A Comparison of Color Metrics

John J. McCann
Consultant, Belmont, Massachusetts

Abstract

Color differences are almost always described by DE in Lab color space. This space, defined by the 1976 CIE report, is calculated using the Tristimulus values X, Y, Z, defined in the CIE 1931 report. Further, a complex image is often evaluated by averaging the individual ΔEs to calculate a Color Metric for the color difference between two images.

The experiments in this paper generate triplets of images: one is defined as "Original" the other two as "Reproductions." Each area in the "Reproduction" differs from the "Original" by a constant ΔE. The goal of the experiments is to see if some choices of colors make better reproductions than others.

The results show that color metrics comparing color differences across edges within the same image predict better reproductions than color metrics comparing absolute Lab values of corresponding areas in different images.

Introduction

CIE Colorimetry[1] provides easy to use equations for wavelength matches, derived from the properties of the retinal light receptors.[2] However, these equations do not predict appearance representing the sensation image after spatial interactions in the human visual system. Physical models of the image at the retina do not predict appearance in the brain's visual cortex. The question we are asking is whether the average ΔE metric,[2] which we all use for color differences, is appropriate for the color appearances.

Four-Area Experiments

The following series of experiments creates triplets of targets: "Original", "Reproduction T" and "Reproduction B". The colors in the "Reproductions" were selected to be close to, but different from, those in the "Original". For every pixel, the ΔEs between the one reproduction and the "Original" was always equal to the ΔE between the other "Reproduction" and the "original". If ΔE is an isotropic Color Metric for color appearance, observers will randomly select each "Reproductions" as the better copy. In other words, ΔE measures the color difference between the "Original" pixel and the Reproduction pixel at each pixel, independent of all the other pixels. If such a pixel-independent of all other pixels strategy works for color appearance, then each "Reproduction" must be as good as the other. If spatial parameters, namely the relationships between different areas within the test target are important, then observers will select the reproduction that best preserves the spatial relationship.

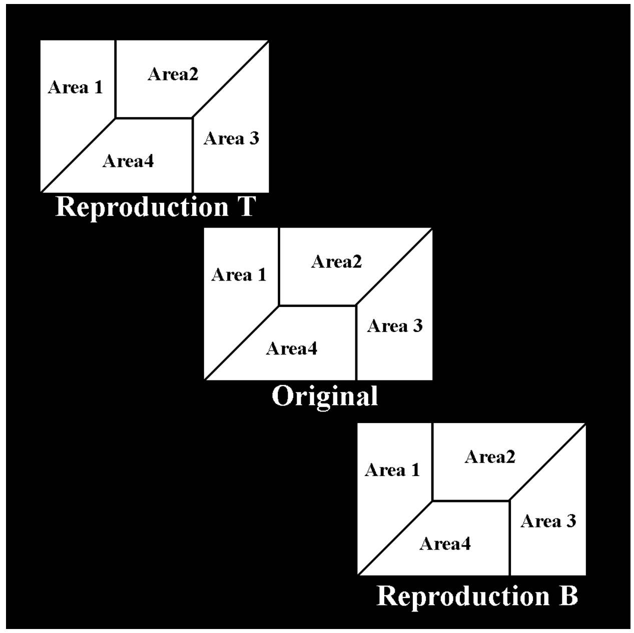


Figure 1. The trio of four-area displays used in these experiments. The central display is described as the "Original". The top(T) and bottom(B) are described as "Reproductions". The observers' task is to select the preferred'"Reproduction".

Observers were shown a trio of similar displays in the same field of view. These experiments used a single media to eliminate properties of materials and calibration variables. These experiments, on identical media were aimed at observer preference of Color Metric without confounding the problem with difference in media. Initially, all experiments were performed on computer displays. Later, these experiments were repeated using print materials, with the same results.

The experimenter asked ten naive observers to select the better "Reproduction" of the "Original" in the middle of the display. The "Reproduction T" showed all four areas as +10 ΔL, 0 Δa, 0 Δb, for a combined ΔE of 10. That is, all areas were 10 L units lighter than the corresponding areas in the "Original". The "Reproduction B" showed each of the four areas as different direction color shift. Each had a ΔE = 10. If the ΔE metric represents the way humans generate color sensations, the two "Reproductions" with identical ΔE errors should represent the original equally well. Since this is a forced choice experiment, we would expect that half of the observers would select the darker image and half would select image with the random shifts.

