Comparing Appearance Models Using Pictorial Images

Taek Gyu Kim, Roy S. Berns, and Mark D. Fairchild
Munsell Color Science Laboratory, Center for Imaging Science
Rochester Institute of Technology, Rochester, New York

Eight different color appearance models were tested using pictorial images. A psychophysical paired comparison experiment was performed where 30 color-normal observers judged reference and test images via successive-Ganzfeld haploscopic viewing such that each eye maintained constant chromatic adaptation and interocular interactions were minimized. It was found that models based on von Kries had best performance, specifically CIELAB, HUNT, RLAB, and von Kries.

Introduction

Color appearance models are necessary to incorporate into the color WYSIWYG chain when images are viewed under dissimilar conditions such as illumination spectral power distribution and luminance, surround relative luminance, and media type where cognition is affected. These differing conditions often occur when comparing CRT and printed images, CRT and projected slides, or rear-illuminated transparencies and CRT or printed images.

A psychophysical experiment was performed to test a variety of color-appearance models described in the literature. Some of these models were developed for only object colors while some were developed for use in many modalities. In practice, different devices have different spatial (resolution and image microstructure) and colorimetric (gamut) properties. It was appropriate, therefore, to first test these appearance models such that these differences were eliminated. This was accomplished by using a single device, a continuous-tone dye-diffusion thermal-transfer printer. Future experiments will add the complexity of comparing different imaging modalities.

Testing color-appearance models involves generating corresponding colors (in this case corresponding images) under a test and reference set of conditions. An appearance model will predict the tristimulus values for a pair of stimuli such that when each is viewed in its respective illuminating and viewing conditions, the stimuli will match in appearance for a CIE standard observer. By colorimetrically characterizing the printer for both conditions, the requisite samples can be generated. The following models were tested: von Kries, CIELAB, CIELUV, LABHNU (Richter), Reilly-Tannenbaum (DuPont), Hunt, Nayatani, and RLAB (Fairchild-Berns).

Appearance-Model Overview

von Kries

$$\begin{aligned} L' &= k_L \cdot L \\ M' &= k_M \cdot M \\ S' &= k_S \cdot S \end{aligned} \tag{1}$$

where L, M, and S represent the excitations of the long-, middle-, and short-wavelength sensitive cones, L', M', and S' represent the post-adaptation cone signals, and k_L, k_M, and k_S are the multiplicative factors, generally taken to be the inverse of the respective maximum cone excitations for the illuminating condition.[3,4] The calculation of the cone fundamentals is a linear transformation of CIE tristimulus values. In this case the Stiles-Estevez-Hunt-Pointer fundamentals were used.[2,9,17] (These are also used in the Hunt, Nayatani, and RLAB models.)

CIELAB

The CIELAB space was recommended by the CIE in 1976 for use as a color-difference metric.[1] While CIELAB was developed to describe color differences, it also incorporates fundamental metrics of color appearance through the cylindrical specification of lightness (L^*), chroma (C^*_{ab}), and hue angle (h_{ab}) and the inclusion of a modified form of the von Kries model of chromatic adaptation (X/X_n, Y/Y_n, Z/Z_n).

CIELUV

The CIELUV space was recommended by the CIE in 1976 at the same time as CIELAB.[1] Although it has similar perceptual metrics to CIELAB, it differs significantly in its chromatic adaptation model ($u'-u'_n$, $v'-v'_n$).

LABHNU

The LABHNU space was developed by Richter.[23] It is similar to CIELUV in that it has an embedded chromaticity diagram and translational chromatic adaptation model:

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16 \tag{2}$$

$$A^* = 500\left(A' - A'_n\right)Y^{\frac{1}{3}} \tag{3}$$

$$B^* = 500\left(B' - B'_n\right)Y^{\frac{1}{3}} \tag{4}$$

where

$$A' = \frac{1}{4}\left(\frac{x}{y} + \frac{1}{6}\right)^{\frac{1}{3}} \qquad RG\text{-chroma} \tag{5}$$

$$B' = \frac{-1}{12}\left(\frac{z}{y} + \frac{1}{6}\right)^{\frac{1}{3}} \qquad JB\text{-chroma} \tag{6}$$

