Received: 19 June 2019 | Revised: 14 October 2019 | Accepted: 15 October 2019
DOI: 10.1002/col.22451

RESEARCH ARTICLE

WILEY

Distance metrics for very large color differences

Saeedeh Abasi[1,2] | **Mohammad Amani Tehran**[1] | **Mark D. Fairchild**[2]

[1]Textile Engineering Department, Amirkabir University of Tehran, Tehran, Iran

[2]Program of Color Science/Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, New York

Correspondence
Mohammad Amani Tehran, Textile Engineering Department, Amirkabir University of Tehran, Tehran, Iran.
Email: amani@aut.ac.ir

Abstract

Small, supra-threshold color differences are typically described with Euclidean distance metrics, or dimension-weighted Euclidean metrics, in color appearance spaces such as CIELAB. This research examines the perception and modeling of very large color differences in the order of 10 CIELAB units or larger, with an aim of describing the salience of color differences between distinct objects in real-world scenes and images. A psychophysical experiment was completed to compare directly large color-difference pairs designed to probe various Euclidean and non-Euclidean distance metrics. The results indicate that very large color differences are best described by HyAB, a combination of a Euclidean metric in hue and chroma with a city-block metric to incorporate lightness differences.

KEYWORDS

color difference formula, distance metric, perceived color difference, very large color difference

1 | INTRODUCTION

Color difference formulas are widely used to evaluate the perceived color difference between two samples. Most color difference formulas, such as CIELAB and CIEDE2000, were developed to fit visual assessment datasets of small color differences, typically under five (ΔE^*_{ab}) units.[1]

However, there is an essential need for color difference formulas applicable to large and very large color differences in many practical applications. For example, there are problems in computer vision and image processing that need large color difference formulas to be solved correctly and accurately. These include color clustering for images, edge detection, object detection, colored feature extraction, texture descriptors for colored images, and so on.[2-4]

Luo et al found that the performance of color difference formulas is significantly different for the estimation of large and small color differences. They extended CIECAM02 to form three suggested uniform color spaces.[5] However, according to their published results, it seems that these proposed formulas do not perform significantly better than CIELAB, ΔE^*_{ab}.

It is concluded from section 16.17 of Reference 6, in general, that it is logical and appropriate to use simple color appearance models such as CIELAB. If this is inadequate, then a combination of CIELAB with a more accurate chromatic adaptation transform would be the next step. For more flexibility, RLAB or RLAB with CAT02 for adaptation transform might be the most appropriate. Then, if further sophistication is needed, CIECAM02 becomes useful.[6] Small, supra-threshold, color differences can be expressed usefully in any of these color spaces.

Wang et al[1] conducted visual experiments for assessing a CIECAM02-based model developed by Luo et al.[5] Considering the average results, CAM02-LCD, CAM02-UCS, and CIEDE2000 gave a similar performance. They concluded that, for predicting color differences with different magnitudes, CIEDE2000 should be used and suggested that it could also be applied in applications involving a large range of color differences.[1]

However, according to our initial investigation of CIEDE2000, discussed in the next section, it does not appear to exhibit a suitable performance over large ranges of color differences.

Guan and Luo conducted an experiment to evaluate the performance of four color difference formulas (CIELAB, CMC, BFD, and CIE94) for large color-difference prediction. The results showed that all formulas except CIE94 exhibit better performance in large color-difference prediction by the introduction of a lightness weighting factor (K_L).[7] It was also found that K_L values of each formula for evaluating large color differences are about 40% lower than those for small color differences. In other words, there is a need to use a lower K_L value (or more significantly incorporate perceived lightness difference) for evaluating large color differences in comparison with small color differences. Thus, in case of large color-difference evaluation, lightness differences are consistently judged to be larger relative to chroma or hue differences.[7]

However, the introduction of a lightness weighting factor does not suitably address the problem. The authors claimed that each formula performance is improved by applying different lightness weighting factors.[7] If lightness differences in large color differences have a different effect on perceived difference, and all formulas have different performance with different weighting function, then, it is implied that there is a need to provide a new approach for predicting large color-difference pairs with special regard to lightness. Fundamentally, large color-difference perception cannot be a function of the equation used to model it. On the other hand, the color differences of pairs in Guan and Luo's study ranged from 6 to 21 ΔE^*_{ab},[7] and pairs with an even larger color difference are important for many applications. Wang et al[1] did conduct a visual experiment for a larger range, but as mentioned before, the results are not consistent with our initial assessments.

The main aim of this study is assessing the performance of different formulas for predicting the perceived color difference of pairs with very large color differences, ranging from 13 to 86 ΔE^*_{ab} units. For this purpose, two datasets were prepared. Each dataset contains different categories of color differences designed for different purposes. A visual experiment was conducted that is explained in detail in the next section. The CIELAB and CIEDE2000 formulas were chosen as the current and most generally utilized color-difference formulas. These two formulas were reported as suitable for large color-difference prediction in Wang[1] and Guan's[7] studies. Other formulas do not perform better than these two formulas.[1,7] In addition, these two well-known formulas are easy to apply, with no complexity and time-consuming implementation for image analysis applications.

In addition to the straightforward CIELAB and CIEDE2000 formulas, four additional formulas were studied, including "City Block in (L^*,a^*,b^*)," "ΔL^*+Euclidean(a^*,b^*)," "City Block in (L^*,C^*,H^*)," and "ΔL^*+Euclidean(C^*,H^*)" as potential novel approaches to this problem. The reasons for developing these new ideas are explained in detail in the next section.

Two datasets were generated for evaluating these formulas. According to the results, the "ΔL^*+Euclidean (a^*,b^*)" has the best fit with visual experiments in the first dataset and best consistency with the second dataset. It is concluded that, for very large color differences, the effect of lightness difference should be considered to be perceptually separable.

2 | DISTANCE METRICS

A space could be a geometric entity in which locations are specified relative to a set of axes. CIE L*a*b* is a color space as a geometric entity (a vector space) with three axes. The distance between two points can be measured in any space using a variety of distance metrics. The Euclidean distance is the most familiar way of measuring distance in a real vector space (such as CIE L*a*b*). The Euclidean distance between two points in a two-dimensional (xy) space is (Chapter 5 of Reference 8):

$$\text{Euclidean} = \left[(x_1-x_2)^2+(y_1-y_2)^2\right]^{1/2} \quad (1)$$

It is concluded from section 5.B of Reference 9 that color difference formulas such as CIELAB are often based on the Euclidean distance in a color space. These formulas calculate the distance between two points in a defined color space such as CIE L*a*b* using a Euclidean metric.[9]

There is a main disadvantage in the calculation of large color differences in multidimensional spaces using Euclidean distance. Assume that there are large differences between some dimensions of two points, but a small difference (or no difference) between other dimensions of the two points. In these cases, when the differences in dimensions with a small difference (or no difference) increase, the increase of overall distance, calculated by a Euclidean metric, is often smaller than the relative perceptual change in the feature dimension.

For example, in a color-difference application, when there is large and dominant difference between two samples in the hue attribute, a significant increase (or decrease) in L* difference results in a relatively small increase (or decrease) in overall calculated Euclidean difference. Computationally, the difference between two samples with very different hues is more fundamentally influenced by changes in hue difference. The effect of lightness (L*) is not taken into account properly with Euclidean metrics, but might be very salient perceptually.

According to initial observations of a few samples that were completed by the authors and colleagues in the RIT, Munsell Color Science Laboratory, it was found that the increase (or decrease) in L* between two pairs with a very large color difference usually causes a significant perceived color difference. However, the calculated color differences by CIELAB and CIEDE2000 do not show such a significant change. An example is presented in Table 1. The second pair has a significantly different perceived color difference in comparison with the first pair, but the calculated color difference with CIELAB and CIEDE2000 for two pairs are very similar. Thus, these two pairs have similar computed color differences, which are not consistent with observations. Regardless of whether these color-difference formulas are an interval scale or a ranking scale, they should give two significantly different results for these stimuli because the differences are visually disparate. As the calculated color difference for these two pairs is very close, these formulas cannot be a suitable interval or ranking scale for samples with large color differences. The *L**, *a**, and *b** of samples in Table 1 are generated as an example of the shortcomings of CIELAB and CIEDE2000 as very large-difference metrics. They are not transformed from another color space. However, in order to display samples, transformations of L*a*b* to XYZ and then XYZ to RGB were performed using a display characterization model explained in Section 3.3.

In order to assess the perceived difference by observers, some collections of color pairs with various properties were generated. If observers' perceptual responses are in agreement with the initial assessments described, then the Euclidean-based formulas are not suitable for large color-difference computations. This would show the disadvantage of Euclidean distance for calculating the distance of color pairs with very large differences.