All the observers selected the image that maintained the ratios across edges, as the better reproduction. The fact that both displays had constant ΔE values was not apparent to the observers. Using this experimental design we looked at a number of variations. Rather than lower L*, a variation increased all areas in the +a* direction, another in the -b* direction, a third in the -L*, -a*,+b* direction. In each case the observer preferred the systematic, constant ratio reproductions. These experiments support the other experiments pointing to the observation that color appearance is a spatial calculation in humans. Obviously, the response of the different spectral sensitivity receptors is very important, but falls far short of explaining the entire color vision calculation. ΔE correlates well with quanta catch of retinal receptors, but cannot be used to evaluate color appearances later in the visual system. Minimizing ΔE searches for best matches independent of spatial information; humans use spatial information to calculate contrast. A different, new metric that uses contrast is needed to calculate best image appearance.

Observers prefer reproductions that are sensitive to the spatial-color relationships within the image. These reproductions are preferred to images made by processes that just minimize average ΔE.

Seventeen-Area Experiments

The second series of experiments uses another triplet of targets: Original, Copy A and Copy B. The colors in the Copy A were selected to be significantly different from, those in the Original. Each area in Copy A is 10 units lighter, 10 units redder and 10 units bluer in Lab space. The combined distance is ΔE = 18. For each area the color difference was ΔE = 18 between the Original and Copy A. Copy B was made with each area ΔE = 18 compared to the original, but it was designed have the color shifts go in many different directions. Copy B significantly changed the local relationships, while Copy A preserved them. If spatial parameters, namely the relationship between different areas within the test target is important, then observers will select Copy A, which preserves the spatial relationship. Observers were shown a trio of similar displays in the same field of view. Figure 2 shows an Original and two Copies. An approach to solving the problem is to use the information learned from the comparison of Copy A with Copy B. The human eye cares more about the relationships of the parts of the image than it does about the absolute value of the match. Relative colorimetry compares the information from a single pixel to the ratio of media white to illuminant; that is information that cannot be derived from the image itself. The media white and the illuminant have to be independently measured. In this experiment the illuminants are the same; relative colorimetry does not differentiate Copy A from Copy B. Spatial comparison within each image are required to find a discriminating Metric.

A Spatial-Color Metric

We can propose a Color Metric more like human vision by comparing X from one area with X' of an adjacent area in the same image (XCopy A/X'Copy A). We can calculate this result with the corresponding data for the Original image (XOriginal/X'Original). We can compare the two with a ratio. [(XCopyA/X'CopyA) / (XOriginal/X'Original)]. The results X, Y, and Z for the 32 edges are plotted in Figures 3a and 3b.

Although it is always dangerous to average numbers into a single figure of merit, we can do so with the Ratio Metric illustrated in Figure 4. ΔR is the square root of the sum of the squares of the (1.0 minus the ratios across edges in the display). The individual ratios for X, Y, Z are plotted in Figure 3. If, instead of Copy A and Copy B we used a perfect copy, all the ratios would be 1.0. ΔR is uses (1.0-ratios), so with a perfect copy its value is ΔR = 0.0. This would mean the closer to 0.0 the value of ΔR, the better the reproduction. The values of DR for Copy A and copy B are listed in Table 1.

The average ΔR for Copy A is 0.17: the average ΔR for Copy B is 1.05, or more than six times greater. The

