Research Article | Vol. 29, No. 15 / 19 July 2021 / Optics Express 24237
Optics EXPRESS

Color-difference evaluation for 3D objects

LAN JIANG,[1,2] GUIHUA CUI,[1,*] MANUEL MELGOSA,[3] KAIDA XIAO,[4] AND SUCHITRA SUEEPRASAN[5]

[1] College of Electrical & Electronic Engineering, Wenzhou University, Wenzhou, China
[2] Yichang Three Gorges Secondary Vocational School, Yichang, China
[3] Department of Optics, University of Granada, 18071 Granada, Spain
[4] School of Design, University of Leeds, Leeds, United Kingdom
[5] Department of Imaging and Printing Technology, Chulalongkorn University, Bangkok, Thailand
[*] guihua.cui@foxmail.com

Abstract: A psychophysical experiment using 3D printed samples was conducted to investigate the change of perceived color differences caused by two different illuminations and two 3D sample shapes. 150 pairs of 3D printed samples around five CIE color centers [Color Res. Appl. 20, 399–403, 1995], consisting of 75 pairs of spherical samples and 75 pairs of flat samples, with a wide range of color differences covering from small to large magnitude, were printed by an Mcor Iris paper-based 3D color printer. Each pair was assessed twice by a panel of 10 observers using a gray-scale psychophysical method in a spectral tunable LED viewing cabinet with two types of light sources: diffuse lighting with and without an additional overhead spotlight. The experimental results confirmed that the lighting conditions had more effect on the perceived color difference between complex 3D shapes than between 2D objects. The results for 3D and 2D objects were more similar under only diffuse lighting. Current 3D results had good correlations with previous ones [Color Res. Appl. 24, 356-368, 1999; J. Opt. Soc. Am. A 36, 789-799, 2019] using 2D samples with large color differences, meaning that color-difference magnitude had more effect on perceived color differences than sample shape and lighting. Considering ten modern color-difference formulas, the best predictions of the current experimental data were found for CAM02-LCD formula [Color Res. Appl. 31, 320-330, 2006]. For current results, it was also found that predictions of current color-difference formulas were below average inter-observer variability, and remarkable improvements were found by adding power corrections [Opt. Express 23, 597-610, 2015].

1. Introduction

3D color printing technology, also known as additive manufacturing technology, is a revolutionary process that has developed during the last decade to produce full-spectrum solid objects utilizing a range of printing materials. With the evolution of various 3D imaging techniques, accurate acquisition and transformation of target object geometrical data into 3D digital models can be achieved [1,2]. By combining 3D image capture and 3D printing techniques, accurate 3D reproduction is possible. Moreover, this technology has the ability to directly interconnect with advanced manufacturing techniques, allowing customization with excellent accuracy, resulting in savings of both time and costs. The process has been extensively utilized in rapid prototyping, successfully applied in medical sciences [3,4], and is gaining popularity in many other multi-disciplinary applications. A well acknowledged technical bottleneck is that 3D color printing has developed without considering the requirements of 3D design. One important reason for this is the inability to measure the absolute color appearance, and the relative color difference between 3D objects generated in the design process and those produced by the manufacturing system. Faithful reproduction of the color appearance of 3D objects is an essential requirement that requires knowledge of 3D metrology, color modelling, human perception and 3D printing.

#432729 https://doi.org/10.1364/OE.432729
Journal © 2021 Received 28 May 2021; revised 7 Jul 2021; accepted 9 Jul 2021; published 15 Jul 2021

Optics EXPRESS

Therefore, collaboration among the 3D printing industry, computer graphics specialists, and color and vision scientists is mandatory.

Color image reproduction techniques based on conventional CIE (Commission Internationale de l'Éclairage) colorimetry [5] have been available for more than 25 years and they perform very well to transform color images from one digital medium to another under various viewing conditions. However, it must be noted that the CIE standard observer and psychophysical data for both color appearance modelling and color-difference evaluation were always developed from flat 2D color samples. The color measurement and color-difference prediction of 3D objects will require more advanced measurement and analysis techniques. Conventional color measurement instruments cannot accurately measure the color of 3D objects because of their 3D shape and non-uniform color appearance. In terms of color-difference evaluation, when a 3D object is viewed by one observer, the viewing distance between his/her eyes to different locations on the 3D object varies. To effectively measure the color of 3D objects and to quantify their color appearance difference is of great importance in several fields of science, medicine and technology. For 3D color printing technology, satisfactory color reproduction is always a critical requirement. Whether the color quality of 3D printed products meets the customers' expectation is of great importance; For the 3D design industry, the color appearance of a 3D printed object largely affects the perception of the overall product design. For all those applications, the color appearance difference between the customer's expectation and the manufactured product needs to be adequately assessed in an objective manner.

The quantitative description of color differences is a prerequisite for color management and accurate color control [5]. However, color-difference evaluation is affected by many factors or viewing conditions, including sample shape, sample size, sample separation, texture, background, color-difference magnitude, etc., generally designated as parametric effects [6]. As mentioned before, it has not been compared the influence of viewing conditions for 3D and 2D objects. Moreover, the translucency, gloss, surface texture and spatial information of 3D objects may affect overall color differences significantly.

CIE has periodically recommended guidelines for color-difference evaluation and requested new reliable experimental datasets [7–9]. The CIELAB [5] color-difference formula was recommended in 1976, CIE94 [10] and CIEDE2000 [11] were developed on the basis of CIELAB color space and recommended in 1995 and 2000, respectively. Currently, CIEDE2000 is a joint CIE/ISO standard color-difference formula [12] for industrial color-difference evaluation. CMC [13], CAM02-UCS [14], CAM16-UCS [15] were series of color-difference formulas or uniform color spaces established on the basis of different visual data obtained under specific viewing conditions. Models for industrial color-difference evaluation are limited to be used under reference conditions. If the viewing conditions change, e.g., the viewing mode changes from 2D to 3D objects, the sensitivity of the human eyes to color differences may change significantly, and correction factors are sometimes used to account for parametric effects. Earlier color-difference studies were only based upon flat (2D) samples, and the influence and mechanisms of color differences using 3D-shaped objects have not been addressed. Compared with 2D objects, the color perception of 3D objects is more complicated and may be affected by more factors, including 3D shape, shadows, lighting conditions [16,17], highlights [18], gloss [19], etc. There is no evidence that the color-difference formulas developed for 2D objects are suitable for 3D objects. The goal of the current study is to investigate the change of perceived color differences caused by two different illuminations and two 3D sample shapes.

2. Experimental method

In this study, 150 pairs of 3D samples were printed by an Mcor Iris paper-based 3D color printer. Specifically, we printed 75 pairs of spherical samples and 75 pairs of flat samples around five CIE color centers recommended in 1978, including gray, red, yellow, green, and blue colors [7,8].

Optics EXPRESS

The spherical samples had 4 *cm* diameter and the flat samples were 4 *cm* in square. Note that the flat samples here were printed in 3D style, i.e., a slice of sample with 0.5 *cm* thickness. A panel of 10 observers performed visual color-difference evaluations on the 3D printed sample pairs using a gray-scale method. Two types of lighting were used in the current experiment: diffuse lighting, and spotlight together with diffuse lighting.

All the 3D samples were measured by an X-Rite Ci64UV spectrophotometer in d:8 measuring geometry, small aperture (4 *mm* diameter), specular component included (SCI) mode, and the corresponding tristimulus values were computed for the CIE 1964 standard observer under the two real light sources employed in our visual experiments. The small measuring aperture was preferred for spectrophotometric measurements because curved 3D surfaces may affect the results if the measuring aperture is too large.

