Accurate Color Reproduction of CRT-Displayed Images as Projected 35-mm Slides

Mark D. Fairchild, Roy S. Berns and Audrey A. Lester*
Rochester Institute of Technology, Center for Imaging Science
Munsell Color Science Laboratory
54 Lomb Memorial Drive, Rochester, New York 14623-5604
E-mail: mdfpph@grace.rit.edu
* Eastman Kodak Company, 1700 Dewey Avenue, Rochester. NY 14650

Abstract

Accurate color reproduction of images presented on a computer-controlled CRT display as projected 35-mm transparencies is a complicated procedure requiring the characterization and control of several imaging processes and the application of appropriate color appearance modeling to account for the changes in viewing conditions. We review a process for image recorder characterization, projection system characterization, and testing of color appearance models for this application. Accurate image recorder characterization was achieved through a combination of empirical modeling of the exposure and processing system and of a physical model of photographic film. The projection system characterization included specification of the spectral properties of the light source, reflectance properties of the viewing screen, and the effects of light exposure and temperature on the photographic transparencies. Color appearance models were used to predict the changes in image color appearance due to changes in media, white point, luminance, and surround. The RLAB model proved to work best in this application. © 1996 SPIE and IS&T

1 Introduction

With the recent and rapid technological growth of affordable, open, color image input and output devices has come a strong need for color management systems to facilitate device-independent color imaging. One application is the development of presentations and other imagery on desktop computer systems with the final intention being the production of 35-mm slides for projection. Ideally, users should be able to select colors and view images on the computer-driven CRT display that are accurate representations of the colors that will ultimately be seen on the projected slides. There are numerous requirements for the successful implementation of such a device-independent, open, color imaging system.[1] This paper reviews the development and evaluation of a system for reproducing CRT-displayed images as projected 35-mm slides.

Particular attention is given to the issues of device characterization and color-appearance transformation. The process of device-independent color imaging has several steps.[2] Assuming one starts with an image, either displayed on or captured using some original device, and desires a reproduction of that image using some other imaging modality (i.e., device), the first step is the colorimetric characterization of the imaging devices. Colorimetric characterization allows the transformation between devicedependent color coordinates (e.g., RGB or CMYK) and device-independent coordinates such as CIE XYZ or CIELAB.[3] The result is the knowledge of just what colors are being produced by the device for a particular set of viewing conditions. Since original images are often observed in viewing conditions substantially different from those of the reproduction, one must also account for the visual impact of these changes in the viewing conditions (e.g., white point, luminance level, surround, medium, etc.). These changes are accounted for by using color appearance models, which allow the transformation from device-independent color coordinates to what has been referred to as *viewing-conditions-independent* color appearance coordinates.[2,4] In such coordinates the appearance of image elements is specified independent of the medium or viewing conditions. These coordinates are the best suited for image manipulations such as preference editing and gamut mapping. Once viewing-conditions-independent coordinates are obtained for the original image and viewing conditions, the entire process is reversed for the reproduction medium and viewing conditions.

While this paper addresses the issues of colorimetric characterization and color appearance modeling, the equally critical issue of gamut mapping was bypassed by limiting the evaluated images to colors common to the gamuts of both image displays. This was not difficult in the current study since the luminance dynamic range of Ektachrome transparency film in general far exceeds that of typical CRT displays. In CRT-to-print reproduction, gamut mapping becomes a more critical issue. In addition to gamut limitations introduced by the imaging devices, the color appearance models sometimes introduce gamut boundary difficulties. It is possible for a given color appearance model to predict a required color on the reproduction that falls outside the device gamut, or, at times, outside the gamut of physically realizable colors. This is not necessarily a limitation of the color appearance model because it is

not always possible to reproduce colors that can be seen on the original, due to a significant change in viewing conditions. This problem was avoided in the current research by limiting the gamuts of the original images where practical, or eliminating a particular color appearance model from the experiments if it introduced significant gamut boundary difficulties.

2 CRT Display Characterization

The calorimetric characterization of CRT displays is of great importance in electronic publishing, computer graphics, and television, and has been the subject of considerable study. Thus it was not necessary to develop new techniques for this research. Instead, carefully developed and tested techniques were used.[5,6] Section 2.1 briefly reviews the CRT characterization techniques that were used in this research to completely specify the system from end to end.

2.1 Colorimetric Characterization

The CRT characterization was carried out using the techniques detailed by Berns, Motta, and Gorzynski.[5,6] A Mitsubishi Diamond Pro 17-in. monitor with a Sony Trinitron CRT driven by a Pixar II image computer was used throughout this study. The tristimulus values of the display primaries were carefully measured with a high accuracy calorimeter (LMT C1200 Colormeter) to derive the 3×3 linear transformation from linearized RGB channel radiometric scalars to CIE XYZ tristimulus values. The XYZ tristimulus values of a series of 11 levels of gray (R = G = B) colors were also measured to characterize the three nonlinear electro-optical transfer functions of the display. These XYZ tristimulus values were transformed to RGB tristimulus values and then a nonlinear curve fitting technique was used to fit a model incorporating gain, offset, and gamma terms to derive the transfer function between digital counts and radiometric scalars for each of the three CRT channels. The typical calorimetric accuracy of this characterization technique is better than 1.0 CIELAB unit color difference between actual and predicted colors. Two CRT setups were used in the experiments, approximating D65 ($x = 0.3124$, $y = 0.2977$, $Y = 53$ cd/m^2) and D93 ($x = 0.2838$, $y = 0.3290$, $Y = 60$ cd/m^2). All colorimetry was performed using the CIE 1931 (2°) Standard Colorimetric Observer.[3]

2.2 Spatial Uniformity

Spatial nonuniformity of the CRT display was assumed to be a negligible issue. However, it is clear that displays vary in luminance across the face of the CRT (by as much as 20%) and are likely to vary slightly in the other parameters of the characterization model as well. The characterization measurements were taken in the center of the display with a middle gray filling the rest of the CRT, and image presentations were limited to this central area. Since the spatial nonuniformities of the display are almost entirely in the luminance dimension and they are slowly varying across the display (low spatial frequency), the visual system is extremely insensitive to the changes. This is illustrated by the bandpass nature of the human luminance spatial contrast sensitivity function.[7] However, since the human visual system is relatively more sensitive to low-frequency changes in color (as illustrated by low-pass chromatic contrast

sensitivity functions[7]), nonuniformities in the other calorimetric dimensions would be of significantly greater concern.