2.1 | New color-difference formulas

As mentioned before, there might be a serious disadvantage in calculating large color differences with a Euclidean metric, but this is not the only way to measure distance. The city-block metric is another distance metric often used in psychological models. It is the sum of absolute differences of each dimension. It is like walking from the first point to the second point in a city. In a city, a person cannot cut through the blocks and so must walk along the streets from one point to the other. The equation for a city-block metric in a two-dimensional space (*xy*) is[8]:

$$\text{City Block} = |x_1 - x_2| + |y_1 - y_2| \tag{2}$$

The Euclidean and city-block metrics reflect different assumptions about the dimensions of the space. City-block metrics are often used when the psychological dimensions are assumed to be separable. In this case, people can easily separate dimensions and attend to one dimension, such as color and size. Euclidean metrics are often used when the psychological dimensions are not separable and are assumed to be integral. In this case, people cannot attend selectively to one of the dimensions.[8]

Zhang and Montag conducted two experiments to explore observers' abilities to control and distinguish different color attributes. The results show that lightness (L*) is a separable color attribute from a*/b* (hue and chroma). If the lightness attribute of two samples is different, it is easy to judge that the different attribute is lightness. However, a*/b* difference is not separable, and observers could not reliably distinguish changes in chroma from changes in hue.[10] So, if lightness is separable from a*/b*, large color differences might be better calculated in a two-dimensional space that treats lightness differently from chromaticness (a*/b*). One of these dimensions is the lightness difference, and the other is a combined a*/b* difference. The distance between two points in this space is considered to be city block as the two dimensions are perceptually separable.

Note that the a*/b* difference is another dimension that is a space itself. The distance between two points in this space is considered Euclidean as the a*/b* (or alternatively C*/H*) are not perceptually separable.

It is like being on a surface where two points exist on the surface. This is a two-dimensional space. If these two dimensions are separable, then the distance between the two points should be calculated by city-block distance. In

TABLE 1 One stimulus with two color pairs in first category

	Pair number	L*	a*	b*	CIELAB	CIEDE2000	Color presentation
Stimulus A	Pair 1	60	−15	6.5	39.28	28.35	
		60	11.5	−22.5			
	Pair 2	60	−15	6.5	40.80	29.74	
		71	11.5	−22.5			

other words, the distance between two points is the sum of the distance of projection of these two points on each axis.

The projection of these two points on one of the axes is the lightness difference. The projection of these two points on the other axis is a combined a*/b* difference. The distance between the projection of two points on the lightness difference axis is simply lightness difference (ΔL*). However, the distance between the projection of these two points on the a*/b* difference axis is the Euclidean distance of a* and b* of these two points $[(a^*_1 - a^*_2)^2 + (b^*_1 - b^*_2)^2]^{1/2}$. Finally, the overall distance between the two points is the sum of two distances. The color difference (CD) between two points using "ΔL*+Euclidean(a*,b*)" is calculated using Equation (3):

$$CD1 = |\Delta L^*| + \left[(a^*_1 - a^*_2)^2 + (b^*_1 - b^*_2)^2\right]^{1/2} \tag{3}$$

The "ΔL*+Euclidean(a*,b*)" is notated as the "HyAB" color difference formula in the rest of this article—"Hy" to indicate that it is a hybrid model, a combination of city block and Euclidean, and the last letters to show that the Euclidean dimensions are a* and b*. The performance of the 3D city-block distance metric in (L*,a*,b*) was also assessed. The "City Block in (L*,a*,b*)" is notated as cbLAB. According to this metric, the color difference between two points is:

$$CD2 = |\Delta L^*| + |\Delta a^*| + |\Delta b^*| \tag{4}$$

As the CIEDE2000 is a recommended and frequently used formula, its performance was also assessed for large color differences. The CIEDE2000 attributes were used for the definition of "City Block in (L*,C*,H*)" and "ΔL*+Euclidean(C*,H*)." The color difference by "City Block in (L*,C*,H*)" and "ΔL*+Euclidean(C*,H*)" are, respectively, described below:

$$CD3 = |\Delta L^*/k_l S_l| + |(\Delta C^*/k_c S_c)| + |(\Delta H^*/k_h S_h)| \tag{5}$$

$$CD4 = |\Delta L^*/k_l S_l| + \left[(\Delta C^*/k_c S_c)^2 + (\Delta H^*/k_h S_h)^2\right]^{1/2} \tag{6}$$

The "City Block in (L*,C*,H*)" and "ΔL*+Euclidean(C*,H*)" are notated as "cbLCH" and "HyCH" in this article. These formulas are not developed as another style of CIEDE2000. These are developed in order to evaluate the performance of the City Block and combination of City Block-Euclidean distances as new metrics. Each attribute is computed using the normal CIEDE2000 weighting functions. Weighting functions are defined to adjust for the perceived color differences between lightness, chroma, and hue in CIELAB space.[11] These weighting functions are calculated according to their definitions in Sharma et al.[12] The rotation term of CIEDE2000 is not considered here as evaluating the performance of the city block in L*, C*, and H* is the main aim in these formulas. Two datasets were generated in order to assess the performance of these formulas. The properties of these datasets are explained in the following section.

3 | EXPERIMENTS

In order to complete a visual experiment, two datasets were developed. Each dataset contains a variety of categories designed for different purposes. Considering the purpose of each category, the stimuli were generated randomly with some restrictions. The first dataset has four categories that are restricted to large differences in different color attributes. They were designed to evaluate large-difference formulas. Stimuli of the second dataset were restricted to smaller differences. They were designed to assess the relationships between the new formulas with current formulas for smaller color difference. The properties of stimuli in each category are explained in the following subsections. The number of categories in each dataset, number of stimuli in each category, and color differences between samples are reported in Table 2. The color difference between two samples in each pair are calculated using CIELAB.

3.1 | Stimuli with very large color differences

The first dataset contains sample pairs with very large color differences. This dataset was designed for evaluating the performance of current and new formulas in perceived color-difference prediction for very large color differences.

For this dataset, 125 stimuli in four categories were prepared. Each stimulus contains two pairs (four samples). The distribution of samples in CIE L*a*b* color space is shown in Figure 1. All samples are distributed very well through the whole space. Different lightness levels and different regions in a* and b* are very well covered by the samples.

3.1.1 | First category

The first category contains 40 stimuli. Each stimulus contains two pairs. The difference between two samples in the first pair are mainly caused by differences along the a* and b* axes, and the L* of two colors in the first pair

TABLE 2 The statistics of samples in each category (calculated color difference in CIELAB units)

	First dataset				Second dataset	
	First category	Second category	Third category	Fourth category	First category	Second category
Number of stimuli	40	50	20	15	265	214
Min. of color difference	17.36	22.14	13.03	13.45	2.83	5
Max. of color difference	58.06	49.80	86.59	46.01	12.53	11.18
Average	30.96	31.24	40.11	24.90	7.11	6.26
SD	9.26	4.44	17.53	6.58	1.61	0.74
Number of observers	17	17	17	17	—	—

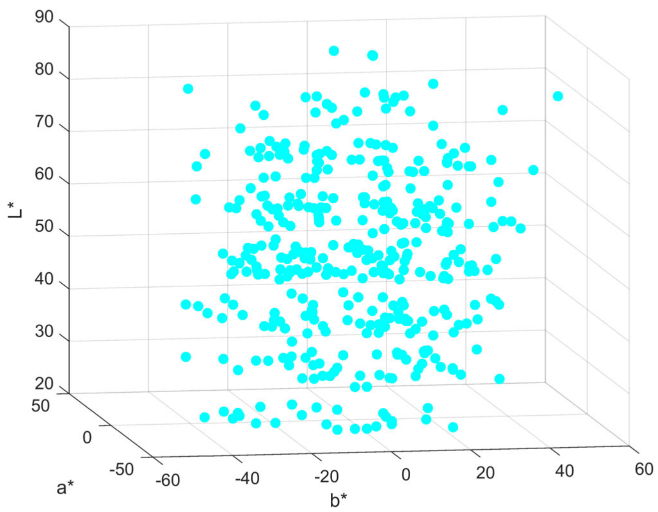

FIGURE 1 Distribution of all samples of first dataset in CIE L*a*b* color space

are equal. In the second pair of each stimulus, a* and b* are the same as in the first pair, but the L* values are significantly different. An example is given in Table 3.