Reilly-Tannenbaum

The Reilly-Tannenbaum model was created at DuPont during the 1970's as a color difference metric. It has been used as a part of their color matching system for automotive colorant formulation and control. It has features of both CIELAB (opponency and cube root) and CIELUV (translational chromatic adaptation model) and has a transformation from CIE tristimulus values to cone fundamentals optimized from color-difference data. It's worth noting that Reilly was one of the key developers of CIELAB; these equations reflect his influence.

$$L = 25G^{1/3} - 16 \tag{7}$$

$$a = a' - (\frac{Y}{100})^{1/3} a_n \tag{8}$$

$$b = b' - (\frac{Y}{100})^{1/3} b_n \tag{9}$$

where

$$a' = 500[(\frac{R}{100})^{1/3} - (\frac{G}{100})^{1/3}] \tag{10}$$

$$b' = 200[(\frac{G}{100})^{1/3} - (\frac{B}{100})^{1/3}] \tag{11}$$

$$R = R'(\frac{100}{G_n}) \tag{12}$$

$$G = G'(\frac{100}{G_n}) \tag{13}$$

$$B = B'(\frac{100}{G_n}) \tag{14}$$

$$R' = 0.7584X + 0.2980Y - 0.1564Z \tag{15}$$

$$G' = -0.4632X + 1.3677Y + 0.0955Z \tag{16}$$

$$B' = -0.1220X + 0.3605Y + 0.7615Z \tag{17}$$

Hunt

Hunt's model[10-12] is diagrammed in Fig. 1. It incorporates many parameters necessary for cross-media color reproduction. However, it is not invertible and in order to use it for color WYSIWYG, a successive-approximation iterative technique is required.

Hunt Appearance Model

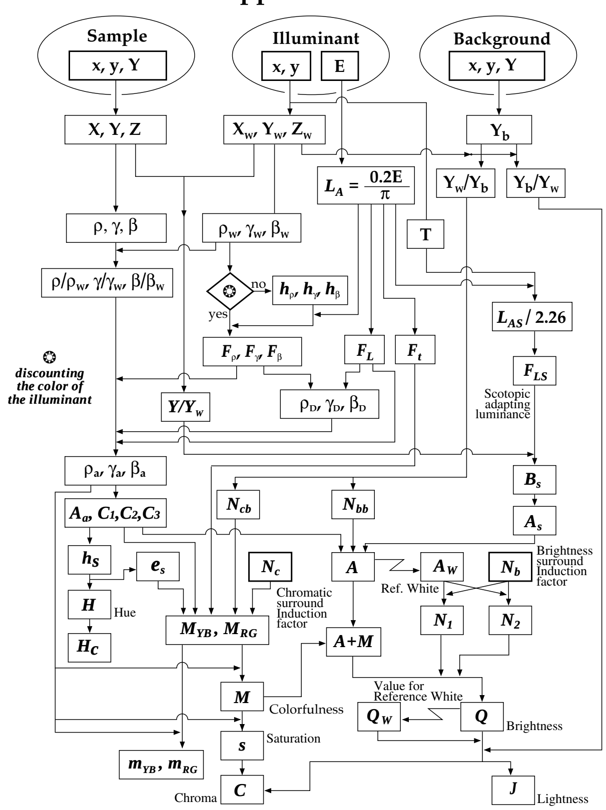

Figure 1. Flow diagram of Hunt appearance model.

Nayatan

Nayatani's color appearance model[13-22,25] is diagrammed in Fig. 2. Although there are many similarities to Hunt's model, the non-linear compression stages are quite different and Nayatani's model is defined only for object colors possibly limiting its use in color WYSIWYG. An advantage of this model over Hunt's model is its relative ease in inversion.

