DOI: 10.2352/issn.2168-3204.2016.1.0.181

Modification of CIEDE2000 for Assessing Color Quality of Image Archives

Roy S. Berns
Rochester Institute of Technology
54 Lomb Memorial Drive, Color Science Hall
Program of Color Science
Studio for Scientific Imaging and Archiving of Cultural Heritage
Rochester, New York, 14623-5604
585-475-2230 (fax) 585-475-4444
berns@cis.rit.edu

Abstract

Evaluating color accuracy of image archives is accomplished using test targets, colorimetric measurements, and total color difference calculations. Both CIELAB and CIEDE2000 are used with the later often preferred being an international standard and having a weighting function for chroma position. Its lightness weighting function is problematic when used for image archiving because it allows for greater lightness errors for dark colors, colors critical in defining image quality. A statistical analysis was performed on data used to derive CIEDE2000. The results do not support the weighting function for lightness. For imaging applications, it is recommended that the lightness weighting function be replaced with unity.

Problem

Color targets are used to evaluate color accuracy where the average camera values for each patch of the target are compared with their reference values obtained using a bi-directional spectrophotometer. Both devices record CIELAB coordinates for illuminant D50 and the 1931 standard observer. Two total color difference metrics are used: ΔE^*_{ab} and ΔE_{00} (CIELAB and CIEDE2000) [1, 2], with average and maximum values defining accuracy. The assumptions are that imaging systems with high color accuracy produce more accurate image archives than systems with lower color accuracy, and that the magnitude of the total color difference predicts perceived color-image quality.

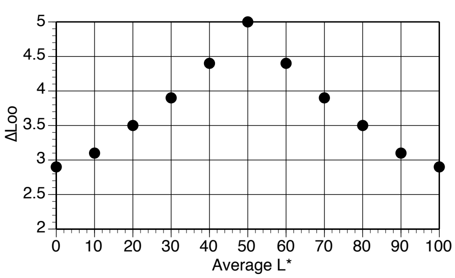

***Figure 1**. ΔE_{00} as a function of the average L* of a color difference pair for a fixed ΔL^* of 5.0.*

The most important criterion of color-image quality is lightness accuracy, particularly for dark colors. CIELAB and CIEDE2000 define lightness accuracy quite differently, where the latter allows for greater differences in ΔL^* for a given total color difference. Nobbs has derived a method of splitting CIEDE2000 into ΔL_{00}, ΔC_{00}, and ΔH_{00} [3]. By plotting ΔL_{00} as a function of the average L* of a color-difference pair, as is shown in Figure 1, the differences between these two metrics are readily seen. In this example for a fixed $\Delta L^* = 5.0$, the average ΔL^* reduces from 5.0 to 3.7 ΔL^*_{00}, and ranges from 2.9 to 5.0.

Because of the critical nature of tone reproduction, it is unclear whether the CIEDE2000 lightness weighting should be used when evaluating the color quality of image archives. To improve clarity, a closer look into the derivation of this weighting was carried out.

Prequel to CIEDE2000: CIE94

CIELAB is recommended "…whenever a three-dimensional spacing perceptually more nearly uniform than that provided by the XYZ system is desired" [4]. This is a very weak recommendation and beginning in the 1970's, the United Kingdom textiles industry derived formulas that improved the total color difference metric (ΔE^*_{ab}), culminating in the CMC formula [5], adopted as both British and American textiles standards [6, 7]. Such formulas use weighting functions (S_L, S_C, S_H) that adjust the total color difference based on the pair's average CIELAB coordinates, and parametric factors (k_L, k_C, k_H) that adjust the relative weightings of lightness, chroma, and hue difference, shown in Eq. 1. During the early 1990's, CIE technical committee 1-29 used a similar approach resulting in CIE94, where S_C and S_H were functions of chroma and S_L was set to unity [8], shown in Eq. 2. Thus, ΔE^*_{ab} and ΔE^*_{94} had the identical lightness difference calculation (but differed from CMC which had a lightness weighting function).