3 Film-Recorder Characterization

Once accurate colorimetric data are available for the original CRT images, it becomes necessary to convert this information into RGB digital counts that can be used to drive the film recorder to produce slides with the desired colorimetry. A model for the film recorder colorimetric characterization and its implementation is described in Sec. 3.1. An MGI Solitaire 8xp image recorder was used. Additional details on the film recorder setup and characterization can be found in Refs. 8 and 9.

3.1 Film Model and Characterization

The first stage of calibrating the filmrecorder system was the characterization of the photographic film. Kodak Ektachrome 100 Plus Professional film was used throughout this project. The developed film was analyzed colorimetrically through spectral transmittance measurements. The spectral transmittance of the film can be treated as a combination of the spectral transmittance of the film base and the variable amounts of the cyan, magenta, and yellow image-forming dyes. This can be accurately modeled using the Beer-Bouger Law as illustrated in Eq. (1):

$$T_\lambda = T_{\lambda,g} \exp[-(c_c D_{\lambda,c} + c_m D_{\lambda,m} + c_y D_{\lambda,y})]. \tag{1}$$

T_λ is the spectral transmittance of the film, $T_{\lambda,g}$ is the spectral transmittance of the base, the c terms are the concentrations, and the D_λ terms are the unit spectral absorptivities of the cyan, magenta, and yellow dyes. Given this model of the film, it remains to determine the spectral absorptivities for the particular film dyes and develop a relationship between the RGB digital counts driving the film recorder exposure and the dye concentrations in the resulting image.

The spectral absorptivities of each of the component dyes were determined through statistical analyses of exposed and developed film colors. A set of 60 different single-color exposures was used. The spectral transmittance measurements of these samples were converted to spectral densities (absorptances) and subjected to a principal components analysis. The first three characteristic vectors were rotated via an equimax rotation to assure that each of the three vectors accounted for approximately equal amounts of the variance in the data set. The first 3 vectors accounted for over 99% of the variance. These vectors provided estimates of the dye spectral absorptivities. However, due to possible interimage effects and a data sampling that might not contain equal variation in each of the three dye concentrations, these estimates might be biased. A second analysis was completed in which ramps of 11 steps in RGB exposure were imaged to produce samples approximately varying in only one dye concentration at a time (i.e., 11 levels of R with G = B = 255, etc.). The first characteristic vector for each of these ramps was obtained as secondary estimates of the dye absorptivities. The first set of vectors was then rotated to the closest least-squares fit to the second set of vectors to produce the most robust global estimate of the dye spectral absorptivities. These were used for the remainder of the modeling. Figure 1 illustrates a typical set of estimates of dye absorptivities for the Ektachrome film.
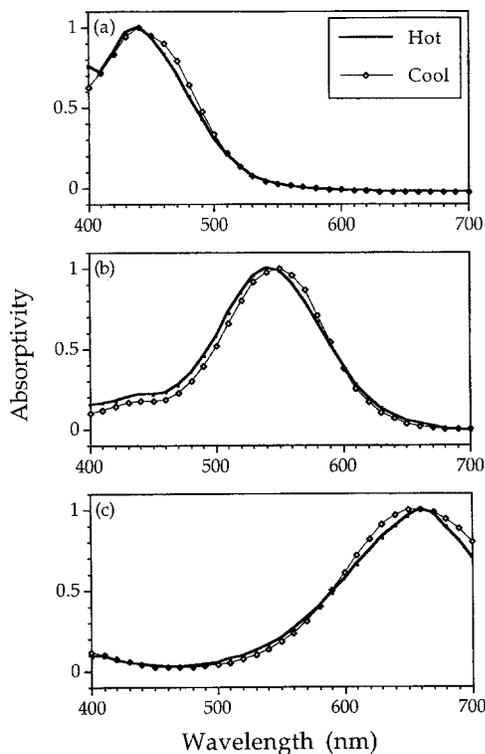
Figure 1. Yellow, magenta, and cyan dye absorptivity curves derived using principal component analyses of spectrophotometric measurements made either in a laboratory spectrophotometer (BYK-Gardner Color Sphere) or in projector with a spectroradiometer system (PhotoResearch PR703A).
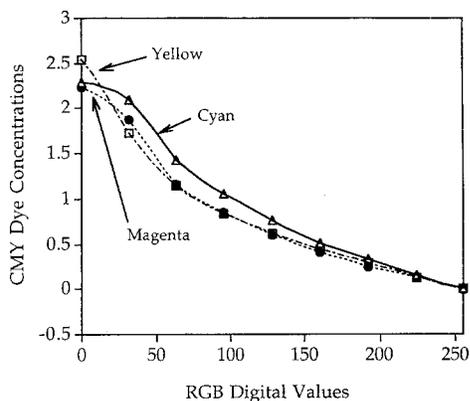
Figure 2. Interpolated tone reproduction curves and data points used to build 1-D LUTs relating RGB exposure digital counts to CMY dye concentrations.

A set of 21 single-color exposures was made to model the dye concentrations. These consisted of ramps of 4 exposures in which one of the RGB digital counts was varied from 64 to 255 in steps of 64, the other two were set at 0, and a fourth ramp of 9 gray exposures with R = G = B varying from 0 to 255 in steps of 32. The spectral transmittances of each of these samples were measured and the dye concentrations were determined iteratively using a modified forml° of the Allen tristimulus matching algorithm described by Bernsll and the dye spectral absorptivities derived earlier. Essentially this involves the solution of Eq.

(1) for the $3c$ terms given in the $3 D_\lambda$ term and the 2 sets of T_λ values through minimization of the colorimetric difference between the measured sample and its estimate. The results were a data set consisting of cyan, magenta, and yellow concentrations and the RGB digital counts used to produce them for the 36 colors.