Nayatani Appearance Model

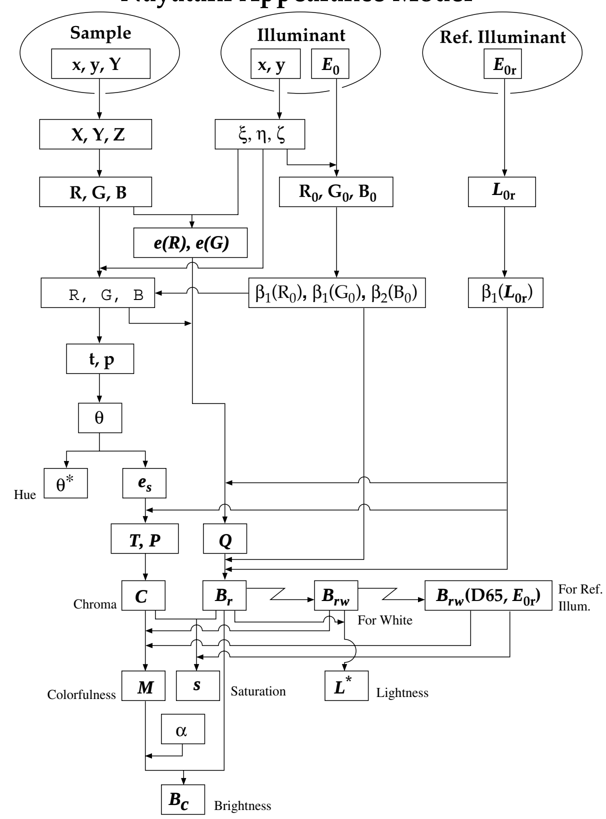

Figure 2. Flow diagram of Nayatani appearance model.

RLAB

The RLAB model developed by Fairchild and Berns[5] can be thought of as a simplification of the Hunt model; it incorporates viewing condition parameters and is mathematically efficient and invertible, all necessary requirements for color WYSIWYG. It is based on Fairchild's model of chromatic adaptation, uses CIELAB for perceptual metrics, and takes in account differences in surround relative luminance.

Experiment

Viewing Booth

A bipartite viewing booth for haploscopic-type viewing was constructed. The interior was painted with an approximately spectrally non-selective gray paint with a luminance factor of 0.2. Diffusing panels were inserted underneath each set of light sources to improve the uniformity of the illumination.

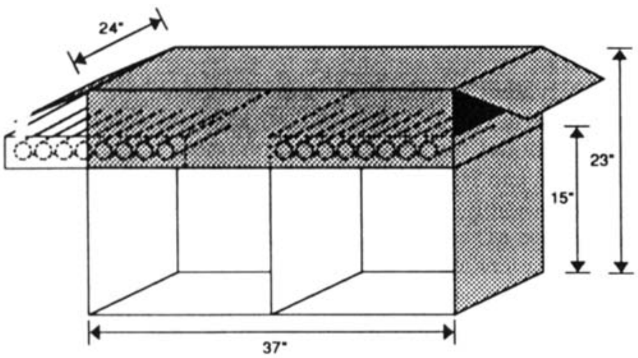

Figure 3. Viewing Booth. The right viewing field is illuminated with simulated illuminated. A and the right field, simulated D65. Switches allow control of each bulb in order to vary luminance.

Illumination

The right side of the booth (reference field) had tungsten bulbs closely simulating CIE illuminant A at 214 cd/m^2. The left side (test field) had high color rendering fluorescent tubes with chromaticities near D65. The daylight test field had three luminance level settings which were equivalent, 1/3 and 3 times the luminance level of the reference illuminant A field (71, 214 and 642 cd/m^2). The each test field setting was named as D65-M, D65-L and D65-H for convenience. The spectral power distributions are shown in Fig. 4.

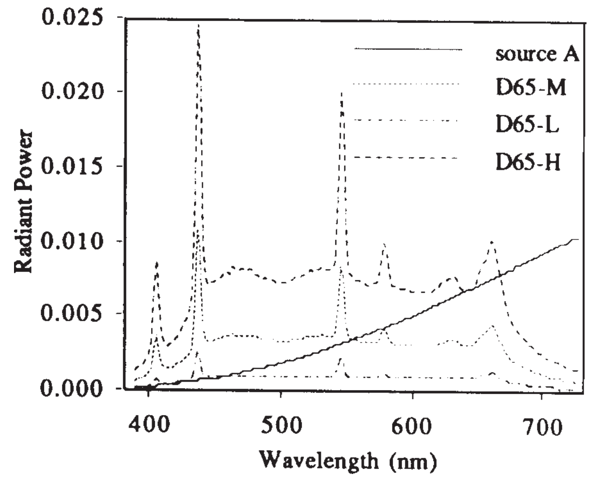

Figure 4. Spectral Distribution of the reference and test field sources.

