Psychophysical Generation of Matching Images for Cross-Media Color Reproduction

Karen M. Braun and Mark D. Fairchild
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology, Rochester, New York

Abstract

A technique was developed for generating images in different media that matched in color appearance. Observers used Adobe Photoshop™ to adjust CRT images to match print originals. Observers produced accurate matches when images were viewed at equal white point. For differing white points, observer-matched images were found to be equal or superior to predictions of color-appearance models.

Introduction

In order to test color appearance models using magnitude estimation or a paired-comparison technique,[1-5] reproductions must be calculated based on the original image data as well as information about the viewing conditions of the original and reproduction. These models are modified and improved as the field of color-appearance modeling continues to mature. Therefore results of a psychophysical experiment to test color appearance models may be quickly outdated. The goal of this phase of research was to develop a technique to generate color-appearance image data *independent* of any color appearance model. The colorimetric data from these matching images can be used to test current models as well as their inevitable future modifications without need for further observations. Other new models can be tested and perhaps derived based on these independent data.

Observers viewed original printed images in one viewing condition and adjusted CRT images to match the originals.The accuracy of this technique was tested by requiring observers to adjust images to match originals on the same CRT with the same viewing conditions. The second phase involved making these adjustments across media and with the constraint of a 60-sec. adaptation period at approximately the same white point chromaticity. Again the colorimetric data were analyzed for accuracy. Two final sessions included a change in white point between the two images.

To be a useful technique, the final adjusted images must be considered accurate matches not only by the observers who created them but by others.Therefore the adjusted images from this experiment were combined with reproductions predicted using various color appearance models in a paired-comparison psychophysical experiment to investigate whether they offered improvement over currently available transformations. A simple linear model was hypothesized for the particular set of viewing conditions used in these experiments and tested on an independent scene.

Experimental Set-up

Observers used Adobe Photoshop™ 3.0.1 to adjust images viewed on a CRT set to the chromaticity coordinates of CIE Illuminant D65 to match original images under the conditions shown in Table I.

Table I. Viewing Conditions of Original Images.

Exp. #	Original Media	CCT (K)	White x	Point y	Lum. (cd/m^2)
1.	CRT	6500	0.3126	0.3287	41
2.	Booth	6500	0.3178	0.3106	41
3.	Booth	9300	0.2886	0.2808	41
4.	Booth	3000	0.4164	0.3659	41
Repro.	CRT	6500	0.3126	0.3287	41

CRT images were surrounded by a gray field consisting of 50% white pixels and 50% black pixels. Also present in the observers' field of view were the menu bar and the tool bar, which both contain full white and black areas. In the first experiment, the original image and the adjustment image were viewed successively on the CRT, such that the two images could never be seen at the same time. For the remaining three experiments, original images were viewed in a light booth, with chromaticities approximating 9300K and CIE Illuminants D65 and A.

All experiments were conducted in a completely darkened room, so that only the print or CRT image occupied the observers' field of view. A divider was constructed from black foam core that prevented observers from viewing the print and CRT image at the same time. Observers moved a slider to reveal the appropriate image. This apparatus, shown in Figure 1 prevented observers from seeing both images at once and prevented stray light from either viewing condition from falling on the other.

Observers sat approximately 35 in. from the printed original images and the CRT screen in all techniques. This was chosen to be a comfortable and realistic viewing distance for an office environment. The images subtended an angle of approximately 17° in the observers' field of view (as measured across the diagonal of the 6" × 8" image.)

Two digital color scenes containing pictorial information were used in the adjustment experiment, a fruit and vegetable basket and an outdoor scene of two golfers. Original images were 6" × 8", printed at 200 dpi using a Fujix Pictrography 3000, a continuous-tone digital printer. These 6" × 8" images included a thin white border that was adjusted and modeled as part of the image.The prints were mounted on spectrally non-selective gray cardboard with a luminance level approximately equal to the gray background on the CRT. A black foam-core aperture was placed in front of the light booth such that the angular subtense of the original print matched that of the CRT.The printed images were digitized before mounting using a Howtek D4000 drum scanner at the resolution of the CRT, 72 dpi, to provide RGB data for preparing the CRT images. The scanner was colorimetrically characterized using a technique by Berns and Shyu[6] before producing the CRT images, so that scanner RGB tristimulus values could be accurately converted to CIE XYZ tristimulus values for the various spectral power distributions used in the light booth.

