COLOR DIFFERENCE EVALUATION FOR DIGITAL IMAGES

Haisong XU*, Qingfen TONG, and Rui GONG

State Key Laboratory of Modern Optical Instrumentation, Department of Optical Engineering, Zhejiang University, Hangzhou 310027, China

*Corresponding author: Haisong XU, chsxu@zju.edu.cn

Keywords: color difference evaluation, psychophysical experiment, digital image

ABSTRACT

A review is firstly presented about the history of color difference formulas and the development of color difference evaluation for digital images. This study involves two phases of psychophysical experiments. Phase I was implemented via a categorical judgment method to test the typical color difference evaluation methods for digital images, including CIELAB, CIEDE2000, CAM02-UCS, S-CIELAB, and iCAM, which indicates that iCAM performs best in the evaluation of image color difference. Two visual experiments were involved in Phase II, of which the first one was conducted to determine the threshold of image color difference and the second one was carried out based on a magnitude estimation method to evaluate the suprathreshold color difference for digital images. The visual data illustrate that the mean acceptable threshold is nearly twice the perceptible one and the different thresholds of lightness, chroma, hue and sharpness should be considered in the future modeling of color difference evaluation for digital images.

INTRODUCTION

With the rapid development and application of various digital imaging devices, color digital image has become an information and communication medium, which is extremely important to industrial production and daily life. Color difference evaluation for digital images is a focus and academic difficulty in the research of color science and imaging technology, and there has not yet been any recommended standard method on the color difference evaluation for digital images. What's more, the studies of existing color difference evaluation methods for digital images have mainly stayed at the threshold level, which urges us to further discuss the suprathreshold color difference evaluation for digital images.

DEVELOPMENT OF COLOR DIFFERENCE EVALUATION

There have been many researchers committed to establish and test new color difference formulas since 1935, and more than 40 color difference formulas were developed, including CIELAB, CIELUV, CMC, BFD, CIE94, CIEDE2000, and etc. These color difference formulas have been widely used in industry such as textiles and printing. Based on the color appearance model CIECAM02, three color difference formulas named CAM02-LCD, CAM02-SCD and CAM02-UCS were proposed to calculate large, small and full range of color difference, respectively, which are superior to previous color difference formulas by considering viewing conditions of background, surround, and etc.

Both the traditional color difference formulas and those from the color appearance model were established based on uniform color patch samples. However, a digital image is formed by a large number of pixels with all kinds of colors, which is quite different from uniform color patches, making its pixels difficult to be measured directly with physical instruments. As well known, human eyes cannot detect an individual pixel within an image at typical viewing distance, and, furthermore, the color perception of images is not merely the simple aggregation of the appearance of individual isolated colors. Hereby a sophisticated color difference model must take into account different factors including the spatial sensitivity.

In 1996, Zhang *et al.* derived the S-CIELAB model by considering the human visual system [1]. The first step of this model is a spatial filtering operation applied to the color image data presented on an opposite color space in order to simulate the spatial blurring by the human visual system. In 2001, Johnson *et al.* followed the earlier work by Zhang and developed a spatial color difference formula based upon CIEDE2000 [2]. Then, they proposed a framework, which refined the CSF (Contrast Sensitivity Function) equations of the S-CIELAB model and added modules for spatial frequency adaptation, spatial localization, and local and global contrast detection [3]. In 2002, Fairchild *et al.* put forward an image color appearance model (iCAM), which provides traditional color appearance capabilities, spatial vision attributes, and color difference metrics in a model convenient enough for practical applications [4].

In 2000, Stokes *et al.* tested the traditional color difference formulas applied on images [5, 6]. They compared the calculated mean image color difference with the perceived image difference, which showed that the mean perceived color difference was roughly 2.5 ΔE^*_{ab}, and that the perceived color differences were obviously influenced by the image contents.

In 2002, Hong *et al.* proposed a metric for evaluating the color difference between images based on colorimetric statistics [7]. This metric agreed well with the visual results, but it was too complicated in practical applications. In 2009, Pedersen *et al.* designed a new color image difference metric named SHAME (Spatial Hue Angle Metric) from the hue angle algorithm, which takes into account the spatial properties of the human visual system [8].