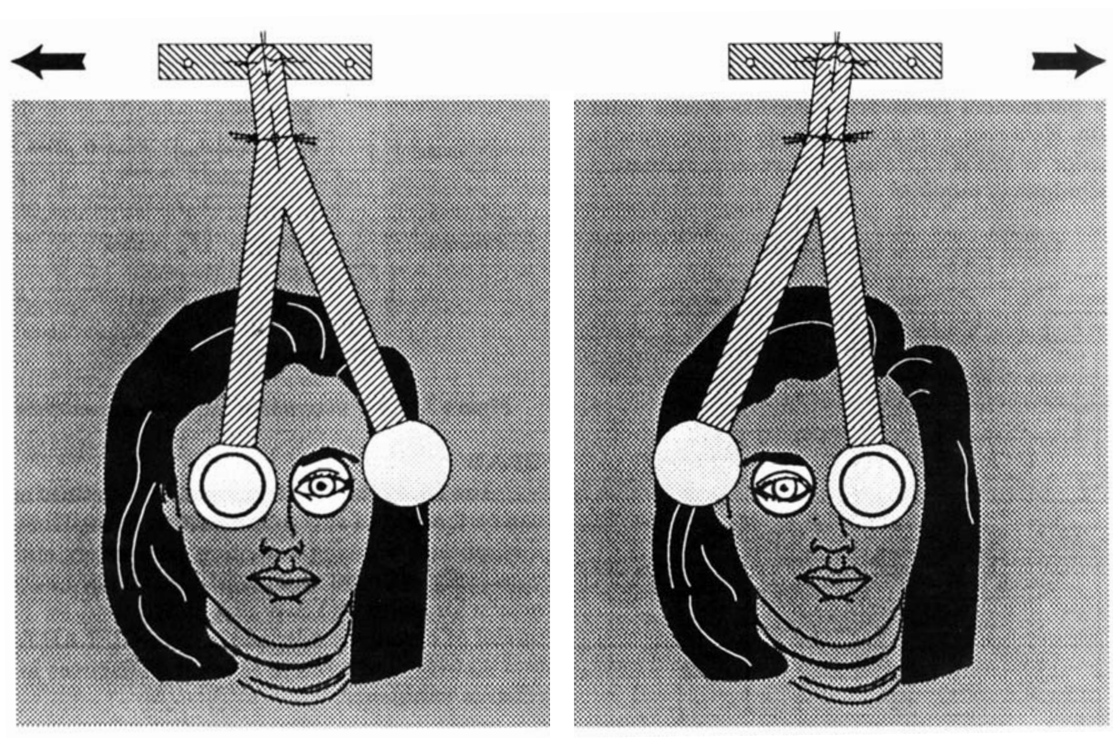

Figure 5. The successive-Ganzfeld haploscopic device will open one of the viewing fields and block the other view with frosted Mylar™. This device will toggle between eyes with a foot switch operation.

Successive-Ganzfeld Haploscopic Device

To achieve successive viewing, a shutter mechanism with diffusers made from frosted Mylar™ was devised as shown in Fig. 5.[6] The observer could control via a foot pedal whether the *Ganzfeld* blocked the test or reference field. The purpose of this alternating viewing was to maintain an eye's state of chromatic adaptation while preventing the simultaneous viewing of images.

Sample Preparation

Four images were acquired from preliminary IT8 standard image sets: "fruit basket," "orchid," "musicians," and "candles." The four-color CMYK 8 bit 2048 by 2560 pixel images were transformed into 3 color CMY 8 bit 1024 by 1280 pixel images using lookup tables based on CIELAB matching of the DuPont 4Cast™ dye-diffusion thermal-transfer printer. The reason for the four to three colors conversion was to avoid technical difficulties of conversion between three channel (XYZ) and CMYK.

When calculating corresponding colors, a significant shift in the color gamut can result for some of the appearance models resulting in many unprintable colors (*e.g.*, LABHNU and CIELUV). Since gamut mapping was excluded from this experiment, the reference images were compressed in CIELAB until all the models predicted corresponding colors that were within the printer's gamut (see Fig. 6). (Because of this gamut reduction, the exclusion of the black printer did not adversely affect the color image quality.) Four sets of the compressed CMY 8 bit images were defined as the original reference images. Visual inspection assured proper color balance and tone rendition.

