-2° (2°)	<u>48</u>	<u>34</u>	<u>35</u>	<u>30</u>
CIE 1931(4°)	48	31	33	<u>28</u>
CIE 1964(4°)	<u>47</u>	<u>30</u>	<u>32</u>	<u>29</u>
CIE 2006-2°(4°)	48	31	33	<u>29</u>
CIE 2006-10°(4°)	<u>47</u>	<u>30</u>	<u>32</u>	<u>28</u>
CIE 2006-4°(4°)	48	32	33	<u>29</u>
CIE 1964(10°)	<u>41</u>	<u>25</u>	<u>25</u>	29
CIE 2006-10° (10°)	<u>41</u>	<u>25</u>	<u>25</u>	<u>28</u>
CIE 1964 (20°)	<u>39</u>	<u>25</u>	25	<u>23</u>
CIE 2006-10° (20°)	<u>39</u>	<u>25</u>	<u>24</u>	<u>22</u>

^aThe best performance of color model is in bold.

5. Modelling the parametric color-difference formulae for sample pairs to have no-separation line

It is well known that more advanced recent color models that include color-difference equations and uniform color spaces were derived using surface colors with a separation line between pairs of samples. So, it is to be expected that the models should perform better with separation conditions than no-separation conditions. As found by Mirjalili *et al.* [19], their sample pairs, with no separation between the two samples in a pair, had a parametric effect due to the embedded lightness-difference in the total color-difference. As shown in Fig. 8, the NS group of sample pairs with a lightness difference had larger perceived color differences than the S group pairs with

a lightness difference, but for those chromatic pairs (without lightness difference), the results showed otherwise. Therefore, the typical color models performed worse for no-separation pairs than separation pairs. Efforts were made to reveal the parametric effect for the NS data and this is described below.

Equation (6) shows a generic color-difference formula.

$$\Delta E_0 = \sqrt{(\Delta L)^2 + (\Delta C)^2 + (\Delta H)^2 + RT}, \quad (6)$$

where RT is the rotation term and is set to zero except for CIEDE2000.

From Section 3.2.2, for no-separation sample pairs having the same color-difference in CIELAB units, the no-separation pairs having lightness differences should have larger predicted differences than those with chromatic differences. This agrees with the finding of Mirjalili *et al.* [19] who proposed a color-difference formula (ΔE_1) given in Eq. (7) by including a lightness weighting function (D_L) to predict lightness difference.

$$\Delta E_1 = \sqrt{\left(\frac{\Delta L}{D_L}\right)^2 + (\Delta C)^2 + (\Delta H)^2 + RT}, \quad (7)$$

where $D_L = a \cdot \Delta E + b$; the a and b coefficients were obtained from the Mirjalili *et al.* dataset, and ΔE was calculated from the original equation, e.g., CIELAB. The function indicates both D_L and ΔE_0 values to have a positive correlation, and implies that for a larger color-difference, the separation line will show more clearly, and would lead to a smaller lightness difference compared with the chromatic difference. It was found that a values of 0.047, 0.079, 0.072, 0.065; and b values of 0.22, 0.27, 0.27, 0.61 for CIELAB, CIEDE2000, CAM16-UCS and ZCAM-QMh, respectively.

The Mirjalili *et al.* D_L factor was derived using surface colors. Taking media into account, a media scaling factor (surface and self-luminous colors), s_M , was introduced to form Eq. (8), to adjust the lightness difference for display colors.

$$\Delta E_2 = \sqrt{\left(\frac{\Delta L}{D_L \cdot s_M}\right)^2 + (\Delta C)^2 + (\Delta H)^2 + RT}, \quad (8)$$

where a value around 0.5 was found to be appropriate for the factor s_M for the self-luminous colors, i.e., values of 0.41, 0.66, 0.65, 0.47 were calculated for CIELAB, CIEDE2000, CAM16-UCS and ZCAM-QMh, respectively [31]. The factor s_M was fixed at 0.5 for all models in the present work. Note that for surface colors, the s_M value is set to one for all color models. This result shows strong evidence that there is a media effect.

Finally, a parametric factor s_S was introduced to account for the physical size (FoV) effect, s_S (see Eq. (9)) to adjust the lightness difference for different FoVs.

$$\Delta E_3 = \sqrt{\left(\frac{\Delta L}{D_L \cdot s_M \cdot s_S}\right)^2 + (\Delta C)^2 + (\Delta H)^2 + RT}, \quad (9)$$

where factor s_S was first optimized for individual physical sizes, and was then fitted to a function of FoV as given in Eq. (10).

$$s_S = a \cdot e^{b \cdot \text{FoV}} + c, \quad (10)$$

where parameters a , b and c were optimized for each color model. These parameters are given in Table 5. Figure 12 shows the FoV scaling factor s_S plotted as a function of FoV.

