Evaluating CATs as Predictors of Observer Adjustments in Softcopy Fine Art Reproduction

Jun Jiang, Franziska Frey, and Susan Farnand. Munsell Color Science Laboratory, Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA

Abstract

A project to evaluate current practices in fine art image reproduction is conducted in which pieces of artwork in various media are being imaged by participating museums. As part of this project, observers were asked to make adjustments in softcopy fine art reproductions. The goal is to see how people working in museums, libraries and archives make color adjustments to artwork presented to them on screen. Observers were led through various interfaces allowing them to adjust the image seen on the screen to better represent the original in a light booth. They were asked to make adjustments until the screen image was 'good enough,' as an exact match may neither be possible nor necessary for us to detect relevant patterns in adjustment among observers. Patterns or trends in adjustments by observers can be used as an indication of how images should be processed to match with the adjustment by observers most closely. The adjustments by observers were compared with the prediction by three chromatic adaptation models. Overall the Fairchild92 model outperforms the Bradford and CAT02 transformations in matching with adjustments by observers more closely.

Introduction

The use of soft proofing in cross-media color reproductions is becoming more and more popular with the wide availability of display and computing technologies. In museums, visual editing and retouching of digital images of the collections are performed by experts for online access by visitors and researchers in the form of catalogs and postcards, for example. The demand to have the softcopy match the appearance of the original hardcopy closely is apparent. To achieve the goal, not only do we need to understand how images of artwork are visually edited, but we also need to learn the difference in perceiving color on self-luminous objects (e.g., display) and hardcopy surfaces (e.g., paper). The use of soft proofing in the hardcopy workflow is not yet well established in the museum world. This experiment will also inform us on what needs to be done to build acceptance for the use of soft proofing in this field.

When looking at a piece of paper under incandescent lighting, the paper appears white. However, if the chromaticity of the paper under the incandescent light is set as the display white point, we will have a difficult time seeing the display color as white. As for hard copies we are more likely to discount the illuminant color, while for self-luminous displays we hardly fully adapt to the white point, if it is further away from that of the natural daylight. ‘Discounting the illuminant’ refers to the cognitive ability of observers to interpret the colors of objects based on the illuminated environment in which they are viewed.[1] While cognitive mechanism relies on the observers’ knowledge of the illuminant, thus being inactive when viewing softcopy, sensory mechanism is always active, as it automatically responds to the stimulus energy.[1] Modern chromatic adaptation models are able to predict appearance matches across different media by accounting for incomplete chromatic adaptation. As a result, cross-media color reproduction is facilitated by using such calculations to predict color matches across different media and illumination conditions.[2]

In the experiment, observers were asked to adjust images off the camera to match with the original artwork in a light booth. Three chromatic adaptation transforms (CATs), Bradford,[3] Fairchild92[4] and CAT02[5, 6] were selected to predict adjustments by observers. Bradford transformation is essentially a von Kries transformation with an additional exponential nonlinearity on the blue channel.[1] In the experiment, the linearized Bradford transformation is included, given that it is the default chromatic adaptation in the latest ICC profile specification (ICC Version 4.2.0.0). The simplified Bradford transformation does not account for incomplete adaptation, while the CAT02 and Fairchild92 models are linear in nature and both can predict incomplete adaptation. Another distinction is that the Bradford and CAT02 models transform from tristimulus values to a ‘spectrally sharpened’ cone space while the Fairchild92 model converts to cone response directly by the Hunt-Pointer-Estevez (HPE) matrix. The von Kries predictions obtained using sharpened responsivities tend to be more color constant than von Kries predictions obtained using cone responsivities.[1] However, negative responsivity at some wavelengths are found in ‘spectrally sharpened’ cone space, thus making it physiologically implausible.[1] It is still under debate whether HPE or CAT02 matrices yield more accurate prediction for chromatic adaptaion.[7]

Experiments

Observers were asked to adjust the softcopy on the display to match with the original in the light booth. Because it was of interest to learn how experts visually edit images, observers were allowed to make adjustments rather than having them provide data for static stimuli. The experimental setup is shown in Figure 1.

A 30” Apple Cinema Display was used for showing softcopy reproductions, and an LMT 1210 colorimeter was used to characterize the display. The display characterization model proposed and detailed by Day, Taplin and Berns[8] was followed to ensure accurate mappings between LCD digital counts and XYZ tristimulus values. Display white point and luminance were adjusted to match with those of the light booth by using a Halon perfect reflecting diffuser (PRD). Additionally, the luminance and chromaticity of the background of the light booth were measured using a PhotoReserach-650 spectroradiometer. The background of the software interface was adjusted to match these settings. The colorimetric performance of the display was evaluated. The mean and max color differences were 0.62 and 1.42, respectively.