The spectrophotometric measurements for each sample were made in five different positions on the surface, and the average reflectance was recorded. These five measurements for each sample were also used to check the color homogeneity of the surface of 3D printed samples. The *MCDM* (mean color difference from the mean) shown in Eq. (1) was used to test the color homogeneity or uniformity.

$$MCDM = \frac{\sum_{i=1}^{N} \sqrt{\left(L_i^* - \overline{L^*}\right)^2 + \left(a_i^* - \overline{a^*}\right)^2 + \left(b_i^* - \overline{b^*}\right)^2}}{N}, \tag{1}$$

where N is the number of measured positions on the surface of each sample (here N=5), L^*_i, a^*_i, b^*_i are CIELAB coordinates for ith measurement, and $\bar{L}^*$, $\bar{a}^*$, $\bar{b}^*$ are the average CIELAB coordinates from N measurements. The *MCDM* was first computed for each printed sample, and then the average *MCDM* for samples in each color center and for all samples were also calculated (see Table 1), assuming D65 illuminant and CIE 1964 standard observer.

From Table 1, the mean *MCDM* values for spherical and flat samples were 0.92 and 0.71 CIELAB units, respectively, which means that the spherical samples are less uniform than the flat samples. Using a paper-based 3D printer, it is hard to produce very uniform colored samples. A psychophysical experiment was conducted using a gray-scale method [20–23]. The gray scale used in our current experiment was manufactured by the SDC Enterprises Limited, and has been widely used in color fastness testing for assessing staining in the textile industry.

Table 1. Color homogeneity for 3D printed samples tested in terms of *MCDM* (CIELAB units), assuming D65 illuminant and CIE 1964 standard observer.

Color centers	Spherical	Flat
Gray	0.84	0.63
Red	0.94	0.82
Yellow	0.97	0.69
Green	0.96	0.62
Blue	0.90	0.81
Mean *MCDM*	0.92	0.71

The visual assessments were first conducted inside a viewing cabinet equipped with a spectrally tunable LED lighting system, provided by Thouslite Inc., China, employed here as a D65 simulator. Next, we used the same viewing cabinet and D65 light source but adding a spotlight source (Lowel Pro-light lamp) with a CCT close to U35. The Lowel Pro-light lamp was erected with a special lamp holder and irradiated the samples overhead of the observers. The relative spectral power distributions (SPDs) for the above two lighting systems were measured at the position of the samples using a JETI spectroradiometer Specbos 1211 UV, and the results are

Optics EXPRESS

plotted in Fig. 1. The illuminances at the position of the samples for the two light sources were around 1207 lx and 6032 lx, and their correlated color temperatures (CCT) were 6697K and 3652K. Thus, the differences between our two lighting conditions include not only geometry of illumination, but also correlated color temperature and illuminance.

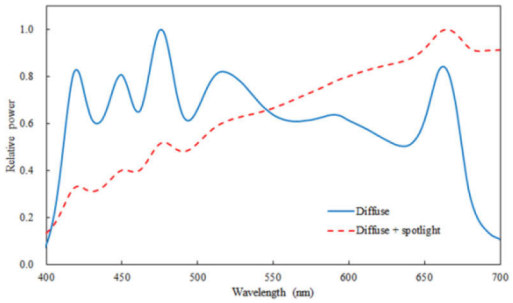

Fig. 1. Relative spectral power distributions (SPDs) of two light sources used in the experiment.

The experiment was carried out in a very quiet and dark room. Ten observers participated in the experiment. They were postgraduate students with normal color vision, and an average age of 24.5 years. All of them repeated the experiment twice to assess observer repeatability. They had little experience in color science and in scaling color differences. Before starting the experiment, the observers were trained for judging color differences using the gray-scale method by carrying out a pilot experiment. During the final visual experiments, the sample pair and the gray-scale samples (with an appropriate mask) were placed in the central part of the floor of the cabinet (see Fig. 2), with the left-right position of samples and sequence of sample pairs presented to the observers being random.

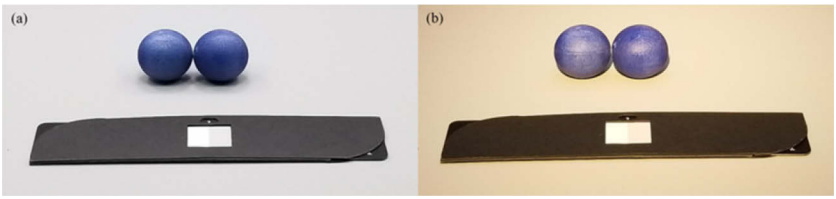

Fig. 2. Configuration of the sample pair and gray-scale samples inside the viewing cabinet under (a) diffuse, and (b) diffuse + spotlight light sources

Before the beginning of the visual experiments, the light sources were warmed up for at least 15 minutes. Each observer performed light adaptation for about 1 minute and then officially started the visual experiment. The gray-scale score (GS) given by each observer for each sample pair was collected [20–23]. Each observer was encouraged to provide the gray-scale scores in terms of grade using at least one decimal. For example, supposing that a sample pair has a visual color difference between Grade '3-5' and Grade '4' being closer to Grade '4', the reported result should be in the range 3.76-3.99.

The visual experiment was divided into four phases, named henceforth DS, DF, SS, and SF, according to the different viewing conditions. The first letter in these abbreviated phase names represents the light source used (with D standing for diffuse lighting, and S for spotlight plus diffuse lighting) and the second letter represents sample shape (with S standing for spherical, and F for flat). To ease the revelation of parametric effects, all possible parametric differences between different two-phase combinations in the current study are listed in Table 2. For example,

Optics EXPRESS

comparing phases DS and DF (first row of Table 2), two different parametric conditions exist, sample shape and color-difference magnitude. So the mixed effect of sample shape and color-difference magnitude on perceived color differences may play a role when we change from DS to DF. However, there is only one parametric difference (lighting) between phases DS and SS (second row of Table 2), or between DF and SF (fifth row of Table 2). Therefore, the isolated lighting effect only exists comparing DS and SS, or DF and SF.

Table 2. Parametric differences between each of two-phase combinations in the current study

1st Phase	2nd Phase	Parametric difference		
		Lighting	Shape	Range of Color-difference Magnitude
		Diffuse vs Spotlight	Spherical vs Flat	Medium-Large vs Small-Large
DS	DF	×	✓	✓
DS	SS	✓	×	×
DS	SF	✓	✓	✓
DF	SS	✓	✓	✓
DF	SF	✓	×	×
SS	SF	×	✓	✓

In the present study, many comparisons are made between two sets of data. For example, experimental visual differences (ΔV) will be compared with color differences (ΔE) computed using different color-difference formulas to test the performance of such formulas. Earlier studies showed that in these comparisons applying different statistical measures sometimes can lead to different conclusions [24,25]. Specifically, we will use the *STRESS* index [25] in Eq. (2), which has been widely used in previous color-difference studies. The *STRESS* index has been recommended by CIE [26] and is defined as follows:

$$STRESS = 100\sqrt{\left(\frac{\sum(\Delta E_i - f\Delta V_i)^2}{\sum \Delta E_i^2}\right)}, \tag{2}$$

where $f = \frac{\sum \Delta E_i \Delta V_i}{\sum \Delta V_i^2}$, and ΔE_i and ΔV_i can be considered as ith data from two compared datasets, e.g., computed color difference by a color-difference formula and its corresponding visual difference for a pair of samples. The smaller the *STRESS* is, the better the agreement between the two compared datasets.