$$\Delta E^*_{weighted} = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C^*_{ab}}{k_C S_C}\right)^2 + \left(\frac{\Delta H^*_{ab}}{k_H S_H}\right)^2} \quad (1)$$

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L^*}{k_L}\right)^2 + \left(\frac{\Delta C^*_{ab}}{k_C\left(1+0.045\bar{C}^*_{ab}\right)}\right)^2 + \left(\frac{\Delta H^*_{ab}}{k_H\left(1+0.015\bar{C}^*_{ab}\right)}\right)^2} \quad (2)$$

The lightness difference data evaluated by TC 1-29 are plotted in Figure 2 [8]. These data were obtained by fitting constrained ellipsoids to visual data about a reference color. Thus

DOI: 10.2352/issn.2168-3204.2016.1.0.181

each point corresponds to a specific color center. The inconsistent trends were considered parametric, that is, specific to each dataset. Appropriately, an S_L function was not derived.

CIEDE2000 lightness weighting function: S_L

During the late 1990's a new CIE technical committee, TC 1-47, was formed to improve CIE94 owing to new experimental data, the result: CIEDE2000 [9]. The CIE94 weighting functions were augmented by functions dependent on L*, a*, and h_{ab}. The main evidence for a lightness weighting were new data by Chou, et al. [10], plotted in Figure 3 (top), and Kim and Nobbs [11]. There is a weak trend where dark and light colors require positive weighting compared with middle gray. The S_L function adopted by the CIE is also plotted in Figure 3 (bottom).

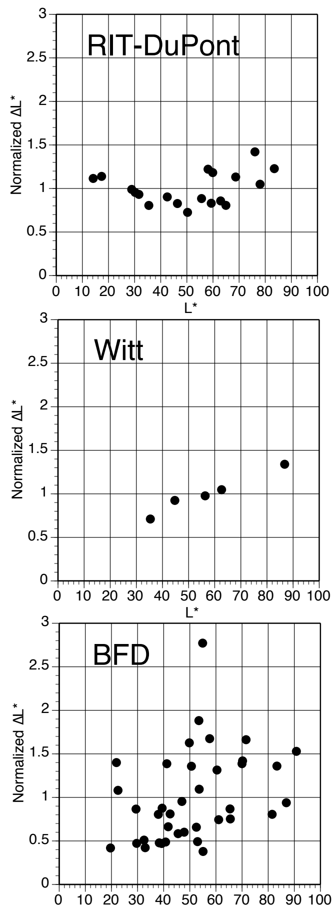

***Figure 2**. Normalized ΔL* vs. L* for sample pairs only varying in lightness, the normalization accounting for differences in average color difference magnitude of each experiment. Each point represents a different color center.*