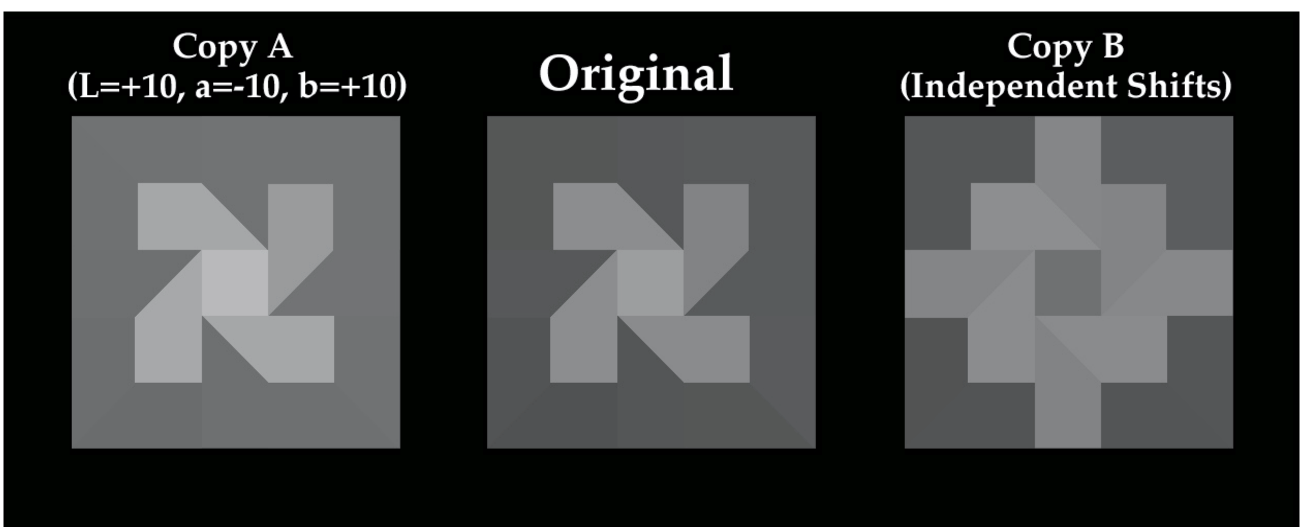


Figure 2. Copy A was made so that each area in Copy A was lighter (ΔL=+10) and redder (Δa=-10) and bluer (Δb=+10) than the Original. The first important observation is that Copy A is a fairly good reproduction, considering that it has a ΔE =18 for each area. Copy B was made so that the ΔE=18 were in many different directions. In this case the ΔE s were chosen to change the appearance of the display. The outer corner patches moved closer in color to each other. The other, mid-side patches moved closer in color to the inner areas. The net effect is that Copy B does not look like the Original. It looks like a different display. Nevertheless, if judged by the DE Color Metric, Copy B is exactly as good a reproduction as Copy A.

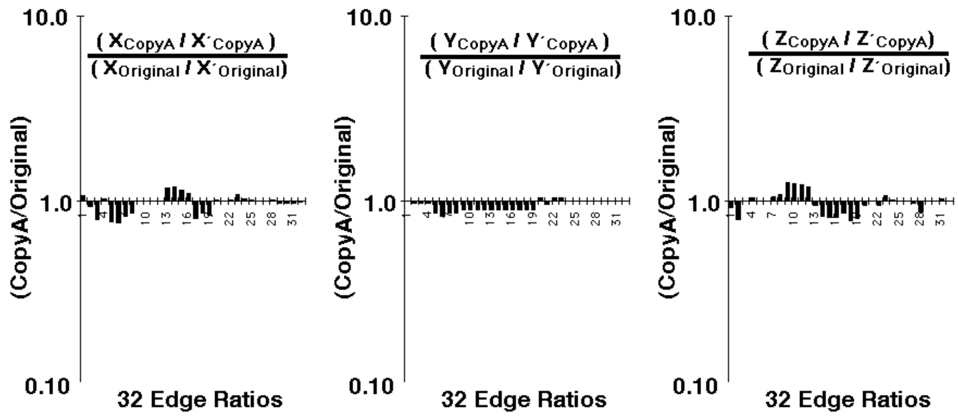


Figure 3a. The three graphs demonstrate a Color Metric for Color Appearance. Results for all the comparisons of X/X' are shown in the left graph, Y/Y' in the middle graph, and Z/Z'' in the right graph). The edge ratios from the comparison of (Copy A / Original) are very close to 1.0, indicating that the two displays are relatively the same. This similarity shows a good reproduction.