The calculated color differences by CIELAB and CIEDE2000 for two pairs in each stimulus are very close. However, the pairs are not visually similar in perceived color difference according to initial observations used to design the experiments. This category was prepared in order to assess the effect of lightness changes on perceived large color differences by observers. The minimum, maximum, and average color difference between two samples in the first category of the first dataset are 17.36, 58.06, and 30.96 CIELAB units, respectively.

3.1.2 | Second category

The second category contains 50 stimuli. Each stimulus has two pairs. One of the pairs has large differences in L* and a* (or b*) and no difference in b* (or a*). The second pair has large differences in all three CIELAB axes. Therefore, the second color of the second pair has different "hue and chroma" in comparison with the second color of the first pair. An example is presented in Table 4.

It is not easy to say which pair has a larger color difference. The calculated color difference for two pairs by CIELAB and CIEDE2000 are also similar. This category was prepared in order to assess the perceived color difference with constant lightness difference but with significant changes in the a*/b* differences. If it is indeed difficult to judge the relative differences between two pairs, then an ideally calculated color difference should be similar. The minimum, maximum, and average color differences between two samples in the second category of the first dataset are 22.14, 49.80, and 31.24 CIELAB units, respectively.

3.1.3 | Third category

The third category contains 20 stimuli. There are two pairs in each stimulus. The first color of both pairs is the same, but the second color of the second pair has different lightness and chroma in comparison with the second color of the first pair. This category was prepared for evaluating the effect of different lightness and chroma on perceived color difference. The minimum, maximum, and average color differences between two samples in the third category of the first dataset are 13.03, 86.59, and 40.11 CIELAB units, respectively.

3.1.4 | Fourth category

The fourth category has 15 stimuli. There are two pairs in each stimulus. The first color of both pairs is the same, but the second color of the second pair has different

TABLE 3 One stimulus with two color pairs in first category

	Pair number	L*	a*	b*	CIELAB	CIEDE2000	Color presentation
Stimulus B	Pair 1	60	−2.25	10.5	31.70	28.97	
		60	21.5	−10.5			
	Pair 2	72	−2.25	10.5	33.90	30.57	
		60	21.5	−10.5			

TABLE 4 One stimulus with two color pairs in second category

	Pair number	L*	a*	b*	CIELAB	CIEDE2000	Color presentation
Stimulus A	Pair 1	70	−17.5	26.5	41.18	29.46	
		50	−17.5	−9.5			
	Pair 2	70	−17.5	26.5	42.76	31.46	
		50	−6	−9.5			

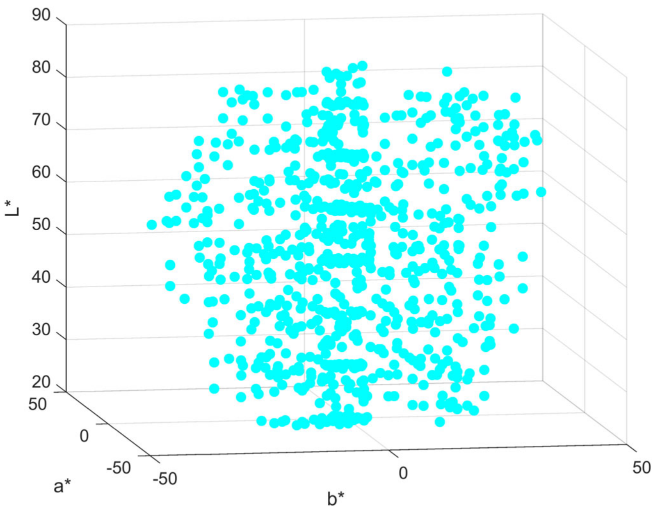

FIGURE 2 Distribution of first category of second dataset in CIE L*a*b* color space

"Lightness, Hue and Chroma" in comparison with the second color of the first pair. This category was prepared in order to evaluate the effect of difference in all three color dimensions on perceived color difference. The minimum, maximum, and average color differences between two samples in the fourth category of the first dataset are 13.45, 46.01, and 24.90 CIELAB units, respectively.

3.2 | Stimuli with small color differences

The calculation of perceived color difference between two adjacent colors in images is a vital capability in image processing and computer vision applications. It is possible that any color pairs with different color difference magnitudes might be adjacent in images and scenes. In addition, color pairs in scenes and images could have color differences ranging from small to large magnitudes. Current color difference formulas are suitable for small color differences. If new formulas are consistent with current formulas, then it could also be used for small color differences. Thus, one formula could be used for the whole range of color difference in images.

In order to assess this issue, a second dataset was prepared. The second dataset has two categories. In the first category, there are 265 stimuli; each stimulus contains two pairs. There is a small difference between a* and b* of two samples in each pair. The a* and b* of the second pair is the same as the first pair, but L* is different. The minimum, maximum, and average color differences between two samples in the first category of the second dataset are 2.83, 12.53, and 7.11 CIELAB units, respectively.

The second category has 214 stimuli. Each stimulus has two pairs. The two samples of the first pair have a small difference in two axes of the CIE L*a*b* color space, but the two samples of the second pair have small differences in all three dimensions. The minimum, maximum, and average color differences between two samples in the second category of the second dataset are 5, 11.18, and 6.26 CIELAB units, respectively.

The distributions of the first and second categories are shown in Figures 2 and 3. As shown, all samples are very well distributed. Different lightness levels and different regions in a* and b* are very well covered by the samples.

3.3 | Visual experiment

A visual experiment was conducted in order to evaluate the performance of different formulas in perceived color-difference prediction for large color differences. The visual experiment was conducted by presenting the stimuli to the observers on a well-calibrated and characterized LCD (Liquid Crystal Display) monitor. For this purpose, a 30-in. Apple Cinema HD LCD Display, controlled by a Macintosh Pro in 30-bit color mode, was used. The LCD display was carefully characterized using the techniques described by Fairchild et al[13] and Day et al.[14]

The measurements of the characterization color patches were completed using a Photo Research PR-655 spectroradiometer. A total of 260 color patches were used for the characterization and model development. These patches were measured twice across 2 days. The color difference between the two measurements were calculated using the CIELAB color difference formula. The minimum, maximum, average, and SD of color differences of 260 patches were 0, 1.589, 0.632, and 0.259 CIELAB units, respectively. The measurements were quite repeatable.

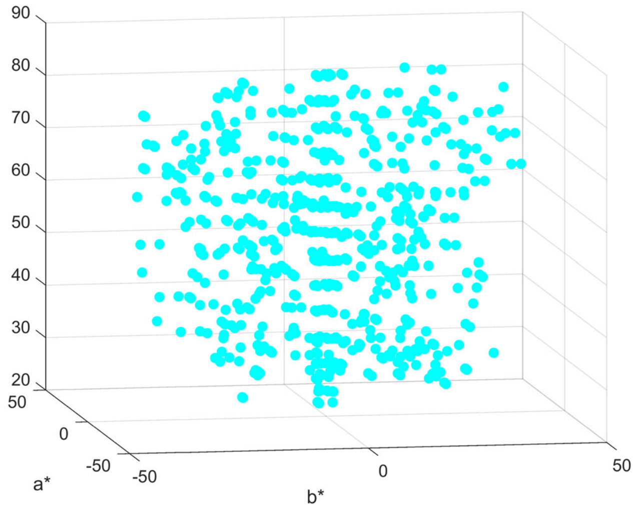

FIGURE 3 Distribution of second category of second dataset in CIE L*a*b* color space

In order to illustrate the stimuli, the "RGB" of color patches were imported to the user interface. The color patches were presented on a gray background. The color of the background was (L* = 50, a* = 0, b* = 0). There was a thin black border around each color patch to aid color discrimination between the patches. Figure 4 is a screenshot of the observer interface.