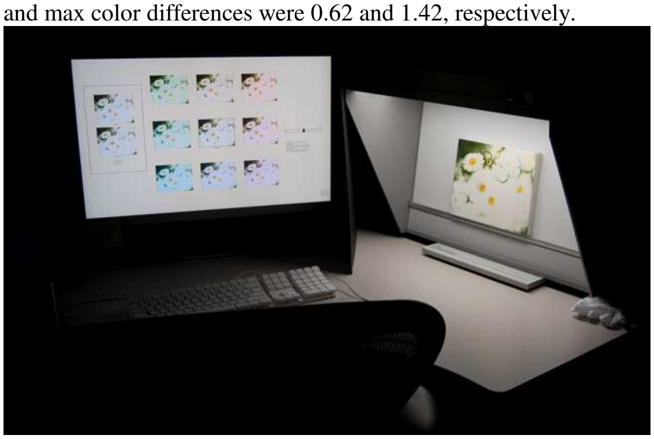

Figure 1. *Lab setup*

Seventeen observers participated in the experiment. Their ages range from early 20s to mid-70s. Most observers are working in the area of artwork reproduction in museums, libraries or archives. Observers were divided into two groups to adjust two different sets of images. The first set included *Daisy*, *Night Sky*, *Orchid*, and *Photo*, and the second set included *Orchid*, *Bridge*, *Daisy*, and *Aquatint*. The first image in each set was used for training, thus being excluded from the analysis. The reproductions of all six paintings from one institution are shown from Figure 14 to 19 in the Appendix as examples. They can also be accessed at http://artimaging.rit.edu/research/images.

Source images from different museums were downsized in Photoshop© to fit the display. A chromatic adaptation was usually needed if the white point of the source color space (AdobeRGB in this case) was different from that of the display. However, in this experiment, the chromatic adaptation was not used. The chromatic adaptation model that predicted the adjustments by observer most closely was to be investigated so that a closer starting point could be determined.

To evaluate the performance of the chromatic adaptation models, a color difference equation, CIEDE2000[9] was used. Color difference equations were derived from comparisons of simple color patches under a controlled environment, and therefore, it might be insufficient to tell the color difference for complex images, such as artwork reproductions. For example, an original and its halftone reproduction would look almost identical to each other, but calculating their color difference pixel-by-pixel would dramatically overestimate the 'perceptual difference' between the two.

To eliminate details in images that could not be differentiated by human eyes due to spatial frequency, a spatial extension to the CIELAB system, S-CIELAB,[10] was used. An input image was initially converted into one luminance and two chrominance color components. Each component image was then passed through a spatial filter that was selected according to the spatial sensitivity of the human eye for that color component. The final filtered images were transformed into XYZ format so that the color difference equation could be applied.[11]

The mixed-effect Analysis of variance (ANOVA[12]) was used to understand the image difference. Several factors were identified in this experiment: *Chromatic Adaptation Transform (CAT)*, *Observer*, and *Image*. *CAT* was a treatment factor and it was fixed, as the three levels (the Bradford, Fairchild92, and CAT02 models) were of special interest. *Observer* was a random factor, because the participants in the experiment were not themselves of interest. Of more interest was how a large population of people in museums did visual editing. Similarly, *Image* was supposed to be a random factor. However, the number of images (three in each group) was not large enough to represent the whole population of images. A more reasonable alternative was to focus on these testing images. All three main factors and their two-way interactions were included in the full mathematical model. A 0.05 confidence level was used to distinguish significant factors from redundant ones. The statistical analysis was done in Minitab©.

User Interface

Software was developed in Matlab© on the extension provided by the high-level Psychophysics Toolbox©. Each image could be edited on a global and local scale. The hue adjustment interface is shown in Figure 2.

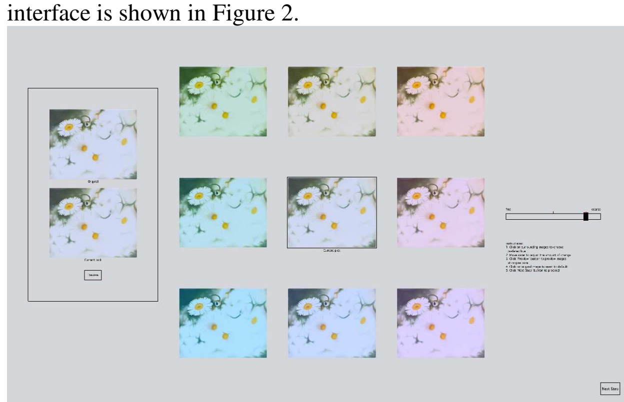

Figure 2. *Image hue adjustment interface*

In Figure 2, eight surrounding images were of the same lightness and same increment in chroma from the central one, but of different hue. The image hue could be adjusted by clicking one of the surrounding images around the central image (current pick). When one of the surrounding images was selected, the selected image appeared in the center and all the other surrounding images shifted in hue based on the central one accordingly.

Once the hue adjustments were complete, observers moved on to the global adjustment interface as shown in Figure 3. Image brightness, contrast, saturation, and sharpness could be adjusted by the sliders to the right of the image.

The global adjustments were indiscriminate to colors, and therefore the local adjustment tools in Figure 4 could be used in order to make certain colors right without affecting other colors in the image. The images after local adjustments were compared with predictions by CATs. CATs were derived by fitting adjustment data by observers, even though what were predicted in the experiment were complex images rather than simple ones.