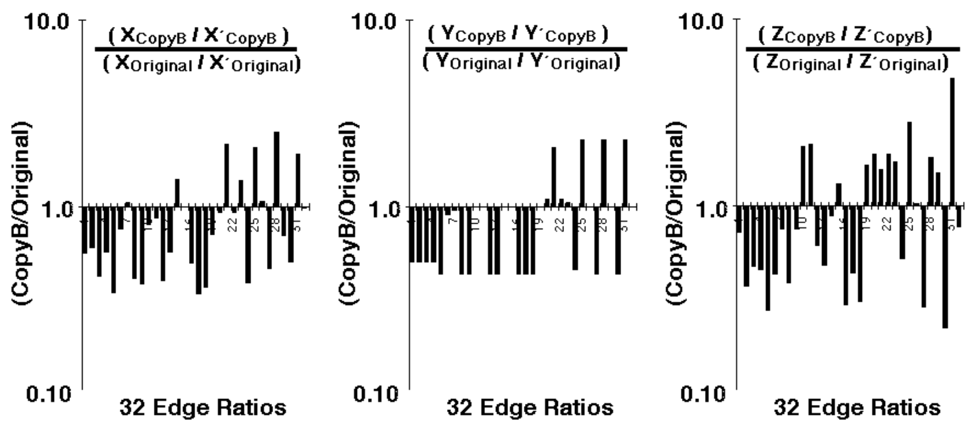
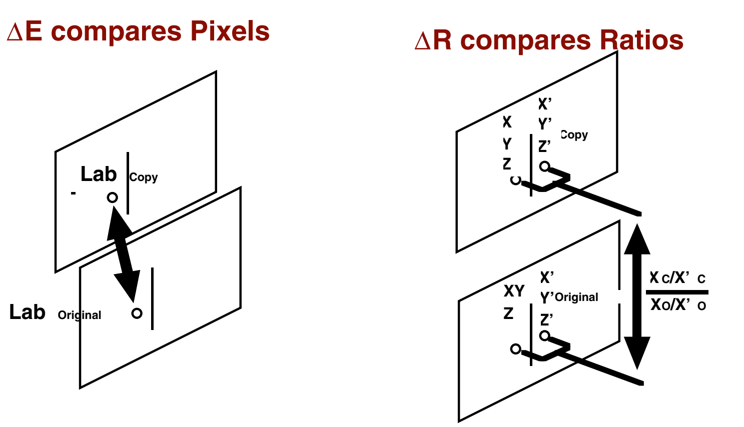


Figure 3b. Results for all the comparisons of X/X' are shown in the left graph, Y/Y' in the middle graph, and Z/Z' in the right graph) The edge ratios from the comparison of Copy B / Original are very far from 1.0, indicating that the two displays are relatively very different. This lack of edge ratio similarity shows a bad reproduction.



$$\Delta E=\sqrt{(L_c-L_o)^2+(a_c-a_o)^2+(b_c-b_o)^2}$$

$$\Delta R=\sqrt{\left(1-\frac{X_c/X'_c}{X_o/X'_o}\right)^2+\left(1-\frac{Y_c/Y'_c}{Y_o/Y'_o}\right)^2+\left(1-\frac{Z_c/Z'_c}{Z_o/Z'_o}\right)^2}$$

Figure 4. The left side illustrates how ΔE is the comparison of single pixels in different images. The equation for ΔE is at the bottom of the figure. The right side illustrates how ΔR is the comparison of ratio of values from differernt pixels in a single image. The equation for ΔR is at the bottom of the figure.

Table 1. List of ratios (Copy/Original) for all 32 edge ratios within the Copy A (left) and Copy B (right).