The experiment was conducted in a darkened room. All four categories of the first dataset, which includes 125 stimuli, were tested together in one session. Completing the experiment took about 30 minutes on average. The distance of the observer from the display was 70 cm. Each color patch was set to be 2° on each side for the 70 cm viewing distance. The stimuli were shown randomly, not by category. Each trial contained two pairs, one pair on the left and the other on the right. Appearance as the left or right pair and top or bottom color was random for each stimulus. There were seven buttons at the bottom of the observer interface. The observer was required to choose one of these buttons. Observer instructions were as follows:

> In each trial you will see two color pairs on a gray background. Please compare the color difference between the pair on the left with the color difference between the pair on the

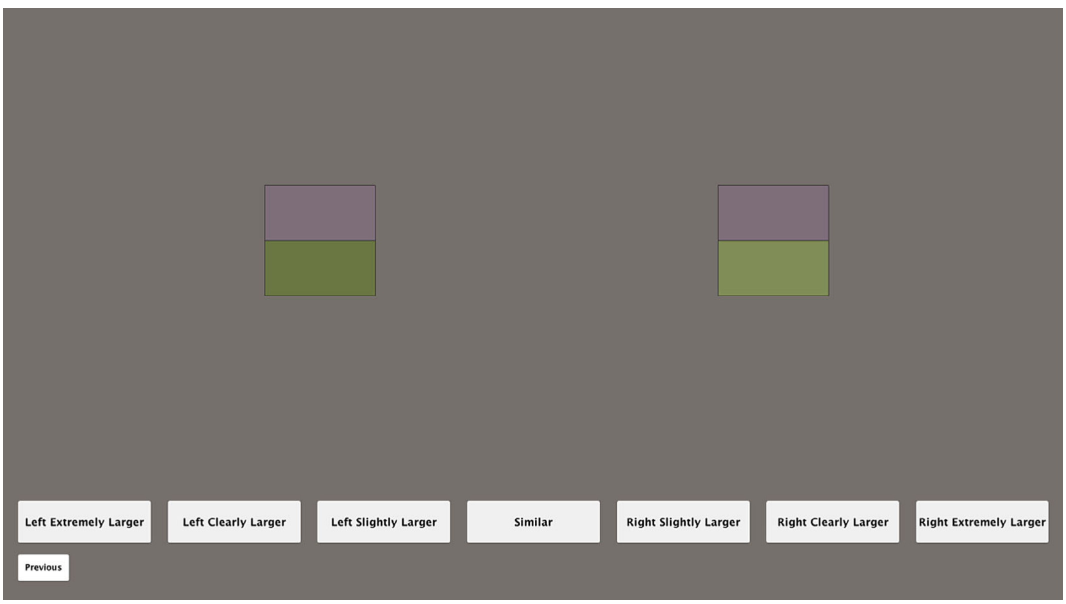

FIGURE 4 Screenshot of observer interface

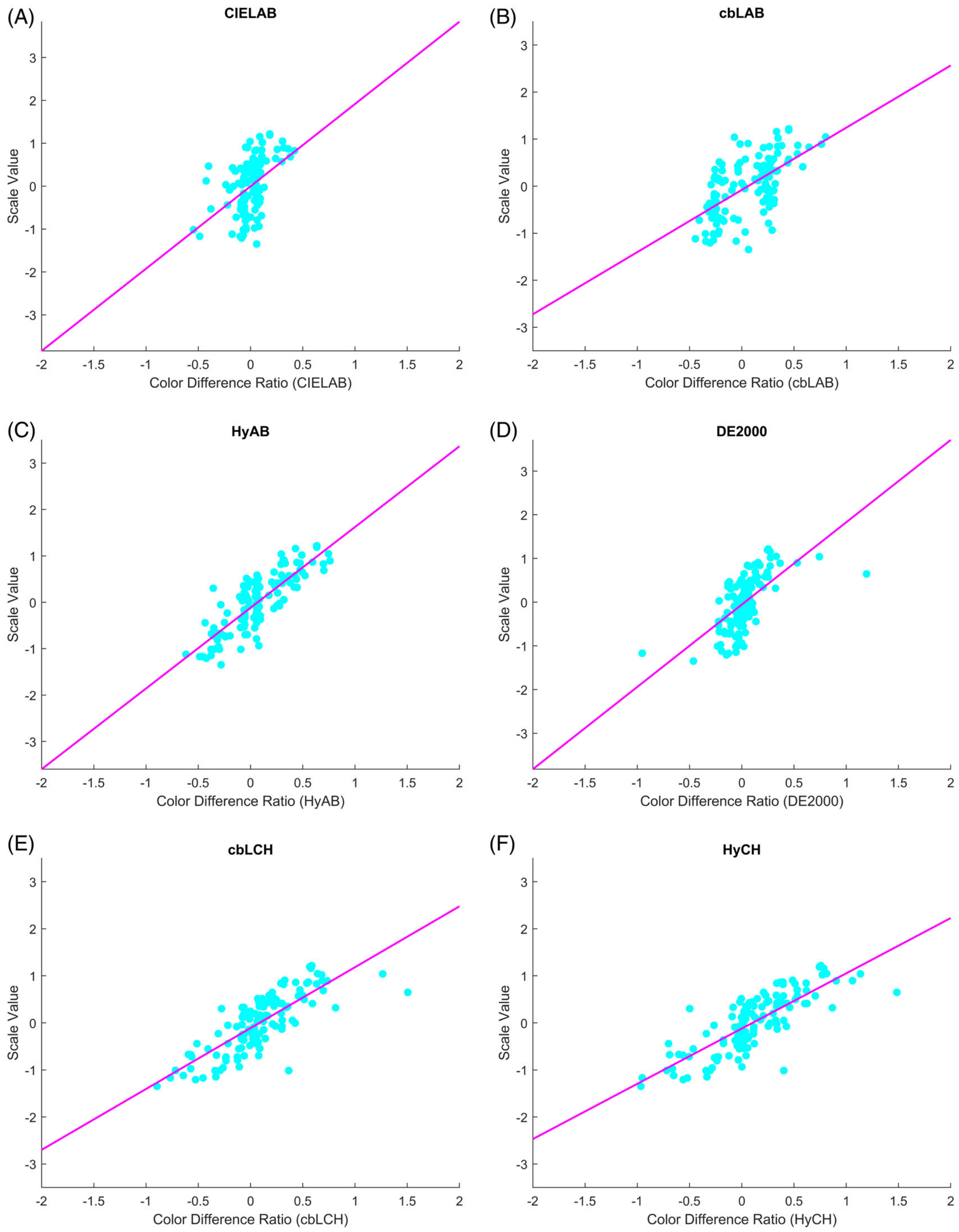

FIGURE 5 The color difference ratio for different formulas vs scale value for all stimuli of first dataset. A, CIELAB; B, cbLAB; C, HyAB; D, CIEDE2000; E, cbLCH; and F, HyCH

right. For each set, you have to make one of seven choices.

Left Extremely Larger
Left Clearly Larger
Left Slightly Larger
Similar
Right Slightly Larger
Right Clearly Larger
Right Extremely Larger

The descriptions for each of the seven possible selections should be self-explanatory. When you make a selection, a new set of colors will appear after a short intermission. If you want to revise your judgment, press "Previous" button to return to the most recently judged set.

TABLE 5 The R^2, intercept, and slope of color difference ratio plots (first dataset)

		CIELAB	cbLAB	HyAB	CIEDE2000	cbLCH	HyCH
First category	R^2	0.81	0.83	0.85	0.67	0.76	0.78
	Intercept	−0.20	−0.18	−0.19	−0.15	−0.15	−0.18
	Slope	8.80	2.48	1.90	4.70	1.54	1.39
Second category	R^2	0.04	0.02	0.05	0.18	0.18	0.22
	Intercept	−0.11	−0.12	−0.11	−0.13	−0.12	−0.13
	Slope	1.12	0.24	1.17	2.16	1.86	2.18
Third category	R^2	0.60	0.58	0.61	0.44	0.54	0.41
	Intercept	0.37	0.16	0.14	0.30	0.03	0.02
	Slope	1.27	1.02	1.04	1.70	1.29	1.14
Fourth category	R^2	0.05	0.02	0.73	0.58	0.77	0.88
	Intercept	0.06	0.04	−0.02	−0.04	−0.05	−0.08
	Slope	1.19	1.78	2.07	1.29	1.08	1.08
All stimuli	R^2	0.23	0.57	0.78	0.60	0.76	0.75
	Intercept	0.01	−0.07	−0.12	−0.06	−0.10	−0.16
	Slope	2.03	1.35	1.76	2.84	1.42	1.25

