Evaluation of Five Color-Appearance Transforms Across Changes in Viewing Conditions and Media

Karen M. Braun and Mark D. Fairchild
Munsell Color Science Lab., Rochester Institute of Technology, Rochester, New York

Abstract

Observers performed psychophysical experiments to compare color-appearance transformations under a variety of viewing conditions, using a memory matching technique. These viewing conditions differed between the original printed images and the CRT reproductions in white point chromaticity, background relative luminance, surround relative luminance, and absolute adapting luminance.

Introduction

CIELAB is often used to produce acceptable color appearance matches across small differences in white point between the original and reproduction.[1] However, color appearance of a patch, area, or image element also depends on adjacent colors (simultaneous contrast), overall luminance level, and surrounding environment, as well as level of adaptation to the light source. Color-appearance models, such as RLAB,[2,3] and models proposed by Hunt[4,5] and Nayatani,[6] attempt to produce matches when these viewing conditions vary between original and reproduction. This research compared these color-appearance models to von Kries adaptation[7] and CIELAB color space[1] transformations. The goal of this research was determine which of these transformations most accurately predicts color matches for many viewing condition changes. Printed images were used as the originals and the color-appearance models were used to produce reproductions on a CRT. Observers performed psychophysical experiments to compare the accuracy of the five models in producing matching color images under a variety of viewing conditions.

Generation of Images

Five digital color images containing pictorial information were used in this study. These images included a thin white border that was adjusted as part of the image content. The original images were printed on the Fujix Pictrography 3000 continuous tone digital printer at 200 dpi. These prints, the original reference images, were mounted on 8" × 10" spectrally non-selective 18% gray cards. The printed images were digitized before mounting using a Howtek D4000 drum scanner at the resolution of the CRT, 99 dpi, to provide *RGB* data for preparing the CRT reproductions. The scanner was colorimetrically characterized before producing the CRT reproductions, so that scanner *RGB* tristimulus values could be accurately converted to CIE *XYZ* tristimulus values for spectral power distributions used in the light booth. The scanner calibration technique was developed by Shyu and Berns.[8]

Five color-appearance models were used to predict matching images for the D65-balanced CRT using the CIE *XYZ* tristimulus values of the print originals. The models used were von Kries adaptation,[7] CIELAB color space,[1] RLAB color-appearance model,[2,3] Hunt's color-appearance model,[4,5] and Nayatani's color-appearance model.[6] To avoid the use of gamut-mapping procedures in producing the CRT reproductions, *RGB* digital counts of the original image data (before printing) were compressed such that all image colors predicted by the appearance transforms remained within the gamut of the CRT.

Viewing Conditions

The capability of the five color-appearance models to accurately predict hardcopy/softcopy matches was investigated under various viewing conditions, as listed in Table 1.

Table 1. Viewing Conditions for Psychophysical Testing of Color-appearance Models.

Print White Point	CRT White Point	Background (Print/CRT)	Surround (Print/CRT)	Luminance Level
D65	D65	Gray/Gray	Dark/Dark	Equal
D50	9300K	Gray/Gray	Dark/Dark	Equal
D65	9300K	Gray/Gray	Dark/Dark	Equal
D50	D65	Gray/Gray	Dark/Dark	Equal
A	D65	Gray/Gray	Dark/Dark	Equal
D65	D65	Gray/Gray	Dark/Dark	Print Higher
D50	9300K	Gray/Gray	Dark/Dark	Print Higher
D50	D65	Gray/Gray	Dark/Dark	Print Higher
D65	D65	Gray/White	Dark/Dark	Equal
D65	D65	Gray/Black	Dark/Dark	Equal
D65	D65	Gray/Gray	Light/Dark	Equal
D65	D65	Gray/Gray	Dim/Dark	Equal

To assure that viewing conditions were properly set up and that the models were working correctly, reproductions were calculated and displayed under viewing conditions identical to the originals (differing only in media.) This is the first experiment shown in Table 1. The remaining experiments are grouped as follows: changes in white point chromaticity, changes in adapting luminance level and white point chromaticity, changes in background, and changes in surround.