3.2 Digital Count to Dye Concentration Model

The first step of the model to relate dye concentrations to exposure digital counts is the characterization of the nonlinear transformation between the two. Since this nonlinearity is a composite of the nonlinear nature of photographic film, the nonlinear characteristics of the CRT used in the film recorder, and the particular nature of the film recorder look-up tables (LUTs) in use, it is extremely difficult to characterize it with an analytical model. The approach taken was to build 1-D LUTs aimed at directly mapping each of the RGB digital counts to the respective CMY dye concentrations. These LUTs were constructed via spline interpolation of the data obtained from the 9 gray samples. If photographic film were a simple tripack with three independent layers and the RGB exposures each produced development in only one layer, then such a model would be sufficient. However, neither of these situations exist. Film is designed such that development in a given layer can affect the amount of dye produced in another layer (interimage effects) and it is difficult (if not impossible) to design and efficient exposing system that assures single-layer exposures. Thus such a model must be enhanced. This was accomplished by adding a 3×3 matrix transformation on the dye concentrations after the linearizing LUTs to produce more accurate estimates of the dye concentrations. This 3×3 was obtained via linear regression while constraining each row to a sum of 1.0 to maintain gray balance. To summarize, RGB digital counts were converted to CMY dye concentrations through 3-D LUTs followed by a 3×3 matrix transformation. These concentrations were used to reconstruct film spectral transmittance curves which were then used to calculate tristimulus values. The 1-D LUTs are illustrated in Fig. 2 and the derived 3×3 matrix is given in Eq. (2).

$$\begin{bmatrix} \hat{c}_c \\ \hat{c}_m \\ \hat{c}_y \end{bmatrix} = \begin{bmatrix} 1.0041 & -0.0081 & 0.0040 \\ -0.1332 & 1.0125 & 0.1207 \\ -0.0087 & -0.0455 & 1.0542 \end{bmatrix} \begin{bmatrix} \hat{c}_{c,\mathrm{LUT}} \\ \hat{c}_{m,\mathrm{LUT}} \\ \hat{c}_{y,\mathrm{LUT}} \end{bmatrix}. \quad (2)$$

It should be noted that the 1-D LUTs are particular to the film, film-recorder LUTs, and processing used in this project; other systems might have significantly different LUT characteristics. The 3×3 matrix is also unique to the complete system used, but it illustrates the significance of the film and exposure interimage effects.

The 3×3 model of the film interimage effects is an extreme simplification of the true chemical process taking place in the film. A more appropriate model would involve deriving a different 1-D LUT to relate, for example, the yellow dye concentration to B digital counts for each different level of R and G exposure. Although such a technique would be more accurate, it would require that an extremely large number of color samples (on the order of thousands) be generated and measured. This would essentially result in the creation of a 3-D LUT relating RGB

digital counts and dye concentrations derived via exhaustive measurement. Such a derivation is not practical given the inherent variability between film batches, processing, and film recorder setup. As soon as the device was characterized, it would be necessary to characterize it again. Thus, a simpler model-based approach that required fewer measurements was derived. This might result in slightly lower accuracy in exchange for significantly more practical characterization procedures. Such an approach has proven extremely valuable in the characterization of CRT displays.[5,6]

3.3 Overall Performance and Implementation

An independent set of test colors was generated to evaluate the overall accuracy of the film recorder model. The average CIELAB ΔE^*_{ab} (2° Observer, Projector Source) was 5.7 with a maximum of 10.8 for the full color set and the average ΔE^*_{ab} was 1.7 with a maximum of 2.1 for the grays. Although this performance is not as good as might be desired, it is probably near the limit of what can be achieved for such a system and was accomplished with a very simple model requiring the generation and measurement of only 21 color samples.

The color variability of the complete system was analyzed by making several series of repeat exposures. The exposure variability for repeated exposures of the same color on a single roll of film showed a maximum ΔE^*_{ab} of 0.5. The processing variability for a similar set of exposures on three different rolls of film showed an average ΔE^*_{ab} of 0.5 with a maximum of 1.9. The degradation of the film after 4 min. of projection showed an average ΔE^*_{ab} of 1.5 and after 16 min. of projection an average ΔE^*_{ab} of 2.6. If the slides are first preconditioned by 4 min. of projection, the change after an additional 16 min. of projection is reduced to an average ΔE^*_{ab} of 1.5. It is clear that the sources of variability in the system quickly add up to an error magnitude similar to the performance of the characterization procedure described earlier. These results are summarized in Table 1. Also note that the processing variability was evaluated over a 3-day period with a single photofinisher. Long-term tracking of control strips showed variation as much as 3 times larger than what was observed during this 3-day period. This more realistic estimate is listed in Table 1 as "Exposure and Processing (2)." In addition, variability across photofinishers would be substantially larger. Thus the estimate of processing variability is conservative.

Table 1. Errors in film recorder system exposure, processing, projection, and characterization. Results are expressed in CIELAB ΔE^*_{ab} for the 1931 CIE Standard Colorimetric Observer and the 3863K projector source.

Source of Error	Mean ΔE^*_{ab}	Maximum ΔE^*_{ab}
Exposure	0.1	0.5
Exposure and Processing (1)	0.5	2.0
Exposure and Processing (2)	1.5	6.0
4 min. Projection	1.5	2.4
16 min. Projection	2.6	3.3
16 min. Proj. after 4 min. Precon.	1.5	2.9
Film Recorder Model	5.7	10.8

The film recorder characterization model described above cannot be analytically inverted to allow the prediction of the RGB digital counts required to produce a desired set of XYZ tristimulus values. Thus the model was implemented using an iterative inversion procedure to build a 3-D LUT that was used to transform image data using tetrahedral interpolation.

4 Projector-System Characterization

The aforementioned film recorder model allows the spectral transmittance of the exposed and developed film to be predicted from the RGB digital counts used to make the exposure. However, the transformation from the spectral transmittance of the film to the tristimulus values measured from a screen when the slides are projected remains to be characterized. Difficulties in this step of the process are reviewed in the following sections. A Leica P-2200 projector was used throughout this study.