3. Results and discussion

3.1. Conversion of gray-scale scores (GS) to visual color difference (ΔV)

In order to convert the gray-scale grade (GS) values into visual color difference (ΔV) values, a regression (curve fitting) technique was used to find the relationship between gray-scale grades and their corresponding CIELAB color differences (ΔE^*_{ab}), as shown in Fig. 3. The CIELAB color difference (ΔE^*_{ab}) for any grade is equivalent to the visual color difference (ΔV). Equation (3) and (4) allow transformations from GS to ΔV for the diffuse lighting, and the diffuse lighting plus spotlight, respectively.

$$\Delta V = -19.95\, ln(GS) + 31.581, \tag{3}$$

$$\Delta V = -20.20\, ln(GS) + 31.956. \tag{4}$$

The coefficients in Eq. (3) and (4) were obtained by fitting logarithmic equations to points indicating the gray-scale grades and their corresponding ΔE^*_{ab} values (Fig. 3). Each observer's

Optics EXPRESS

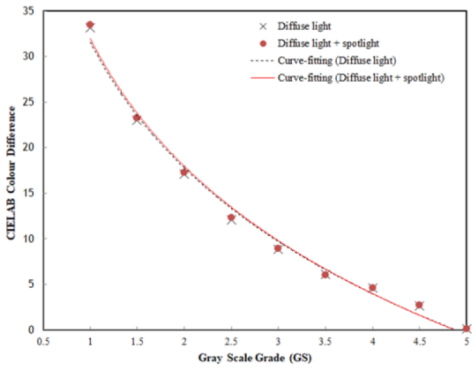

Fig. 3. The relationship between the gray scale grades (GS) and corresponding CIELAB color differences in our two lighting conditions (i.e. diffuse light, and diffuse light plus spotlight)

gray-scale scores were transformed using Eq. (3) or (4) depending on the lighting condition, and for each pair of samples the arithmetical mean of the ΔV values from all observers was calculated. These mean values for each of the four phases will be used as the experimental data to test various color-difference formulas and to reveal the influence of the different parametric effects involved in the current study.

3.2. Observer variability

The observer variability was evaluated in terms of intra- and inter-observer variations by means of the *STRESS* index. [27]

As mentioned earlier, each phase was assessed by a panel of ten observers, and each observer assessed each sample pair twice at different times. The *STRESS* index was calculated between each individual observer's two repeated sessions to represent one observer's intra-observer variability. Next, the mean values for the ten observers for each phase were calculated (average intra-observer variation or typical repeatability performance). The mean *STRESS* values for intra-observer variability in each phase are listed in Table 3, and ranged from 11.5 to 14.9 units.

Table 3. Intra-observer variability and inter-observer variability in terms of STRESS for four phases in current experiment

Phase	DS	DF	SS	SF
Intra-observer variability	14.9	12.6	11.5	12.2
Inter-observer variability	23.5	19.4	21.1	20.2

The *STRESS* index was also calculated between each individual observer's mean and the mean visual results of the ten observers (ΔV) to represent one observer's inter-observer variability. Next, the mean of ten observers' inter-observer variability for each phase was calculated to represent the group's inter-observer variability. The results found are also given in Table 3, and ranged from 19.4 to 23.5 units. From Table 3, in the four phases of our experiment the inter-observer variability was larger than the intra-observer variability. Comparing with other previous color-difference studies [20–23], the observers' variability listed in Table 3 are similar to those found for flat samples, and represent the typical performances in visual evaluations using the gray-scale method. In Table 3, the differences between *STRESS* values for different sample shapes under the same lighting condition (e. g., DS and DF, or SS and SF) were small. For example, the maximum difference of 4.1 was observed for inter-observer variability between

Optics EXPRESS

DS and DF. This means that there is no particular difficulty for the observers to judge color differences for spheres comparing with the difficulty to judge flat objects.

3.3. Chromaticity ellipses

Chromaticity ellipses and ellipsoids calculated in CIELAB color space were usually used for visualization of experimental data, and comparison between results from different studies. The next ellipsoid equation with 4 coefficients (Eq. (5)) was used to fit experimental results for each color center,

$$\Delta E_{ab}^{*\,2} = k_1\,\Delta L^{*2} + k_2\,\Delta a^{*2} + k_3\,\Delta b^{*2} + k_4\,\Delta a^{*}\,\Delta b^{*}, \tag{5}$$

where the terms with $\Delta a^*\Delta L^*$ and $\Delta b^*\Delta L^*$ were disregarded because it was previously reported that they have very small effect [28,29]. Thus, k_1 to k_4 were optimized to give the best fit for each color center (Table 4). Setting ΔL^* to zero allows the corresponding ellipse to be calculated in CIELAB a^*b^* plane. The ellipses for the four phases were plotted in Fig. 4. The sizes of the five ellipses plotted in each phase were adjusted by a single scaling factor to bring the average area of different phases on the same scale than those in phase DF. From Fig. 4 the worst agreement between the four ellipses in each of the five centers was found for the yellow and red centers. For example, we can see that for these two centers the orientations of the major axes of ellipses for spherical samples (DS and SS) are very different to those from flat samples (DF and SF).

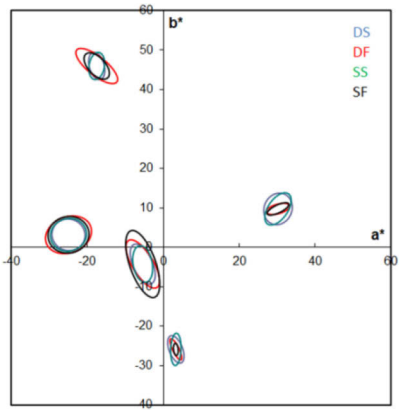

Fig. 4. The a^*b^* ellipses for the four phases of the current experiment in the 5 CIE-recommended centers [7,8]

3.4. Performance of color-difference formulas

3.4.1. Performance of original formulas

For each of the four phases, the average visual color differences (ΔV) were plotted against corresponding CIELAB color differences (ΔE^*_{ab}) in Fig. 5. Power functions were fitted to results in the four plots in Fig. 5 as well.