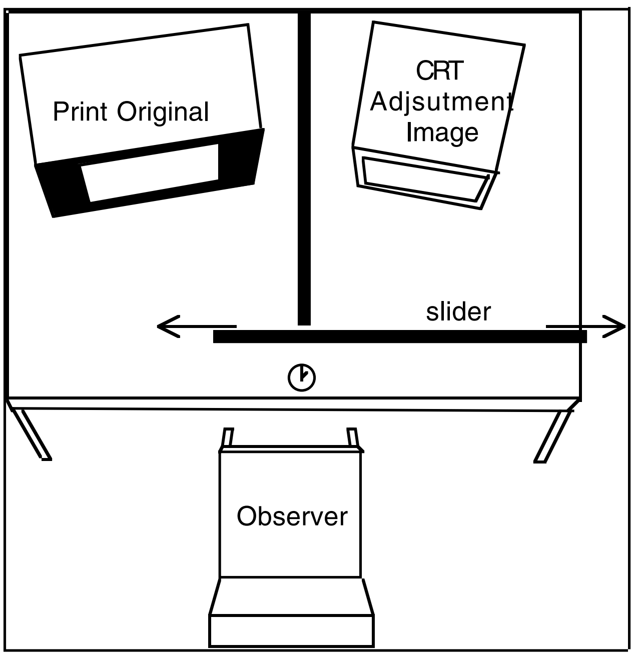

Figure 1. Experimental set-up for observers adjusting images to match originals. The shutter was moved to reveal the image of interest.

Printed original images were illuminated and viewed using fluorescent tubes in a Soft-View Transparency/Print Viewer light booth (made by Graphic Technology, Inc., model SOFV-1). Sources used in this experiment were measured with the Photo Research Spectra Scan® PR-650 spectroradiometer, and included a Macbeth 6500 bulb, a Graphlite® D5000 Color Viewing Lamp, and a General Electric Soft White Home Fluorescent bulb (approximately 3000K.) Adjustment images were displayed on an Apple Multiple Scan 20 CRT Display with white point chromaticities approximating those of CIE Standard Illuminant D65. This CRT was controlled by a Power Macintosh 8100/110. The CRT characterization technique of Berns, Motta, and Gorzynski[7] was used to allow a desired set of tristimulus values to be produced with the appropriate digital counts on the CRT.

Photoshop is somewhat limited in that images are quantized to 256 levels at each adjustment step. This limitation was reduced by using initial adjustment images that were not too different from the final adjusted images, by using well-trained observers who did not require excessive numbers of adjustments to make matches, and by averaging ten adjusted images per scene (two adjusted images per observer, five observers) to mask quantization errors that may have occurred in a particular adjusted image.

Psychophysics

Adjustment of Images

Five observers performed this adjustment experiment for the viewing conditions given in Table I. These included the author, her advisor, and three R.I.T. students with experience in image manipulation using Adobe Photoshop. Observers were initially given the Farnsworth-Munsell 100-Hue Test to evaluate their color-discrimination ability. No observer had more than 3 two-cap inversions. Observers were instructed to view the original then adjust the other image to match it. They were not permitted to view both images at the same time. They could use any of the controls in Photoshop except Zoom, Image Size, Crop, Canvas, Size, Eye Dropper, Hand Tool, or Show Info and were instructed to avoid spatial operations such as blurring and sharpening. They could adjust both the whole image and specific objects in the image, but were instructed to adjust specific objects last. They were permitted to return to the initial image if they felt it was necessary. Based on the results of a study of the time-course of chromatic adaptation by Fairchild and Reniff,[8] observers adapted to the original printed image for 60 sec., and viewed the CRT image for 60 sec. before making any adjustments. They were advised to shift their focus around the scene such that they would not locally adapt to colors in the scene. Each observer made adjustments for the two scenes, *Fruit* and *Golfer*. They repeated these matches beginning with different CRT images such that each observer made two matches to the same original. Therefore there were ten CRT image matches for each original printed image in a given experiment.

Observers' actions in Photoshop were recorded using a script recorder called DayStar Digital's PhotoMatic™. These scripts were examined to determine which tools observers found most useful. This software records the physical coordinates and actions of the observers such as screen coordinates, mouse-clicking, and keyboard commands. It does not keep track of the detailed Photoshop commands that the observer performs. A video recorder was used to capture the screen as observers made their adjustments in order to cross-reference the software scripts.