To compare the performance of the parametric formulae in predicting the visual results, the STRESS values were calculated and the F -test was applied to show the improvement between

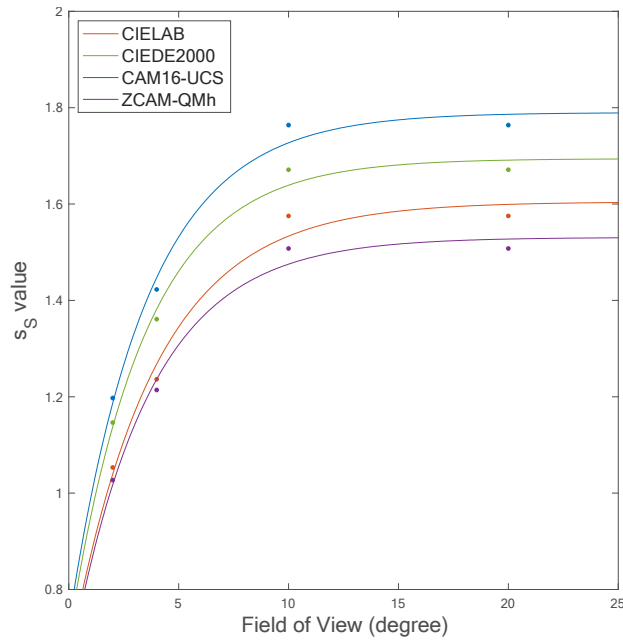


Fig. 12. The FoV scaling factor s_S plotted as a function of FoV.

Table 5. The parameters of the FoV scaling factor s_S in Eq. (10)

	a	b	c
CIELAB	-0.95	0.26	1.60
CIEDE2000	-0.99	0.29	1.69
CAM16-UCS	-1.06	0.28	1.79
ZCAM-QMh	-0.89	0.28	1.53

different stages for each formula ($F_C = 0.79$, $1/F_C = 1.27$). Table 6 lists the performance of the original (ΔE_0), ΔE_1 , ΔE_2 and ΔE_3 formulae in terms of STRESS values. From the table above, some conclusions can be drawn below:

- All the improved formulae from each color model performed better than their original formula.
- All models predicted the results of larger FoV (10° and 20°) more accurately than those of smaller FoV pairs (2° and 4°).
- Comparing ΔE_0 and ΔE_1 for all models, the lightness-difference parametric factor D_L gave noticeable improvements except ZCAM-QMh. According to the F -test, all models showed significant improvements for all FoVs and the combined data ($F = 1.46 \sim 2.25$) except ZCAM-QMh.
- The ΔE_2 formulae including the s_M parametric factor was found only slight improvement on ΔE_1 formulae for all color models in terms of STRESS values. According to F -test, these differences between ΔE_1 and ΔE_2 formulae are significant only for 2° (all models) and 4° (ZCAM-QMh) ($F = 1.44 \sim 2.19$).

- Including the factor s_S , led to ΔE_3 formula giving smaller STRESS values over the ΔE_2 formula. This indicates the performance of all models consistently improved. There were significant improvements for 10° (CIEDE2000, CAM16-UCS and ZCAM-QMh) and 20° (CIEDE2000, CAM16-UCS and ZCAM-QMh) according to the F -test ($F = 1.36 \sim 1.88$).
- For CIELAB, CIEDE2000 and CAM16-UCS, the ΔE_3 formula outperformed ΔE_0 significantly from the F -test ($F = 2.19 \sim 2.56$), except for the ZCAM-QMh model, which only gave a small improvement ($F = 1.19$). When predicting color-difference using the ΔE_3 formula, CIEDE2000 and ZCAM-QMh performed the best, followed by CAM16-UCS, with CIELAB the worst.
- Note that the ΔE_3 formula for the NS group gave similar STRESS value to that of the ΔE_0 formula for the S group. Considering the combined dataset, CIEDE2000 performed the best, followed by CAM16-UCS and ZCAM-QMh, with CIELAB the worst.