4.1 Theory and Measurements

In theory, tristimulus values of a projected image would be calculated by multiplying the transmittance of the film by the spectral power distribution of the projector source and the spectral reflectance factor of the screen and then integrating with each of the CIE color matching functions. The one variable still to be addressed would be the proper measurement geometry for the spectral transmittance. The two geometry choices were diffuse/0°, measuring total transmittance, or 0°/0° measuring regular transmittance. This theory was evaluated by making measurements of the spectral transmittances of slides in the projector system and comparing them to measurements made in a spectrophotometer using both total and regular geometries. Discrepancies were discovered that were spectrally selective, functions of the slide color, and not correlated with geometric differences. An example is illustrated in Fig. 3, which shows the spectral transmittance of a green slide measured in the projector system as well as in a spectrophotometer using two different geometries. The cause of these discrepancies were determined and addressed as described in Sec. 4.2.
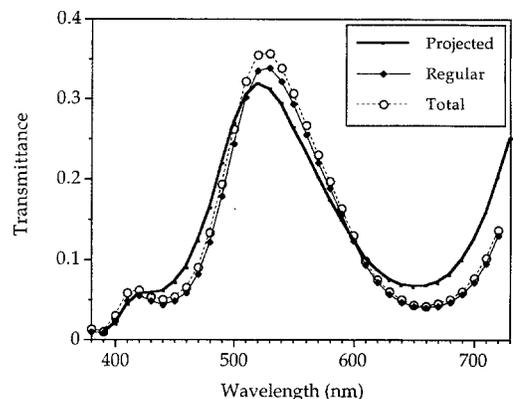
Figure 3. Comparison of spectral transmittance measurements for a green Ektachrome slide by means of total and regular measurement geometries in a laboratory spectrophotometer and the in-projector measurement.

4.2 Thermochromism

The major cause of discrepancies between the spectral transmittance measurements made in the projector and those made in a spectrophotometer was determined to be thermochromism—a change in color with temperature. The measurements made in the spectrophotometer were at room temperature (20°) while the slide temperature in the projector was approximately 52°C. Significant color shifts are expected in any colored material on such large temperature fluctuations. A series of colored slides were measured both in the spectrophotometer (cool) and in the projector (hot). The color shifts were found to be significantly color dependent and typically resulted in a decrease in chroma with increasing temperature. The changes averaged a ΔE^*_{ab} of 4.6 with a maximum of 10.3. The color shifts were reversible on cooling. Figure 4 illustrates the reversibility and magnitude of these color shifts for a magenta slide after 5, 10, 15, and 20 min. of projection. Details of the analysis of thermochromism in slide film can be found in Ref. 12.

This complication was addressed by measuring the color samples used to characterize the film through principal components analysis in the projector system. This resulted in dye spectral absorptivity curves for hot film that differed from those for cool film. Both sets of dye absorptivity curves are illustrated in Fig. 1. Heating the film resulted in a narrowing of the yellow-dye absorptivity, a broadening and short-wavelength shift in the magenta-dye absorptivity, and a narrowing of the cyan-dye absorptivity. The "hot film" curves were used when modeling the system. In addition, all of the measurements used for the construction of the digital count to dye concentration model were made in the projector system. Had this procedure not been followed, the overall accuracy of the film recorder characterization would have decreased by a factor of 2. The requirement that measurements be made in the projector system further strengthens the advantage of a model-based characterization technique over the exhaustive measurement procedure for populating a 3-D LUT.
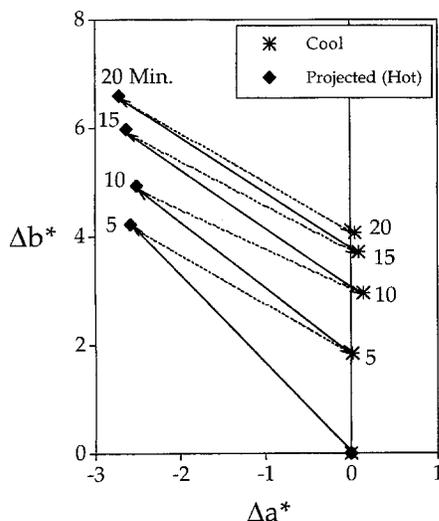
Figure 4. Color shift (CIELAB Δa^*-Δb^*) as the magenta slide is repeatedly heated in the projector, then allowed to cool.

4.3 Staining

Slide film also is degraded on projection through dye fading and the production of a yellow stain.[12,13] This was evaluated by measuring color samples after various durations of projection. After the first 4 min. of projection the average color shift was 1.5 CIELAB units. After an additional 8, 16, and 32 min., the average color differences increased to 2.0, 2.6, and 3.2 units, respectively. These results are summarized in Table 1. Figure 4 also shows the effects of the development of yellow staining products after prolonged projection. This is illustrated by the systematic shift in the Δb^* measurement in the yellow direction for both the hot and cool measurements as a function of projection time. The yellow-stain formation is not reversible, thus limiting the useful projection time of a given slide if color accuracy is a concern. Figure 5 summarizes the effects of thermochromism, a shift toward lower chroma with increased temperature, and staining, an overall shift toward yellow with prolonged projection. The effects of projection on the colors of photographic transparencies are significant and must be carefully considered (even with typical projectors such as the one used in these studies).
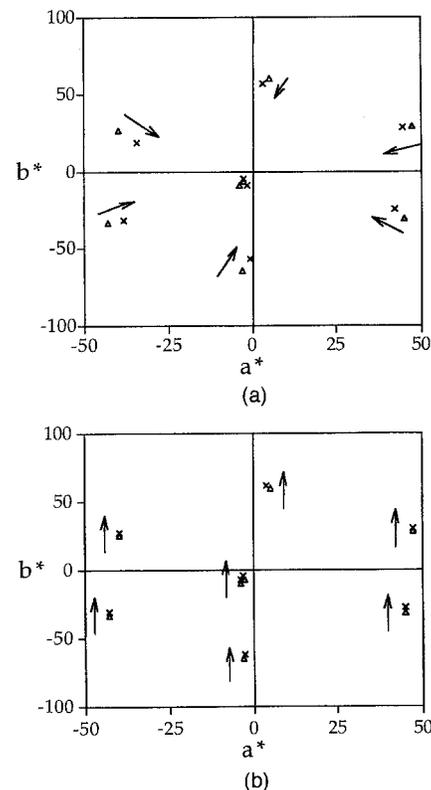
Figure 5. Comparison of color changes after 20 min of projection. With heating on projection (a) the dominant color shifts are toward lower chroma, and the degradation of the images due to projection (b) results in an overall shift toward yellow.