In addition, while CATs were global operators, they by no means changed colors in the same amount. The refinement on a local scale became a useful complement to the global adjustments to ensure the most accurate results by observers.

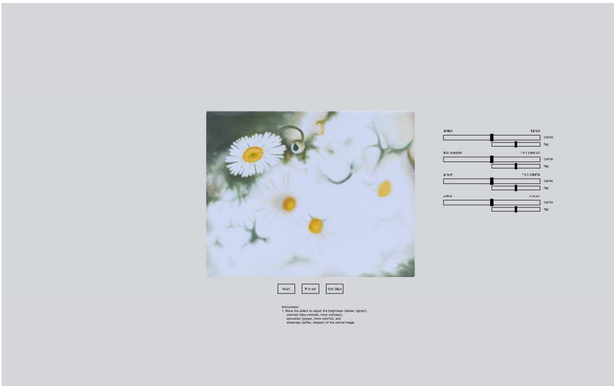

Figure 3. *Global color adjustment interface*

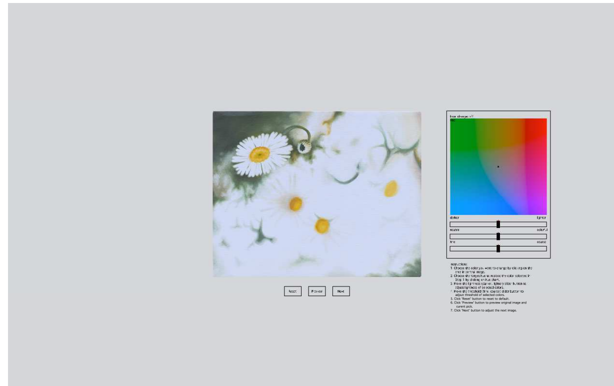

Figure 4. *Local color adjustment interface*

Results

In Figure 5, hue adjustments in a* and b*, and global adjustments in lightness, contrast, chroma and sharpness for each image are shown. The value of 1 on the y-axis meant average adjustment was zero across observers on that dimension. While the adjustments were image-dependent, certain general trends could still be detected. In the top central plot, observers increased b* by a noticeable margin for all images except for *Night sky*. As the images were encoded in XYZ with D65 as default white point, the images appeared bluish on the D50 display compared with the original in the light booth. Observers were trying to make images more yellowish. From the plot on the bottom right, sharpness was enhanced for all images except for *Aquatint*. The increase in sharpness might result from the loss in details when the images were captured or during downsizing of images. Different interpolation methods to downsize images in Photoshop© could even complicate the workflow in preparing images.

Image-dependent information could also be learned from Figure 5. Heavy impasto could be found in the sky in the hardcopy of *Night sky*, which was difficult for reproduction. The sharpness plot in Figure 5 shows that observers increased the sharpness of *Night Sky* much more than they did to all the other images. Additionally, the increases in lightness and contrast were most evident for *Photo* in Figure 5, indicating a loss of lightness dynamic range in the softcopy of *Photo*.