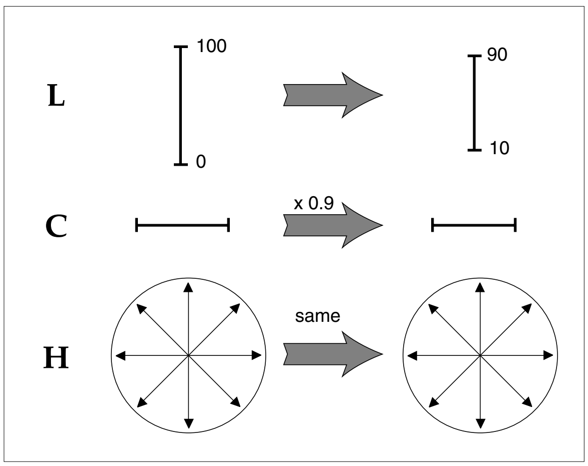

Figure 6. Gamut compression of original images. Pixel information was transformed into LCH and L and C were used to calculate new color coordinates.

To simplify the process of running many images through two successive tetrahedral interpolation lookup tables, CMY to XYZ for the reference condition and XYZ to CMY for each of three test conditions, $CMY_{reference}$ to CMY_{test} lookup tables were built each with 33×33×33 entries based on the method of Hung.[8]

Once corresponding images were computed, image pairs were printed for each pairwise combination of the eight appearance models corresponding to 28 pairs. This was repeated for each test field condition.

Psychophysics

Thirty color-normal observers with varying imaging experience took part in the experiment. Observers were instructed to select one of the images from the test pair that most closely matched the reference image. Three separate observering sessions were used corresponding to the three test field conditions (D65-M, D65-L and D65-H). A total of 336 observations resulted per observer.

Observer data were collected and summed on tally sheets as in TABLE I. Each cell represents preference of one model (column) over another model (row). For example, the column 3 and row 4 cell which has 116 means that CIELAB was preferred 116 times out of 120 over CIELUV.

Table I. Count tally sheet at the same luminance level for the four images combined.

	VK	CLAB	CLUV	LABH	RT	Hunt	Nay	RLAB
VK		67	9	9	25	69	28	61
CLAB	53		4	20	12	47	29	58
CLUV	111	116		112	113	113	101	114
LABH	111	100	8		62	105	66	107
RT	95	108	7	58		103	65	97
Hunt	51	73	7	15	17		27	59
Nay	92	91	19	54	55	93		89
RLAB	59	62	6	13	23	61	31	

Table II. Normalized matrix at the same luminance level for the four images combined.

	VK	CLAB	CLUV	LABH	RT	Hunt	Nay	RLAB
VK		0.56	0.08	0.08	0.21	0.58	0.23	0.51
CLAB	0.44		0.03	0.17	0.10	0.39	0.24	0.48
CLUV	0.93	0.97		0.93	0.94	0.94	0.84	0.95
LABH	0.93	0.83	0.07		0.52	0.88	0.55	0.89
RT	0.79	0.90	0.06	0.48		0.86	0.54	0.81
Hunt	0.43	0.61	0.06	0.13	0.14		0.23	0.49
Nay	0.77	0.76	0.16	0.45	0.46	0.78		0.74
RLAB	0.49	0.52	0.05	0.11	0.19	0.51	0.26	

Each cell was divided by the total number of judgments to calculate proportions (TABLE II). Using Thurstone's law of comparative judgments,[7] Z-scores were calculated (TABLE III). The column sum results in an interval scale where the larger the number, the more accurate a model was in predicting appearance matches. Dif-ferences between models greater than 1.39 are statistically significant at a 95% confidence interval.

TABLE III. Z-score matrix at the same luminance level for the four images combined.

	VK	CLAB	CLUV	LABH	RT	Hunt	Nay	RLAB
VK		0.13	-1.48	-1.48	-0.84	0.18	-0.74	0
CLAB	-0.15		-1.88	-0.99	-1.28	-0.28	-0.71	-0.05
CLUV	1.41	1.75		1.48	1.56	1.56	0.99	1.65
LABH	1.41	0.95	-1.56		0.03	1.13	0.13	1.23
RT	0.81	1.28	-1.65	-0.05		1.04	0.1	0.84
Hunt	-0.2	0.25	-1.65	-1.18	-1.08		-0.77	-0.03
Nay	0.71	0.67	-1.04	-0.13	-0.13	0.74		0.64
RLAB	-0.03	0.03	-1.65	-1.28	-0.88	0	-0.67	
SUM	3.96	5.06	-10.91	-3.63	-2.62	4.37	-1.67	4.28