It can be seen from Fig. 5 that the range of CIELAB color differences (x-axis) in this study was very wide, including small, medium and large color differences, especially for the two phases involving flat samples. Specifically, for phases DS, DF, SS and SF, the average color differences of the sample pairs were 10.3, 6.7, 10.2, and 6.3 CIELAB units, respectively, with ranges 2.9-23.6, 0.3-24.3, 2.9-21.0, and 0.3-20.6 CIELAB units, respectively. It is worth mentioning that the lower ends of color-difference ranges and the mean values of color differences for the two phases involving flat samples, DF and SF, were smaller than those of two phases involving spherical samples, DS and SS. There were about 50% of pairs with CIELAB color differences lower than 5

Optics EXPRESS

Table 4. The ellipsoids' coefficients of the current study in CIELAB space (see Eq. (5))

Phase	Color Center	k_1	k_2	k_3	k_4
DS	Gray	1.447	0.515	0.592	3.192
	Red	0.568	-0.057	0.496	1.581
	Yellow	1.042	0.217	0.418	0.100
	Green	0.485	0.070	0.820	0.043
	Blue	2.603	0.804	1.004	0.100
DF	Gray	1.310	0.605	0.732	5.673
	Red	0.821	-1.815	7.343	8.800
	Yellow	0.877	0.719	1.332	2.812
	Green	1.589	0.395	0.601	0.883
	Blue	12.231	5.712	3.648	2.284
SS	Gray	1.227	0.176	0.385	1.863
	Red	0.846	-0.423	0.711	0.947
	Yellow	0.837	-0.094	0.322	0.419
	Green	0.846	0.119	0.284	0.590
	Blue	4.731	-0.139	0.516	0.100
SF	Gray	1.196	0.310	0.463	6.032
	Red	1.559	-2.526	6.202	5.717
	Yellow	0.857	0.506	0.855	2.905
	Green	2.588	0.306	0.643	1.950
	Blue	17.334	1.541	2.650	2.038

CIELAB units in phases DF (Fig. 5(b)) and SF (Fig. 5(d)), but only about 15% of pairs have CIELAB color differences lower than 5 CIELAB units in phases DS (Fig. 5(a)) and SS (Fig. 5(c)). Most pairs in two phases with spherical samples were in the category of large color-difference magnitude, while only half of pairs in two phases with flat samples were in such category. This may have an impact on the performance of color-difference formulas predicting the current data.

The scatter plots in Fig. 5 were usually employed for visually evaluating the performance of a color-difference formula. For instance, if a color-difference formula exactly predicts the average visual data from a particular group of observers, the points in the plots of Fig. 5 will fall on the 45° (broken) line. The faraway the points departing from this line, the worse the performance of the formula. In the current case, reasonable linear trends can be found from the distribution of data points for the two phases using spherical samples, DS and SS, as shown in Fig. 5(a) and Fig. 5(c), respectively. However, apparent nonlinear trends can be observed for two phases using flat samples, DF and SF, as shown in Fig. 5(b) and Fig. 5(d), respectively. This means that the CIELAB color-difference formula may perform better in predicting visual results from two phases involving spherical samples than results from two phases involving flat samples. Potential reasons for this result will be discussed later in detail. Similar scatter plots to those shown in Fig. 5 were also found for other color-difference formulas different from CIELAB.

In general, a Euclidean color-difference formula can be expressed in a generic form as given in Eq. (6), where ΔL, ΔC, and ΔH are the lightness, chroma and hue difference, respectively, and k_{L}, k_{C}, and k_{H} are the parametric factors for the lightness, chroma and hue difference, respectively,

$$\Delta E = \sqrt{\left(\frac{\Delta L}{k_L}\right)^2 + \left(\frac{\Delta C}{k_C}\right)^2 + \left(\frac{\Delta H}{k_H}\right)^2}. \tag{6}$$

Optics EXPRESS

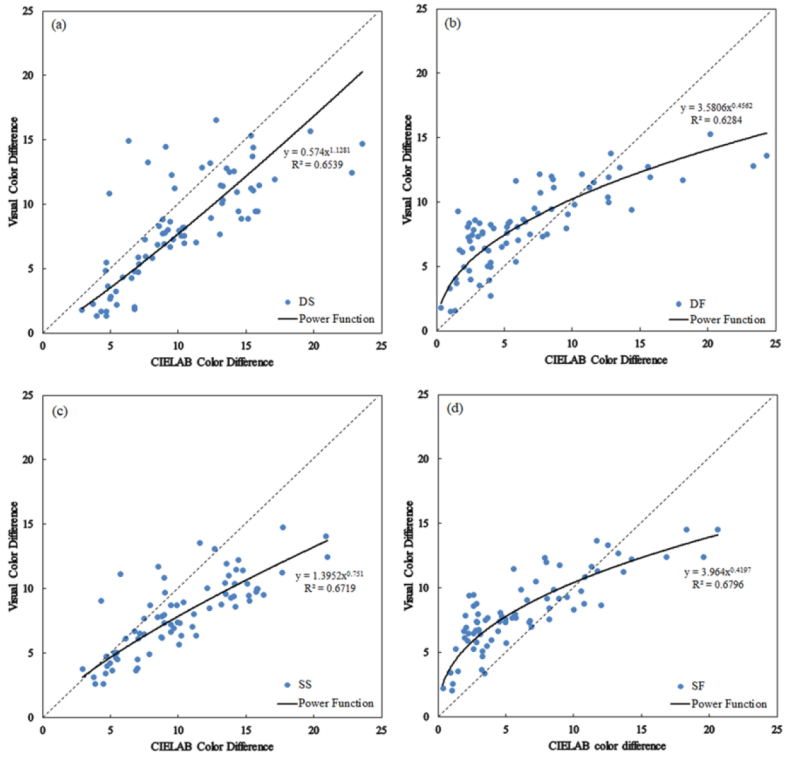

Fig. 5. The visual color differences (ΔV) plotted against corresponding CIELAB color differences for phase (a) DS, (b) DF, (c) SS, and (d) SF. Fits of points by specific power functions are also shown.

From current visual data, we have tested the performances of ten modern color-difference formulas, CIELAB [5], CIE94 [10], CIEDE2000 [11,12], CMC [13], DIN99d [30], OSA [31], OSA_GP_Eu [32], CAM02-UCS [14], CAM02-SCD [14] and CAM02-LCD [14], using the *STRESS* index (Eq. (2)). Initially, from the original forms of all these color-difference formulas (i.e., $k_L = k_C = k_H = 1$), the results found are shown in Table 5.

Results in Table 5 indicate that all color-difference formulas predicted the results under diffuse plus spotlight slightly better than the results under only diffuse light. For example, the *STRESS* values for SS are smaller than those for DS, and the *STRESS* values for SF are also slightly smaller than those for DF (except for the CMC formula).

From Table 5, it can be seen that the average *STRESS* values of all color-difference formulas are in a short range from 30.0 to 37.9 (with average of 33.7 and standard deviation of 2.3), with CAM02-LCD performing the best, and CMC the worst. This is not surprising because the CAM02-LCD was specially developed for evaluating large color differences. Overall, Table 5 also indicates that average predictions of all color-difference formulas under both light sources were better (i.e., smaller *STRESS* values) for spherical samples (DS and SS) than for flat samples (DF and SF). This result is highly surprising because all these color-difference formulas were developed using flat samples. In our opinion, in addition to sample shape, this result is affected by the different ranges and average values of color differences for flat and spherical samples (see Fig. 5). There are many evidences [22,33–34] that the magnitude of color differences is a factor influencing the performance of modern color-difference formulas, which were mainly developed for industrial applications involving color differences with small-medium size. Only one color-difference formula cannot be very accurate in the full range of color differences. For

Optics EXPRESS

Table 5. Performance of ten color-difference formulas in terms of *STRESS* for phases DS, DF, SS, and SF plus average and standard deviation (SD) results. The best for each phase is indicated in bold and worst in underline.