Printed original images were illuminated and viewed in a Macbeth SpectraLite II light booth. Light sources under which the original images were viewed simulated CIE standard illuminants D65, D50 and A. To vary the luminance level in the light booth, neutral tint screens of appropriate transmittance attenuated the sources. Sources were measured with these screens in place. Reproductions were displayed on a Sony GDM-1950 CRT. The CRT white point approximated the white point chromaticity coordinates of CIE standard illuminant D65 and 9300K. All experiments were conducted in a completely darkened room, so that only the print or CRT image occupied the observers' field of view. Observers sat far enough from the images to assure that they could not resolve the 99 dots per inch of the CRT. The images subtended an angle of approximately 13° in the observers' field of view (as measured across the diagonal of the 6" × 8" image.)

Psychophysics

In the psychophysical experiment, observers compared CRT reproductions to print originals, for each of the viewing conditions under investigation. A memory viewing technique was employed based on the results of a previous study[9] that showed that the results obtained using this technique were most consistent with practical viewing environments. The print original and the CRT reproduction were placed at ninety degrees from each other with respect to observers. Observers adapted to an 18% gray card in the booth for 60 seconds.[10] Then observers examined a print original in the light booth. When they were confident that they could remember the colors in the image, the original was covered and they turned their attention to the CRT. After adapting to a field of 20% maximum luminance at the white point of the CRT, observers compared pairs of CRT reproductions and decided which of the pair looked most like the original in color content. Only one of the two reproductions was displayed at a time and observers toggled between the two reproductions. Observers compared each reproduction derived using one of the five models to every other reproduction, which resulted in ten paired comparisons. The CRT reproduction shown first in each pair was randomized as was the order in which the ten pairs were presented. Observers made decisions for the ten pairs of reproductions of the first print original then repeated the entire procedure for the remaining four images.

Data Analysis

Using Thurstone's Law of Comparative Judgments,[11] the choices of CRT reproduction were converted to an interval scale, which indicated the model performance in producing a visually matching CRT reproduction of the print original. This technique is explained in detail by Torgerson.[12] Confidence intervals on the scale values were calculated in terms of the scale units. One unit on the interval scale equals $\sqrt{2}\,\sigma$. Therefore, the standard deviation, σ, of a given value is $1/\sqrt{2}$, or 0.707 units. A 95% confidence interval is calculated in Eq. 2.

$$R \pm 1.96\ \frac{\sigma}{\sqrt{N}} = R \pm 1.96\ \frac{0.707}{\sqrt{N}} = R \pm \frac{139}{\sqrt{N}} \qquad (1)$$

N is the number of observations for a sample. Two models were equivalent if the interval scale values were within ($1.39/\sqrt{N}$) units of each other. When image results were averaged, the confidence interval in Eq. (1) was divided by the number of images.

Acceptability of Model Reproductions

A difficulty with the paired-comparison technique is that it does not reveal whether the models produce *good* color-appearance matches. Therefore, after observers completed the paired comparison experiment, they were shown each of the model reproductions individually and asked to rate the quality of the reproduction on a category scale. The categories included: excellent, good, just acceptable, just unacceptable, poor, and terrible match.

Results

White Point Changes

Fifteen observers performed this experiment as part of a previous study on viewing techniques. The images used for this section are described in a paper on that study.[9] Figure 1 shows results when the original print was viewed with a source approximating the white point chromaticity approximating either CIE standard illuminant A or D50 and the reproduction was displayed on a CRT with white point chromaticity approximating illuminant D65. For the averaged results, the confidence interval in Equation 1 was divided by $\sqrt{5}$, giving a confidence interval equal to 0.160.

