Color Research & Application

WILEY

COLOR RESEARCH AND APPLICATION

RESEARCH ARTICLE OPEN ACCESS

Acceptable Color Differences for Printed Color Samples

Min Huang[1] | Xiaoyu Shang[1] | Xuping Gong[1] | Minchen Wei[2,3,4] | Dan Wang[1] | Yu Liu[1] | Xiu Li[1]

[1]Printing and Packaging Engineering, Beijing Institute of Graphic Communication, Beijing, China | [2]Color, Imaging, and Illumination Laboratory, The Hong Kong Polytechnic University, Kowloon, Hong Kong | [3]Color Imaging and Metaverse Research Center, The Hong Kong Polytechnic University, Kowloon, Hong Kong | [4]PolyU-Daya Bay Technology and Innovation Research Institute, Guangdong, China

Correspondence: Minchen Wei (minchen.wei@polyu.edu.hk)

Received: 11 March 2025 | **Revised:** 9 April 2025 | **Accepted:** 16 April 2025

Funding: This work was supported by the Research Grants Council, University Grants Committee (PolyU R5006-21) and College's Scientific Research Project (BIGCEc202302).

ABSTRACT

Three experiments were carried out to investigate how different parameters affected the color difference evaluations. In each experiment, 450 pairs of printed samples, with the chromaticities surrounding nine CIE recommended color centers, were prepared, and the color difference of each pair was evaluated by a group of observers. The results suggested that the experienced observers were more sensitive to color differences, and the observers were more sensitive to color differences in the standard viewing booth with the directional illumination. The performance of two widely used color difference metrics (i.e., CIELAB and CIEDE2000) was evaluated using the experiment results, and optimizations were then performed on the k_L, k_C, and k_H factors, respectively. It was found that an optimized value of k_H in the CIELAB formula and those of k_L and k_H in the CIEDE2000 formula can significantly improve the performance, suggesting that the observers were much more sensitive to differences in the hue dimension. Threshold values for color difference and hue difference were proposed based on the experiment results.

1 | Introduction

Color measurement and color quality evaluation are critically important to a wide range of industries related to color, such as display, printing, and textile. The evaluations on product quality are typically performed based on perceived color difference [1–4]. Though color difference metrics and formulas have been developed, it is commonly found that the threshold of calculated color difference does not also match the perceived color difference. For example, a pair of color samples may have the calculated color difference below a certain value, but the color difference is visually unacceptable. Some past studies (e.g., [5–7]) suggested that such phenomena could be due to the different tolerance of color difference in different regions in a color space. It could also be due to the different viewing conditions, in comparison to those recommended by CIE [8] and CY/T3 1999 [9].

Color perception involves three different factors—the observer, the light source, and the object. Any change in one of these three factors could introduce differences in color perception and color difference judgments. Most studies only focused on one of these three factors, such as experienced versus inexperienced observers [3], different light sources [10–12], and different light levels [13]. Some studies also proposed color difference thresholds for different types of samples [3, 12, 14–17].

In this study, we focused on printed color samples, but we investigated how experienced and inexperienced observers and types of light source affect the judgment of color difference. A

https://doi.org/10.1002/col.22988

total of 450 pairs of printed color samples were carefully prepared, with the color difference ranging between 1.0 and 3.0 CIELAB units.

2 | Methods

2.1 | Preparation of Color Samples

An EPSON Stylus Pro 7980 inkjet printer on a substrate with a gloss level of 40 units measured at an angle of 75° by a TC-108DPA gloss meter was used to prepare color samples with a size of 5.0 cm × 5.0 cm. The spectral reflectance distributions of the samples were measured 2 weeks after the preparation, in the 0°:45° geometry using an X-Rite eXact spectrophotometer, so that the colors were stable. A total of 450 samples, with 50 samples around each of the nine color centers recommended by the CIE [18], were prepared. A total of three experimental sessions were carried out. In order to avoid color fading and rubbing during the experiment, we prepared three sets of the 450 color samples, with each set used for an experimental session.

The color samples were designed and prepared to have a color difference from the corresponding color center, ranging from 1.0 to 3.0 units in CIELAB (ΔE^*_{ab}) using the CIE Illuminant D65 and the CIE 1964 10°Color Matching Functions, with the difference from lightness (ΔL^*), hue (ΔH^*_{ab}), or chroma (ΔC^*_{ab}). Figure 1 shows the distribution of the nine color centers for the three sets of samples, and Figure 2 shows the distribution of the color differences for the 150 color samples for the red color center (i.e., Center No. 2). Table 1 summarizes the minimum, maximum, and average values of ΔL^*, ΔH^*_{ab}, and ΔC^*_{ab}. Figure 3 shows the histogram of the color difference values. The percentages of the samples with the differences in the three attributes, such as ΔL^*, ΔH^*_{ab}, and ΔC^*_{ab} in the three experiments were 33%, 33%, and 34%; 30%, 39%, and 31%; and 37%, 32%, and 31%, respectively, suggesting the samples were uniformly distributed.

2.2 | Experiment Design

The experiment was designed into three sessions. Exps. I and II were designed to compare the differences between two groups of observers under daylight illumination, with Exp. I carried out at Beijing Institute of Graphic Communication (BIGC) with student observers and Exp. II carried out at Suining Kuanzhai Printing Co. Ltd. with the professional employees having at least one-year experience in quality inspection. Exps. I and III were designed