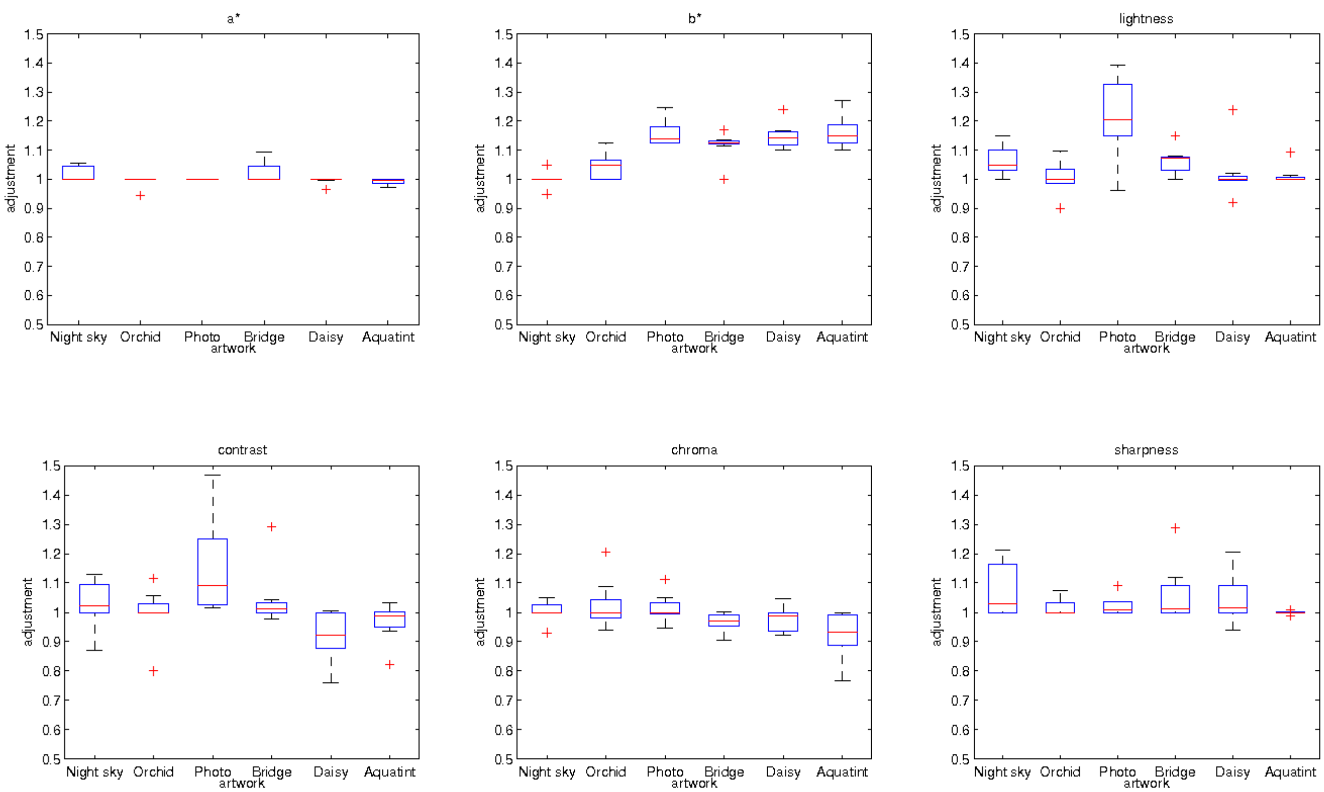

Figure 5. *Global color adjustments of each image*

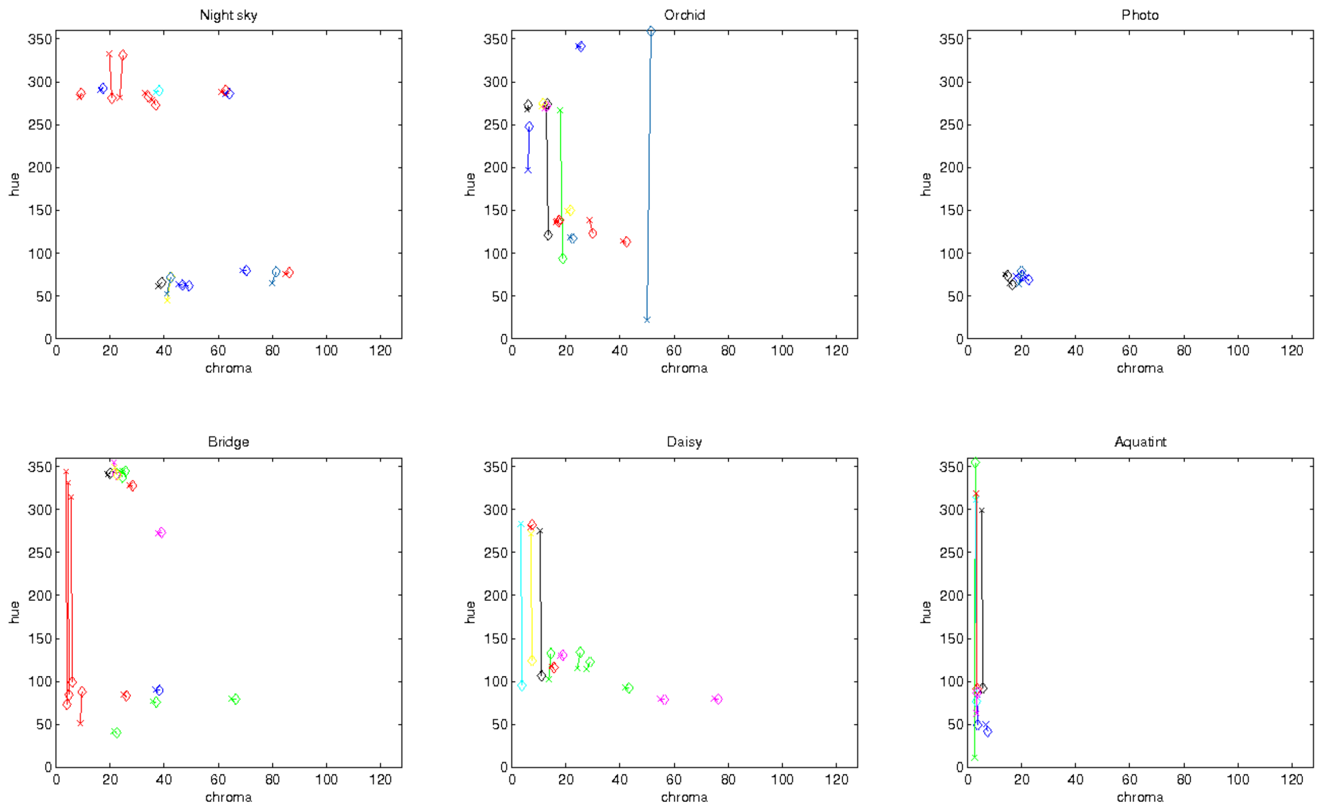

Figure 6. *Local color adjustments for each image*

Local adjustments are shown in Figure 6. The y-axis is hue angle (in degree unit) and the x-axis is chroma. The cross 'x' at one end of each line was the color selected for adjustments, and the diamond was the target color. Lines of the same color in Figure 6 were local changes made by one observer.