		CopyA vs. Original				CopyB vs. Original			
		X/X'	Y/Y'	Z/Z'	delta R	X/X'	Y/Y'	Z/Z'	delta R
1	Ratio 1/2	1.07	1.01	0.91	**0.11**	0.56	0.50	0.72	**0.72**
2	Ratio 1/3	0.93	0.97	0.79	**0.23**	0.60	0.50	0.37	**0.90**
3	Ratio 1/4	0.79	0.97	1.00	**0.21**	0.42	0.50	0.47	**0.93**
4	Ratio 1/5	1.03	0.97	1.05	**0.07**	0.57	0.50	0.45	**0.86**
5	Ratio 2/6	0.76	0.85	0.99	**0.28**	0.35	0.43	0.27	**1.13**
6	Ratio 2/7	0.76	0.82	0.99	**0.30**	0.76	0.90	0.43	**0.63**
7	Ratio 2/8	0.83	0.85	1.07	**0.24**	1.06	0.95	0.74	**0.27**
8	Ratio 2/9	0.85	0.85	1.09	**0.22**	0.41	0.43	0.38	**1.02**
9	Ratio 3/9	0.99	0.89	1.27	**0.29**	0.39	0.43	0.74	**0.87**
10	Ratio 3/10	1.00	0.89	1.25	**0.28**	0.80	1.00	2.07	**1.09**
11	Ratio 3/11	1.01	0.89	1.24	**0.26**	0.87	1.00	2.15	**1.15**
12	Ratio 3/12	1.00	0.89	1.20	**0.23**	0.40	0.43	0.61	**0.91**
13	Ratio 4/12	1.18	0.89	0.94	**0.22**	0.57	0.43	0.48	**0.88**
14	Ratio 4/13	1.20	0.89	0.82	**0.29**	1.42	1.00	0.87	**0.44**
15	Ratio 4/14	1.16	0.89	0.81	**0.27**	0.99	1.00	1.32	**0.32**
16	Ratio 4/15	1.11	0.89	0.81	**0.24**	0.50	0.43	0.29	**1.04**
17	Ratio 5/6	0.80	0.89	0.86	**0.27**	0.34	0.43	0.43	**1.04**
18	Ratio 5/15	0.85	0.89	0.77	**0.30**	0.37	0.43	0.31	**1.10**
19	Ratio 5/16	0.83	0.89	0.80	**0.29**	0.71	1.00	1.67	**0.73**
20	Ratio 5/17	1.02	1.05	0.93	**0.08**	0.93	1.11	1.89	**0.90**
21	Ratio 6/7	0.99	0.96	1.00	**0.04**	2.18	2.08	1.58	**1.70**
22	Ratio 6/17	1.02	1.05	0.93	**0.08**	0.93	1.11	1.89	**0.90**
23	Ratio 7/8	1.09	1.05	1.08	**0.12**	1.39	1.05	1.73	**0.83**
24	Ratio 8/9	1.04	1.00	1.02	**0.04**	0.39	0.46	0.52	**0.95**
25	Ratio 9/10	1.01	1.00	0.99	**0.02**	2.08	2.31	2.79	**2.47**
26	Ratio 10/11	1.01	1.00	0.99	**0.02**	1.08	1.00	1.03	**0.09**
27	Ratio 11/12	0.99	1.00	0.97	**0.03**	0.46	0.43	0.29	**1.06**
28	Ratio 12/13	1.02	1.00	0.87	**0.13**	2.50	2.31	1.82	**2.16**
29	Ratio 13/14	0.96	1.00	0.99	**0.04**	0.70	1.00	1.51	**0.59**
30	Ratio 14/15	0.96	1.00	1.00	**0.04**	0.50	0.43	0.22	**1.08**
31	Ratio 15/16	0.97	1.00	1.03	**0.04**	1.92	2.31	5.45	**4.73**
32	Ratio 16/17	0.98	1.00	1.00	**0.02**	0.97	1.00	0.77	**0.23**
				Average	**0.17**				**1.05**
				Standard Deviation	**0.11**				**0.83**

standard deviation of the mean of ΔR for Copy A is 0.11: for Copy B is 0.83. The ΔR values for Copies A and B are significantly different. Copy A has a much better score than Copy B. Copy A is a better reproduction than Copy B.

Obviously this is an oversimplified test case, since the displays were designed to have the same ΔEs. A proper spatially designed Color Metric must consider the color space used for the calculation, the degree that this space is isotropic with regard to color appearance, and the mechanism that is used to make spatial calculations and overall color shifts. Each question requires deliberate study to make a suitable Spatial-Color Metric. The results shown in this paper support the idea that such a Spatial-Color Metric has considerable promise.

Summary

This poster displays some Four-Area Color Displays and some Seventeen-Area Displays. Both sets illustrate the importance of making comparisons within images before make comparisons between images. Such an analysis of Copy A and Copy B shows that we can calculate significantly different color metric values from the displays. When the ratio of edge ratios were nearly 1.0, as in Copy A, it is a fairly good reproduction, despite the ΔE=18 difference from the "Original". When the ratio of edge ratios were much greater than 1.0, as in Copy B, the appearance of the display changed. The comparison of Copy A and "Original" shows an additional important point. Although Copy A has the same edge ratios as the Original, it does not match the Original. Edge ratios are an important tool for successful metric, but not a complete metric by itself. Human vision normalization depends on both spatial and absolute quanta catch information.

Acknowledgments

The author wishes to thank Norbert Herzer for his collaboration in designing, creating and evaluating the targets used in these experiments.

References

1. CIE Proceedings 1931, p. 19, Cambridge University Press, Cambridge, 1932.
2. V. C. Smith and J. Pokorney, "Spectral Sensitivity of the Foveal Cone Photopigments between 400 and 500 nm," *Vision Res.*, **15**, p. 161, 175.
3. CIE "Recommendations on Uniform Color Spaces, Color Difference Equations, Psychometric Color Terms," Supplement No. 2 of *CIE Publ.* No. **15** (E-1.3.1) 1971, Bureau Central de la CIE, Paris, 1978.

published previously in the IS&T 1996 Color Imaging Conference Proceedings, page 155