Results and Discussion

The z scores along with their 95% confidence limits for each image and the four images combined (total) are shown in Figs. 7-9 for the three luminance levels. Each model is shown along the ordinate in order of the combined scale value (z-score). The model order changed slightly with differences in adapting luminance level of the test field; these differences, however, were not statistically significant. This was a somewhat surprising result. The models of von Kries, CIELAB, CIELUV, Reilly-Tannenbaum, and Richter are all luminance invariant while the remaining models take into account adapting luminance. This suggests that for these experimental conditions, luminance did not have a perceptual effect and that observers were judging the relative appearance attributes of hue, lightness, and chroma and not the absolute attributes of hue, brightness, and colorfulness. This result may not persist for different experimental conditions or when viewing single stimuli rather than images.

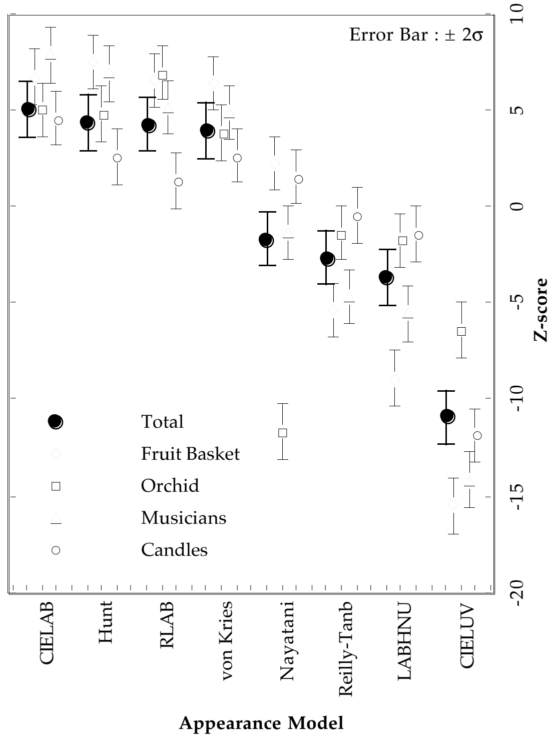

Figure 7. Plot of z-scores for the same luminance level.

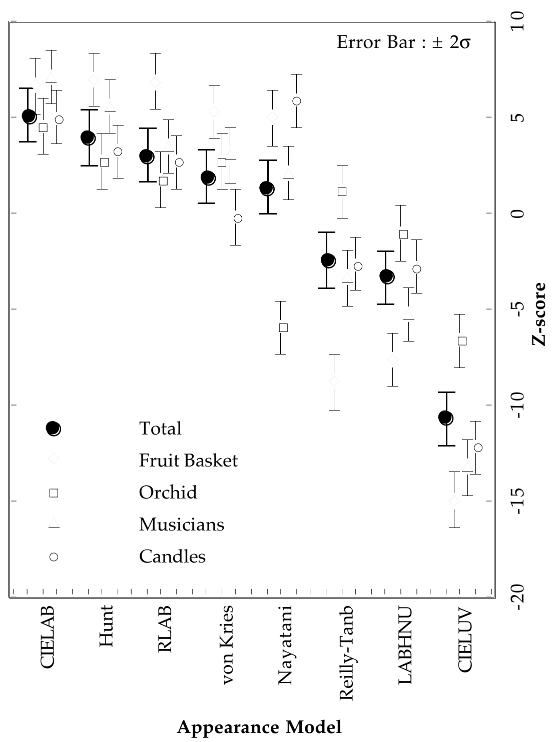

Figure 8. Plot of z-scores for the 1/3 luminance level.