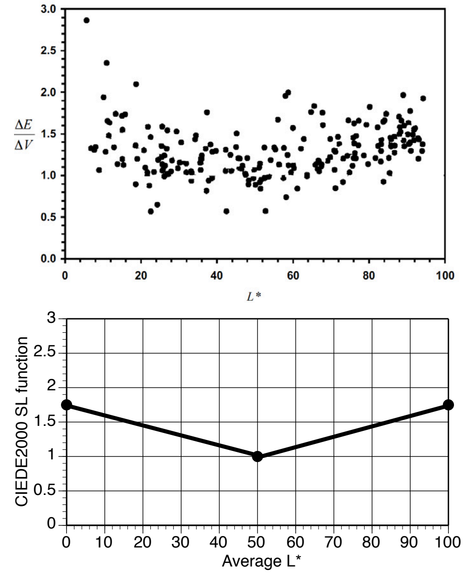

***Figure 3**. (Top) $\Delta E^*_{ab}/\Delta V$ vs. L* for Chou, et al. [2]; (Bottom) CIEDE2000 S_L function.*

The visual data evaluated by TC 1-47 was dubbed the "COM" set (combined) [12]; it included individual color difference pairs with visual differences (BFD-D65, BFD-C, BFD-M, Leeds-GS, Leeds-PC, and Witt) and color-difference data based on derived tolerances from individual pairs (RIT-DuPont T_{50}). A significant difference between the approaches used by TC 1-29 and TC 1-47 was the type of data used to evaluate performance. TC 1-29 used ellipsoid-based data that filtered much of the observer uncertainty whereas TC 1-47 used individual color-difference pair data, with the exception of the RIT-DuPont dataset. RIT-DuPont data for individual difference pairs were derived and published in 2010 [13] and for these analyses, they replaced the T_{50} data. Thus all datasets were individual color-difference pair data. The 4641 pairs were reduced to 454 pairs by including pairs with $|\Delta L^*|/\Delta E^*_{ab} \geq 0.9$; that is, analyzed pairs varied predominantly in lightness. For each dataset, the visual data were scaled so that their average was unity. The visual data, scaled-$\Delta E^*_{ab}/\Delta V$, are plotted against the average L* of each pair along with the CIEDE2000 S_L function (shifted to minimize RMS error between the visual data and S_L) and a line fit, shown in Figure 4. The R^2 between the visual data and S_L was 0.135 and for the line fit, 0.142. Clearly, the lightness data are poorly fit in either case questioning both the precision of this type of visual data and TC 1-47's decision to include a S_L function other than unity.

 © 2016 SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY

DOI: 10.2352/issn.2168-3204.2016.1.0.181

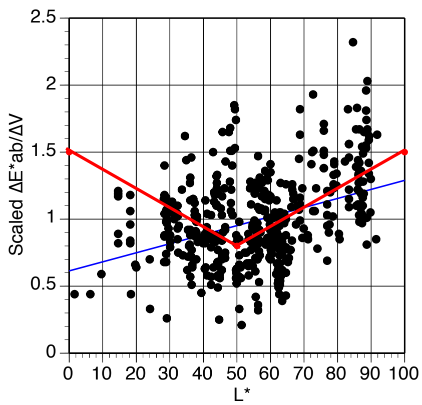

***Figure 4**. Scaled $\Delta E^*_{ab}/\Delta V$ vs. L* for the lightness dataset, shifted CIEDE2000 S_L function (red line) and linear fit to the data (blue line).*