Paired Comparison Verification

The viewing conditions were identical to those used by the expert observers in experiment 4. The software to control the paired-comparison technique was available on a unix system connected to a Pixar imaging computer. Digital counts of the adjusted images on the Apple CRT were converted to CIE tristimulus values and then to digital counts for the Pixar CRT, based on the colorimetric characterizations of both CRTs. These adjusted images were compared to images generated using various color appearance models, including RLAB,[9,10] Hunt's model,[11-13] CIELAB,[14] and von Kries chromatic adaptation.[15]

A paired-comparison experiment was performed by 32 naïve observers to test whether the adjusted images matched the original image at least as well as was predicted by the various color appearance models. Three-by-three linear matrix transformations were derived to predict the adjusted images from the original (as will be described in the Data Analysis section.) Images derived using these matrices were also compared to the averaged adjusted images and model reproductions. In total, seven reproductions were compared resulting in 21 pairs for each original image. Table II lists the transformations that were performed on each of the three scenes to produce reproductions for the paired-comparison experiment.

Table II. Techniques Used to Predict Matching Reproductions for the Three Scenes Used in the Paired-Comparison Experiment

Reproduction Technique	Scene: Fruit	Golfer	Barn
Adjustment	x	x	
Fruit Matrix	x		x
Golfer Matrix		x	x
Average Matrix	x	x	x
von Kries	x	x	x
CIELAB	x	x	x
RLAB	x	x	x
Hunt	x	x	x

The order in which pairs of reproductions were shown was randomized within the experiment. To further test the ability of the models and matrices to predict matches, a third scene, a barn on a sunny day, was introduced that was not used in the adjustment experiment. The order in which the three scenes were shown was varied for the 32 observers. Observers received the following instructions:

In the following experiment, you will view an original print image in the light booth. Study the color information in the image, including hue, saturation, contrast, lightness, etc. You will then view a pair of computer reproductions that you may toggle between using the [1] and [2] keys on the keyboard.

"Which of the two computer reproductions most closely matches the original image?"

When you have made your decision, toggle to the image you have chosen and press the space bar. A second pair of reproductions will appear on the monitor and you will repeat the above procedure.

Base your decisions on the accuracy of the match, NOT on your personal preference between the reproductions. If you have and questions or are not sure of your task, please ask me. ALSO, if you feel you have made an error (for example, accidentally pressed the space bar,) please tell me immediately.

Observers in the paired-comparison experiment adapted to a gray card in the light booth for 60 sec. before viewing the original printed image. They were also shown a gray field on the CRT at a luminance level approximately equal to the background of the print for 60 sec. before viewing the pairs of reproductions. They were required to look back at the original once during each session, after having made 10 or 11 of the 21 comparisons. Thirty-two observers performed this paired-comparison experiment. Most were in the field of imaging science but had little experience judging color images. The average observer age was 31.4 with a range from 21 to 51.

Analysis and Results

Observer Consistency and Ability

Results from the first adjustment experiment, where both originals and adjustment images were viewed on the CRT at the same white point, were analyzed to determine how well observers could adjust the image to match the original. The digital counts for the ten *Fruit* images were averaged on a pixel-by-pixel basis to give an average adjusted image for the *Fruit* scene. This was also done for the *Golfer* scene to give an average *Golfer* image. The image was divided into object regions to avoid excessive weighting for large image areas of a single color. The digital count values were averaged for each of these regions. Seventeen object regions were used in the Fruit scene and sixteen regions in the *Golfer* scene. These average digital counts were converted to tristimulus values by applying the inverse CRT calibration model. CIELAB color differences between the original and adjusted images were calculated for these regions.

Experiment 1: CRT to CRT

Linear regression was performed between the CIE L* and C* values for the average adjusted Fruit image and for the original Fruit image. The coefficient of determination, R^2, for the Fruit image was 0.9989 for L* and 0.9931 for C*. For the Golfer image, R^2 was 0.9982 for L* and 0.9862 for C*. Observers were most accurate in adjusting lightness to match the original, but still quite accurate in chroma. The average CIELAB color differences for these images are shown in Table III.

Table III. CIE ΔE_{ab}^* between original images and average adjusted images.