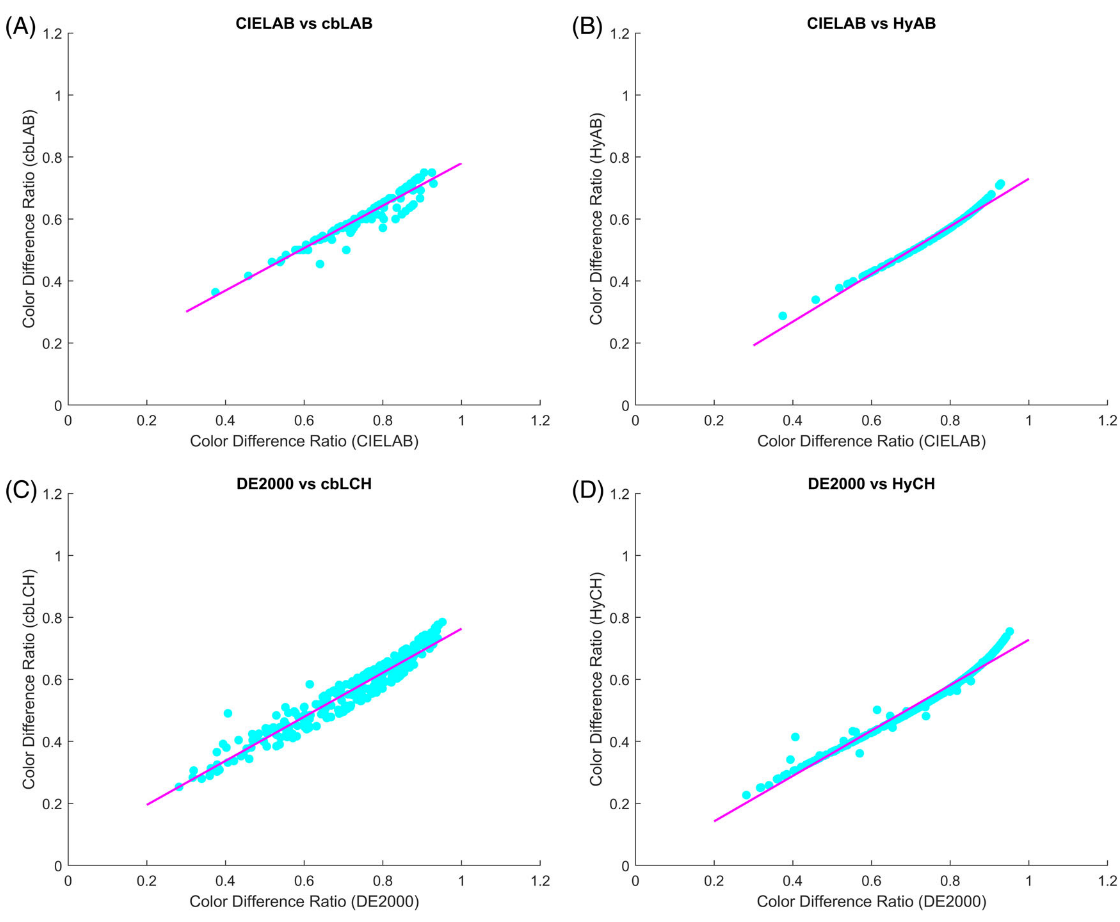

FIGURE 6 The color difference ratio between two pairs in first category of second dataset: A, CIELAB vs cbLAB; B, CIELAB vs HyAB; C, CIEDE2000 vs cbLCH; and D, CIEDE2000 vs HyCH

Before starting the main experiments, each observer completed a training experiment. The training contained four stimulus sets, which were a good representation of the range of color differences found in all stimuli. There were 17 observers. They were all faculty members and students of the RIT "Munsell

Color Science Laboratory" and tested positively for normal color vision.

4 | RESULTS

The results of experiments are presented in this section. The visual experiment was conducted using the first dataset for assessing large color differences. The second dataset was used only for computations and comparing the results of different formulas for smaller color differences.

4.1 | Large color difference

All 125 stimuli of the first dataset were judged by all observers. The Z-score for each response of all observers was calculated. As mentioned in Section 3.3, there were seven choices or seven categories for each stimulus. These choices or categories were assumed simply to have an equal interval. The first choice was assumed as "1," second choice as "2," third choice as "3," and so on. The "Scale Value" of each stimulus was calculated by averaging of the Z-score of each stimulus across all observers.[15]

$$Z_{ij} = \left(R_{ij} - MR_j\right)/SDR_j \tag{8}$$

The Z_{ij} is the Z-score for stimulus i by observer j. The R_{ij} is rating assigned to stimulus i by observer j. For example, $R_{15,4}$ is 6. This means that observer 4 chose "6" or "Right Clearly Larger" for stimulus 15. The MR_j is the average rating assigned to all stimuli by observer j. The SDR_j is the SD of ratings assigned by observer j.[15]

The scale values of all stimuli vs color difference ratio, which are calculated by different formulas, are shown in Figure 5. The color difference ratio was calculated by Equation (9):

$$CDR2 = (CD1 - CD2)/(\min\{CD1, CD2\}) \tag{9}$$

The CD1 is "Color Difference between two samples in 1st pair." The CD2 is "Color Difference between two samples in 2nd pair." If the calculated color-difference ratio

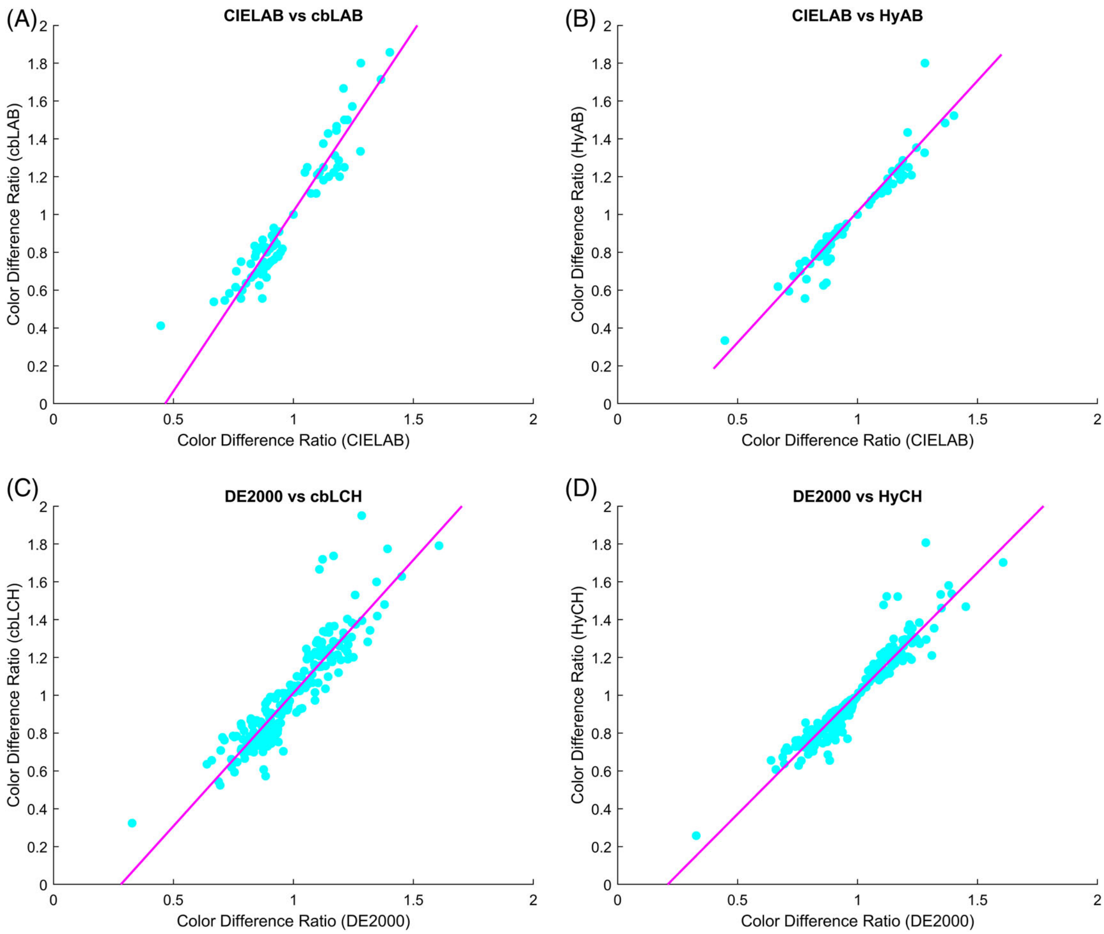

FIGURE 7 The color difference ratio between two pairs in second category of second dataset: A, CIELAB vs cbLAB; B, CIELAB vs HyAB; C, CIEDE2000 vs cbLCH; and D, CIEDE2000 vs HyCH

TABLE 6 The R^2, intercept, and slope of color difference ratio plots (second dataset)

		CIELAB/cbLAB	CIELAB/HyAB	CIEDE2000/cbLCH	CIEDE2000/HyCH
First category	R^2	0.985	0.993	0.926	0.987
	Intercept	0.072	−0.040	0.046	0.003
	Slope	0.724	0.768	0.720	0.717
Second category	R^2	0.905	0.983	0.866	0.939
	Intercept	−0.923	−0.263	−0.340	−0.221
	Slope	1.938	1.272	1.340	1.226