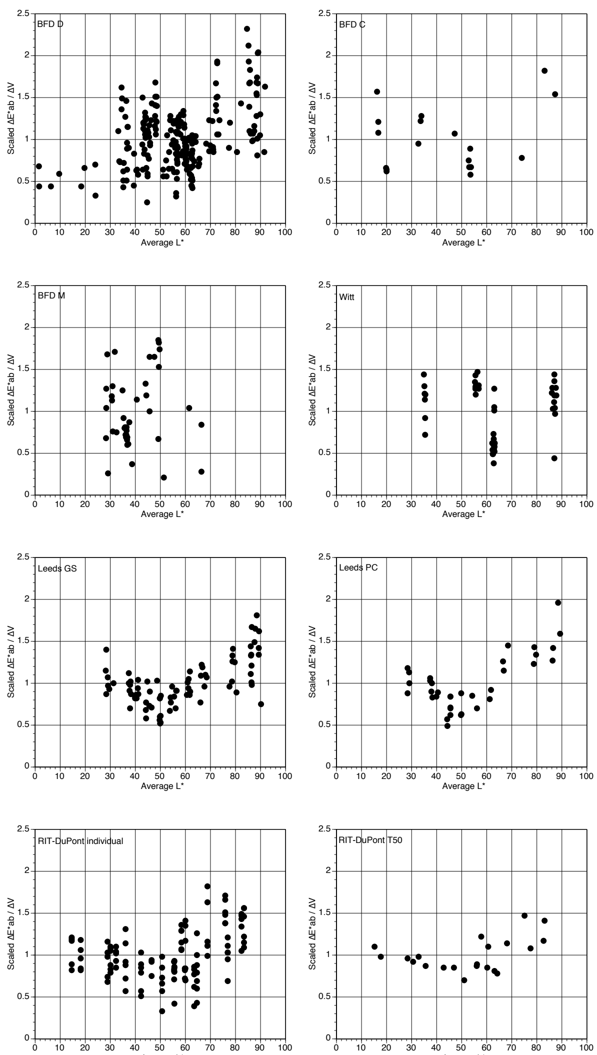

***Figure 5**. Scaled $\Delta E^*_{ab}/\Delta V$ vs. L* for the each listed dataset.*

The individual datasets are plotted in Figure 5. The Leeds' datasets, used to derive S_L, do not contain the range of lightness common to image archives. Its minimum L* is only near 30. Datasets with darker colors, RIT-DuPont and BFD D, tend to flatten out rather than increase. The RIT-DuPont T_{50} and individual data have both been plotted to reveal the lack of precision of individual color-difference pairs.

Each dataset (excluding RIT-DuPont T_{50}) was analyzed for statistical significance using the multi-dimensional metric STRESS and a two-tailed F test at a 95% confidence level [14]. Both TC 1-29 and TC 1-47 used this type of significance testing. The visual data were compared with the CIEDE2000 S_L function (Eq. 3) and with unity (S_L = 1, Eq. 4), the results given in Table I. STRESS decreases with increasing correlation. The S_L function improved performance only for the Leeds' data. For the entire dataset (All), performance was equivalent.

These analyses were repeated by further filtering using only pairs with average $C^*_{ab} \leq 5.0$. The identical statistical results were obtained.

Thus, CIEDE2000's S_L function is a parametric function for the Leeds' datasets. (The same is likely for the Chou, et al. dataset; the data could not be obtained for these analyses.) It is not a weighting function that accounts for a fundamental limitation in CIELAB's ΔE^*_{ab}. Melgosa, Huertas, and Berns found similar results [15, 16].

Table I. STRESS for each listed dataset and S_L. A statistically significant difference is shown as bold.

Dataset	n	SL function	SL = 1
All	454	29.09	31.36
BFD-D65	172	33.17	33.39
BFD-C	11	29.34	42.25
BFD-M	22	46.63	40.38
Leeds_GS	76	**16.16**	25.72
Leeds_PC	35	**19.38**	30.39
Witt	30	30.53	31.91
RIT-DuPont	108	25.53	28.45

Modification of CIEDE2000

Having established that CIEDE2000 does not generalize for all the datasets when the analysis is limited to color differences that vary predominantly in lightness, the validity of CIEDE2000 comes into question, particularly when applied to quantifying an imaging system's color accuracy because accurate lightness (tone) reproduction is critical for quality image archives. The solution is simple: modify the CIEDE2000 S_L (Eq. 3) by setting it to unity, as shown in Eq. 4.

$$S_L = 1 + \frac{0.015\left(\bar{L} - 50\right)^2}{\sqrt{20 + \left(\bar{L} - 50\right)^2}} \quad (3)$$

$$S_{L,\text{modified}} = 1 \quad (4)$$

DOI: 10.2352/issn.2168-3204.2016.1.0.181

Imaging Example

A color-managed ProPhotoRGB image of the Xrite Digital Color Checker SG was transformed to an LAB encoded image, Figure 6. The L* channel was adjusted using Photoshop's curves. In one case the midtones were lightened (Figure 6 top), in the second, the shadows were lightened (Figure 6 bottom). The average CIEDE2000 was 1.0 between the original and each manipulated image. Using CIEDE2000 with S_L = 1, the midtone-adjusted image was 1.0 and the shadow-adjusted image was 1.5. Visually, CIEDE2000 with S_L = 1 has better correlation. The shadow-lightened image has poorer quality than the midtone-lightened image.