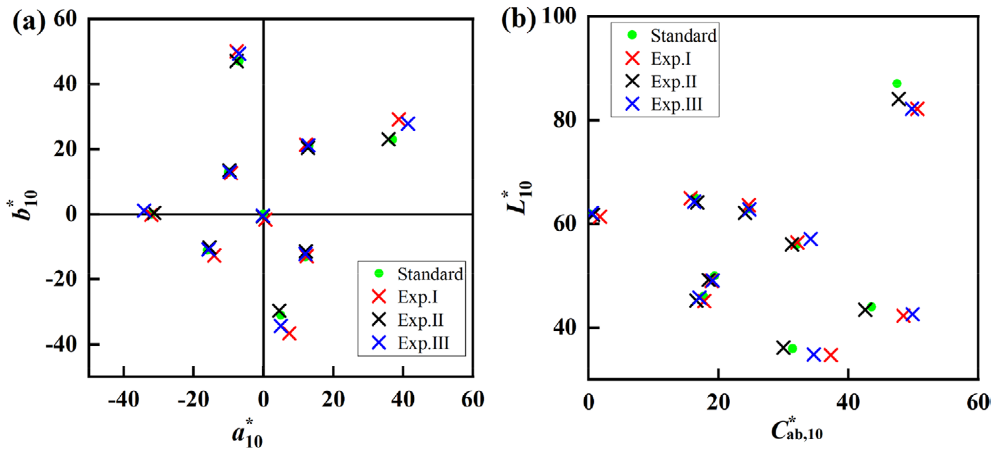

FIGURE 1 | Distributions of the nine color centers of the three sets of color samples, and the CIE recommended color centers, in the CIELAB color space. (a) a^*_{10}–b^*_{10} plane. (b) L^*_{10}–$C^*_{ab,10}$ plane.

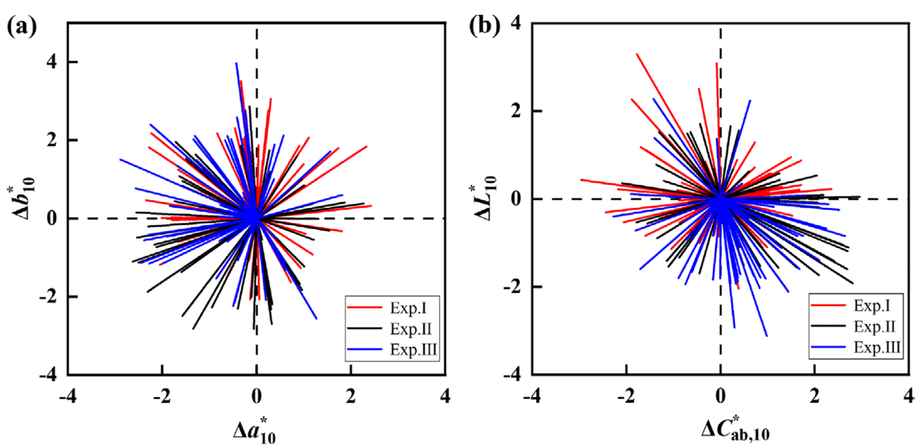

FIGURE 2 | Distributions of color differences for the 150 color samples for the red color center (i.e., Center No. 2) for the three sets of color samples in the CIELAB color space. (a) Δa^*_{10}–Δb^*_{10} plane. (b) ΔL^*_{10}–$\Delta C^*_{ab,10}$ plane.

TABLE 1 | Summary of the color differences of the color samples in each set.

	Exp. I			Exp. II			Exp. III		
	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
$\triangle E^*_{ab}$	3.88	0.27	1.91	3.92	0.21	2.04	4.33	0.13	2.00
$\triangle L^*$	3.33	0.00	0.94	3.17	0.00	1.02	3.53	0.00	1.05
$\triangle C^*_{ab}$	3.35	0.01	0.98	3.90	0.00	1.08	3.93	0.00	1.00
$\triangle H^*_{ab}$	3.16	0.00	0.92	2.94	0.01	0.95	3.50	0.00	0.90

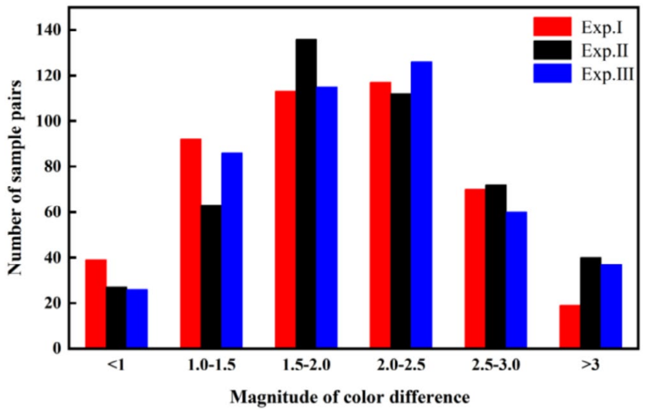

FIGURE 3 | Histogram of the magnitude of color differences of the three sets of color samples.

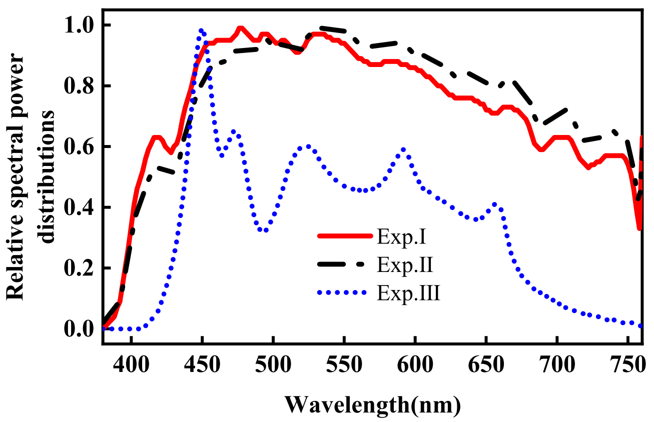

FIGURE 4 | Relative spectral power distributions (SPDs) of the illumination in the three experiment sessions.

to compare the differences between two different types of light sources.

Figure 4 shows the average spectral power distribution of the illumination of the three experiment sessions, as measured using a PhotoResearch PR-670 spectroradiometer through the period of the experiment. In particular, Exps. I and II were carried out in a daylight space under sunny days, with a horizontal illuminance between 570.8 and 980.5 lx, a correlated color temperature (CCT) between 5972 and 6433 K, and a color rendering index (CRI) between 95.0 and 99.5. Exp. III was carried out in a multichannel LED viewing booth with an illuminance of 640 lx, a CCT of 6407 K, and a CRI of 95.1. The viewing conditions were designed according to the recommendations in CY/T3 1999 [9].