Image	Fruit	Golfer
Average ΔE_{ab}^*	2.909	3.240
Minimum ΔE_{ab}^*	0.795	0.333
Maximum ΔE_{ab}^*	5.871	8.583

Stokes et al.[16] demonstrated that the average perceptibility tolerance of observers for complex pictorial images was 2.15 CIELAB color difference units, with a range of 1.57 to 2.56 units. In that experiment, observers viewed images sequentially with a 0.2-sec. time delay between the original and reproductions. The color differences found in the present paper experiment were about 3.0 color difference units. This indicates that observers would still be able to perceive a color difference between the original and their adjusted images. This is perhaps due to the fact that in the Stokes study, the color differences were systematic while in the adjusted images the differences were more random. Presumably random color difference among the pixels in two images would be less detectable than systematic errors.

Experiment 2: Booth to CRT

Observers made accurate tristimulus matches between the CRT and the printed originals viewed in the light booth under D65. The coefficient of determination values, R^2, were 0.9943 and 0.9892 for the Fruit and Golfer images, respectively, when L* of the average adjusted image was regressed against the L* value of the original. The values of R^2 were 0.9298 and 0.9895 for the Fruit and Golfer images respectively for C*(original) versus C*(reproduction). The average CIE color differences were calculated for the various object regions between the original printed image and the average adjusted CRT image. The average color differences for the *Fruit* and *Golfer* images for experiment 2, as well as the minimum and maximum color difference are given in Table IV.

Table IV. CIE ΔE_{ab}^* between original images and average adjusted images.

Image	Fruit	Golfer
Average ΔE_{ab}^*	2.909	3.240
Minimum ΔE_{ab}^*	0.795	0.333
Maximum ΔE_{ab}^*	5.871	8.583

The color differences were slightly higher than those given in Table III. This increase in color difference was surprisingly small considering the added constraints of media change and of remembering images over the 60-sec. adaptation period that were not present in experiment 1. Also a small difference in white point chromaticity between the two conditions is shown in Table I. This may require some color-appearance modeling to predict the match more accurately. Experiment 2 demonstrated observers' ability to make matches over the change in viewing conditions and the time delay.

Calculation of Matrices

For the *Fruit* and the *Golfer* scene in this experiment, the 3×3 matrix was determined that best converted tristimulus values of the original printed image to those of the average adjusted image. Multiple linear regression was used to determine the best-fitting matrix between the average tristimulus values of the print and adjusted images for the various object regions. Regression was performed using (1) just the *Fruit* data, (2) just the *Golfer* data, and (3) both sets of data. Systat was used for this analysis. These matrices were calculated for the final two viewing conditions listed in Table I, 3000K -> 6500K and 9300K -> 6500K. The adjusted R^2 value for the fit of the predicted matrix model was greater than 0.993 for experiment 3 and greater than 0.998 for experiment 4. The resulting matrices found for experiment 4 viewing conditions are given in Figure 2.

The Average matrix in Figure 2 was calculated by including all object regions from both scenes in the regression and this matrix is referred to as AveMatrix in Figures 7 through 9.

Color Appearance Model Performance

The tristimulus values of the adjusted image were calculated for the object regions described above. Using the tristimulus values of the original images for the object regions, predicted reproductions were also calculated using various color appearance models including RLAB, Hunt's model, von Kries chromatic adaptation, and CIELAB color space, and the 3×3 matrices calculated for the images. The advantage of producing color-appearance-matching images that are independent of any model or other transformation is that new models or revisions to existing models can be tested without any further psychophysical experimentation. In this vein, three other models were also tested using the adjusted image data, LLAB,[17] Finlayson's spectral sharpening model, and Nayatani's model.[18,19] The average CIELAB color-difference, ΔE_{ab}^*, was calculated between the adjusted image data and data from images predicted using each model. The minimum and maximum ΔE_{ab}^* were also recorded as well as the standard deviation of the color-differences among the regions.

Experiment 3 Results

Fruit Matrix

1.2235	0.3825	-0.1200
-0.1656	1.8115	-0.1001
0.0439	0.0503	0.9904

Fruit Matrix

1.2235	0.3825	-0.1200
-0.1656	1.8115	-0.1001
0.0439	0.0503	0.9904

Avg. Matrix

1.2235	0.3825	-0.1200
-0.1656	1.8115	-0.1001
0.0439	0.0503	0.9904

Experiment 4 Results

Fruit Matrix

1.2235	0.3825	-0.1200
-0.1656	1.8115	-0.1001
0.0439	0.0503	0.9904

Golfer Matrix

1.2235	0.3825	-0.1200
-0.1656	1.8115	-0.1001
0.0439	0.0503	0.9904

Avg. Matrix

1.2235	0.3825	-0.1200
-0.1656	1.8115	-0.1001
0.0439	0.0503	0.9904

Figure 2. Matrices for experiments 3 and 4. Average Matrix was found by performing linear regression on the tristimulus data from both the Fruit and the Golfer.