For *Night sky* (top left plot in Figure 6), most adjustments were made at two hue angles, around 50 degrees and 270 degrees. The adjustments made at around 50 degrees were generally pointing toward more chromatic direction, indicating the reproduced yellowish color in the softcopy was not chromatic enough. It might result from the corresponding color matching from D65 to D50. Or the color of the sky (blue) was adjusted to be right during global hue adjustments, while the yellowish color needed more refinement. The colors adjusted at around 270 degrees were mostly of lower chroma.

For *Orchid* (top central plot in Figure 6), local changes were made to colors of chroma lower than 30. The longest line in the plot was misleading, as the hue was not changed that much. (Actually, its hue angle was changed from somewhere positive to negative around 0 degree.)

For *Photo* (top right plot in Figure 6), almost all adjustments were made at around 70 degrees on the hue axis, pointing towards more chromatic directions, because *Photo* was relatively neutral and its dominant hue was around 70 degrees.

Across all images, almost all big local changes in hue occurred in colors of low chroma. This may result from the fact that during global adjustments, observers might have concentrated more on the central object in the scene (such as the bridge or orchid) while areas of lower chroma (usually the background) underwent more changes during local adjustments.

Chromatic Adaptation Models

The Fairchild92, CAT02, and Bradford models were implemented and evaluated to determine which model better predicted the adjustments made by observers. *Daisy* adjusted by one observer and predicted by all three models are shown in Figure 7 and 8 as an example.

***Figure 7**. Daisy (source image in D65 XYZ)*

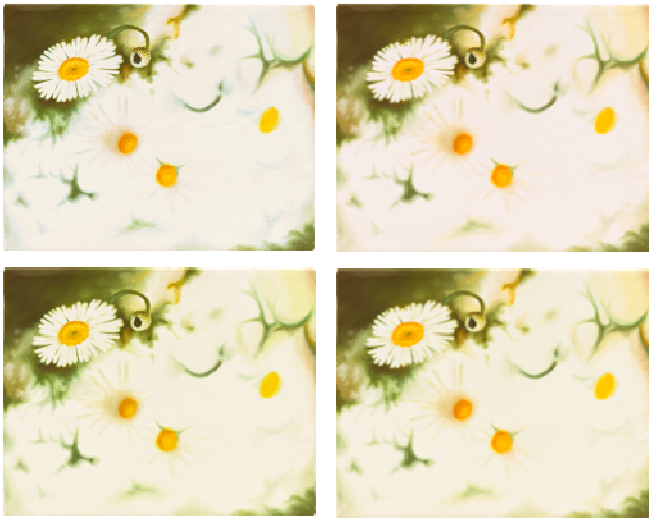

Figure 8. *(Left top) Adjusted image by one observer (Right top) Fairchild92 model (Left bottom) CAT02 model (Right bottom) Bradford transformation*

The original image in Figure 7 appeared noticeably more bluish on D50 display than the hardcopy in the D50 light booth, given the absence of chromatic adaptation before adjustments by observers. The adjusted image by one observer matches with the predictions by three chromatic adaptation models more closely in Figure 8 than with the original in Figure 7.

The S-CIELAB model was evaluated by comparing *Daisy* predicted by the Fairchild92 model before and after the spatial filtering. The details, such as the canvas in the top right plot in Figure 8, became unnoticeable in Figure 9. Color difference was calculated then between the adjusted image by observers and the image through chromatic adaptation models both after the spatial filtering by the S-CIELAB model.

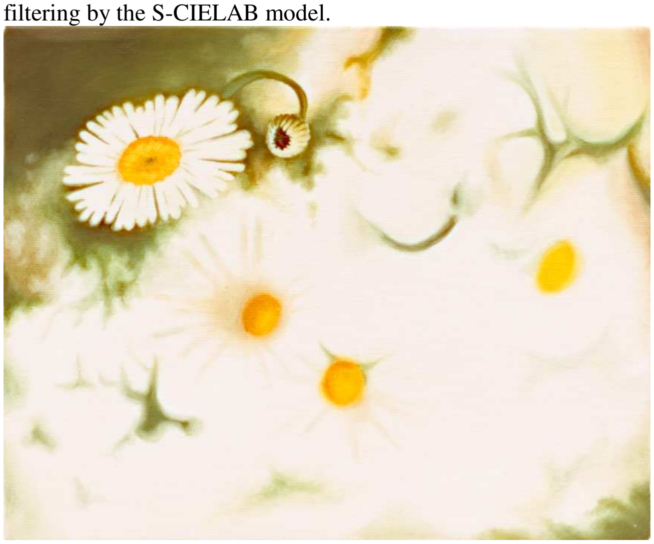

Figure 9. *Daisy by Fairchild92 model processed by S-CIELAB*

Color difference was also calculated when spatial filtering was absent. The mean color difference between adjusted images and outputs from the three chromatic adaptation models increased only by a small margin. One important reason was that the viewing angle assumed to be constant in the S-CIELAB model did not remain unchanged in the experiment, as observers were likely to lean ahead and sat really close to the screen when making adjustments.