4.4 Spatial Uniformity

The spatial uniformity of the projection system was not addressed in the characterization. All measurements were made in the central portion of the image area and the image sizes were restricted to this area to minimize variation due to spatial nonuniformity. The uniformity of the system was evaluated and found to be mainly a luminance falloff from the center to the corners of approximately 10%—less than that typically found on CRT displays. Interestingly, the dark surround in which projected transparencies are typically viewed has the visual effect of increasing the per-

ceived brightness of image areas, and this effect is greater at the corners of the image than in the center.[4] Thus the visual effects of the dark surround serve to mask the nonuniformity of the projected image caused by lens falloff. This works in addition to the spatial properties of the visual system previously discussed[7] with respect to CRT displays to allow the spatial nonuniformity in the luminance of the system to be considered a negligible cause of error for this application.

4.5 Summary Recommendations

Several procedures were followed to assure the most accurate colorimetric presentation of the projected slide images. All slides were preconditioned in the projection system for 4 min. prior to any measurements or visual evaluations. The largest portion of the yellow-stain formation took place in the first 4 min. By preconditioning the slides, these large initial shifts were avoided in the measurements and observations. Each slide was placed in the projector for 75 s prior to being viewed by any observer or to being measured. This allowed the temperature, and therefore thermochromic shifts, to stabilize. After any slide was visually evaluated 8 times (16 min. projection), it was discarded and replaced with a new slide. After this duration of projection, the colors have shifted enough to be perceptibly different than when the slide was first viewed.[14] In addition, all the slides used in the experiments described later were exposed and processed on a single day to minimize processing variability.

The entire characterization process is dependent on the slide projector being used. Thus accurate characterization for projected slides must include characterization of the projection system, not just the film recorder. However, characterization of a film recorder system for typical slide projectors would still result in a substantial improvement in color reproduction quality for practical situations.

5 Color-Appearance Transformations

Once the CRT and film recorder systems are colorimetrically characterized, the issue of color appearance modeling must be addressed since the CRT images and projected slides are typically viewed with different white points, luminance levels, and surrounds. Simple reproduction of tristimulus values results in unsatisfactory color reproduction because these variables are not taken into consideration. Color appearance models must be used. Several models were implemented and psychophysically evaluated as described in the following sections.

5.1 Implementation

The models evaluated were RLAB[2,15] Hunt[16] CIELAB,[3] and von Kries.[17] The model published by Nayatani *et al.*[18] was not included since it was previously shown to produce unacceptable images; furthermore, because of the many complications described earlier, the number of experimental variables had to be minimized.

Two CRT setups were evaluated. The first had whitepoint chromaticities approximating CIE illuminant D65 at a luminance (white) of 53 cd/m², and the second had whitepoint chromaticities approximating CIE illuminant D93 at a luminance of 60 cd/m². The CRT images were always viewed with a gray background filling the remainder of the CRT display and a dim surround consisting of a room illuminated with cool-white fluorescent lighting at an illuminance of 345 lx. None of the ambient light was allowed to reflect off the CRT face. The slide projector system had a white-point correlated color temperature of 3863 K. The luminance of white in the projected slides was 109 cd/m² using the projection configuration described here. The slide images had a gray background and were viewed with a dark surround. The CRT and projected images were arranged such that they subtended the same visual angle (approx. 7 deg) from the observers' position.

A room was specifically designed for the viewing and comparison of CRT and projected images. The observer was seated in a rotating chair situated such that the viewing distance was equal to approximately 4 image heights for both the CRT image and the projected transparency. The observer was seated at a desk 70 cm in front of the CRT display and 290 cm from the projection screen. In addition, the CRT was arranged such that it could be viewed simultaneously with a projected image if desired. A 150-cm square, highly Lambertian projection screen (Clarion M1300) was mounted on the wall directly in front of the observers. Two slide projectors, a Kodak Carousel 4400 and a Leica P-2200, with 90-mm flat-field lenses, were mounted directly behind and above the observers at a distance of 366 cm from the screen. All image slides were projected using the Leica projector and were placed in the projector for 75 s prior to allowing observers to view them (to assure thermochromic shifts had stabilized). During this period, observers viewed a uniforrn gray image projected with the Kodak projector to preserve adaptation. The image source was alternated between the two projectors using a manual shutter. The walls of the room in which the experiments were performed were painted a medium gray. The configuration of the images on both the CRT display and projected slides is illustrated, along with image dimensions, in Fig. 6.
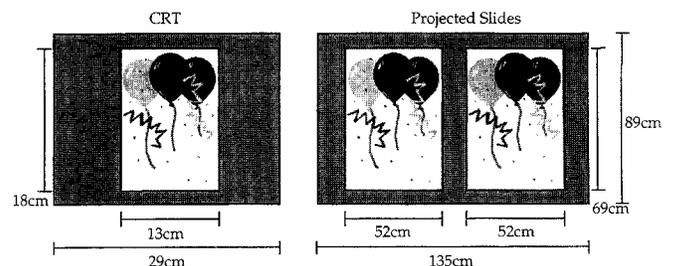
Figure 6. Arrangement and dimension of images in both media during the matching and preference experiments. Images were presented on medium-gray (luminance = 20% of white) backgrounds.

Each model was implemented as published with no specific optimization for the viewing conditions. The RLAB and Hunt models were implemented for a dim surround on the CRT and dark surround for the slides, and it was assumed that there was no cognitive "discounting the illuminant" for either display. (This also means that the Helson-Judd effect was predicted to occur according to the Hunt model.) A medium-gray background (20%) was used and implemented in the Hunt model. The von Kries and

CIELAB models only account for changes in the white point. The von Kries model was implemented using the same fundamentals used in the RLAB and Hunt models and coefficients defined as the inverse of the white-point cone excitations.