Table 6. The performance of the new color models in terms of STRESS units for each FoV (CIE 1931 CMFs for 2° FoV and CIE 1964 CMFs for the other FoVs).^a

	FoV	ΔE_0	ΔE_1	ΔE_2	ΔE_3
CIELAB	2°	53	37	<u>29</u>	<u>29</u>
	4°	50	35	32	<u>31</u>
	10°	44	<u>33</u>	36	<u>33</u>
	20°	44	30	31	<u>28</u>
	mean	48	34	32	<u>30</u>
CIEDE2000	2°	42	32	<u>24</u>	<u>24</u>
	4°	37	28	26	<u>23</u>
	10°	29	<u>22</u>	30	<u>22</u>
	20°	31	21	25	<u>19</u>
	mean	35	26	26	<u>22</u>
CAM16-UCS	2°	41	31	25	<u>24</u>
	4°	37	28	28	<u>25</u>
	10°	29	<u>24</u>	32	<u>24</u>
	20°	30	20	26	<u>19</u>
	mean	34	26	28	<u>23</u>
ZCAM-QMh	2°	28	34	<u>23</u>	<u>23</u>
	4°	26	30	25	<u>24</u>
	10°	23	25	28	<u>23</u>
	20°	20	22	21	<u>18</u>
	mean	24	28	24	<u>22</u>

^aThe best performing FoV data for each parametric formula is in bold, and the best performing formula for each FoV data is underlined.

Figures 13(a) and 13(b) show the performance of the parametric formula for each color model, and for the individual FoV subsets of CIEDE2000, respectively. CIEDE2000 is the present ISO/CIE standard, and is used here to represent the typical effect for all models. The results in Fig. 13 support the conclusion in Table 4. Comparing different models in Fig. 13(a), all models

outperformed CIELAB and in general ΔE_3 performed the best, followed by ΔE_2 , E_1 , and the ΔE_0 the worst. In Fig. 13(b), the larger the FoVs, CIEDE2000 performed the better.

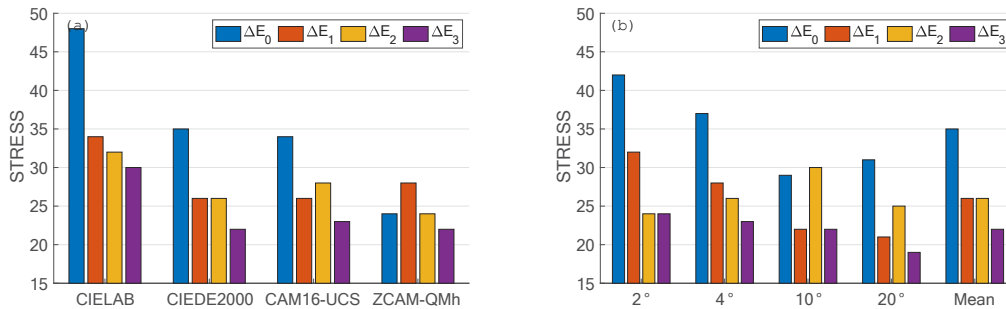


Fig. 13. The performance of the parametric formula (a) for different color models and (b) for individual FoVs using CIEDE2000.

6. Conclusion

An experiment was conducted using display colors to investigate parametric color-difference effects, including separation, color-difference magnitude and field of view. The results revealed various parametric effects.

Different CMFs were tested to calculate color-differences using the visual results for different fields-of-view. The results showed that there was little difference between the data calculated using the different CMFs. A parametric formula was successfully developed to predict three parametric effects, the field-of-view, the media (surface or self-luminous colors) and the color-difference magnitude, for sample pairs that to have no-separation line.

Funding. National Natural Science Foundation of China (61775190).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. CIE 015: 2018, "Colourimetry," (CIE, Vienna, 2018).
2. F. Clarke, "Modification to the JPC79 color-difference formula," *J.soc.dyers Colourists* **100**(1984).
3. M. R. Luo and B. Rigg, "BFD (l:c) colour-difference formula Part 1-Development of the formula," *J. Soc. Dyers Colour.* **103**(2), 86–94 (1987).
4. M. R. Luo and B. Rigg, "BFD (l:c) colour-difference formula Part 2-Performance of the formula," *J. Soc. Dyers Colour.* **103**(3), 126–132 (1987).
5. CIE, "Industrial Colour-Difference Evaluation," (1995).
6. D. H. Kim and J. H. Nobbs, "New weighting functions for the weighted CIELAB colour difference formula," in *Proceedings of the AIC*, 1997, 446–449.
7. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.* **26**(5), 340–350 (2001).
8. CIE, "Improvement to Industrial Colour Difference Equation," (2001).
9. M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," *Color Res. Appl.* **31**(4), 320–330 (2006).
10. CIE 159: 2004, "A colour appearance model for colour management systems: CIECAM02," (CIE, Vienna, 2004).
11. C. Li, Z. Li, Z. Wang, Y. Xu, M. R. Luo, G. Cui, M. Melgosa, M. H. Brill, and M. Pointer, "Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS," *Color Res Appl* **42**(6), 703–718 (2017).
12. M. Safdar, J. Y. Hardeberg, and M. Ronnier Luo, "ZCAM, a colour appearance model based on a high dynamic range uniform colour space," *Opt. Express* **29**(4), 6036–6052 (2021).
13. CIE 101-1993, "Parametric effects in colour-difference evaluation," (1993).
14. K. M. R. Ho, G. Cui, M. R. Luo, and B. Rigg, "A Lightness Difference Formula for Predicting Crispening Effects," in *The Tenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications, CIC 2002, November 12, 2002, Scottsdale, Arizona, USA, 2002*,