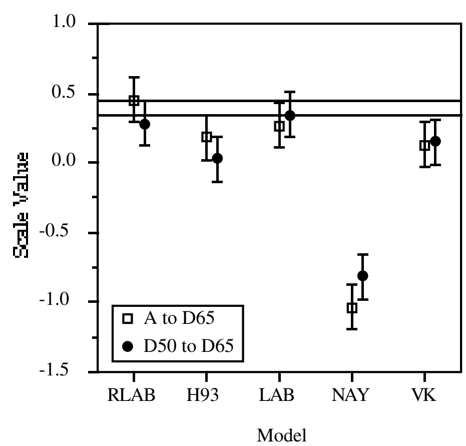

Figure 1. Interval scale for white-point experiments. Error bars equal 0.16.

RLAB performed best for both conditions on average, and for booth D50 → CRT D65, CIELAB performed statistically as well as RLAB. RLAB and CIELAB showed good performance for many of the individual images under both conditions. CIELAB showed outstanding performance for a hybrid image containing patches and pictorial information, under both sets of conditions. Perhaps CIELAB reproduces the specific colors in this image better than other models, or perhaps there is a fundamental difference between matching colored patches and matching complex images.

The von Kries and Hunt models performed reasonably well, predicting best matches for about half of the images. Hunt's model appeared to perform better for graphics than for pictorial images, perhaps because it was developed using simple patch experiments. Nayatani's model had poor performance for all five images under both sets of viewing conditions, possibly because of the strong Helson-Judd effect predicted by this model.

Background Changes

Twenty observers performed a paired-comparison experiment for changes in background from original print to CRT reproduction. Both the light booth and the CRT had white point chromaticities approximating D65. Because some of the frequency-of-seeing matrices for this pair of experiments contained elements of zero and one, Torgerson's method for incomplete matrices was used to form the scale.[12] For one image in each experiment, none of the observers chose Hunt's reproduction in any pair. For these cases, it was not possible to use Torgerson's method so these images were removed from the analysis. Figure 2 shows the interval scale data with their 95% confidence intervals for the background experiments, where the print original was surrounded by a one inch gray border and the CRT reproductions by a one-inch white or black border. Since four images were averaged, the confidence interval in Equation 1 was divided by $\sqrt{4}$ and therefore equals 0.155.

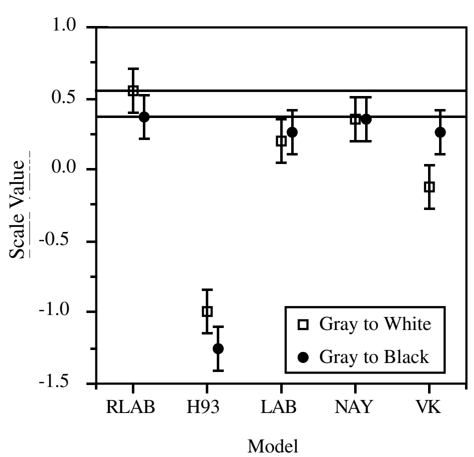

Figure 2. Interval scale for background experiments. Confidence intervals equal 0.155

For the experiment where the CRT reproduction was surrounded by a white border, RLAB gave the best reproductions for three of four images. Nayatani gave the best reproduction for three images, CIELAB for one image, and von Kries and Hunt never gave the best reproductions. For the average of the five images, RLAB produced the most accurate color-appearance reproductions. It is unclear why the von Kries model performs significantly worse than RLAB since the two models were quite similar under these very similar viewing conditions. For the experiment where the CRT reproduction was surrounded by a black border, Nayatani produced the best reproductions for all four images, and CIELAB, von Kries, and RLAB gave best results for three images. Again, Hunt's model never produced reproductions that were considered the most accurate match to the original. On average, RLAB, CIELAB, Nayatani's model, and von Kries's model were not considered significantly different in their ability to produce accurately matching reproductions to the printed original.