	DS	DF	SS	SF	Average	SD
CIELAB	29.2	41.6	23.3	39.1	34.7	8.6
CIEDE2000	27.1	46.6	25.9	40.3	35.0	10.1
CIE94	28.4	44.8	26.0	37.8	34.3	8.7
CMC	28.4	47.7	27.5	48.0	37.9	11.5
DIN99d	27.4	46.2	26.6	38.5	34.7	9.4
OSA	29.2	**39.2**	21.1	37.2	31.7	8.3
OSA_GP_Eu	29.3	49.8	21.0	41.6	35.4	12.8
CAM02-UCS	**25.6**	44.5	21.5	34.6	31.6	10.2
CAM02-SCD	26.5	46.7	23.8	35.6	33.2	10.3
CAM02-LCD	25.9	40.8	**19.8**	**33.6**	**30.0**	9.1
Average	27.7	44.8	23.7	38.6	33.7	
SD	1.4	3.3	2.7	4.1	2.3	

example, the data points with color differences lower than 5 CIELAB units in Fig. 5 may locate on one straight line, but the points with color differences larger than 5 CIELAB units may locate on another different line. To analyze this point, results in each of our four phases were divided into two groups, considering color pairs with color-difference magnitudes below and above 5 CIELAB units. The performance of each of the ten previous color-difference formulas has been tested for each of these two groups, and the results found are listed in Table 6. It can be seen that, as expected from reference conditions, for color differences below 5 CIELAB units, most modern color-difference formulas, CIELAB included, performed better for flat samples than for spherical samples under each lighting condition (i.e., DS vs. DF, or SS vs. SF), the opposite being true for color differences above 5 CIELAB units. It cannot be discarded that the visual testing method employed in the current study, e.g., 2D conventional gray scales, may have had some influence in our current results.

Table 6. Performance of ten color-difference formulas in terms of *STRESS* for phases DS, DF, SS, and SF, distinguishing results for pairs below and above 5 CIELAB units. The best for each phase is indicated in bold and worst in underline.

Magnitude	< 5 CIELAB units				>5 CIELAB units			
Phase	DS	DF	SS	SF	DS	DF	SS	SF
CIELAB	57.1	39.2	40.4	33.7	27.7	31.6	22.0	26.2
CIEDE2000	43.0	35.6	28.4	26.5	26.3	36.5	25.1	29.8
CIE94	50.4	39.8	34.8	30.3	27.3	34.7	25.2	25.8
CMC	38.4	36.3	22.3	**26.3**	27.8	38.8	26.9	40.6
DIN99d	44.7	35.4	26.2	28.5	26.5	36.4	25.9	27.8
OSA	63.5	38.1	47.8	35.0	27.2	**29.5**	19.3	24.5
OSA_GP_Eu	**36.0**	37.0	29.1	27.7	28.8	40.4	19.8	31.4
CAM02-UCS	48.2	35.7	26.9	28.8	24.4	34.8	20.7	25.4
CAM02-SCD	43.6	35.9	**21.5**	29.5	25.7	37.2	23.2	26.6
CAM02-LCD	54.4	**35.2**	34.5	28.6	**24.3**	31.1	**18.7**	**24.2**
Average	47.9	36.8	31.2	29.5	26.6	35.1	22.7	28.2

Optics EXPRESS

3.4.2. F-test for original color-difference formulas

The statistical significance of the difference between two color-difference formulas predicting the current 3D color-difference datasets was tested using the F-test method recommended by CIE [26] and widely used in recent color-difference evaluation [14,25,33–35]. The *STRESS* index can be used to ascertain whether or not two color-difference formulas significantly differ with respect to a given visual dataset. A parameter F, defined as the square of the ratio of the *STRESS* values from two different color-difference formulas A and B, is compared with a specific confidence interval $[F_C, 1/F_C]$, where F_C is a critical value from a two-tailed F-distribution which depends on the assumed confidence level (95% in this study) and the number of color pairs (N) [25,26]. The testing result can be divided into 5 categories, as follows:

- Formula A is significantly better than formula B when $F < F_C$;
- Formula A is significantly worse than formula B when $F > 1/F_C$;
- Formula A is insignificantly better than formula B when $F_C \leq F < 1$;
- Formula A is insignificantly worse than formula B when $1 < F \leq 1/F_C$;
- Formula A is equal to formula B when $F = 1$.

The critical value F_C of the two-tailed F-distribution, assuming 95% confidence level and N-1 degrees of freedom, was in the current case 0.63 ($1/F_C = 1.58$), because the number of color pairs is N=75 in each phase. As an example, the F-test results (F values) for phase DS are listed in Table 7. Similar tables to the results shown in Table 7 can be found for DF, SS, and SF, but not listed here. In Table 7, the smaller the F value, the better the formula A (in the first column) comparing to the formula B (in the first row). It can be seen from Table 7, that most F values were close to one, which means that the performance of the ten tested color-difference formulas were in general very close. This result is consistent with the small value of the ranges of *STRESS* values shown previously in Table 5. Only a few pairs of color-difference formulas were found statistically significant differences, e.g., for phase DF, there is only statistically significant differences between OSA and OSA_GP_Eu.

Table 7. Results of *F*-test [25,26] comparing the performance of any two color-difference formulas in predicting the visual data of phase DS

Formula A \ Formula B	CIELAB	CIEDE2000	CIE94	CMC	DIN99d	OSA	OSA_GP_Eu	CAM02-UCS	CAM02-SCD	CAM02-LCD
CIELAB		1.16	1.06	1.06	1.14	1.00	0.99	1.30	1.21	1.27
CIEDE2000	0.86		0.91	0.91	0.98	0.86	0.86	1.12	1.05	1.09
CIE94	0.95	1.10		1.00	1.07	0.95	0.94	1.23	1.15	1.20
CMC	0.95	1.10	1.00		1.07	0.95	0.94	1.23	1.15	1.20
DIN99d	0.88	1.02	0.93	0.93		0.88	0.87	1.15	1.07	1.12
OSA	1.00	1.16	1.06	1.06	1.14		0.99	1.30	1.21	1.27
OSA_GP_Eu	1.01	1.17	1.06	1.06	1.14	1.01		1.31	1.22	1.28
CAM02-UCS	0.77	0.89	0.81	0.81	0.87	0.77	0.76		0.93	0.98
CAM02-SCD	0.82	0.96	0.87	0.87	0.94	0.82	0.82	1.07		1.05
CAM02-LCD	0.79	0.91	0.83	0.83	0.89	0.79	0.78	1.02	0.96	

Optics EXPRESS

3.4.3. Performance of optimized formulas

The performance of each of ten color-difference formulas was also tested computing the optimum k_L value which minimized *STRESS* values (k_C and k_H were set to one). These minimum *STRESS* values were then averaged, to represent the best performance of each color-difference formula for four phases, and the results found are given in Table 8, together with the corresponding optimum k_L values in brackets.

Comparing the equations' performance with optimized k_L for each phase, we can note that the CAM02-LCD formula again outperformed the others. The CMC formula gave the poorest performance. It can be seen that the *STRESS* values after optimizing k_L for each phase are slightly smaller than those with k_L set to one (Table 5), as expected. It can also be seen in Table 8 that the optimum k_L values are very close to 1.0 for spheres, but always below 1.0 for the flat samples. This means that all current formulas predicted the lightness difference in total color difference in a different way for spherical and flat samples. Specifically, to improve the prediction of visual perception, the computed lightness differences must be about two times higher for flat than for spherical samples. This result may be related to CIE recommendation of k_L=2 for textile samples against k_L=1 for homogeneous samples [10,12], which has been also reported as a consequence of simulated textures [36]. The lighting of spheres, particularly using a spotlight, produced lightness gradients that are not present in flat samples and can be assimilated in some way to textures in the sense they both lead to non-uniformly colored surfaces.