Figure 3 shows the calculated color difference between the average adjusted image and reproductions predicted by various models for the fourth set of viewing conditions listed in Table I.

As expected, the models derived from the adjusted images, namely the three matrix images, had a smaller color difference from the adjusted images than any of the color appearance models. The image derived from the Golfer matrix (GoMatrix) had higher color differences when applied to the Fruit scene than the Golfer scene, and the reverse was true for the Fruit matrix (FrMatrix). The Average matrix (AveMatrix) was a good compromise between these matrices. RLAB, CIELAB, and LLAB gave color differences on the order of 6 units, von Kries gave differences from 6 to 8 units, and spectral sharpening, Hunt's model and Nayatani's model gave differences greater than 10 units.

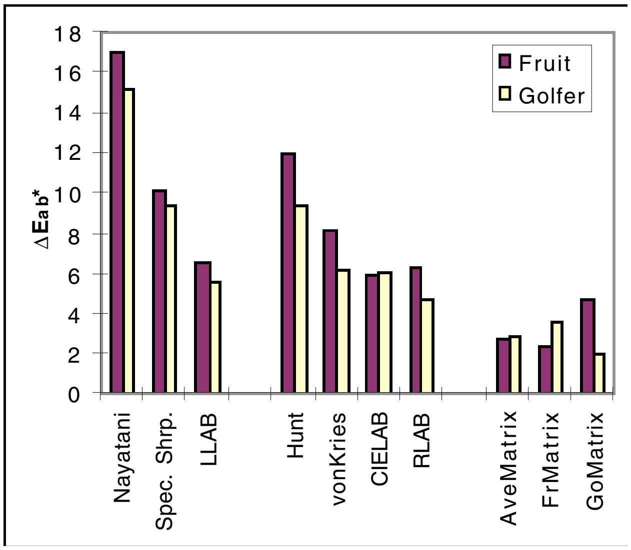

Figure 3. ΔE_{ab}^* between average adjusted images and color appearance model predictions for experiment 4 conditions. The first three models were not tested in the paired-comparison experiment, whereas the second set of four models was. The final three models used matrices calculated from the average adjusted image data.

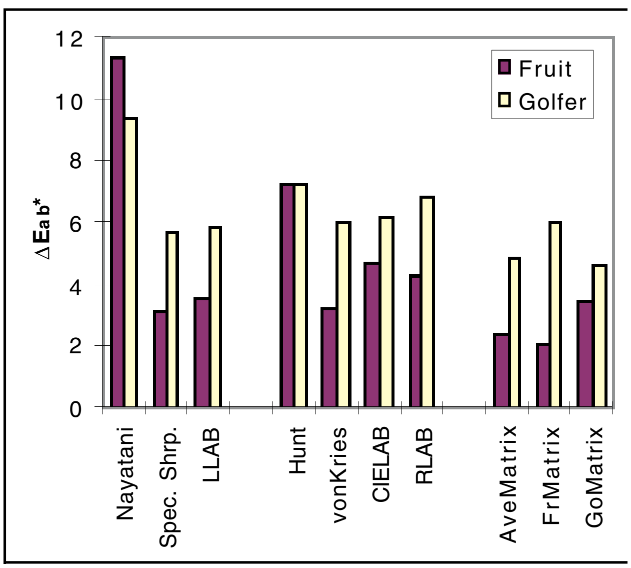

Figure 4. ΔE_{ab}^* between average adjusted images and color appearance model predictions for experiment 3 conditions. No paired comparison experiment was performed for this set of viewing conditions. The last three models used matrices calculated from the average adjusted image data.

Interestingly, for the viewing conditions in experiment 3, most color appearance models performed significantly better for the *Fruit* image than the *Golfer* image, including the *Golfer* matrix. Because the viewing condition change was not as great for this experiment, most color appearance models performed approximately equal, with color differences around 3 to 5 units for the *Fruit* image and 5 to 7 units for the *Golfer* image. Hunt's model gave a difference of around 7 for the *Fruit* image and Nayatani's model gave differences from 9 to 11 units.