2.3 | Evaluation of Color Difference

The evaluation of the color difference between a pair of color samples was performed with a dark gray background having a relative luminance Y_{10} of 5.04. The samples were placed next to each other without a gap between them. The viewing distance was around 25–30 cm, so that each sample had a field of view (FOV) greater than 10°.

Before the formal experiment, several pairs of color samples having different magnitudes of color difference were presented to the observers to let them have an idea about the range of the differences they would experience. The evaluations were made on a 3-point rating scale, with the following instructions provided to the observers. Such descriptions were developed based on the practical rules used for quality inspection of products.

- Rating 1—*No color difference*: The color appearance of the two samples is the same.
- Rating 2—*Acceptable color difference*: The color difference between the two samples is perceptible, but the difference is very small and can be considered acceptable if it appears on products.
- Rating 3—*Unacceptable color difference*: The color difference between the two samples is too large, and it cannot be accepted if it appears on products.

In data analyses, the judgments of Rating 1 and Rating 2 were classified as "acceptable color difference" and those of Rating 3 were classified as "unacceptable color difference."

2.4 | Observers

Table 2 summarizes the information of the observers in the three experiment sessions, with all the observers passing the Ishihara Color Vision Test. Each pair of samples was evaluated 51 times in each experiment session. In Exp. I, all the 17 observers made the evaluations three times; in Exp. II, 25 observers made the evaluations twice; in Exp. III, 7 observers made the evaluations three times and the other 15 observers made the evaluations twice.

3 | Results

3.1 | Observer Variations

The observer variations, in terms of inter- and intra-observer variations, were characterized based on the wrong decision (WD%) [3, 12, 19, 20]. The inter-observer variations were characterized by comparing the judgments made by each observer and by an average observer (i.e., the judgments made by more than 50% of the observers), with the different judgments classified as "wrong decisions." The intra-observer variations were characterized based on the repeated judgments made by each observer on the same pair of color samples, with the different judgments classified as "wrong decisions." Table 3 summarizes the intra- and inter-observer variations for the three experiment sessions, with Figure 5 showing the boxplots of each color center. In general, the observer variations were comparable to those in previous studies [19, 20]. In particular, the variations in Exp. II were smaller than those in Exp. I and Exp. III, which were due to the involvement of the experienced observers.

3.2 | Comparisons Among the Three Sessions

The results from the three experiment sessions were compared. As shown in Figure 6, the color difference values ΔE^*_{ab} of the 450 pairs of samples were linearly correlated in the three sessions, but the judgments made by the observers, as characterized using the probability $P\%$ of unacceptable color differences, were not linearly correlated. In particular, the $P\%$ values in Exp. II were the highest, suggesting the higher sensitivity of the experienced observers to the color differences.

Moreover, the comparisons among the three sessions can also be made based on the chromaticity ellipses for acceptable color differences. The judgments made by the observers were converted to the probability P of "unacceptable color differences" and then into a z-score using $Z = \left(P_i - \overline{P}\right) / S$, where P_i is the probability of "unacceptable color difference" for each of the 450 pairs of samples, $\overline{P}$ is the average of the 450 pairs, and S is the standard deviation. The perceived color difference ΔV was calculated as $\Delta V = Z + 3.1$, with all the ΔV values greater than 0 [21, 22]. An optimization was performed to minimize the *STRESS* value between the perceived color difference ΔV and the calculated color difference $\Delta E^{*2}_{ab} = b_{11}\,\Delta a^{*2} + 2b_{12}\,\Delta a^*\,\Delta b + b_{22}\,\Delta b^{*2} + b_{33}\,\Delta L^{*2}$ for each color center. The chromaticity ellipses for acceptable color differences were then plotted for the $P\%$ value of 50%, which is considered as the threshold of acceptable color differences. Figure 7a shows the derived ellipses using all the 450 pairs of samples ("chromaticity ellipses"); Figure 7b shows the ellipses only using the pairs that were judged as acceptable color difference by more than 50% of the observers ("acceptable ellipses"), with the parameters of the ellipses, such as the semi-major axis (A), semi-minor axis (B), orientation angle (θ), and ellipse area (S), summarized in Table 4.

It can be observed that all the ellipses had the major axes toward the origin of the a^*–b^* plane, suggesting that the observers generally had a greater tolerance for chroma difference. The chromaticity ellipses were similar for the three experiment sessions. The acceptable ellipses for Exp. II, however, were different from those for the other two sessions, suggesting that the experienced observers were more sensitive to hue differences.

4 | Discussion

4.1 | Performance of CIELAB and CIEDE2000

Both CIELAB and CIEDE2000 are widely used in the surface color industry to characterize the color differences. They can be expressed in the format of $\Delta E = a \times \left(\sqrt{\left(\frac{\Delta L^*}{k_L}\right)^2 + \left(\frac{\Delta C^*_{ab}}{k_C}\right)^2 + \left(\frac{\Delta H^*_{ab}}{k_H}\right)^2 + \Delta R} \right)^b$, where ΔL^*, ΔC^*_{ab}, and ΔH^*_{ab} are the lightness, chroma, and hue differences; k_L, k_C, and k_H are the lightness, chroma, and hue parametric factors; and ΔR is the rotation term to characterize the chroma-hue interaction in CIEDE2000. For better characterizing the contributions of the lightness, chroma, and hue differences, Nobbs used a splitting method and revised the format to

TABLE 2 | The information of the observers in the three experiments.

Experiment session	Number of observers	Age range	Average age	SD of age
I (students at BIGC)	17 (12 females and 5 males)	22–27	23.4	1.5
II (quality inspection employees)	26 (24 females and 2 males)	23–41	29.9	5.3
III (students at BIGC)	22 (12 females and 10 males)	19–27	21.0	2.4

TABLE 3 | Summary of the intra- and inter-observer variations in the three experiment sessions.