5.2 Psychophysical Tests

The CRT images were treated as originals and the desired reproductions for each set of viewing conditions were produced according to each of the 4 models. Three different pictorial images were used. These included a pictorial image of a woman holding balloons with a cityscape in the background ("Balloon"), a 2-D business graphic image with a wide variety of subtle colors including simple drawings of two people and colored text ("Japan"), and a 3-D synthetic image of a Kodak film box and a reflecting surface with a gray-scale test target appended ("EK Box"). The images were chosen to minimize the number used (due to experimental complexities) and incorporate the range of image types likely to be encountered in CRT-to-slide reproduction applications. Problems of gamut mapping were avoided by gamut compressing the original images such that no attempts were made to produce out-of-gamut colors. For one image ("EK Box," D93 CRT white point), the prediction of the Hunt model did not allow a reproduction to be made without excessive gamut compression; thus the Hunt model was not used for that particular condition.

Slides were produced that each contained a pair of reproductions. Each possible pair was produced. A total of 15 color-normal observers performed 4 phases of observations. Observers ranged in age from 21 to 43 years and had a variety of experience in judgement of color reproduction quality. For each CRT white point, a preference and a matching experiment were completed. The preference experiment was performed first. Observers were asked, for each pair, which of the two images they preferred. This was done with no knowledge of the original images. In the matching experiment, observers first studied the CRT original, and then made judgements of the slides by choosing which of each pair was the best color match to the original. Observations were carried out in two sessions, each of approximately 30-min. duration. The experimental technique, short-term memory matching, has been determined to be the most appropriate experimental technique for crossmedia image comparisons in an extensive study of viewing techniques carried out for CRT-to-print reproduction.[19] The preference and matching data were analyzed using Thurstone's law of comparative judgements[19,20] to derive interval scales and uncertainties of image preference and match quality for each CRT white point and image. Lastly, observers were asked to respond on whether or not each preferred image was an acceptable reproduction of the CRT original. The context for acceptability judgements was defined as the reproduction of computer-generated slides for a professional presentation.

6 Results

Figure 7 shows a comparison between the scale values obtained for the preference and matching experiments for both CRT white points. It is clear that the results are not highly correlated ($R^2 = 0.063$ for D93 and $R^2 = 0.358$ for D65). One would expect a linear relationship with unit slope if observers were using the same criteria for both the preference and matching responses. Thus it can be concluded that the observers were indeed making different judgments when asked to choose images that were color appearance matches to the originals than when they were simply asked to choose their preferred image with no knowledge of the original.
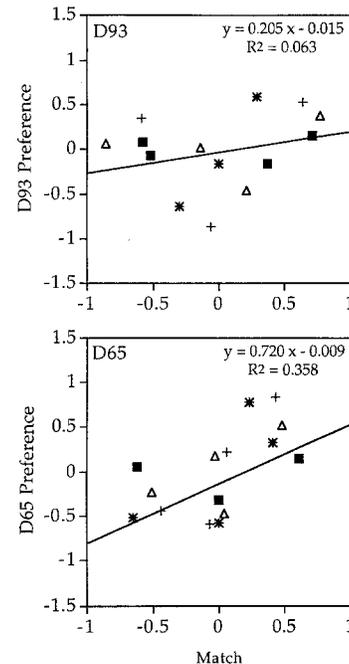
Figure 7. Comparison of preference scaling results with image matching results for both CRT white points.