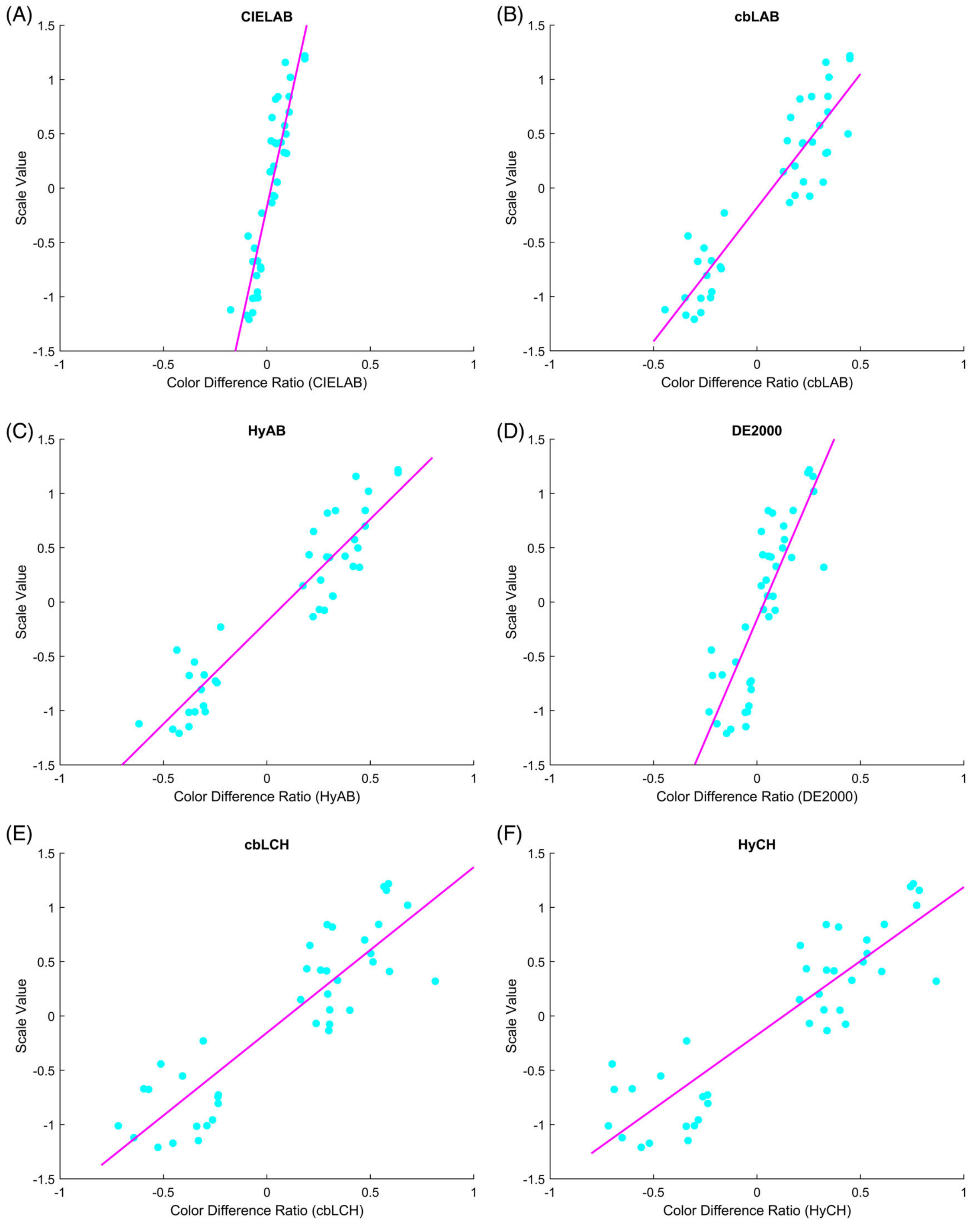

FIGURE 8 The color difference ratio for different formulas vs scale value for first category of first dataset

is negative, it means the calculated color difference of the second pair (left pair on screen) is larger than the first pair (right pair on screen). If it is positive, the difference of the first pair (right pair on screen) is larger than the second pair (left pair on screen). The color difference between two colors was calculated by each of the investigated formulas. The R^2, slope, and y-intercept of each plot is reported in Table 5. It should be mentioned that the statistical properties of plots are all calculated using "fitlm" function in MATLAB with the robust fitting method as the regression method. Using "fitlm" with the "RobustOpts" method creates a model that is not significantly affected by outliers.[16]

4.2 | Small color difference

The color difference of two samples in each pair of the second dataset was calculated using the CIELAB, CIEDE2000, cbLAB, HyAB, cbLCH, and HyCH formulas.

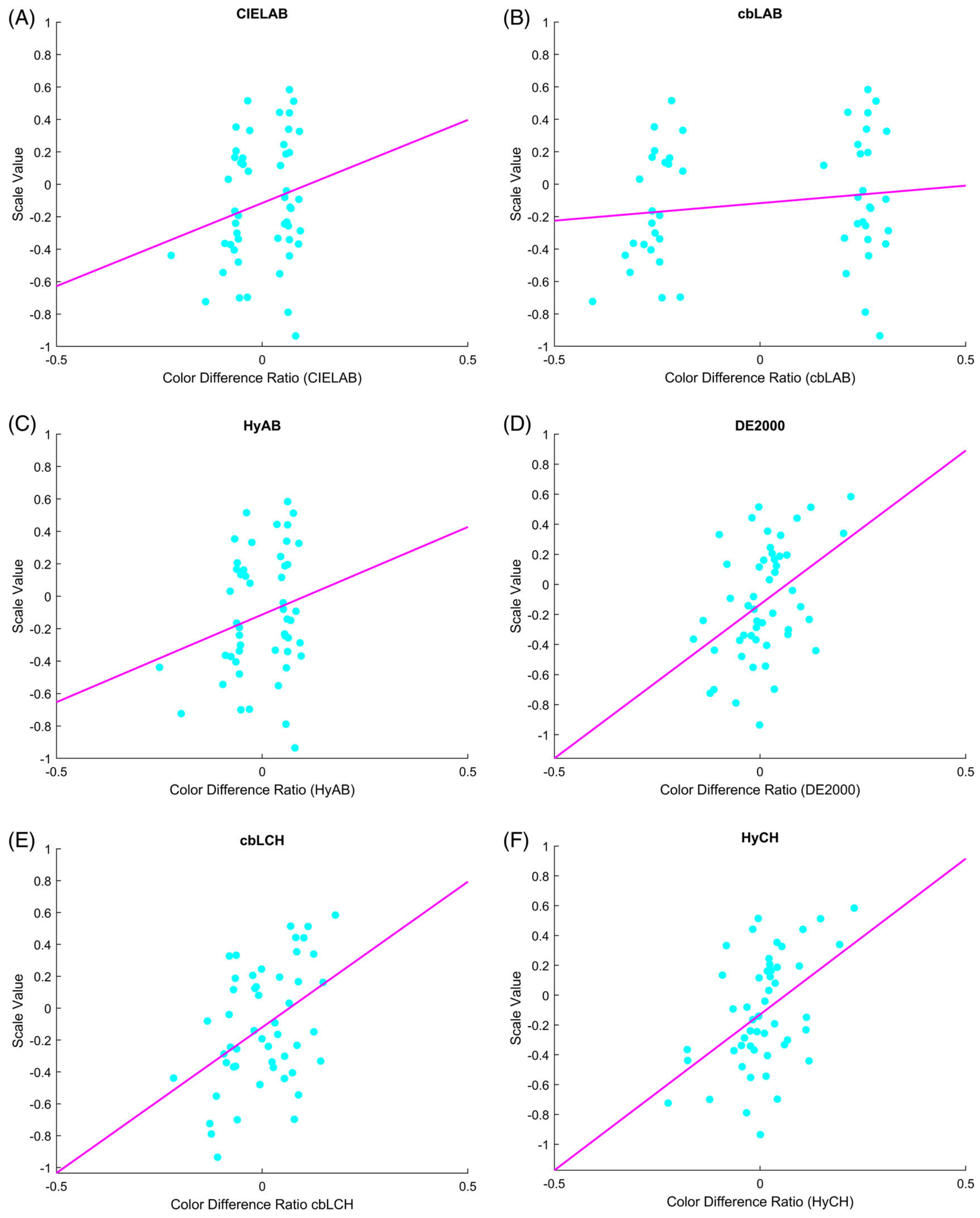

FIGURE 9 The color difference ratio for different formulas vs scale value for second category of first dataset