Conclusions

When TC 1-47 evaluated CIEDE2000, it was statistically significantly better than CIE94 for both the COM set and each individual dataset, leading to the inclusion of the S_L function derived by Nobbs. Using all individual color-difference pair data and sorting for $|\Delta L^*|/\Delta E^*_{ab} \geq 0.9$ (or adding a second sorting of $C^*_{ab} \leq 5.0$), only the Leeds data are improved using the CIEDE2000 S_L function rather than unity. This leads to the conclusion that CIEDE2000 is biased for the Leeds data and the S_L function should not be included in CIEDE2000. Its inclusion reduces correlation between calculated colorimetric performance and visual evaluation, particularly for imaging applications. It is recommended that CIEDE2000 should be modified by using S_L = 1.0.

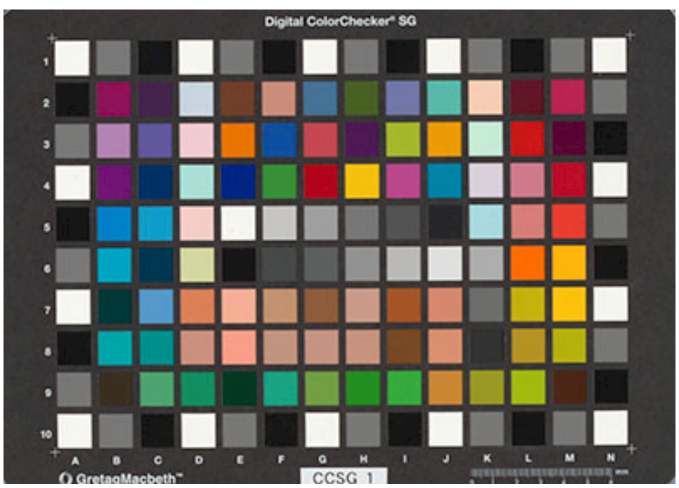

***Figure 6**. ProPhoto RGB image of Xrite Digital ColorChecker SG.*

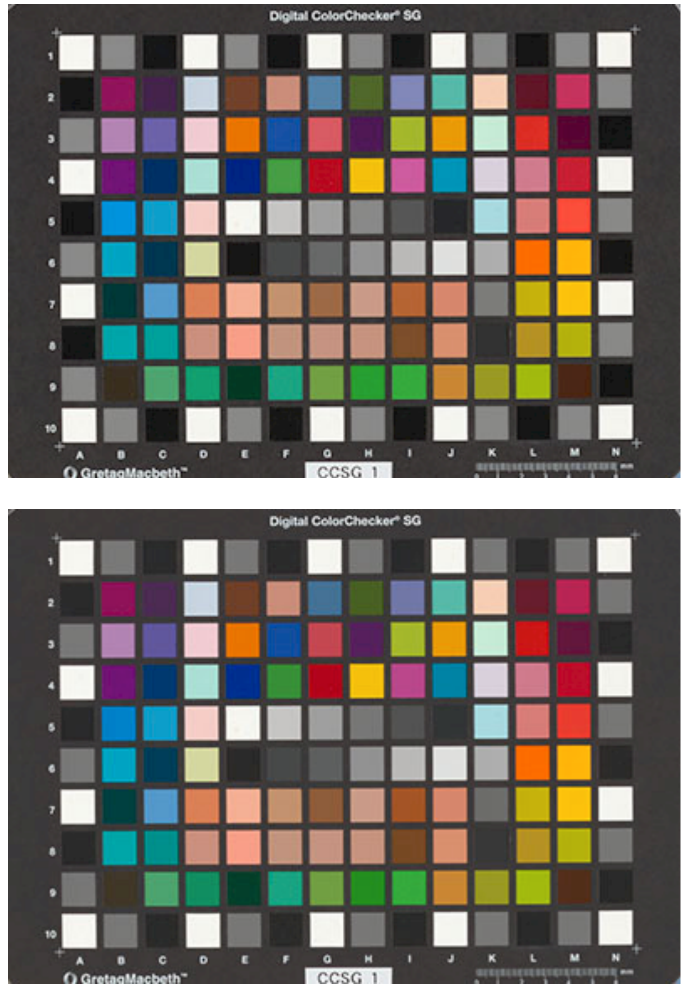

***Figure 7**. Manipulated image shown in Figure 6 by either lightening midtones (top) or lightening shadows (bottom). The Average CIEDE2000 for top and bottom images is 1.0. Average CIEDE2000 with modified S_L = 1.0 for top and bottom images is 1.0 and 1.5, respectively.*