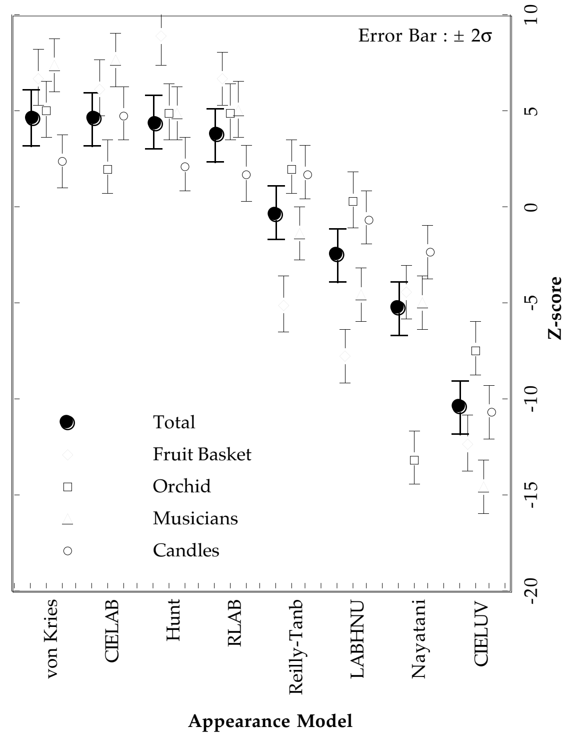

Figure 9. Plot of z-scores for the 3 times luminance level.

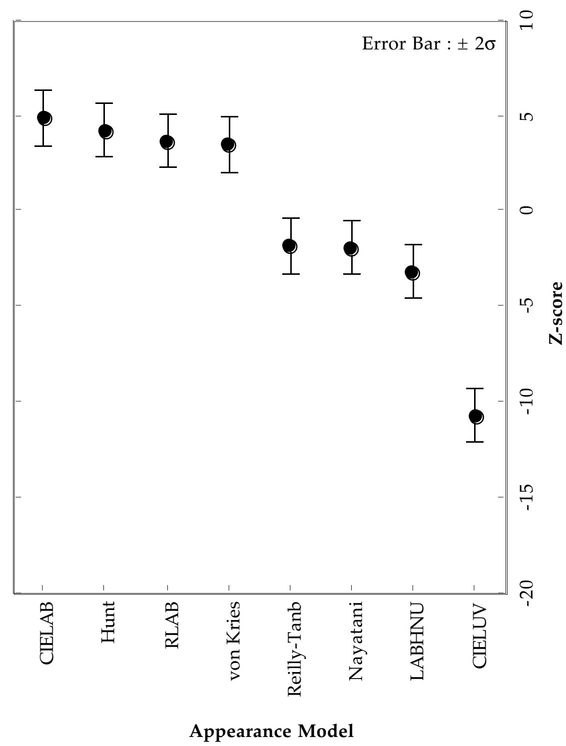

Figure 10. Plot of z-scores for average of all three luminance levels.

The models of low to poor performance had wide ranges of performance that depended on the image. This was particularly notable for the Nayatani model results. Image “orchid” always yielded very poor performance because of its predominant dark-bluish background. The Nayatani model predicts a large change in hue and lightness of the background to take into account the Helson-

Judd and Steven's effects. The effects were not observed causing the large discrepancy between the observed and predicted results. The LABHNU model provides another image-dependent example where the image "fruit basket" always had the poorest performance; in this case the hues of the high-chroma fruit colors were incorrectly predicted.

The average for all four images and three test field conditions is shown in Fig. 10. The appearance models could be divided into three statistical categories. The first category consisted of von Kries, CIELAB, Hunt, and RLAB; these models produced images that most closely matched the reference images. The differences between these models were not statistically significant. The second category consisted of LABHNU, Reilly-Tannenbaum, and Nayatani producing significantly poorer results. The third category consisted of CIELUV producing the worst results. For these viewing conditions, von Kries-based appearance models with the exception of Nayatani were the most effective models in yielding color appearance matches. The Nayatani model could be improved to the level of performance of the other von Kries-type models by reducing the Helson-Judd and Steven's effects. Both effects adversely altered the tone reproduction by changing the gray balance (light grays became yellowish and dark grays became bluish) and reducing contrast. Changes in tone reproduction and gray balance are very noticeable in images.

Conclusions

The color appearance models of von Kries, CIELAB, Hunt, and RLAB were the most effective models in predicting color appearance matches under the conditions studied: successive-Ganzfeld haploscopic viewing, fluorescent-daylight and tungsten sources, and pictorial stimuli. The psychophysics, however, generated interval scales of matching effectiveness, not scales of acceptability. In order to determine acceptability, further testing is required under typical WYSIWYG conditions.

Acknowledgments

We thank Toru Hoshino for his calibration software and data. Special thanks are also given to those observers who have spent many hours judging images. This work was supported by the E. I. DuPont Company.