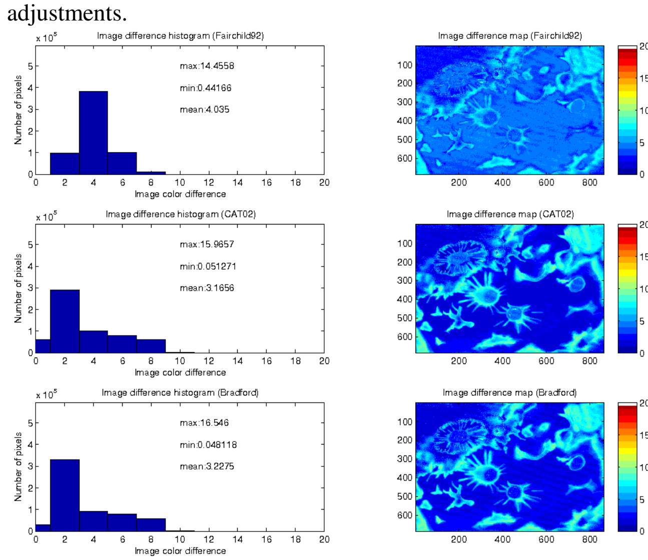

Figure 10. *Image difference map for Daisy adjusted by one observer*

The error distribution map of *Daisy* (when the S-CIELAB model was used) by one observer is shown in Figure 10. Daisy adjusted by the observer was predicted well by all three models, given mean image difference between 3 and 4. The pedals of daisy were predicted better than other parts of the painting, as indicated by the deep bluish color in the image difference map in Figure 10.

The ANOVA analysis was performed to discover main effects and interactions that contributed significantly to the image difference. The analysis was performed on two sets of images separately. For Group 1 (*Night Sky*, *Orchid*, and *Photo*), the interaction between *Observer* and *Model* was insignificant (p-value=0.376), indicating that no model predicted certain observer's adjustments significantly better than the adjustments by other observers. It was reasonable because these models were designed to match with the cone responsivities (or color matching functions) of average observers. The main effect, *CAT*, was significant (p-value=0.014), together with the interaction between *Image* and *Observer* (p-value=0.001), and that between *Image* and *model* (p-value=0.001). The *Image* and *Observer* main effects were insignificant (p-value=0.345 and 0.229 respectively), but they were retained in the model due to hierarchy. The adjusted R-Sq of the final model was over 95%. Given the significance of interaction terms in the model, an interaction plot was made to better understand the data as shown in Figure 11.

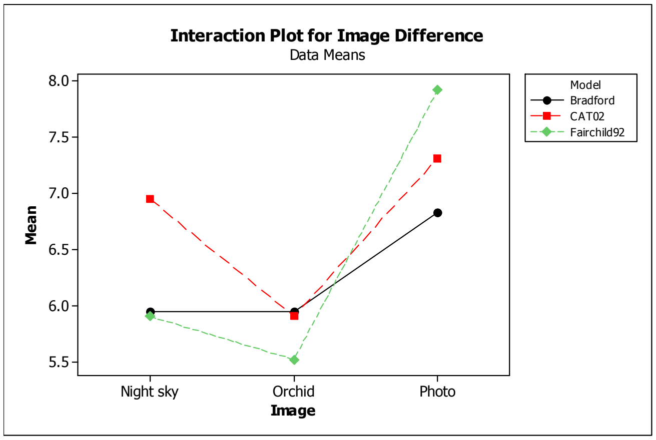

Figure 11. *Interaction plot between Image and CAT for 1st set of images*

In Figure 11, the performance of the Fairchild92 model was better than that of the other two models for *Night sky* and *Orchid* but not for *Photo*, as indicated by the lower mean image difference of the first two paintings by the Fairchild92 model.