15. G. H. Cui, M. R. Luo, B. Rigg, and W. Li, "Colour-difference evaluation using CRT colours. Part I: Data gathering and testing colour difference formulae," *Color Res. Appl.* **26**(5), 394–402 (2001).
16. G. Cui, M. R. Luo, B. Rigg, and W. Li, "Colour-difference evaluation using CRT colours. Part II: Parametric effects," *Color Res. Appl.* **26**(5), 403–412 (2001).
17. R. Huertas, M. Melgosa, and E. Hita, "Influence of random-dot textures on perception of suprathreshold color differences," *J. Opt. Soc. Am. A* **23**(9), 2067–2076 (2006).
18. H. Wang, G. Cui, M. R. Luo, and H. Xu, "Evaluation of colour-difference formulae for different colour-difference magnitudes," *Color Res. Appl.* **37**(5), 316–325 (2012).
19. F. Mirjalili, M. R. Luo, G. Cui, and J. Morovic, "Color-difference formula for evaluating color pairs with no separation: ΔE_{NS} ," *J. Opt. Soc. Am. A* **36**(5), 789–799 (2019).
20. CIE 199:2011, "Methods for evaluating colour differences in images," (Commission Internationale de l'Éclairage, Vienna, Austria, 2011).
21. M. R. Luo and B. Rigg, "A colour-difference formula for surface colours under illuminant A," *J. Soc. Dyers Colour.* **103**, 161–167 (1987).
22. CIE 170-1:2006, "Fundamental chromaticity diagram with physiological axes - Part 1," (Commission Internationale de l'Éclairage, Vienna, Austria, 2006).
23. CIE 170-2:2015, "Fundamental Chromaticity Diagram with Physiological Axes – Part 2: Spectral Luminous Efficiency Functions and Chromaticity Diagrams," (CIE, 2016).
24. R. S. Berns, "Methods for characterizing CRT displays," *Displays* **16**(4), 173–182 (1996).
25. ISO/CIE, "Colorimetry – Part 6: CIEDE2000 colour-difference formula," in *ISO/CIE 11664-6:2014(E) (ISO/CIE, 2014)*, (2014), pp. 11664–11666.
26. M. Melgosa, "Request for existing experimental datasets on color differences," *Color Res. Appl.* **32**(2), 159 (2007).
27. A. Robertson, "CIE guidelines for coordinated research on color-difference evaluation," *Color Res Appl* **3**, 33–39 (1978).
28. 105-A02, "Textiles - Tests for colour fastness - Part A02: Grey scale for assessing change in colour," (ISO, Geneva, 1993).
29. CIE 217:2016, "Recommended Method for Evaluating the Performance of Colour-Difference Formulae," (2016).
30. B. Zhao, Q. Xu, and M. R. Luo, "Color difference evaluation for wide-color-gamut displays," *J. Opt. Soc. Am. A* **37**(8), 1257–1265 (2020).
31. Q. Xu, B. Zhao, G. Cui, and M. R. Luo, "Testing uniform colour spaces using colour differences of wide colour gamut," *Opt. Express* **29**(5), 7778 (2021).
32. R. S. Berns, D. H. Alman, L. Reniff, G. D. Snyder, and M. R. Balonon-Rosen, "Visual determination of suprathreshold color-difference tolerances using probit analysis," *Color Res. Appl.* **16**(5), 297–316 (1991).
33. K. Witt, "Geometric relations between scales of small colour differences," *Color Res. Appl.* **24**(2), 78–92 (1999).
34. K. Witt, "Parametric effects on surface color-difference evaluation at threshold," *Color Res. Appl.* **15**(4), 189–199 (1990).
35. M. Cheung and B. Rigg, "Colour-difference ellipsoids for five CIE colour centres," *Color Res. Appl.* **11**(3), 185–195 (1986).