References

[1] Federal Agencies Digitization Initiative (FADGI) - Still Image Working Group, Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files. http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf (2010).

[2] H. van Dormolen, 2012. Metamorfoze Preservation Imaging Guidelines: Image Quality, Version 1.0, The Hague: National Library of the Netherlands, 2012.

[3] J. H. Nobbs, "A lightness, chroma and hue splitting approach to CIEDE2000 colour differences," Advances in Color Science and Technology, 5(2), pp. 46-53, 2002.

[4] CIE, 15:2004 Colorimetry. 3rd ed. Vienna: Commission Internationale de L'Eclairage. 2004.

[5] F. J. J. Clarke, R. McDonald, and B. Rigg, "Modification to the JPC79 colour-difference formula," J. Soc. Dyers Colour. 100, pp. 128-132, 1984.

[6] British Standards, BS 6923-1988 Method for Calculation of Small Colour Differences. 1988.

 © 2016 SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY

DOI: 10.2352/issn.2168-3204.2016.1.0.181

[7] AATCC, TM173-TM, Test Method: CMC: Calculation of Small Color Differences for Acceptability. Research Triangle Park : American Association of Textile Chemists and Colorists, 1989.
[8] R. S. Berns, "The mathematical development of CIE TC 1-29 proposed color-difference equation: CIELCH," in Proceedings of AIC Colour 93 (AIC, 1993), Vol. B, pp. 189–192, http://www.aic-colour.org.
[9] CIE, S 014-6/E:2013, Colorimetry – Part 6: CIEDE2000 colour-difference-formula, Vienna: Commission Internationale de L'Eclairage, 2013.
[10] W. Chou, H. Lin, M.R. Luo, S. Westland, B. Rigg, J. Nobbs, "Performance of lightness difference formulae," Coloration Technology 117 pp. 19-29, 2001.
[11] D. H. Kim and J. H. Nobbs, ''New weighting functions for the weighted CIELAB colour difference formula,'' in Proceedings of AIC Colour 97, Kyoto, 1997, Vol. 1, pp. 446–449; http://www.aic-colour.org.
[12] Data provided by M Melgosa, personal communication.
[13] R. S. Berns and B. Hou, "RIT-DuPont supra-threshold color-tolerance individual color-difference pair dataset," Color Research Application 35, pp. 274-283, 2010.
[14] 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, pp. 1823–1829, 2007.
[15] M. Melgosa, R. Huertas, and R. S. Berns, "Relative significance of the terms in the CIEDE2000 and CIE94 color-difference formulas," J. Opt. Soc. Am. A 21, pp. 2269–2275, 2004.
[16] M. Melgosa, R. Huertas, RS Berns, Performance of recent advanced color-difference formulas using the standardized residual sum of squares index, J. Opt. Soc. Am. A, 25, pp. 1828-1834, 2008.

Acknowledgements

This research was supported by a grant from the Andrew W. Mellon Foundation.

Author Biography

Roy S. Berns holds the Richard S. Hunter Professorship in Color Science, Appearance, and Technology at Rochester Institute of Technology (RIT). He has directed the Munsell Color Science Laboratory and graduate programs in Color Science at RIT. He is the director of the Andrew W. Mellon sponsored Studio for Scientific Imaging and Archiving of Cultural Heritage. Berns is an IS&T Fellow with lifetime achievement awards from the International Association of Colour, the U.S. Inter-Society Color Council, and the Colour Group of Great Britain.