	Exp. I		Exp. II		Exp. III	
WD%	Intra-variation	Inter-variation	Intra-variation	Inter-variation	Intra-variation	Inter-variation
Mean	23.1	21.6	16.5	15.7	19.9	18.0
Max	32.3	41.6	30.7	34.4	27.3	32.4
Min	16.1	15.9	7.6	8.3	13.6	12.4

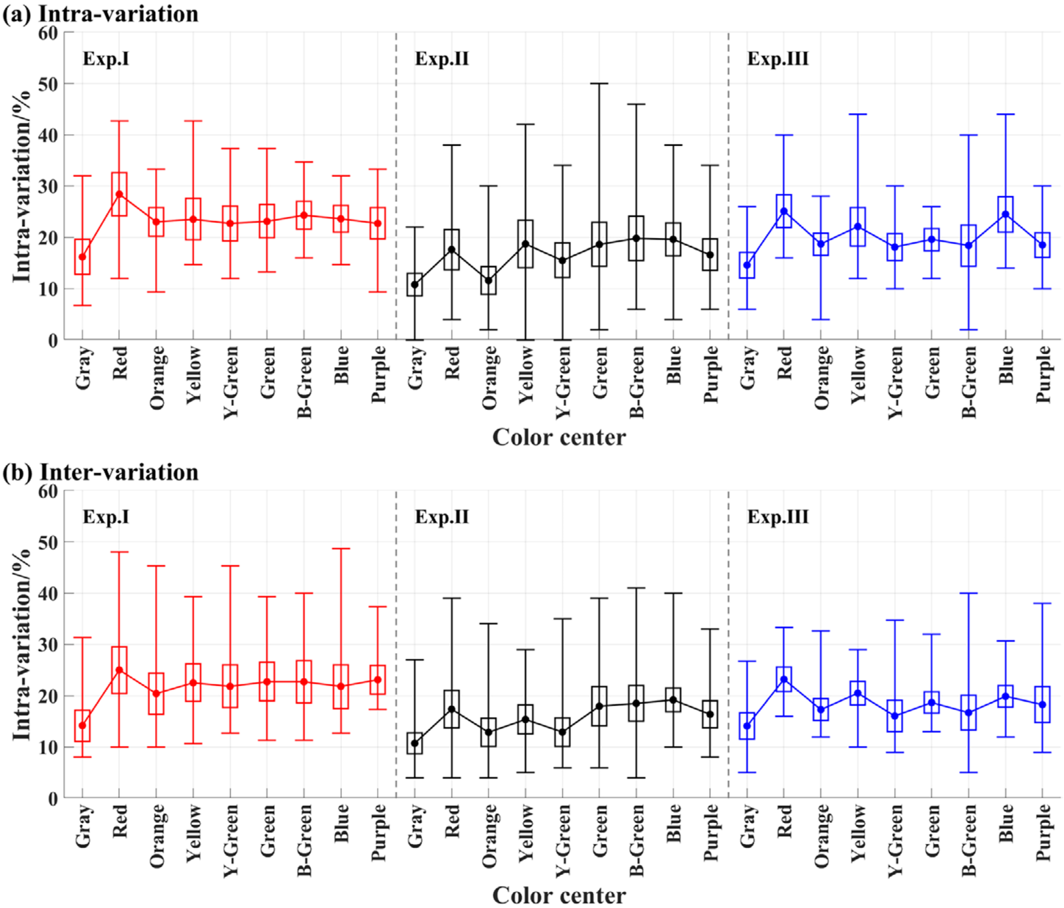

FIGURE 5 | Boxplots of the intra- and inter-observer variations for each color center in the three experiment sessions. (a) Intra-observer variations. (b) Inter-observer variations.

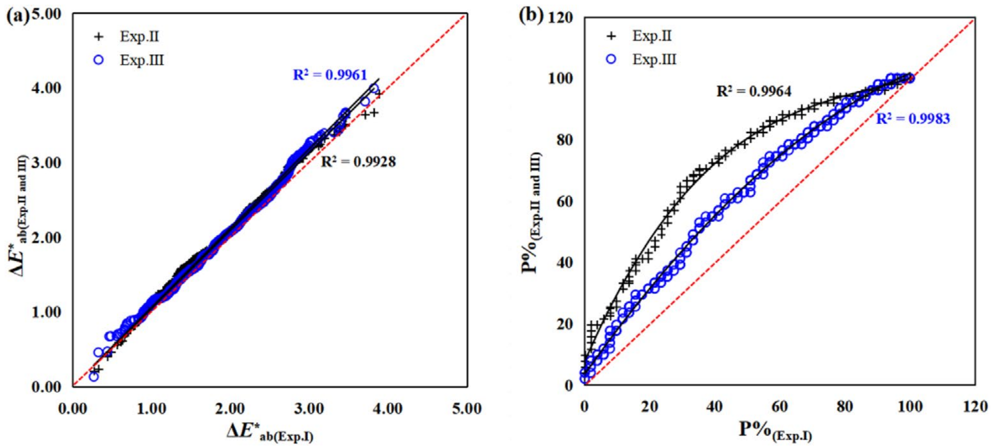

FIGURE 6 | Correlation of the measured color differences and perceived color differences for the three experiment sessions. (a) Measured color differences as characterized using ΔE^*_{ab} in CIELAB. (b) Perceived color differences.

$\Delta E_{00} = a \times \left(\sqrt{ \left(\frac{\Delta L_{00}}{k_L} \right)^2 + \left(\frac{\Delta C_{00}}{k_C} \right)^2 + \left(\frac{\Delta H_{00}}{k_H} \right)^2 } \right)^b$, with the values of k_L, k_C, k_H, a, and b recommended to be set to 1 under the reference conditions.

Here, we used the standardized residual sum of squares (*STRESS*) and F values to evaluate the performance of CIELAB and CIEDE2000 in characterizing the perceived color differences, with a smaller value suggesting a better performance. In addition, optimizations were also performed on k_L, k_C, and k_H individually, with the value of a simultaneously optimized for scaling purposes, to minimize the *STRESS* and F values (*note:* the *STRESS* value does not change with the value of a). Table 5 summarizes the results of the *STRESS* and F values, and the corresponding parameters.