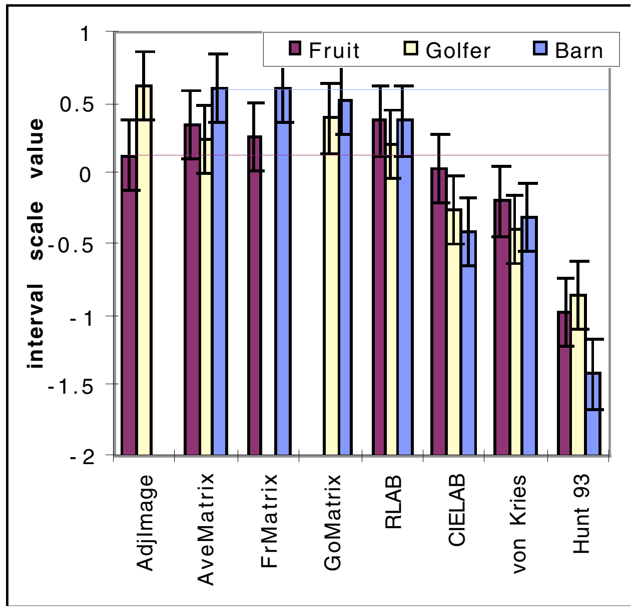

Figure 5. Results of paired-comparison experiment using average adjusted image, images determined from matrices, and color appearance models (Hunt, von Kries, CIELAB, RLAB). Viewing conditions were equivalent to experiment 4 listed in Table I. Error bars represent 95% confidence intervals around the mean.

Paired Comparison Experiment

Using Thurstone's Law of Comparative Judgments,20 the choices of reproduction were converted to an interval scale of color reproduction quality for the various models. This analysis technique is described in detail by Torgerson.[21] Results of paired-comparison experiments using the viewing condition of experiment 4 are given in Figure 5. The average adjusted image for the *Fruit* scene produced a match that was as close to the original as reproductions calculated from the matrices, RLAB, and CIELAB. Observers found the match between the *Golfer* adjusted image and the original to be superior to images produced with any other transformation. The matrix found using both images gave good results on all scenes including the independent *Barn* image.This indicates that the adjustments that observers made were not strictly image dependent and that they were able to account for attributes of appearance that were not accurately predicted by the color appearance models. It also shows the potential for a new model to be developed based on this adjustment technique.

Scripts and Videos

Scripts were recorded for fifty of the eighty trials in the four adjustment experiments. Table V shows the average number of times observers selected a particular tool to adjust the CRT image. Most of the tools listed can be used to adjust the achromatic and chromatic content of the image. The tools listed in the top section of the table are listed from most popular to least popular.

Table V. Average Number of Times per Experiment Each Photoshop Tool was used by Observers for all Eight Trails of Each Scene (2 Trials. 4 Experiments).

Observer	Fruit	Golfer
Adjust Color Balance	3.73	3.75
Adjust Hue/Saturation	3.35	3.11
Adjust Curves	2.55	2.93
Adjust Levels	2.35	2.04
Adjust Brightness/Contrast	1.82	1.98
Selective Color	1.35	1.02
Adjust Variations	1.20	4.05
Replace Color	0.09	0.13
Select Object	5.42	4.79
Viewed Original (exp. 1)	37.88	30.17

Selective Color and Replace Color were the least popular tools, while other choices depended on the observer and the color distortions in the initial adjustment image. Observers found the sliders of the first two tools in Table V most useful for changing chromatic information, while the flexibility of histogram reshaping (*Adjust Curves, Adjust Levels*) was more useful when adjusting achromatic information. Certain effects can be achieved with various tools. For example, a change in contrast in the image could be achieved by altering the shape of the histogram using *Adjust Levels*, changing the look-up table between initial and adjusted pixels using *Adjust Curves*, using the contrast adjustment of *Adjust Brightness/Contrast* or using *Adjust Variations*. In experiment 1 where both images were viewed on the CRT, observers viewed the original approximately 30 times per session, or once a minute.