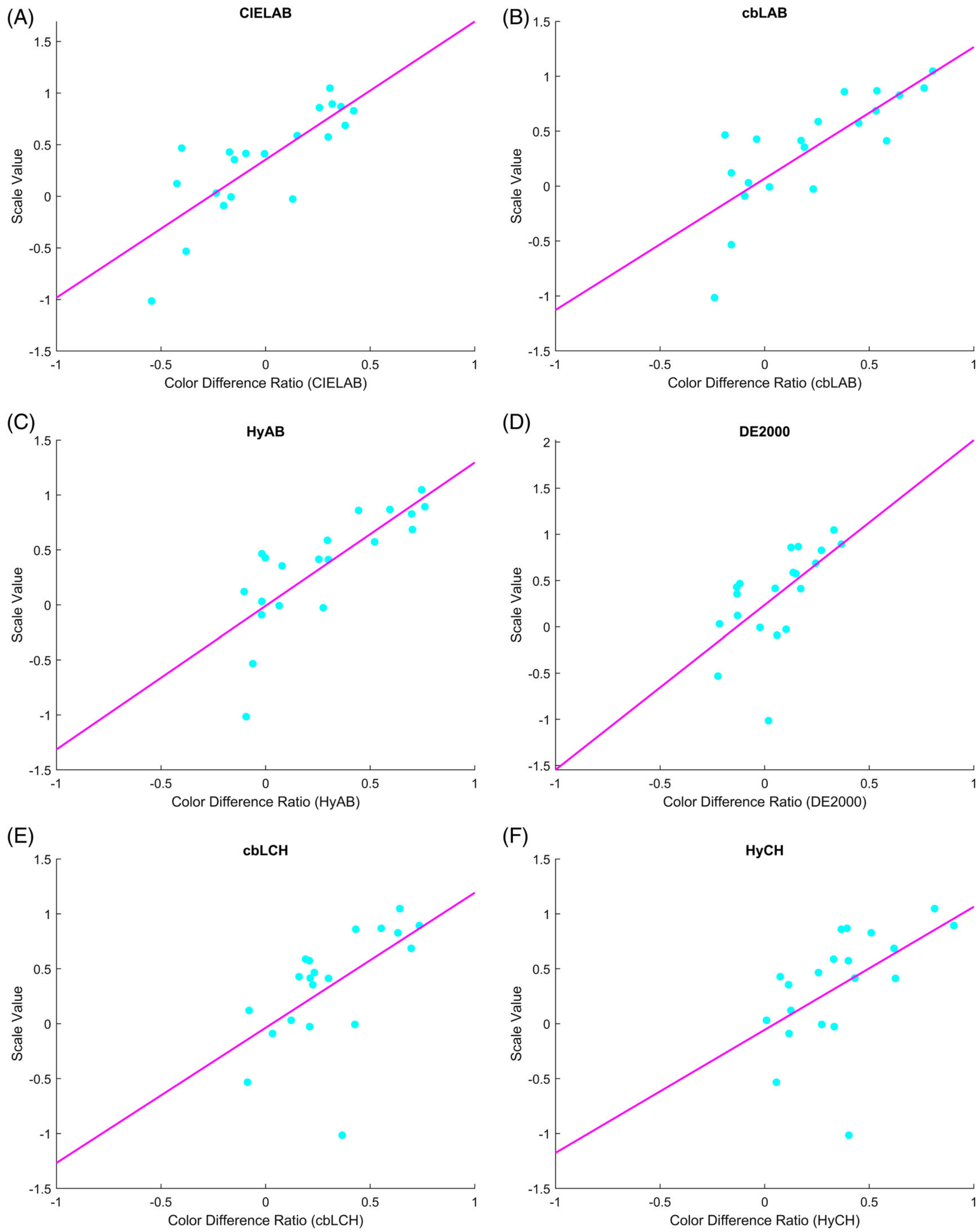

FIGURE 10 The color difference ratio for different formulas vs scale value for third category of first dataset

The color difference ratio of two pairs in each stimulus was calculated for each formula using Equation (7):

$$\mathrm{CDR1} = \mathrm{CD1}/\mathrm{CD2} \tag{7}$$

The CDR1 is for "Color Difference Ratio 1." The CD1 and CD2 are "Color Difference" of the first and second pair in each stimulus, respectively. The CDR1 for different formulas was assessed. The results are discussed in the next sections. The performances of cbLAB and HyAB are compared directly with CIELAB because their dimensions are the same. The performances of cbLCH and HyCH are also compared with CIEDE2000 because their dimensions are the same.

For this purpose, the ratio of color difference of the first pair to the color difference of the second pair for each stimulus is calculated using the different formulas. The results are shown in Figures 6 and 7 for first and

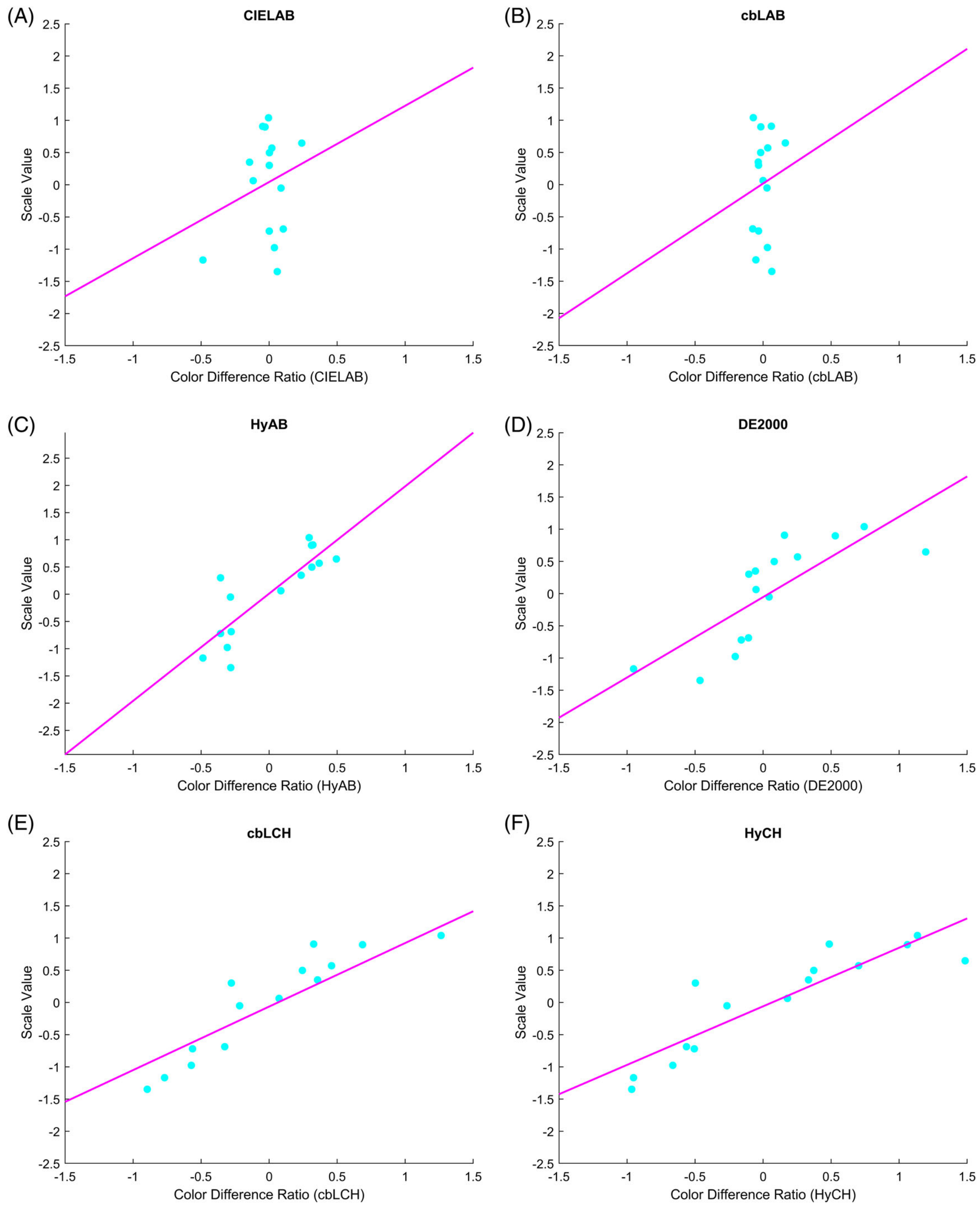

FIGURE 11 The color difference ratio for different formulas vs scale value for fourth category of first dataset

second categories of second dataset, respectively. The R^2, slope, and y-intercept of plots are reported in Table 6.