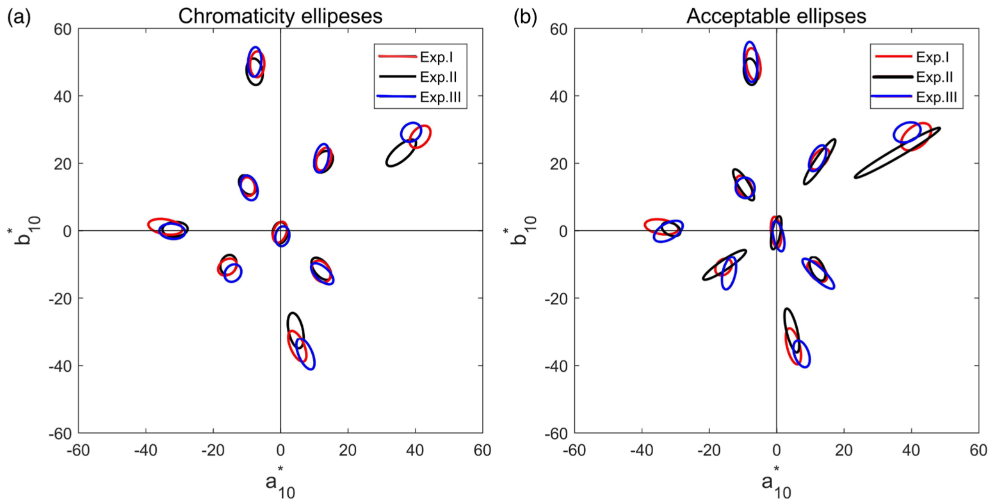

FIGURE 7 | Threshold of acceptable color differences for the three experiment sessions. (a) Chromaticity ellipses derived based on all the judgments. (b) Acceptable ellipses derived based on the judgments of acceptable color differences.

TABLE 4 | Parameters of the ellipses for the different color centers in the three experiment sessions.

		Chromaticity ellipses				Acceptable ellipses			
Exps.	**Colors**	***A***	***A/B***	***θ***	***S***	***A***	***A/B***	***θ***	***S***
I	Gray	1.07	1.67	71.71	2.14	1.53	3.12	95.79	2.37
	Red	1.27	1.57	49.23	3.23	1.72	1.67	39.38	5.55
	Orange	1.18	1.74	68.61	2.53	1.22	1.85	50.58	2.54
	Yellow	1.28	1.80	88.22	2.84	1.61	2.37	98.66	3.45
	Y-green	0.93	1.33	101.90	2.04	1.17	1.70	117.88	2.54
	Green	1.71	2.31	170.51	3.97	1.66	2.31	171.26	3.74
	B-green	1.00	1.35	35.22	2.32	0.93	1.33	46.95	2.06
	Blue	1.63	2.40	113.52	3.48	1.83	2.90	103.98	3.60
	Purple	1.15	1.39	121.72	3.01	1.22	1.72	132.02	2.72
II	Gray	1.06	1.43	83.64	2.45	1.66	3.95	78.93	2.20
	Red	1.83	2.58	41.10	4.10	4.86	11.30	30.21	6.60
	Orange	1.16	1.45	57.25	2.93	2.67	5.68	55.78	3.94
	Yellow	1.32	1.67	98.23	3.26	1.31	1.96	96.32	2.77
	Y-green	1.05	1.44	123.26	2.42	1.78	3.49	121.97	2.85
	Green	1.25	1.74	2.85	2.84	0.98	1.63	155.22	1.86
	B-green	1.00	1.28	73.08	2.44	2.53	5.27	33.95	3.83
	Blue	1.78	2.51	102.16	3.95	2.23	4.21	103.44	3.72
	Purple	1.29	1.84	128.85	2.82	1.27	2.05	115.83	2.48
III	Gray	0.99	1.43	79.35	2.14	1.52	2.98	102.30	2.44
	Red	1.05	1.22	37.92	2.85	1.37	1.49	21.89	3.96
	Orange	1.43	2.13	77.61	3.02	1.39	2.11	63.99	2.90
	Yellow	1.46	2.21	86.08	3.01	1.99	3.02	93.63	4.11
	Y-green	1.26	1.59	108.54	3.14	1.02	1.09	103.11	3.00
	Green	1.28	1.71	174.42	3.01	1.57	1.96	29.70	3.94
	B-green	0.93	1.21	50.94	2.23	1.65	2.62	77.73	3.24
	Blue	1.65	2.43	113.01	3.51	1.38	1.97	109.89	3.05
	Purple	1.37	2.08	137.82	2.83	2.07	3.57	136.82	3.80

TABLE 5 | Summary of CIELAB and CIEDE2000 in characterizing the perceived color difference, with the optimized values of a, b, k_L, k_C, and k_H listed. The critical value of F is 0.83, with the confidence level of 97.5%, which is labeled with *.

Experiment session	Evaluations		a	b	k_L	k_C	k_H	*STRESS*	F
(a) CIELAB									
Exp. I	Reference	—	1	1	1	1	1	34.0	1.00
	Optimization	*A*	1.13	1	1.60	1	1	31.8	0.87
		B	1.05	1	1	1.16	1	33.8	0.99
		C	0.79	1	1	1	0.57	29.3	0.74*
Exp. II	Reference	—	1	1	1	1	1	33.1	1.00
	Optimization	*A*	1.09	1	1.36	1	1	32.2	0.94
		B	1.09	1	1	1.29	1	32.4	0.95
		C	0.80	1	1	1	0.57	28.7	0.75*
Exp. III	Reference	—	1	1	1	1	1	35.9	1.00
	Optimization	*A*	1.15	1	1.60	1	1	33.5	0.87
		B	1.04	1	1	1.12	1	35.8	0.99
		C	0.80	1	1	1	0.56	31.1	0.75*
(b) CIEDE2000									
Exp. I	Reference	—	1	1	1	1	1	33.6	1.00
	Optimization	*A*	1.27	1	1.49	1	1	24.9	0.55*
		B	0.90	1	1	0.81	1	32.0	0.91
		C	0.77	1	1	1	0.76	28.8	0.73*
Exp. II	Reference	—	1	1	1	1	1	32.7	1.00
	Optimization	*A*	1.25	1	1.42	1	1	26.5	0.66*
		B	0.91	1	1	0.83	1	31.4	0.92
		C	0.80	1	1	1	0.78	29.4	0.81*
Exp. III	Reference	—	1	1	1	1	1	37.7	1.00
	Optimization	*A*	1.34	1	1.57	1	1	26.5	0.49*
		B	0.88	1	1	0.77	1	35.6	0.89
		C	0.73	1	1	1	0.71	30.7	0.66*