Figures 8 and 9 show the color appearance match quality scales for each of the models for the D93 and D65 CRT white points, respectively. These figures include results averaged over all of the images as well as the individual image results. The overall average results were obtained by combining the raw data prior to performing the analysis rather than by averaging the individual image scale values. The error bars in the plots represent 95% confidence intervals on the scale values. Thus if the mean scale value for one model prediction falls within the error bars for another model, their performance is not statistically significantly different. When averaged over all images, the RLAB model performed statistically significantly ($\alpha = 0.05$) better for the image matching task than all the other models. The general trend, although not the statistical significance, holds up for each of the images. There is a significant image dependency that seems to be correlated with the dominant colors in the images and the tendency for certain models to perform better or worse than others for particular colors. There is no data point for the Hunt model for the "Balloon" image, D93, and the "Japan" image, D65, situations because the prediction of this model was never chosen as the best match in any paired comparison, thus eliminating the possibility of deriving scale values. The "EK Box" image, D93, combination was not included for the Hunt model due to gamut restriction difficulties.
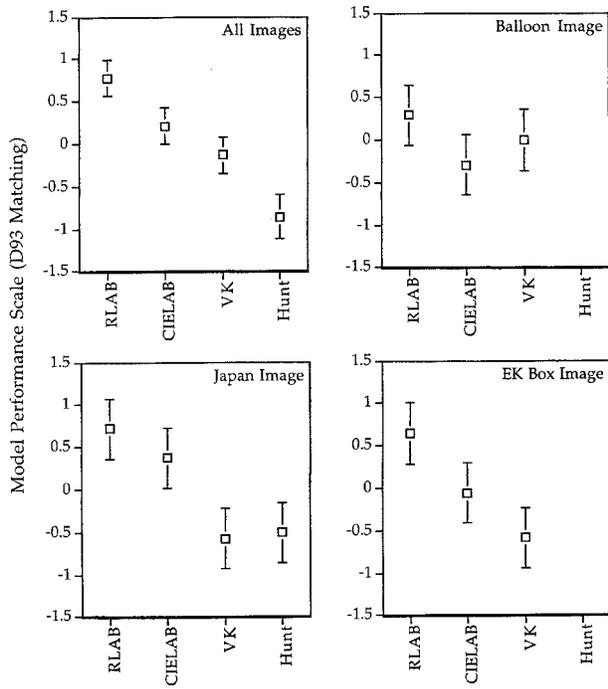
Figure 8. Model performance scales for the image matching experiment with D93 CRT white point.
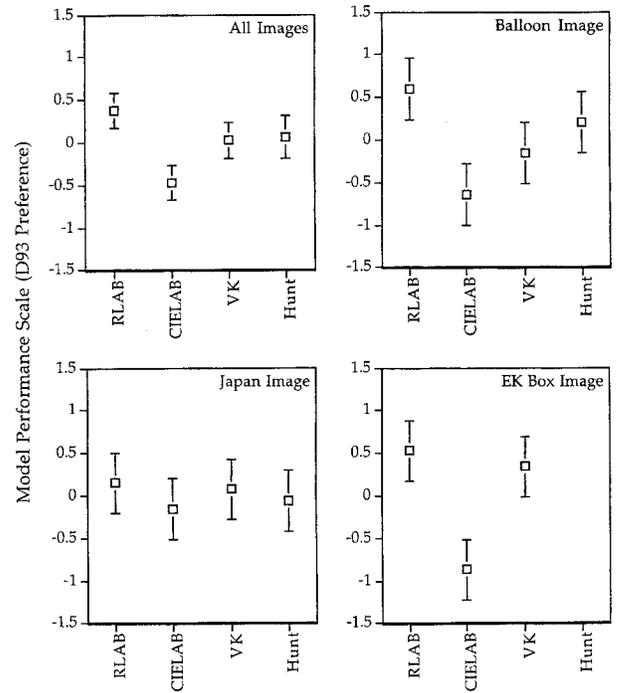
Figure 10. Model performance scales for the ment with D93 CRT white point.
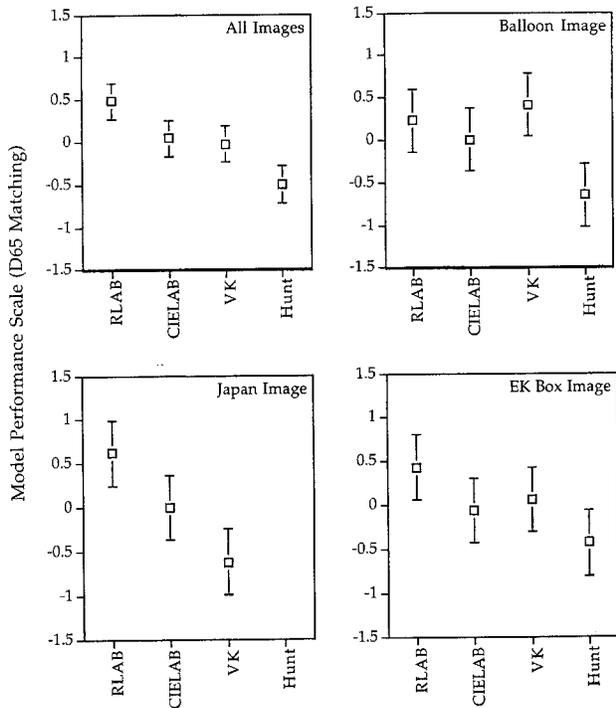
Figure 9. Model performance scales for the image matching experiment with D65 CRT white point.
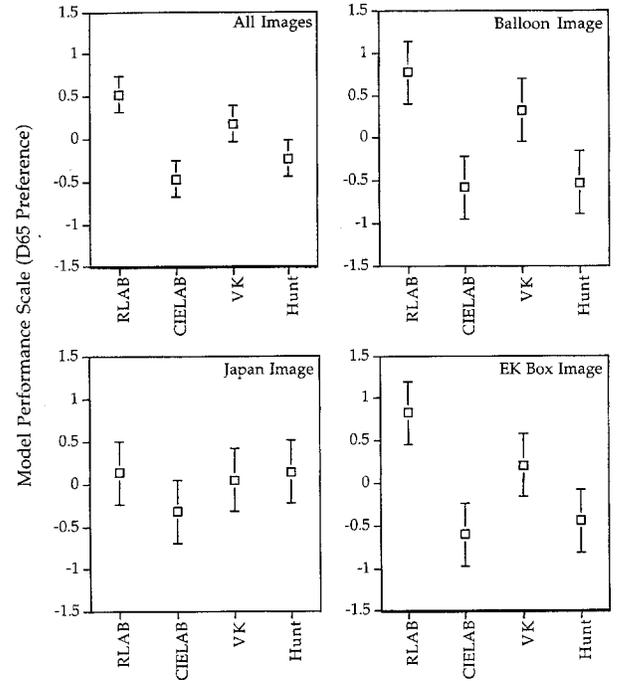
Figure 11. Model performance scales for the image preference experiment with D65 CRT white point.

Figures 10 and 11 show similar results for the image preference experiment. In the preference experiment, there is a greater degree of image dependence and significantly different results. However, the RLAB model still performed best overall. The image preference results must be carefully interpreted, because they include aesthetic judgements and the results could depend substantially on the choice of original image setup as well as content. For

example, if the original image was of very low contrast, a good color appearance model would produce a reproduction of equally low contrast, and a poor color appearance model might produce a higher contrast reproduction. Thus, the reproduction with higher contrast might be preferred, although it remains a poor match to the original, simply because a poor original was used. This situation was avoided as much as possible by trying to optimize the originals for

Table 2. Rank order of model performance in each of the 4 experimental phases and percent acceptability results.

Model	D65 Match	D65 Pref.	D93 Match	D93 Pref.	Acceptable Percent
RLAB	1	1	1	1	78%
CIELAB	2	4	2	4	51%
von Kries	2	2	3	2	54%
Hunt	4	3	4	2	30%

color quality; however, this goal had to be compromised in the process of gamut limiting the originals to avoid gamut mapping difficulties. The result of the RLAB model performing best in both image matching and image preference experiments (while the other models performed very differently in the two experiments) lends greater support to the utility of the RLAB model in crossmedia image reproduction applications.

Table 2 summarizes the experimental results showing the rank order of each model's performance in the four experimental phases averaged over all three images. Scale values that are statistically distinct with 95% confidence are given different ranks. In each experiment, the RLAB model performed best. The Hunt model performed worst for the color matching experiments, but better for the image preference experiments. The results were significantly image dependent. However, the RLAB model performed best in all cases. The image preference judgements are significantly different than the image matching judgements, affirming that observers are indeed making distinct judgements for the two tasks. Again, the RLAB model was best in both circumstances. The observers were also asked to respond on whether or not the best matching image was an acceptable reproduction of CRT original. For the RLAB model, observers thought the reproductions were acceptable 78% of the time. The acceptability of the other models was substantially lower, as illustrated in Table 2.