In this study, several above-mentioned methods were tested for color difference evaluation of digital images, in which the color difference of each image pair was calculated using CIELAB, CIEDE2000, CAM02-UCS, S-CIELAB, and iCAM, respectively, and then was compared with the subjective assessment results from observers. Moreover, a deep discussion of suprathreshold color difference evaluation for digital images was conducted via two psychophysical experiments to lay the foundation for modeling the image color difference.

EXPERIMENTAL

Two phases of visual experiments were performed in this study. Phase I was designed to test and compare the existing color difference evaluation methods, and two psychophysical experiments involved in Phase II were carried out to assess the threshold and suprathreshold color difference for digital images.

In Phase I, six original images, i.e. four ISO SCID 300 images, one CIE TC8-03 sRGB image, and an additional image containing sky and plants, were selected and clipped to the same size of 15 cm × 20 cm. To generate test images similar to their originals under a limited extent of color difference, the six original images were manipulated in terms of four attributes, i.e. lightness, chroma, hue angle (corresponding to L^*, C, and h of CIELAB color space, respectively), and sharpness. A total of 216 test images (6 images × 36 manipulations) were produced. The visual experiment was conducted in a dark room on a 24-inch EIZO LCD, which was characterized using the GOG (Gain-Offset-Gamma) [9] model under the D65 white point and with the colorimetric accuracy of 0.92 ΔE^*_{ab}. In each session, the image pairs composed by the original images and their modulated versions were presented on the display in a random order, and the position of the two images in a pair on the left or right was also randomized. The horizontal viewing angle was 11.2° with an interval of 1°, and the vertical viewing angle was 6.8°. A panel of 10 observers with normal color vision took part in the experiment to assess the color difference sensation of the image pairs via the psychophysical method of category judgment, which contains seven grades of color difference.

In the first visual experiment of Phase II, the original images, modulation types and the monitor settings were the same as Phase I, except for the modulating parameter settings, which led to more manipulations, resulting in a total number of 540 test images (6 images × 90 manipulations). The same panel of 10 observers were invited to estimate the color differences of the displayed image pairs by clicking the different buttons on the screen as the visual responses of 'no difference', 'just perceptible difference' or 'just not acceptable difference', respectively. The whole assessments were divided into 2 sessions so that any one session lasted about 20 minutes to avoid fatigue.

The second psychophysical experiment of Phase II was carried out based on a magnitude estimation method to evaluate the suprathreshold color difference of digital images. Each modulation of the original images corresponded to its belonging reference scale, which was modulated out according to the perceptible color difference threshold parameter determined in the first experiment of Phase II. Each reference scale had six levels, namely, the original image (no difference), the threshold image (modulated by the perceptible threshold parameter), double (2 times) threshold image (modulated by the perceptible threshold parameter multiplied by 2), then triple (3 times), quadruple (4 times), and quintuple (5 times) threshold images, which corresponded to the grades of 0 to 5, respectively. To ensure the color difference range of the test images being between 0 and 5 times the threshold value, the test image modulation parameters were adjusted slightly based on the results from the first experiment of Phase II. And the total number of test images was still 540 (6 images × 90 manipulations). The visual task of observers was to estimate the color differences between the test images and their corresponding original images based on the reference scales. Altogether, 5400 (6 images × 90 modulations × 10 observers) sets of visual data were collected.

RESULTS AND DISCUSSION

All the categorical data obtained in the experiment of Phase I were transformed into the equal-interval scale values, which was equivalent to the visual judgment of the observer. Then the color differences were calculated between the original images and their modulated versions pixel by pixel using models of CIELAB, CIEDE2000, CAM02-UCS, S-CIELAB, and iCAM, respectively. The correlations between the calculated color differences and their corresponding visual evaluation values were analyzed by standardized residual sum of squares (STRESS) [10], as shown in Fig. 1, with the standard deviations demonstrating the impact of image contents on the correlation. It can be seen that iCAM performs best in terms of the overall prediction accuracy for all the manipulation attributes, followed by S-CIELAB, CIELAB and CIEDE2000, while CAM02-UCS is the poorest.

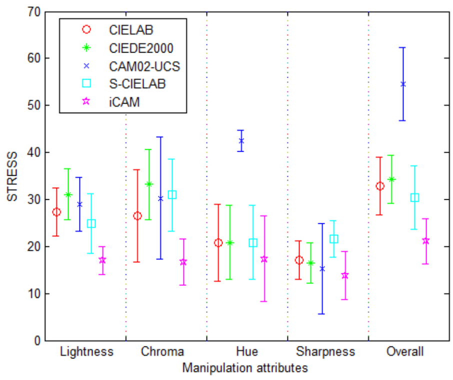

Fig. 1 The average values of STRESS in Phase I.