It can be found that both CIELAB and CIEDE2000 with the parameters as the reference condition had similar performance for the three experiment sessions. More importantly, the performance of the two formulas can always be improved by optimizing the value of k_H to be smaller than 1, but not by optimizing k_C and k_L, which was never performed in past studies [3, 23–26]. This clearly suggested that in CIELAB formula, the observers were more sensitive to the differences in the hue dimension and a greater weighting factor should be placed to the hue difference. Moreover, in CIEDE2000 formula, the observers were less sensitive to the differences in lightness and more sensitive to the differences in hue, and the results can be significantly improved through an optimization.

4.2 | Improving the Consistency Between Calculated and Perceived Color Differences

In order to develop a method for industries to effectively evaluate the color difference of printed color samples, the thresholds of the color differences that were calculated using the different methods listed in Table 6 were analyzed. The consistency was defined as the calculated color difference beyond the ΔE_T (or ΔH_T), with the $P\%$ larger than 50%, or the calculated color difference less than the ΔE_T, with the $P\%$ smaller than 50%. The optimizations were performed by varying the threshold values of the color differences for maximizing the percentages of the consistency (Cons.%), with the results listed in Tables 6 and 7. It can be found that the results in Tables 5 and 6 were similar, with the optimizations made on hue (for CIELAB) or lightness (for CIEDE2000) leading to the best performance.

The comparisons among the original methods and the various optimized methods were made for color difference, as shown in Figure 8, and hue difference, as shown in Figure 9. It can be observed that the distributions of the color difference of Exp. II and Exp. III were similar, which were different from that of Exp. I. This clearly suggested the difference between the experienced and inexperienced observers.

Moreover, the results of the 1350 pairs of samples included in the three experiments were used to find the upper and lower limits of the color difference and hue difference thresholds using a Z-test method, as expressed in $X = \mu - Z\sigma / \sqrt{n}$, where μ is the maximum consistency result, σ is the standard deviation of all the consistency results, n is the number of consistency results, and Z is the critical value with the significance level of 0.05. In the calculation, the values of ΔE^*_{ab}, ΔE_{00}, and ΔH values ranged between 0.3 and 2.5, 0.3 and 2.0, and 0.3 and 1.0 respectively, with an interval of 0.01, and the corresponding n was 221, 171, and 71. The upper and lower limits can then be calculated, as

TABLE 6 | Color difference thresholds of the original CIELAB, CIEDE2000 formulas, and the different optimized methods, together with the corresponding maximum consistency (%).

	ΔE^*_{ab}				ΔE_{00}			
Exp.	**Original**	***A***	***B***	***C***	**Original**	***A***	***B***	***C***
(a) Color difference threshold								
I	1.46	1.43	1.35	1.33	1.12	1.29	1.08	1.07
II	0.82	1.10	0.76	1.02	0.76	0.86	0.77	0.68
III	0.89	1.14	0.83	1.34	0.72	0.83	0.77	0.87
Combined	1.09	1.14	0.94	1.33	0.76	0.96	0.91	0.95
(b) Maximum consistency (%)								
I	70.9	74.2	70.7	75.1	72.4	77.1	71.8	76.2
II	82.2	82.9	82.2	84.0	82.9	84.0	82.0	83.6
III	75.3	77.3	75.1	77.3	75.8	80.0	76.4	75.6
Combined	74.8	77.0	74.7	78.7	75.1	79.3	75.6	77.0

TABLE 7 | Performances of Hue differences ΔH^*_{ab} and ΔH_{00} in terms of the maximum consistency (%).

Exp.	**Threshold**	**Cons. (%)**	**Threshold**	**Cons. (%)**
I	$(\Delta H^*_{ab})_T = 0.65$	76.4	$(\Delta H_{00})_T = 0.45$	77.3
II	$(\Delta H^*_{ab})_T = 0.53$	82.4	$(\Delta H_{00})_T = 0.34$	82.9
III	$(\Delta H^*_{ab})_T = 0.55$	74.2	$(\Delta H_{00})_T = 0.39$	76.0
Combined	$(\Delta H^*_{ab})_T = 0.37$	78.1	$(\Delta H_{00})_T = 0.42$	77.6%

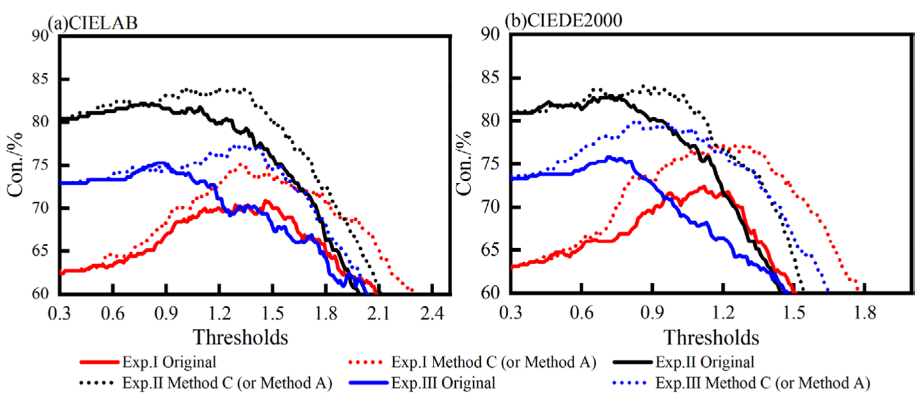

FIGURE 8 | Relationship between the percentage of consistency versus the threshold of color difference calculated using different methods for the three experiment sessions. (a) CIELAB color difference. (b) CIEDE2000 color difference.