Table 8. Optimized performance of ten color-difference formulas in terms of *STRESS* index, using the optimal k_L values shown in parentheses. The best for each phase is indicated in bold and worst in underline.

	DS	DF	SS	SF	Average
CIELAB	29.2 (1.0)	39.1 (0.4)	23.2 (1.1)	36.6 (0.5)	32.0
CIEDE2000	26.9 (0.9)	42.9 (0.4)	25.5 (0.8)	38.0 (0.6)	33.3
CIE94	28.2 (1.2)	42.6 (0.5)	25.8 (1.2)	37.4 (0.8)	33.5
CMC	28.4 (1.0)	43.0 (0.4)	27.5 (0.9)	42.5 (0.4)	35.4
DIN99d	27.4 (0.9)	42.1 (0.4)	26.5 (0.9)	36.9 (0.6)	33.2
OSA	28.5 (1.4)	39.0 (0.8)	20.4 (1.3)	37.1 (0.9)	31.3
OSA_GP_Eu	28.2 (0.7)	43.7 (0.3)	20.9 (1.1)	38.3 (0.4)	32.8
CAM02-UCS	25.4 (1.2)	41.0 (0.5)	21.5 (1.1)	34.0 (0.8)	30.5
CAM02-SCD	26.5 (0.9)	42.1 (0.4)	23.6 (0.9)	34.4 (0.7)	31.7
CAM02-LCD	**25.0 (1.5)**	**38.7 (0.6)**	**18.9 (1.4)**	**33.2 (0.8)**	**29.0**

The *STRESS* values with optimum k_L values for each phase and all formulas are larger than the value for observer variability listed in Table 3. This indicates that k_L-optimized color-difference formulas are less accurate than average observers in the present study. The statistical significance of the improvements of the ten color-difference formulas by adding optimum k_L were also tested using the *F*-test method, but the results are not listed here. It is found that all the *F* values are in the range of confidence interval [0.63, 1.58]. This means that the improvement achieved by adding an optimized k_L in the original color-difference formula is statistically insignificant.

3.4.4. Performance of power-function color-difference formulas

To improve the performance of current color-difference formulas in predicting visual color differences in a wide range of magnitude (e.g., from threshold to very large color differences), a simple power correction for advanced color-difference formulas was proposed by Huang *et al* in the form of $\Delta E' = a\ \Delta E^{\ b}$ [34], where ΔE is the calculated color difference from a given color-difference formula, a and b are coefficients fitted for each color-difference formula, and

Optics EXPRESS

ΔE' is the corresponding power corrected color difference. As mentioned in Subsection 3.4.1, the color-difference magnitude in this study covered a wide range, from small to large color differences, especially for the flat samples. Therefore, the performances of ten power corrected color-difference formulas (with a and b proposed in ref. 34) in predicting the current experimental data has been tested and the results found are listed in Table 9. The statistically significant improvement of each of the ten power corrected formulas against the original one has been also tested using the F-test method, and the results found are listed in Table 10.

Table 9. Performance of ten power corrected color-difference formulas [34] in terms of *STRESS*. The best for each phase is indicated in bold and worst in underline.

	DS	DF	SS	SF	Average
CIELAB	29.8	**22.5**	21.0	**20.9**	23.5
CIEDE2000	**21.9**	30.8	16.8	24.4	23.5
CIE94	24.8	30.3	19.2	23.9	24.6
CMC	22.8	29.5	17.2	26.3	23.9
DIN99d	23.0	33.0	18.8	25.8	25.1
OSA	28.1	23.4	18.8	21.7	**23.0**
OSA_GP_Eu	23.2	37.5	**15.7**	28.9	26.3
CAM02-UCS	23.6	32.3	16.2	23.8	24.0
CAM02-SCD	23.0	34.3	16.8	24.3	24.6
CAM02-LCD	25.2	34.1	17.6	27.5	26.1
Average	24.5	30.8	17.8	24.8	24.5

Table 10. *F*-test results on statistical significance of the differences between power corrected color-difference formula and its original form for each phase. Statistically significant differences are indicated in bold

	DS	DF	SS	SF
CIELAB	1.04	**0.29**	0.81	**0.29**
CIEDE2000	0.65	**0.44**	**0.42**	**0.37**
CIE94	0.76	**0.46**	**0.55**	**0.40**
CMC	0.64	**0.38**	**0.39**	**0.30**
DIN99d	0.70	**0.51**	**0.50**	**0.45**
OSA	0.93	**0.36**	0.80	**0.34**
OSA_GP_Eu	**0.63**	**0.57**	**0.56**	**0.48**
CAM02-UCS	0.85	**0.53**	**0.57**	**0.47**
CAM02-SCD	0.76	**0.54**	**0.50**	**0.46**
CAM02-LCD	0.95	0.70	0.79	0.67

Comparing Table 9 and Table 5, it can be seen that the power-function color-difference formulas developed from normal 2D samples in ref. 34 are working well in predicting the current data for 3D samples, because they achieve smaller *STRESS* values in most cases. Specifically, this result is confirmed by the fact that all F values in Table 10 are lower than 1.0, except for CIELAB prediction of results for phase DS. Furthermore, the differences between power corrected formulas and their corresponding original formulas for most cases of DF, SS, and SF were statistically significant (i.e., F values lower than the critical value of 0.63, printed in bold in Table 10). The power correction to CIELAB did not work in predicting data of phase DS, which is connected with the scatter diagram shown in Fig. 5(a), where a reasonable linear distribution for all the

Optics EXPRESS

points can be observed, and perhaps the power function with $b = 0.55$ proposed in ref. 34 makes an over-correction of the current data.

3.5. Parametric effects

One way to study a parametric effect (e.g., lighting effect), is to directly compare experimental visual color differences (ΔV) obtained in different phases, using the *STRESS* index. For example, the influence of the change of lighting for spherical (DS vs. SS), and flat (DF vs. SF) in this study can be investigated using this method, and *STRESS* values of 21.9 and 12.8 were found between the two light sources used in the current study (i.e., diffuse lighting and spotlight plus diffuse lighting) for spherical and flat samples, respectively. This means that the change in lighting did affect more the perceived color differences between spheres than between flat samples. This result may be understood considering that complex 3D objects irradiated by an additional spotlight may show much more apparent highlights, shadows etc. than flat objects. In addition, these *STRESS* values were very close to the values for intra- and inter-observer variability (from 11.5 to 23.5, see Table 3) of the current study, which means that the effect of change in lighting on visual color differences was similar to observer's uncertainty.

Another way to study a parametric effect is to compare ellipses from different phases if it is not possible to directly compare their visual color differences, e.g., from different experimental samples or different experiments. A quantitative comparison between the ellipses of the present four phases was carried out using the Monte Carlo method developed by Strocka *et al*. [37], where the ΔE values from two ellipse's equations using 1000 pairs of randomly generated color samples were compared using the *STRESS* index. The results found are given in Table 11 for the different phases. The average variation using this method was 21.5 *STRESS* units for the five color centers. The agreement (i.e. low *STRESS* value) is generally good between different ellipses in the same color center, in particular for the gray and green centers. By far the worst agreements were found for the red and yellow centers, which is consistent with the qualitative impression gained from Fig. 4.