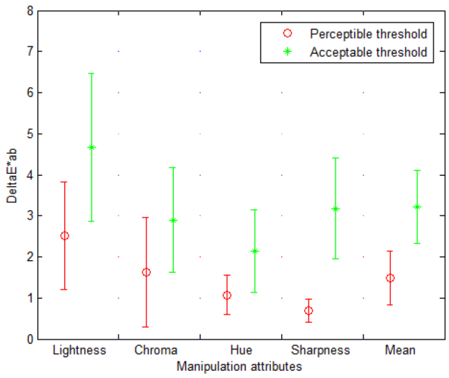

Fig. 2 The perceptible and acceptable thresholds in Phase II.

The visual data gathered in the first experiment of Phase II, as shown in Fig. 2, indicate that the mean perceptible threshold of all the four color parameters is about 1.85 ΔE^*_{ab}, and the mean acceptable threshold is about 3.60 ΔE^*_{ab}, nearly twice the perceptible threshold. The perceptible threshold of lightness difference is 2.51 ΔE^*_{ab}, greater than those of chroma (1.63 ΔE^*_{ab}), hue (1.07 ΔE^*_{ab}) and sharpness (0.69 ΔE^*_{ab}), which implies that it is important to consider the magnitude relationship of lightness, chroma, and hue differences when calculating the color difference of digital images.

The observers' evaluation data collected in the second experiment of Phase II were transformed into the visual evaluation images. Therefore, the threshold and suprathreshold visual color difference data for digital images were combined, which would be used to establish a color difference evaluation model for digital images in the future work.

CONCLUSIONS

By reviewing the development of color difference evaluation for digital images, two phases of psychophysical experiments were carried out to test the typical color difference evaluation methods and to collect the visual data at the levels of threshold and suprathreshold color difference for digital images, respectively. The detailed analysis indicates that iCAM performs the best among the existing color difference models, and that the acceptable threshold is nearly twice the perceptible one and the thresholds of lightness, chroma, hue and sharpness are quite different. Hereby, the image color difference evaluation should consider not only the human visual characteristics and image spatial properties, but also the different contributions of lightness, chroma and hue. Moreover, the suprathreshold visual results from this study would be involved in the further modeling of color difference evaluation for digital images.

REFERENCES

1. Zhang, X. M., & Wandell, B. A. (1997). A spatial extension of CIELAB for digital color image reproduction. *Journal of the SID*, 5(1), 61-63.
2. Johnson, G. M., & Fairchild, M. D. (2003). A top down description of S-CIELAB and CIEDE2000. *Color Res. Appl.*, 28(6), 425-435.
3. Johnson, G. M., & Fairchild, M. D. (2001). Darwinism of color image difference models. *The 9th IS&T/SID Color Imaging Conference*, 108-112.
4. Fairchild, M. D., & Johnson, G. M. (2004). iCAM framework for image appearance, differences, and quality. *J. Elec. Imaging*, 13(1), 126-138.
5. Song, T., & Luo, M. R. (2000). Testing color difference formulae on complex images using a CRT monitor. *The 8th IS&T/SID Color Imaging Conference,* 44-48.
6. Gibson, J. E., Fairchild, M. D., & Wright, S. L. (2000). Colorimetric tolerances of various digital image displays. *The 8th IS&T/SID Color Imaging Conference*, 295-300.
7. Hong, G., & Luo, M. R. (2001). Perceptually based color difference for complex images. *The 9th Congress of the International Color Association*, 618-621.
8. Pedersen, M., & Hardeberg, J. Y. (2009). A New Spatial Hue Angle Metric for Perceptual Image Difference. *Computation Color Imaging*, 5646, 81-90.
9. Berns, R. S., Motta, R. J., & Gorzynski, M. E. (1993). CRT colorimetry. Part I: Theory and practice, *Color Res. Appl.*, 18(5), 299-314.
10. García, P. A., Huertas, R., Melgosa, M., & Cui. G. H. (2007). Measurement of the relationship between perceived and computed color differences. *JOSA A*, 24(7), 1823-1829.