The results in Figure 2 show that Hunt's model overpredicted the effect of background on the images. The visual effect was to produce an image too high in lightness, saturation, and contrast for the white background, and too low in lightness, saturation, and contrast for the black background. The application of Hunt's model in this experiment was not strictly justified. This background parameter was designed using patches of approximately 2° angular subtense on the retina with a border extending to 10° angular subtense.[4] The images in this experiment subtended an angle of 13° with a background subtending 16°. The effect of background on color appearance may be decreased for images and is certainly decreased as the area of the background decreases and its distance from the center of the image increases. It was clear from viewing the images that colors near the border *were* affected by background, appearing slightly too dark in the other model reproductions when viewed with a white background and appearing too light with the black background. Perhaps a smaller value for the background parameter would have improved the prediction ability of Hunt's model in this experiment.

For both background experiments, observers judged the reproductions predicted by Hunt's model as "just unacceptable" and all other model reproductions as "good". Although this category scale was not as sensitive to small differences between the reproductions, it generally agrees with the results found using the paired comparison technique.

Other Viewing Conditions

Results of other experiments described in Table 1 were not available at the time of printing and may be picked up at the site of this poster presentation.

Conclusions

The results shown here demonstrate that simple models of color appearance (von Kries, CIELAB, and RLAB) give results as accurate or more accurate than more complicated models (Hunt '94 and Nayatani) over a wide range of viewing conditions. The visual experiments were performed on images containing a large sampling of the color space. Although sometimes the more complicated models did not yield as favorable results, they attempt to model the physiology of vision and may, in the future, lead to a greater understanding of the visual system.

Acknowledgment

This research was supported by the Eastman Kodak Company and the Munsell Color Science Laboratory. Additional funding was provided through the NSF-NYS/IUCRC and NYSSTF-CAT Center for Electronic Imaging Systems. The time and patience of the observers who performed the psychophysical experiments are greatly appreciated.

References

1. *Colorimetry,* 2nd ed., CIE Publ. No. **15**.2, Central Bureau of the CIE, Paris, 1986.
2. M. D. Fairchild, "Formulation and testing of an incomplete-chromatic adaptation model", *Color Res. Appl.* **16**, 243-250 (1991).
3. M. D. Fairchild and R. S. Berns, "Image color appearance specification through extension of CIELAB", *Color Res. Appl.* **18**, 178-190 (1993).
4. R. W. G. Hunt, "Revised colour-appearance model for related and unrelated colours", *Color Res. Appl.* **16**, 146-165 (1991).
5. R. W. G. Hunt, "An improved predictor of colourfulness in a model of colour vision", *Color Res. Appl.* **19**, 23-26 (1994).
6. Y. Nayatani, K. Takahama, and H. Sobagaki, "Color-appearance model and chromatic-adaptation transform", *Color Res. Appl.* **15**, 210-221 (1990).
7. J. von Kries, "Chromatic adaptation", *Festschrift der Albrecht-Ludwig-Universität*, (1902).
8. M. J. Shyu, *Colorimetric Characterization of a Desktop Drum Scanner Using a Spectral Model*, Rochester Institute of Technology, M.S. Thesis, July 1994.
9. K. M. Braun, M. D. Fairchild, P. J. Alessi, "Viewing tech niques for cross-media image comparisons", *Color Res. Appl.* **21,** 6–17 (1996).
10. M. D. Fairchild and L. Reniff, "Time-course of chromatic adaptation for color-appearance judgments", *J. Opt. Soc. Am. A* **12**, 824-833 (1995).
11. L. L. Thurstone, "A law of comparative judgment", *Psych. Rev.* **34**, 273-286 (1927).
12. W. S. Torgerson, Chapter 9: "The Law of Comparative Judgment", *Theory and Methods of Scaling,* Wiley, New York, 1967.

published previously in the IS&T 1995 Color Imaging Conference Proceedings, page 93