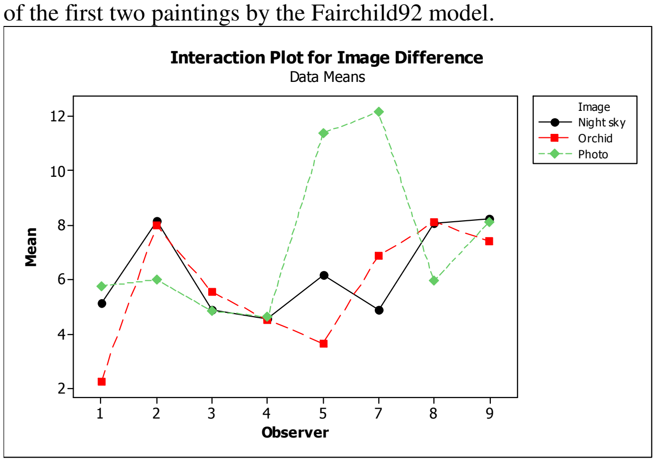

Figure 12. *Interaction plot between Image and Observer for 1st set of images*

In Figure 12, the interaction between *Image* and *Observer* is examined. No image was predicted to have the least mean colorimetric errors, thus being consistently closer to the adjusted images across all observers. It indicated the large variability in adjustments among observers, and the confounding effect between observers and images. The variability might result from a few factors. First, observers were asked to adjust three images in 30 minutes, while it could usually take them hours to adjust one image in their work. Secondly, not all observers were familiar with image editing software, which was confirmed from their feedback that the software had a learning curve.

The ANOVA analysis of the second set of images (*Bridge*, *Daisy* and *Aquatint*) was performed, and the interaction between *Image* and *Observer* was significant (p-value=0.001), in agreement with the result from the first image set.

The interaction between *Image* and *CAT* was significant (p-value=0.002), and the interaction plot is shown in Figure 13. The Fairchild92 model outperformed the other two models for *Bridge* and *Daisy* but not for *Aquatint*. *Aquatint* was similar to *Photo* in appearance in that neither had strong chromatic colors.

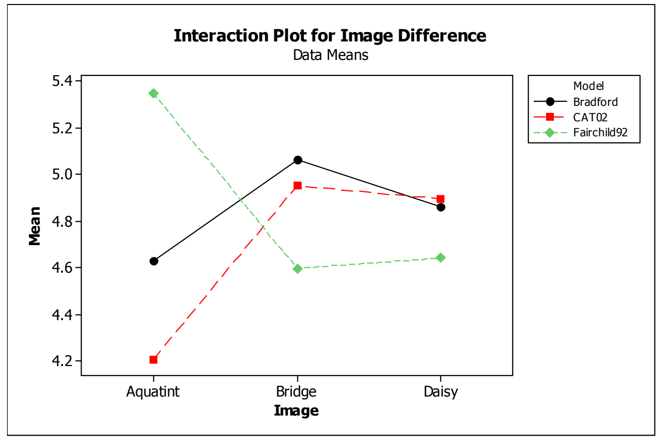

Figure 13*. Interaction plot between Image and CAT for 2nd set of images*

From the above analysis, no chromatic adaptation model was found to better predict adjustments by observers across all images used in the experiment. By comparing Figure 11 and 13, the Fairchild92 model generally matched with the visual editing by observers more closely than the Bradford or CAT02 model except for images with near-neutral appearance. The predictions by the Bradford transformation were not much worse than the other two. Given its inability to discount the illuminant and simplicity in implementation, the performance of the Bradford model was better than expected. However, common practice of comparing displays and reflection prints side-by-side produces unpredictable color appearance,[1] as viewing softcopy and hardcopy simultaneously might have caused the state of adaptation to be unstable. It has been noted that short-term memory matching technique produced more reliable results.[1]

Conclusions

An experiment investigating current soft proofing of artwork reproductions was made, during which observers were asked to make appearance match of artworks across two different media. While some general features could be extracted from adjustments by observers, the adjustments were also image dependent. The significance of the interaction between *Observer* and *Image* highlighted their confounded effect. Color adjustments by observers were compared with the predictions by the Bradford, Fairchild92 and CAT02 chromatic adaptation models. Based on the mixed-effect ANOVA analysis, the Fairchild92 model outperformed the Bradford and CAT02 model for all the testing images except those with neutral appearance. The S-CIELAB model was used to remove unperceivable details at certain spatial frequencies, but its effect was limited by the fact that the viewing distance did not always remain constant during adjustments.

Visual editing made by experts in museums sometimes involves more complex lighting, such as daylight from outside and fluorescent light in the office. To account for different viewing conditions, such as changes in illuminance level, background or surround, a color appearance model should be included in the future work. The lighting condition in a local museum was measured, and more realistic parameters would be used to serve accurate cross-media reproductions by color appearance models.