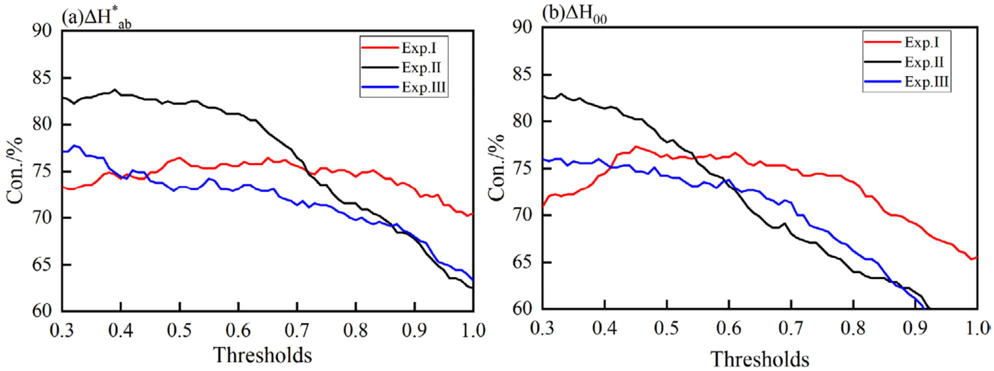

FIGURE 9 | Relationship between the percentage of consistency versus the threshold of hue difference calculated using different methods for the three experiment sessions. (a) CIELAB hue difference. (b) CIEDE2000 hue difference.

TABLE 8 | Summary of the upper and lower limits of the thresholds for the different methods calculated using a *Z*-test method.

	ΔE^*_{ab}		ΔE_{00}			
Methods	***A***	***C***	***A***	***B***	ΔH^*_{ab}	ΔH_{00}
μ	74.8	78.7	75.1	79.4	78.1	77.6
n	221	221	171	171	71	71
σ	9.12	7.96	11.40	10.24	3.78	5.65
X	73.6	77.7	73.4	77.7	77.3	76.3
Lower	0.75	1.24	0.56	0.77	0.30	0.30
Upper	1.20	1.39	1.00	1.15	0.50	0.49

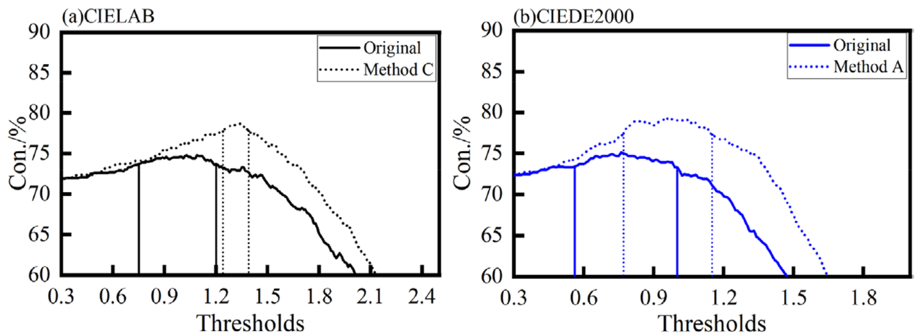

FIGURE 10 | Relationship between the percentage of consistency versus threshold of color difference calculated using different methods for the three experiment sessions. (a) CIELAB color difference. (b) CIEDE2000 color difference. *Note:* The vertical lines show the lower and upper limits.

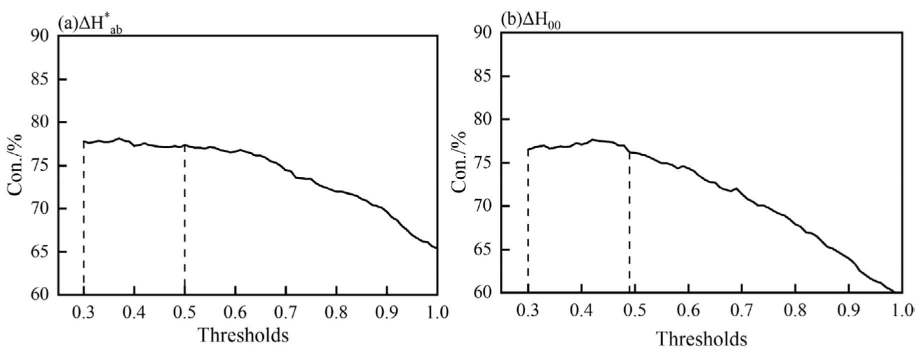

FIGURE 11 | Relationship between the percentage of consistency versus threshold of hue difference calculated using different methods for the three experiment sessions. (a) CIELAB hue difference. (b) CIEDE2000 hue difference. *Note:* The vertical lines show the lower and upper limits.

summarized in Table 8, with the changes of percentage of consistency with color and hue differences shown in Figures 10 and 11.

5 | Conclusions

A total of 1350 pairs of printed color samples were prepared for color difference evaluations, whose chromaticities were around the nine color centers recommended by the CIE and had color differences ranging between 0.13 and 4.33 (mean = 1.98) CIELAB units. Three experiments were carried out to investigate how experienced versus inexperienced observers and how different light sources affected the color difference evaluations.

It was found that the experienced observers were more sensitive to the color differences, especially to the hue differences for the color centers of red, orange, blue–green, and blue. Optimizations were performed on the CIELAB and CIEDE2000 color difference formulas, with an optimized value of k_H in the CIELAB formula and an optimized value of k_L in the CIEDE2000 formula introducing the most significant improvements to predict the perceived color differences. For practical applications, thresholds of color difference and hue difference in the CIELAB and CIEDE2000 formulas were also proposed to classify whether the color difference between two printed samples is acceptable or not.

Author Contributions

Min Huang: conceptualization and data analysis. **Xiaoyu Shang:** data collection, data analysis, and draft preparation. **Xuping Gong:** conceptualization. **Minchen Wei:** conceptualization and manuscript preparation. **Dan Wang:** data collection. **Yu Liu:** data collection and analysis. **Xiu Li:** data collection.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. P. F. Guo, D. Min, W. Lv, et al., "Color Measurement and Chromatic Aberration Evaluation of Photobar Holographic Prints," *Digital Printing* (in Chinese) 1 (2021): 22–28.