References

1. Commission Internationale de l'Eclairage (CIE), *Colorimetry Publication CIE No 15.2,* Central Bureau of the CIE, Paris, 1978.
2. O. Estevez, *On the fundamental data base of normal and dichromatic colour vision*, Ph.D. Thesis, University of Amsterdam, 1979.
3. M. D. Fairchild, *Chromatic adaptation and color appearance*, Ph.D. Dissertation, University of Rochester, (1990).
4. M. D. Fairchild, Formulation and testing of an incomplete-chromatic-adaptation model, *Color Res. Appl.*, **16**, 243-250 (1991).
5. M. D. Fairchild and Roy S. Berns, Image Color-Appearance Specification Through Extension of CIELAB, *Color Res. Appl.*, **18**, 178-190 (1993).
6. M. D. Fairchild, E. Pirrotta, T. Kim, On the Successive Ganzfield Technique, *Color Res. Appl.*, in press, 1993.
7. G. A. Gescheider, *Psychophysics Method, Theory, and Application*, Lawrence Erlbaum Associates, 1985, pp. 264.
8. P. Hung, Tetrahedral division technique applied to colorimetric calibration for imaging media, *IS&T Final Program and Advance Printing of Paper Summaries*, 419-422 (1992).
9. R. W. G. Hunt and M. R. Pointer, A colour-appearance transform for the CIE 1931 standard colorimetric observer, *Color Res. Appl.*, **10**, 165-179 (1985).
10. R. W. G. Hunt, A model of colour vision for predicting colour appearance in various viewing conditions, *Color Res. Appl.*, **12**, 297-314 (1987).
11. R. W. G. Hunt, *Measuring Colour,* 2nd Ed. Chichester, 1990, pp. 213-258.
12. R. W. G. Hunt, Revised Colour-Appearance Model for Related and Unrelated Colours, *Color Res. Appl.* **16**, 25-43 (1991).
13. Y. Nayatani, K. Takahama, and H. Sobagaki, A nonlinear model of chromatic adaptation, *Color Res. Appl.*, **6**, 161-171 (1981).
14. Y. Nayatani, K. Takahama, H. Sobagaki, and J. Hirono, On exponents of a nonlinear model of chromatic adaptation, *Color Res. Appl.*, **7**, 34-45 (1982).
15. Y. Nayatani, K. Takahama, and H. Sobagaki, Prediction of color appearance under various adapting conditions, *Color Res. Appl.*, **11**, 62-71 (1986).
16. Y. Nayatani, K. Hashimoto, K. Takahama, and H. Sobagaki, Whiteness-blackness and brightness response in a nonlinear color-appearance model, *Color Res. Appl.*, **12**, 121-127 (1987).
17. Y. Nayatani, K. Hashimoto, K. Takahama, and H. Sobagaki, A nonlinear color-appearance model using Estévez-Hunt-Pointer primaries, *Color Res. Appl.*, **12**, 231-242 (1987).
18. Y. Nayatani, K. Takahama, and H. Sobagaki, Field trials on color appearance of chromatic colors under various light sources, *Color Res. Appl.*, **13**, 307-317 (1988).
19. Y. Nayatani, Y. Umemura, K. Hashimoto, K. Takahama, and H. Sobagaki, Analyzing the natural color system's color-order system by using a nonlinear color-appearance model, *Color Res. Appl.*, **14**, 69-78 (1989).
20. Y. Nayatani, K. Takahama, and H. Sobagaki, Field trials on color appearance and brightness of chromatic object colors under differenct adapting-illuminance levels, *Color Res. Appl.*, **13**, 307-317 (1988).
21. Y. Nayatani, K. Takahama, H. Sobagaki and K. Hashimoto, Color-appearance model and chromatic-adaptation transform, *Color Res. Appl.*, **15**, 210-221 (1990).
22. Y. Nayatani, Y. Umemura, H. Sobagaki, K. Takahama, and K. Hashimoto, Lightness perception of chromatic object colors, *Color Res. Appl.*, **16**, 15-25 (1991).
23. K. Richter, Cube-Root Color Spaces and Chromatic Adaptation, *Color Res. Appl.*, **1**, 25-43 (1980).
24. K. Takahama, H. Sobagaki and Y. Nayatani, Analysis of chromatic-adaptation effect by a linkage model, *J. Opt. Soc. Am.,* **67**, 651-656 (1977).
25. K. Takahama, H. Sobagaki and Y. Nayatani, Formulation of a nonlinear model of chromatic adaptation for a light-gray background, *Color Res. Appl.*, **9**, 106-115 (1984).