Appendix

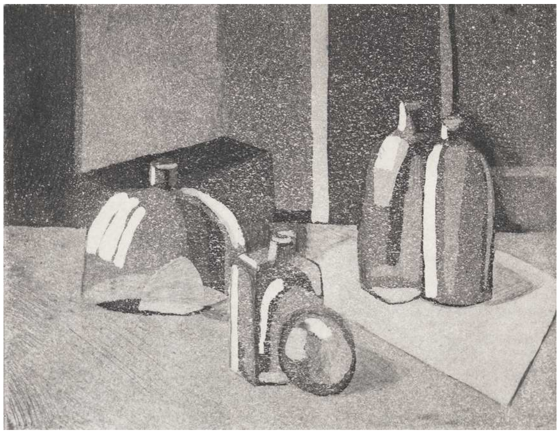

Figure 14. *Aquatint*

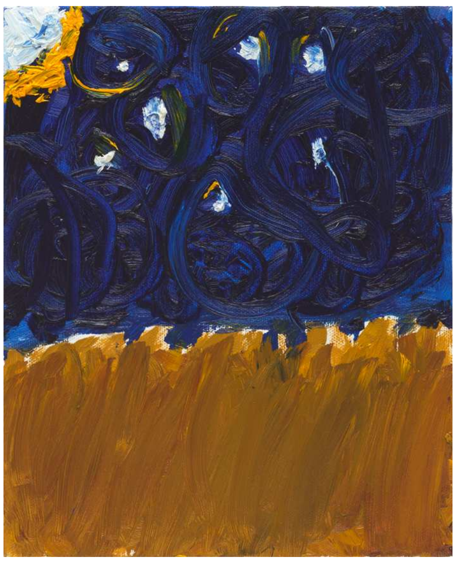

Figure 15. *Night Sky*

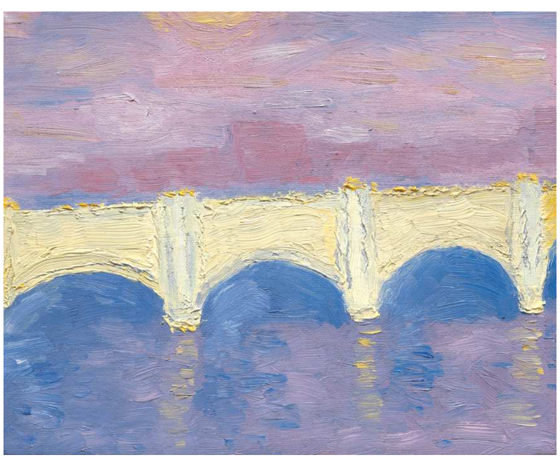

Figure 16. *Bridge*

Figure 17. *Daisy*

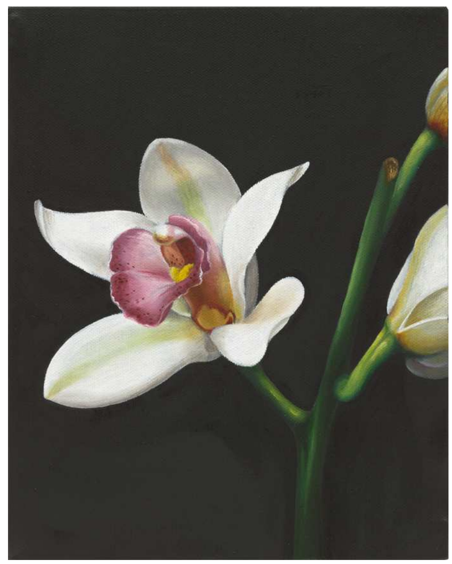

Figure 18. *Orchid*

Figure 19. *Photo*

References

[1] Mark D. Fairchild. Color Appearance Models 2nd edition.

[2] Mark D. Fairchild. Formulation and Testing of an incomplete-Chromatic-Adaptation Model. Color Research and Application Volume 16, Number 4, August 1991.

[3] International Color Consortium. Specification ICC.1: 2004-10 (Profile version 4.2.0.0)

[4] Karen M. Braun and Mark D. Fairchild. Testing Five Color-Appearance Models for Changes in Viewing Conditions. Color Research and Application Volume 22, Number 3, June 1997.

[5] Nathan Moroney, Mark D. Fairchild, Robert W.G. Hunt, Changjun Li, M. Ronnier Luo and Todd Newman. The CIECAM02 Color Appearance Model. IS&T/SID Tenth Color Imaging Conference

[6] CIE. A Color Appearance Model for Color Management System: CIECAM02. CIE 15x:2003.

[7] Changjun Li, Elisabet Chorro-Calderon M Ronnier Luo and Michael R Pointer. 17th Color Imaging Conference Final Program and Proceedings. 2009 Society for Imaging Science and Technology.

[8] Ellen A. Day, Lawrence Taplin, Roy S. Berns. Colorimetric Characterization of a Computer-Controlled Liquid Crystal Display. Color Research and Application. Volume 29, Number 5, October 2004.

[9] M. R. Luo, G. Cui, B. Rigg. The Development of the CIE 2000 Color-Difference Formula: CIEDE2000. Color Research and Application. Volume 26, Number 5, October 2001.

[10] Xuemei Zhang and Brian A. Wandell (1997). A Spatial Extension of CIELAB For Digital Color Image Reproduction. SID Journal.

[11] Xuemei Zhang, D. Amnon Silversein, Joyce E. Farrell. Color Image Quality Metrics S-CIELAB and Its Application on Halftone Texture Visibility. IEEE COMPCON97 Symposium Digest.

[12] Angela Dean and Daniel Voss. Design and Analysis of Experiments.

Author Biography

Jun Jiang is a second-year graduate student in the Munsell Color Science Laboratory in the Chester F. Carlson Center for Imaging Science at Rochester Institute of Technology. He is currently doing research on visual editing and cross-media color reproduction of artworks in museums.