Table 11. Comparing the present results with different phases in terms of *STRESS* using the ellipse-equation. (the best for each phase is indicated in bold and worst in underline)

1st phase	2nd phase	Gray	Red	Yellow	Green	Blue	Average
DS	DF	9.6	33.3	29.0	27.5	11.0	22.1
DS	SS	12.0	14.2	33.0	27.5	23.0	21.9
DS	SF	**6.7**	30.5	19.1	31.5	15.7	20.7
DF	SS	21.5	43.0	49.2	6.0	32.2	30.4
DF	SF	12.1	**10.7**	**10.3**	8.7	22.3	**12.8**
SS	SF	10.6	39.3	42.2	**4.3**	**8.7**	21.0
Average		12.1	28.5	30.5	17.6	18.8	21.5

It also can be seen from Table 11 that some parametric effects exist. For example, the light source effect can be observed by comparing DS and SS (row 2), as well as DF and SF (row 5), with average values of 21.9 and 12.8 *STRESS* units, respectively. These results agree with those obtained by direct comparison of visual color differences (ΔV). It is worth mentioning that the largest average *STRESS* value in Table 11 (30.4) was found between phases DF and SS, where there were apparent changes of three parametric factors: lighting, sample shape, and color-difference magnitude. Furthermore, the smallest average *STRESS* value in Table 11 (12.8) was found between DF and SF phases, where only the lighting difference for the flat samples was considered.

Optics EXPRESS

3.6. Comparisons with other studies

The current experimental results for each color center in four phases were compared with those found in two previous experiments from Guan and Luo [22], and Mirjalili *et al.* [38], where the same five color centers, large color-differences, and flat samples were studied. The Monte Carlo method proposed by Strocka *et al.* [37] was again used here to analyze the relationships between the ellipses from current and previous studies.

3.6.1. Comparing with Guan and Luo's study

The main information about Guan and Luo's experiment [22] is summarized in Table 12. The color-difference magnitude and visual experimental method in our current experiment were similar to the ones in Guan and Luo's experiment [22]. For example, the average size of color difference in Guan and Luo's study was 13 CIELAB units, which is close to the average color difference in DS and SS phases, and about twice larger than the one in DF and SF phases. The most obvious difference between the two studies under comparison is that the current study involves 3D printed spheres. It can be said that the closest experiments to our current ones in Guan and Luo's study were those using gray background and hairline gap (i.e. GHM_1, GHH_5 and GHH_6). The *STRESS* values found from ellipse's equations in Guan and Luo's and current phases are shown in Table 13.

Table 12. Main information about Guan and Luo's experiment [22].

Phase	Viewing background	Gap	Luminance
GHM_1	Gray	Hairline	Medium
WHM_2	White	Hairline	Medium
BHM_3	Black	Hairline	Medium
GGM_4	Gray	Large	Medium
GHH_5	Gray	Hairline	High
GHL_6	Gray	Hairline	Low

Table 13. *STRESS* values between current results and those from Guan & Luo's [22]. The best for each phase is indicated in bold and worst in underline.

	DS	DF	SS	SF	Average
GHM_1	13.9	29.0	27.3	27.4	24.4
WHM_2	19.7	27.3	24.5	24.5	24.0
BHM_3	19.3	33.7	24.1	29.3	26.6
GGM_4	17.5	32.3	28.3	28.8	26.7
GHH_5	**11.5**	**25.9**	**22.6**	**22.8**	**20.7**
GHL_6	20.5	30.1	29.1	27.4	26.8
Average	17.1	29.7	26.0	26.7	24.9

From Table 13, it can be seen that among the six Guan and Luo's subsets, GHH_5 had the best correlation with the results of this study, as shown by the smallest *STRESS* values of 11.5, 25.9, 22.6, and 22.8 for DS, DF, SS, and SF, respectively. It is surprising to see that our results for flat samples (phases DF and SF) had the worse correlation with Guan and Luo's GHH_5 subset than those for spherical samples (phases DS and SS). A possible explanation of this result may be the small size of color differences in phases DF and SF (average of 6.7 and 6.3 CIELAB units) in comparison with Guan and Luo's experiment (average 13 CIELAB units). The different ranges of color-difference magnitude may have affected these comparisons (see Table 6). The lowest

Optics EXPRESS

STRESS value between DS and GHH_5 may be a consequence of particularly similar parametric factors, because both considered large color differences, gray background, hairline gap, and high luminance.

An obvious illuminant effect can be found from Table 13. For example, the *STRESS* values for phase DS (from 11.5 to 20.5) were smaller than those for phase SS (from 22.6 to 29.1). This suggests that agreement with classical color-difference evaluation using 2D objects (i.e. Guan and Luo [22]) is better judging color differences between 3D objects under diffuse lighting (DS), than under diffuse plus spotlight lighting (SS). It also can be seen that the *STRESS* values in phase DF (from 25.9 to 33.7) were very close to those in phase SF (from 22.8 to 29.3). This means that changing lighting conditions in this study did not affect too much the perceived color differences between flat samples, but the same cannot be stated for the perceived color differences between spherical samples. We feel that this result may be related to the fact that the spotlight source introduces higher lightness gradients in the spherical than in the flat samples. For agreement with predictions provided by current color-difference formulas, it is suggested evaluation of color differences between 3D objects under only diffuse lighting.

3.6.2. Comparing with Mirjalili *et al*'s study

The Mirjalili *et al.*'s study [38] considered 11 color centers, and 1012 pairs of gapless 2D printed objects, with average color differences of 1, 2, 4, and 8 CIELAB units (named ZJU_1, ZJU-2, ZJU_4, and ZJU_8, respectively), which were visually evaluated using a gray-scale method. The *STRESS* values found from the ellipses' equations from Mirjalili *et al.*'s and current study were computed, and the results are shown in Table 14.

Table 14. *STRESS* values between current results and those in four subsets from Mirjalili *et al.*'s [38]. (the best for each phase is indicated in bold and worst in underline)

	DS	DF	SS	SF	Average
ZJU_1	17.8	27.5	28.9	29.0	25.8
ZJU_2	14.8	29.2	26.7	30.2	25.2
ZJU_4	11.9	27.2	20.3	26.7	21.5
ZJU_8	**11.8**	**25.5**	**18.2**	**24.3**	**20.0**
Average	14.1	27.4	23.5	27.6	23.1

From Table 14, it can be seen that current experimental ellipses had a good correlation with Mirjalili *et al.*'s ellipses, especially for the large color-difference magnitude subset ZJU_8, which had the closest color-difference magnitude to the current study. The minimum *STRESS* of 11.8 was observed between the ellipse-equation data from DS and subset ZJU_8, and the maximum *STRESS* of 30.2 was found between the ellipse-equation data from SF and subset ZJU_2. Comparing with small-medium color-difference magnitudes in ZJU_1, ZJU_2, and ZJU_4, we can note that the subset ZJU_8 with largest color-difference magnitude always had the best agreement (lowest *STRESS* value) with four phases in the current experiment (bold font in Table 14), which indicates that the color-difference magnitude has an important influence on the visual characteristics of the color pairs, regardless of object shape. Furthermore, the agreement with Mirjalili *et al.*'s results was worse for the two phases using flat samples (DF and SF) than for the two phases using spherical samples (DS and SS), indicating that color-difference magnitude has an important impact on visual color differences, as already mentioned.

Table 14 shows similar results to the above-mentioned illuminant effect on perceived color differences between 3D objects, e.g., the *STRESS* values for all the comparisons involving phase DS were smaller than those involving phase SS. It is confirmed that diffuse lighting is the most appropriate one for color-difference evaluation between complex 3D objects, if we want to obtain

Optics EXPRESS

results in agreement with those found for color differences between 2D objects under diffuse lighting.