Summary

It has been shown that, using the adjustment technique described in this paper, observers can produce consistent matches over the required 60-sec. adaptation period and across a change in media. This adjustment technique produced images that matched at least as well to printed originals as reproductions created using color appearance models, as shown by the paired-comparison experimental results. Matrices derived using the adjusted images also predicted matching reproductions for an independent scene better than color appearance models. Some restrictions of the technique are that observers must be proficient in Adobe Photoshop and must spend about a half an hour per image to make an accurate match.The most useful tools in Photoshop were shown in Table V. In order to derive a new model of color-appearance, this experiment must be repeated for a wide range of viewing conditions and image content.

RLAB color appearance model consistently produced good cross-media matches for images. There is evidence that LLAB would also produce good matches. The color-difference calculations showed inaccuracies in Hunt's and Nayatani's model in producing matches, alhtough ΔE_{ab}^* alone does not seem to be optimum for quantifying visual differences among images. A metric is needed that would correlate the results of the adjustment experiment to results of the paired-comaprison technique so that future model modifications could be tested with no further psychophysical experimentation.

References

1. P. J. Alessi, CIE Guidelines for coordinated research on evaluation of colour appearance models for reflection print and self-luminous display image comparisons, *Color Res. Appl.* **19**, 48-58 (1994).
2. M. D. Fairchild, R. S. Berns, A. A. Lester, and H. K. Shin, Accurate color reproduction of CRT-displayed images as projected 35mm slides, *IS&T/SID 2nd Color Imaging Conference*, Scottsdale, 69-73 (1994a); (see pg. 248, this publication).
3. T. G. Kim, R. S. Berns, and Fairchild, M. D., Comparing appearance models using pictorial images, *IS&T/SID Color Imaging Conference*, 72-77 (1993); (See pg. 49, this pub.).
4. M. C. Lo, M. R. Luo, P. A. Rhodes, Evaluating colour models' performance between monitor and print images, *Color Res. Appl.* **21**, 18–28 (1996).
5. E. Pirotta, and M. D. Fairchild, Directly testing chromatic-adaptation models using object colors, *Proceedings of the 23rd Session of the CIE*, (New Dehli) 1, 77-78 (1995).
6. R. S. Berns, and M. J. Shyu, Colorimetric characterization of a desktop drum scanner using a spectral model, *J. Electronic Imaging* **4**, 360-372 (1995).
7. R. S. Berns, R. J. Motta, and M. E. Gorzynski, CRT colorimetry. Part I: Theory and practice, *Color Res. Appl.* **18**, 299-314 (1993a).
8. M. D. Fairchild, and L. Reniff, Time course of chromatic adaptation for color-appearance judgments, *J. Opt. Soc. Am. A* **12**, 824-833 (1995).
9. M. D. Fairchild, Refinement of the RLAB Color Space, *Color Res. Appl.*, **21**, 338–346 (1996).
10. M. D. Fairchild, and R. S. Berns, Image color appearance specification through extension of CIELAB, *Color Res. Appl.* **18**, 178-190 (1993).
11. R. W. G. Hunt, Evaluation of a model of colour vision by magnitude scalings: discussion of collected results, *Color Res. Appl.* **19**, 27-33 (1994a).
12. R. W. G. Hunt, An improved predictor of colourfulness in a model of colour vision, *Color Res. Appl.* **19**, 23-26 (1994b).
13. R. W. G. Hunt, Revised colour-appearance model for related and unrelated colours, *Color Res. Appl.* **16**, 146-165 (1991).
14. *Colorimetry*, 2nd ed., CIE Publ. No. 15.2, Central Bureau of the CIE, Paris, 1986.
15. J. von Kries, Chromatic adaptation, *Festschrift der Albrecht-Ludwig-Universität*, (1902).
16. M. Stokes, M., D. Fairchild, R. S. Berns, Precision requirements for digital color reproduction, *ACM Transactions of Graphics* **11**, 406-422 (1992).
17. M. R. Luo, M.-C. Lo, and W.-G. Kuo, The LLAB(l:c) color model, submitted for publication to *Color Res. Appl.* (1996).
18. Y. Nayatani, H. Sobagaki, K. Haskimoto, and T. Yano, Lightness dependency of chroma scales of a nonlinear color-appearance model and its latest formulation, *Color Res. Appl.* **20**, 156-167 (1995a).
19. Y. Nayatani, Revision of the chroma and hue scales of a nonlinear color-appearance model, *Color Res. Appl.* **20**, 143-155 (1995b).
20. L. L. Thurstone, A law of comparative judgment, *Psych. Rev.* **34**, 273-286 (1927).
21. 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 1996 Color Imaging Conference Proceedings, page 214