The rather poor performance of the Hunt model can be traced to two factors. The first is the inclusion of the Helson-Judd effect prediction, which causes a hue shift in the gray scale as a function of lightness. This prediction was included since the model was implemented strictly according to procedures that were originally published. In the Hunt model, when the cognitive "discounting-the-illuminant" does not occur, adaptation is incomplete and the Helson-Judd effect is present. The RLAB model does not predict the Helson-Judd effect (in fact it is never observed in practical viewing conditions), but does predict incomplete chromatic adaptation. Recently, Hunt and Luo[21] have suggested that for projected slides the Hunt model should be implemented with incomplete adaptation, but with no Helson-Judd effect. This would result in predictions much more similar to the RLAB predictions. Also, the surround and luminance-level compensation in the Hunt model is apparently more severe than that in RLAB. This combined with the HelsonJudd effect prediction cause the Hunt model to produce images that cause much more severe gamut mapping challenges, because the model requires reproduction of a greater luminance dynamic range than the imaging system is capable of producing.

7 Conclusion

The accurate colorimetric characterization of a CRT-to-projected-slide color reproduction system provides significant challenges. However, the results obtained are quite striking and far superior to the reproductions typically obtained with film recorders that have not been colorimetrically calibrated and characterized.

The model-based approach outlined in this paper is both accurate and efficient. The CRT display can be calibrated with only 14 calorimetric measurements and the film recorder system requires only 21 measurements after the film has been characterized. This allows for efficient recharacterization of the system when various system parameters are changed. Exhaustive measurement characterization techniques cannot be used to update characterizations as efficiently and have no implementation advantage since the models can be used to build 3-D LUTs to implement highspeed image transformations. In fact, an exhaustive measurement characterization of the film recorder and projection system would be impossible within the practical constraints of film batch variability, processing variability, measurement time, and projector bulb life.

The RLAB color appearance model proved to be a satisfactory mechanism for transforming image color appearance data from a CRT display viewed in a dim surround to a projected slide viewed in a dark surround. The more complicated Hunt model could make similar predictions if its parameters were optimized. However, the added complexity of the Hunt model might not be necessary in image reproduction applications for which the precision of color appearance judgements seems to be reduced and the accuracy of colorimetric device characterization is limited.

Acknowledgments

References

1. M. D. Fairchild, "Some hidden requirements for device-independent color imaging," *SID International Symposium*, 865-868 (1994).
2. M. D. Fairchild and R. S. Berns, "Image color-appearance specification through extension of CIELAB," *Color Res. Appl.* **18**, 178-190 (1993).
3. *Colorimetry*, CIE Publication No. **15**.2, Vienna, Austria (1986).
4. M. D. Fairchild, "Considering the surround in device-independent color imaging," *Color Res. Appl.* **21** (in press).
5. R. S. Berns, R. J. Motta, and M. E. Gorzynski, "CRT colorimetry. Part I: theory and practice," *Color Res. Appl.* **18**, 299-314 (1993).
6. R. S. Berns, M. E. Gorzynski, and R. J. Motta, "CRT colorimetry, Part II. metrology," *Color Res. Appl.* **18**, 315-325 (1993).
7. D. H. Kelly, Ed., *Visual Science and Engineering: Models and Applications*, Chap. 3, Marcel Dekker, New York (1994).
8. R. S. Berns, "Colorimetric characterization of the Solitaire 16 film recorder for Kodak Ektachrome 100 Plus Professional: a pilot study," *Munsell Color Science Laboratory Technical Report* (**May** 1993).
9. A. A. Lester, "Color reproduction of CRTdisplayed images

as projected transparencies," *MS Thesis*, Rochester Institute of Technology (1994) .

10. F. Grum and C. J. Bartleson, *Optical Radiation Measurements*, Vol. **2**, Color Measurement, Chap. 1, Academic Press, Orlando, FL (1984).

11. R. S. Berns, "Spectral modeling of a dye-diffusion thermal-transfer printer," *J. Elec. Imaging* **2**, 359-370 (1993).

12. A. A. Lester and M. D. Fairchild, "Thermochromism of Ektachrome l00 Plus transparencies upon projection," *J. Im. Sci. Tech.* **38**, 332-338 (1994).

13. H. Wilhelm and C. Brower, *The Permanence and Care of Color Photographs: Traditional and Digital Color Prints, Color Negatives, Slides, and Motion Pictures*, Chap. 6, Preservation Publishing Co., Grinell, Iowa (1993).

14. M. Stokes, M. D. Fairchild, and R. S. Berns, "Precision requirements for digital color reproduction," *ACM Trans. Graphics* **11**, 406 422 (1992).

15. M. D. Fairchild, "Visual evaluation and evolution of the RLAB color space*," IS&T/SID 2nd Color Imaging Conf.* 9-13 (1994); (see page 85, this publication).

16. R. W. G. Hunt, "An improved predictor of colourfulness in a model of color vision" *Color Res. Appl.* **19**, 23-26 (1994).

17. J. von Kries, "Chromatic adaptation." in *Festschrift der Albrecht-Ludwig Universitat*, Fribourg, Germany (1902); translation: D. L. MacAdam, *Sources of Color Science*, MIT Press, Cambridge (1970).

18. Y. Nayatani, K. Takahama, H. Sobagaki, and K. Hashimoto, "Color appearance model and chromatic adaptation transform," *Color Res. Appl.* **15**, 210-221 (1990).

19. K. Braun, M. D. Fairchild, and P. J. Alessi, "Viewing environments for cross-media image comparisons " *Color Res. Appl.* **21** (in press).

20. W. S. Torgerson, *Theory and Methods of Scaling*, Chap. 9, Wiley, New York (1967).

21. R. W. G. Hunt and M. R. Luo, "Evaluation of a model of color vision by magnitude scalings: discussion of collected results," *Color Res. Appl.* **19**, 27-33 (1994).