4. Conclusions

A gray-scale psychophysical experiment was conducted to collect visual color-difference data between 3D printed objects in two different shapes (sphere and flat), under two different illuminations (diffuse and diffuse plus spotlight). The effects of light source and sample shape on perceived color difference and on the performance of ten color-difference formulas (both, the original formulas and those corrected by specific power functions) were studied. The effect of color-difference magnitude on the perceived color differences was also investigated by comparing with two previous studies. The experimental results show that the range of color-difference magnitude, light source, and 3D shape had more or less influence on the perceived color differences between 3D objects and on the performance of color-difference formulas. It is found that the best illuminant condition for color-difference evaluation of 3D objects in agreement with results found for 2D objects is diffuse lighting. It is also found that all available color-difference formulas were less accurate than the average observer, which claim for the development of new color-difference formulas. However, a remarkable improvement in predictions of wide range color differences between 3D objects was found by adding a power correction to current color-difference formulas. In future research, different 3D shapes of samples and even different 3D shaped gray scales could be used to test the current methodology employing 2D gray scales.

Funding. National Natural Science Foundation of China (61775170); Ministry of Science and Innovation (National Government of Spain) with support of the European Research Development Fund (European Union) (PID2019-107816GB-I00/SRA/10.13039/501100011033).

Acknowledgment. The authors would like to thank many observers from Wenzhou University for participating in this study.

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. K. Xiao, P. L. Sun, and M. R. Pointer, "Review of colour image reproduction using 3D printing technology," Digital Printing **3**, 1–13 (2019).
2. A. Brunton, C. A. Arikan, and P. Urban, "Pushing the limits of 3D color printing: error diffusion with translucent materials," ACM Trans. Graph. **35**(1), 1–13 (2015).
3. K. Xiao, F. Zardawi, R. V. Noort, and J. M. Yates, "Color reproduction for advanced manufacture of soft tissue prostheses," J. Dent. **41**, e15–e23 (2013).
4. K. Xiao, F. Zardawi, R. Noort, and J. M. Yates, "Developing a 3D colour image reproduction system for additive manufacturing of facial prostheses," Int. J. Adv. Manuf. Technol. **70**(9-12), 2043–2049 (2014).
5. CIE 015:2018, *Colorimetry*, 4th Edition (CIE Central Bureau, 2018).
6. CIE 101-1993, *Parametric effects in colour-difference evaluation* (CIE Central Bureau, 1993).
7. A. R. Robertson, "CIE guidelines for coordinated research on colour-difference evaluation," Color Res. Appl. **20**(3), 399–403 (1995).
8. K. Witt, "CIE guidelines for coordinated future work on industrial colour-difference evaluation," Color Res. Appl. **20**(6), 399–403 (1995).
9. M. Melgosa, "Request for existing experimental datasets on color differences," Color Res. Appl. **32**(2), 159 (2007).
10. CIE 116-1995, *Industrial Colour-Difference Evaluation* (CIE Central Bureau, 1995).
11. 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).
12. ISO/CIE 11664-6:2014(E), *Colorimetry – Part 6: CIEDE2000 colour-difference formula* (CIE Central Bureau, 2014).
13. F. J. J. Clarke, R. McDonald, and B. Rigg, "Modification to the JPC79 colour-difference formula," J. Soc. Dyers Colour. **100**(4), 128–132 (1984).
14. 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).
15. 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).

Optics EXPRESS

16. M. G. Bloj, D. Kersten, and A. C. Hurlbert, "Perception of three-dimensional shape influences colour perception through mutual illumination," Nature **402**(6764), 877–879 (1999).
17. M. Bloj, D. Weiß, and K. R. Gegenfurtner, "Bias effects of short- and long-term color memory for unique objects," J. Opt. Soc. Am. A **33**(4), 492–500 (2016).
18. R. J. Lee and H. E. Smithson, "Low levels of specularity support operational color constancy, particularly when surface and illumination geometry can be inferred," J. Opt. Soc. Am. A **33**(3), A306–A318 (2016).
19. B. Xiao and D. H. Brainard, "Surface gloss and color perception of 3D objects," Vis. Neurosci. **25**(3), 371–385 (2008).
20. G. 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).
21. S. Guan and M. R. Luo, "Investigation of parametric effects using small colour differences," Color Res. Appl. **24**(5), 331–343 (1999).
22. S. Guan and M. R. Luo, "Investigation of parametric effects using large colour differences," Color Res. Appl. **24**(5), 356–368 (1999).
23. M. Melgosa, J. Martínez-García, L. Gómez-Robledo, E. Perales, F. M. Martínez-Verdu, and T. Dauser, "Measuring color differences in automotive samples with lightness flop: A test of the AUDI2000 color-difference formula," Opt. Express **22**(3), 3458–3467 (2014).
24. E. Kirchner and N. Dekker, "Performance measures of color-difference equations: correlation coefficient versus standardized residual sum of squares," J. Opt. Soc. Am. A **28**(9), 1841–1848 (2011).
25. P. A. García, R. Huertas, M. Melgosa, and G. Cui, "Measurement of the relationship between perceived and computed color differences," J. Opt. Soc. Am. A **24**(7), 1823–1829 (2007).
26. CIE 217:2016, *Recommended method for evaluating the performance of colour-difference formulae* (CIE Central Bureau, 2016).
27. M. Melgosa, P. A. García, L. Gómez-Robledo, R. Shamey, D. Hinks, G. Cui, and M. R. Luo, "Notes on the application of the standardized residual sum of squares index to the assessment of intra- and inter-observer variability in color-difference experiments," J. Opt. Soc. Am. A **28**(5), 949–953 (2011).
28. M. R. Luo and B. Rigg, "Chromaticity-discrimination ellipses for surface colours," Color Res. Appl. **11**(1), 25–42 (1986).
29. M. Melgosa, E. Hita, A. J. Poza, D. H. Alman, and R. S. Berns, "Suprathreshold color-difference ellipsoids for surface colors," Color Res. Appl. **22**(3), 148–155 (1997).
30. G. Cui, M. R. Luo, B. Rigg, G. Roesler, and K. Witt, "Uniform colour spaces based on the DIN99 colour-difference Formula," Color Res. Appl. **27**(4), 282–290 (2002).
31. D. L. MacAdam, "Uniform color scales," J. Opt. Soc. Am. **64**(12), 1691–1702 (1974).
32. C. Oleari, M. Melgosa, and R. Huertas, "Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space," J. Opt. Soc. Am. A **26**(1), 121–134 (2009).
33. 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).
34. M. Huang, G. Cui, M. Melgosa, M. Sánchez-Marañón, C. Li, M. R. Luo, and H. Liu, "Power functions improving the performance of color-difference formulas," Opt. Express **23**(1), 597–610 (2015).
35. Q. Xu, B. Zhao, G. Cui, and M. R. Luo, "Testing uniform colour spaces using colour differences of a wide colour gamut," Opt. Express **29**(5), 7778–7793 (2021).
36. 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).
37. D. Strocka, A. Brockes, and W. Paffhausen, "Influence of experimental parameters on the evaluation of color-difference ellipsoids," Color Res. Appl. **8**(3), 169–175 (1983).
38. F. Mirjalili, M. R. Luo, G. Cui, and J. Morovic, "Color-difference formula for evaluating color pairs with no separation: ΔENS," J. Opt. Soc. Am. A **36**(5), 789–799 (2019).