2. B. Zhao, Q. Xu, and M. Luo, "Color Difference Evaluation for Wide-Color-Gamut Displays," *Journal of the Optical Society of America A* 3 (2020): 1257–1265.

3. M. Huang, Y. Xi, J. Pan, Y. Li, Y. Wang, and X. Li, "Optimization of the Method for Color Measurement and Color-Difference Calculation of Holographic Prints With Light Pillars," *Applied Optics* 60 (2021): 6989–6999.

4. I. Cabral, A. Schuch, and F. Steffens, "Color Biomimetics in Textile Design: Reproduction of Natural Plant Colors Through Instrumental Colorant Formulation," *Journal of Imaging* 10, no. 7 (2024): 150.

5. M. R. Luo, Q. Xu, M. Pointer, et al., "A Comprehensive Test of Colour-Difference Formulae and Uniform Colour Spaces Using Available Visual Datasets," *Color Research & Application* 48, no. 3 (2023): 267–282.

6. M. Melgosa, G. Cui, C. Oleari, et al., "Is It Possible to Improve the Weighting Function for Lightness in the CIEDE2000 Color-Difference Formula?," *Cultura E Scienza Del Colore-Color Culture and Science* 10, no. 10 (2018): 59–65, https://doi.org/10.23738/ccsj.i102018.07.

7. X. P. Gong, M. Huang, X. Y. Shang, et al., "Improving the Color Quality Evaluation of Holographic Prints With Light Pillars," *Optics Express* 32, no. 20 (2024): 35095–35109, https://doi.org/10.1364/OE.531522.

8. CIE, *Parametric Effects in Colour-Difference Evaluation*, vol. 101 (CIE, 1993).

9. CY Publication, "Illumination and Observation Conditions of Color Evaluation," CY/T3-1999, 1999 (in Chinese).

10. Q. Xu, G. Cui, M. Safdar, L. Xu, M. R. Luo, and B. Zhao, "Assessing Colour Differences Under a Wide Range of Luminance Levels Using Surface and Display Colours," *Color and Imaging Conference* 27 (2019): 355–359.

11. Y. L. Zheng, S. S. Zhou, Y. H. Qi, et al., "Performance of CIE LAB-Based Color Difference Formulae Under Different Viewing Conditions," *Advanced Materials Research* 457 (2012): 1294–1298.

12. X. Z. Lin, Z. W. Wei, Z. M. Liao, et al., "Evaluation of Visual Color Difference and Color Difference Threshold Based on Cigarette Holographic Prints," *Digital Printing* (in Chinese) 6 (2022): 45–52.

13. S. K. Kulappurath and R. Shamey, "The Effect of Luminance on the Perception of Small Color Differences," *Color Research & Application* 46, no. 5 (2021): 929–942, https://doi.org/10.1002/col.22637.

14. X. Wang, J. Zhang, Y. Jiang, J. Du, D. Miao, and C. Xu, "Color Difference of Yarn-Dyed Fabrics Woven From Warp and Weft Yarns in Different Color Depths," *Pigment & Resin Technology* 53, no. 1 (2024): 28–35.

15. L. D. McHutchion, L. McHutchion, L. Zhao, S. Dixon, and R. Seelaus, "Thresholds of Perceivable Color Difference in Silicone: A Pilot Study," *International Journal of Maxillofacial Prosthetics* 3, no. 1 (2020): 12–22, https://doi.org/10.26629/ijmp.2020.04.

16. M. P. María, G. Razvan, J. H. Luis, et al., "Color Difference Thresholds for Computer-Simulated Human Gingiva," *Journal of Esthetic and Restorative Dentistry* 30 (2018): E24–E30.

17. N. Na, K. Choi, J. Lee, and H. J. Suk, "Color Tolerance Study on White in Practical Aspect: Perceptibility Versus Acceptability," *Color Research & Application* 39, no. 6 (2014): 582–588.

18. K. Witt, "CIE Guidelines for Coordinated Future Work on Industrial Colour-Difference Evaluation," *Color Research and Application* 20 (1995): 399–403.

19. M. R. Luo, C. Minchew, P. Kenyon, et al., "Verification of CIEDE2000 Using Industrial Data//AIC 2004 Interim Meeting," in *Proceedings of Interim Meeting of the International Color Association* (International Color Association, 2004), 97–102.

20. H. Yang, M. Qi, G. Zhang, et al., "Investigation on the Equi-Depth of Color Depth Formulas Based on Color Discrimination Threshold and Acceptable Tolerance," *Color Research and Application* 48, no. 4 (2023): 736–751.

21. E. Coates, S. Day, J. R. Provost, et al., "The Measurement and Assessment of Colour Differences for Industrial Use—The Accuracy of Colour-Difference Equations," *Journal of the Society of Dyers and Colourists* 88 (1972): 69–75.

22. E. Coates, S. Day, J. R. Provost, et al., "The Measurement and Assessment of Colour Differences for Industrial Use III. Method of Scaling Visual Assessments," *Journal of the Society of Dyers and Colourists* 88 (1972): 186–190.

23. R. He, K. Xiao, M. Pointer, M. Melgosa, and Y. Bressler, "Optimizing Parametric Factors in CIELAB and CIEDE2000 Color-Difference Formulas for 3D Printed Spherical Objects," *Materials* 15 (2022): 4055.

24. Q. Xu, K. Shi, and M. R. Luo, "Parametric Effects in Color-Difference Evaluation," *Optics Express* 30 (2022): 33302–33319.

25. M. Melgosa, G. Cui, C. Oleari, et al., "Revisiting the Weighting Function for Lightness in the CIEDE2000 Colour-Difference Formula," *Coloration Technology* 133, no. 4 (2017): 273–282, https://doi.org/10.1111/cote.12294.

26. H. Wang, G. Cui, M. R. Luo, and H. Xu, "Evaluation of Colour-Difference Formulae for Different Colour-Difference Magnitudes," *Color Research and Application* 37 (2012): 316–325.