5 | ANALYSIS OF RESULTS

5.1 | Analysis of very large color difference results

According to Figure 5A,D, the CIELAB and CIEDE2000 formulas do not correlate well with perceived large color difference. The calculated color difference ratio by these two formulas are very close to zero, but the perceived color differences for most of the stimuli are significant according to observers' judgments. This shows that CIELAB and CIEDE2000 are not consistent with the visual experiment on very large color differences.

In addition, the R^2 of CIELAB and CIEDE2000 for all stimuli are 0.23 and 0.60, respectively. Their plots have a higher slope in comparison with the other formulas. So, CIELAB and CIEDE2000 are not suitable for the calculation of very large color differences.

According to Figure 5C, and considering R^2, it could be concluded that HyAB is the best formula for the calculation of large color difference. It has a good correlation with perceived color difference. It shows the L* can indeed be considered a separable attribute. As mentioned earlier, the relative effect of difference in the L* attribute is reduced by Euclidean distance when there is a large difference between hue and chroma or between a* and b*. For more details, it is better if the results of visual experiments are analyzed for each category separately.

In the first category, the lightness difference of the second pair is significantly larger than for the first pair for each stimulus. According to Figure 8A,D, the calculated color difference ratio with CIELAB and CIEDE2000 are very close to zero. The slope of two plots are also larger than the other plots. This means that perceived color difference of first pair should be similar to perceived color difference of second pair, but the result of visual observations shows that they are significantly different. For the first category, HyAB has the best correlation (R^2) with the scale value. cbLAB, cbLCH, and HyCH also demonstrate good performances.

The results show that "difference in Lightness" has a significant effect on the perceived color difference for large color differences. For the utilized large color differences, the difference between a* and b* or between "Hue and Chroma" are large, so the magnitudes of calculated color difference by formulas that use a Euclidean distance metric are mostly affected by a* and b* or "Hue and Chroma." Therefore, the effect of difference in lightness is almost neglected. It can be concluded that the color difference formulas that use Euclidean distance as the distance metric (for all dimensions) are not suitable for large color differences.

In the second category, there is some ambiguity in the visual assessments. If the calculated color difference ratio is negative, it means the calculated color difference of the left pair is larger than the right pair. If it is positive, the difference of the right pair is larger than the left pair. According to Figure 9, for all color difference formulas, when the calculated color difference ratio is negative, some of the scale values are negative and some of them are positive. In addition, when the calculated color difference ratio is positive, some of scale values are positive and some of them are negative.

This means there is often poor agreement between color difference formulas and visual assessment for this category. The observers' judgment are in agreement with color difference formulas for some stimuli but are in ordinal disagreement for other stimuli. So, it could be claimed for this category that the differences between two pairs are perceptually similar or at least very difficult to judge in comparison to one another.

The cbLAB has a weak performance for this category. It has a significant calculated ratio for all stimuli. The HyAB has a small range around zero. It means, calculated color difference by this formula is similar for both pairs, so it could be a good formula for this category.

The results for third and fourth categories are shown in Figures 10 and 11, respectively. According to Figure 10, the HyAB has the best performance for the third category stimuli. The other color difference formulas have a positive ratio for negative scale value for some stimuli or vice versa. For the fourth category, the HyAB, cbLCH, and HyCH have good correlation with visual observations.

5.2 | Analysis of small color difference results

As shown in Figure 6, the HyAB has perfect agreement with CIELAB for the first category of the second dataset. The HyCH also has a good consistency with CIEDE2000 for this category.

For the second category of the second dataset (Figure 7), the HyAB has the best agreement with CIELAB. The HyCH is also consistent with CIEDE2000. So, the HyAB and HyCH could be used instead of CIELAB and CIEDE2000 for small and medium ranges of color differences, as well as the large color differences that were the focus of this research.

6 | CONCLUSION

It is reasonable to conclude that the HyAB formula has the best correlation with visual observations for very large color differences. It has the best performance for the first category and third category of the first dataset. It also holds up well for smaller color differences and thus is a good candidate formula for image processing and computer vision applications. Finally, it has the advantage of being based on the computationally simpler CIELAB model rather than the weighted CIEDE2000 dimensions.

The HyAB, cbLCH, and HyCH all have good performance for the fourth category of the first dataset. This does not change the above conclusion.

For stimuli in the second category of the first dataset, it is hard to judge between pairs. If the differences between pairs are similar, then HyAB could be a good formula for this category as well. Because it has a small range around zero for this category, the calculated color difference by this formula is similar for both pairs.

The HyAB has the best consistency with CIELAB for all categories of the second dataset. The HyCH has a good consistency with CIEDE2000 for the second dataset as well. So, the HyAB and HyCH could be used for small to large color difference calculation. However, the HyAB has a slightly better performance and the advantage of simpler computation.

It is thus concluded that, for very large color differences, the effect of lightness difference should be considered perceptually separable, and the HyAB color difference formula is the most practical and accurate solution.

ACKNOWLEDGEMENTS

The authors thank Prof. Roy S. Berns for his advice and the observers who took part in this study. Prof. Michael Murdoch kindly helped us with LCD characterization and visual experiment setup.

ORCID

Saeedeh Abasi https://orcid.org/0000-0002-8240-3238

REFERENCES

[1] Wang H, Cui G, Luo MR, Xu H. Evaluation of color-difference formula for different color-difference magnitudes. *Color Res Appl*. 2012;37(5):316-325.

[2] Forsyth D, Ponce J. *Computer Vision: A Modern Approach*. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2003.

[3] Gonzalez RC, Woods RE. *Digital Image Processing*. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2002.

[4] Manjunath BS, Ohm JR, Vasudevan VV, Yamada A. Color and texture descriptors. *IEEE Trans Circuits Syst Video Technol*. 2001;11(6):703-715.

[5] Luo MR, Cui G, Li C. Uniform color spaces based on CIECAM02 color appearance model. *Color Res Appl*. 2006;31(4):320-330.

[6] Fairchild MD. *Color Appearance Models*. 3rd ed. New York, NY: Wiley; 2013.

[7] Guan SS, Luo MR. A colour-difference formula for assessing large colour differences. *Color Res Appl*. 1999;24(5):344-355.

[8] Pashler H. *Steven's Handbook of Experimental Psychology*. 3rd ed. New York, NY: Wiley; 2003.

[9] Berns RS. *Billmeyer and Saltzman's Principles of Color Technology*. 4th ed. New York, NY: Wiley; 2019.

[10] Zhang H, Montag ED. How well can people use different color attributes? *Color Res Appl*. 2006;31(6):445-457.

[11] Johnson GM, Fairchild MD. A top down description of S-CIELAB and CIEDE2000. *Color Res Appl*. 2003;28(6):425-435.

[12] Sharma G, Wu W, Dalal EN. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Res Appl*. 2005;30(1):21-30.

[13] Fairchild MD, Wyble D. *Colorimetric Characterization of the Apple Studio Display (Flat Panel LCD)*. Munsell Color Science Laboratory Technical Report; 1998.

[14] Day E, Taplin L, Berns R. Colorimetric characterization of a computer-controlled liquid crystal display. *Color Res Appl*. 2004;29(5):365-373.

[15] Brown TC, Daniel TC. *Scaling of Ratings: Concepts and Methods*. Res. Pap. RM-293. Fort Collins, CO: U.S. Department of Agriculture, Forest Service; 1990:1-24.

[16] MathWorks, (R2018a). https://www.mathworks.com/help/stats/fitlm.html. Retrieved June 18, 2019.

AUTHOR BIOGRAPHIES

Saeedeh Abasi received a B.S. (2011) and M.S. (2015) from Amirkabir University of Technology. She is a PhD student at Amirkabir University of Technology and a visiting scholar at Munsell Color Science Laboratory, Rochester Institute of Technology. Her research interests are color science and image processing.

Mohammad Amani Tehran received a B.Sc. from Isfahan University of Technology (1987), an M.S. from Sharif University of Technology (1990), and a Ph.D. from Amirkabir University of Technology (2001). He is a professor in the Textile Engineering Department at the Amirkabir University of Technology. His research interests are color technology, modeling, image processing, and nanofibers.

Mark D. Fairchild is Professor and Founding Head of the Integrated Sciences Academy in RIT's College of Science and Director of the Program of Color Science and Munsell Color Science Laboratory. He received B.S. and M.S. degrees in imaging science from R.I.T. and a Ph.D. in vision science from the University of Rochester.

How to cite this article: Abasi S, Amani Tehran M, Fairchild MD. Distance metrics for very large color differences. *Color Res Appl*. 2019;1–16. https://doi.org/10.1002/col.22451