Evaluating the Perceived Quality of Soft-copy Reproductions of Fine Art Images With and Without the Original Present

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 the perceived image quality of fine art reproductions was conducted in which pieces of artwork were imaged by participating institutions. As part of this project, observers were asked to rank fine art reproductions on a characterized display either with or without the presence of the originals. The goal was to see whether the availability of the original artwork influenced how people appreciated the reproductions. A low correlation was found between the ranking results with and without the originals, indicating a shift in the criterion employed by observers in evaluating perceived image quality from color accuracy to preference. A web-based experiment was designed in order to better understand the necessity of a controlled environment when evaluating image quality based on preference. A significantly high correlation was found between the results from the experiment without the original and from the one conducted online. Therefore, the preference judgments of perceived image quality were stable regardless of changes in viewing conditions. In addition, the areas that were considered most important by observers in making ranking decisions were identifiable by observers' clicks on the image in the web-based experiment. By understanding the part of the paintings to which more attention was drawn, information regarding the image saliency could be learned.

Introduction

The proliferation of digital images of works of art created in cultural heritage institutions has dramatically changed the access to the collections. The availability of the Internet further enables patrons from almost anywhere in the world to browse the holdings of cultural heritage institutions. In addition, a permanent visual record of a work of art can be made and sent to other institutions to allow comparison between objects to be made, even if the object itself cannot travel.[1]

Research shows that when the original artwork was present, an accurate reproduction of the original was usually desired. Research on improving color accuracy for museum applications is abundant. Advanced capturing systems were used in order to obtain high-resolution images. Algorithms were designed to simplify the process of visual editing.[2] In the past decade, spectral imaging techniques[3] were developed and introduced into the field of art conservation. With the aid of spectral imaging, the spectral reflectance factor could be reconstructed accurately while simultaneously achieving high colorimetric performance for a defined illuminant and observer.[4]

However, it is generally the situation that people experience fine art reproductions without the original artwork available for direct comparion.[5] As a result, the criterion used to evaluate the image quality might not always be color accuracy. It was of interest then to investigate the shift in criterion used by people judging soft copy reproductions with or without the presence of the original artwork. Experiments were therefore designed to understand the effect of the original on how the reproductions were judged.

Moreover, museum websites are generally visited by people from their own digital devices. Given the differences in display configurations and lighting conditions, the colorimetric accuracy achievable on the characterized display in a university laboratory became hardly relevant. Therefore, in order to learn the perceived quality of soft copy reproductions, a web-based experiment was conducted, in which reproductions of fine art would be evaluated by observers through the Internet.

Given more and more fine art images that can be accessed not only by conventional desktops but also from mobile devices, guidelines based on which soft copy reproductions of high quality could be attained became needed. Therefore, another goal of the project was to help create guidelines to improve the fine art reproductions in museums.

Experimental methodology

Soft copy reproductions of six images (*Aquatint*, *Bridge*, *Daisy*, *Firelight*, *Mountain* and *Photo*) from sixteen institutions were collected for the experiments. These images varied widely in color, texture and content. The reproductions from one institution are shown from Figure 14 to 19 in the Appendix as examples. They may also be accessed at http://artimaging.rit.edu/research/images. This site also provides more details on the participating museums, and how the reproductions were processed.[6]

Ranking experiments were designed and conducted to examine the perceived image quality of the reproductions from different institutions. Observers were asked to make pairwise comparisons of images on the display. The reason to choose the method of paired comparison over the method of magnitude estimation was to ensure consistency of image evaluations during the whole experiment. Given that the number of images that could fit on the interface was limited by both the display and image size, and that the difference between images might be small, observers were liable to forget how the images on previous interfaces looked, thus making it difficult to maintain consistency using magnitude estimation in the experiment.

On the other hand, it would take $m \cdot \binom{n}{2}$ comparisons to do a complete pairwise comparison between n reproductions of m images (each reproduction being compared to all the other ones only once but not to itself). With sixteen reproductions and six test images, the number of judgments became $\binom{16}{2} \cdot 6 = 720$ for one observer. Usually it took ten seconds to make a pairwise comparison decision. Given that it would take about two hours to finish the experiment, it was considered difficult for observers to stay engaged during the whole session, thus diminishing the accuracy of the collected data.

An adaptive form of paired comparison was therefore used to improve efficiency.[7, 8] For example, to rank reproductions one through four in Figure 1, comparisons between *1* and *2*, and between *3* and *4* were made. Based on the rules on the right in Figure 1, *1* was considered better than *2*, and *4* better than *3*. Therefore, both *1* and *4* were moved to the front, with *2* and *3* near the end of the rank. Similarly, based on the next rule on the second row in Figure 1, the positions of *1* and *4*, and those of *2* and *3* were interchanged. So far all the neighbors in the ranking had been compared except for *1* and *3*. Therefore, they were compared and their ranks were shifted according to the third rule (*3>1*). While reproduction *2* and *4* had not been compared yet, their relative positions were determined by the transitive relations based on the existing rules (*4>1* and *1>2*). The idea could be generalized when more reproductions were included.

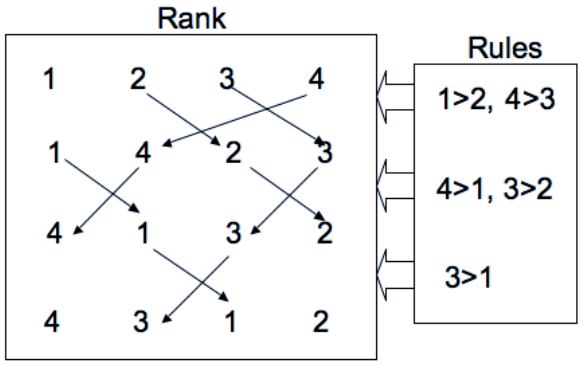

Figure 1. *An adaptive form of paired comparison used in the ranking experiment*

In the experiment, the preliminary ranks of the sixteen reproductions could be found after a few rounds of comparisons. Further adjustments were made by comparing the neighboring reproductions and moving them upper or lower in ranks. The process of comparisons would not end until the ranks of all the reproductions were determined either by the comparisons made by an observer or by the transitive relations based on the existing rules. After using the adaptive form of paired comparison, the number of comparisons was reduced to around 40 for each image, and the whole experiment would take one-third of its original time, about 30 to 40 minutes for all test images. While efficient, the method might be less accurate due to observer inconsistency.[8] A large pool of participants was therefore used to stabilize the ranking obtained from the partial paired comparison.

Experiment setup

Ranking experiments were conducted under three different conditions. The first two were performed on a characterized display in Munsell Color Science Laboratory (MCSL). The main difference between the first two was that the original artwork was available in the first experiment but not in the second one. The third experiment was run online, so that data could be gathered in an uncontrolled environment.

Ranking experiment with the original in MCSL

The experimental setup is shown in Figure 2. Both the soft copy images on the display and the original artwork in the light booth could be viewed side-by-side simultaneously by observers. During the experiment, observers were instructed to click on the image on the display that represented the original better.

A 30" Apple Cinema Display was used for showing softcopy reproductions, and an LMT 1210 colorimeter was used to characterize the display. The display was adjusted to have D65 white point and native gamma (2.2). The display characterization model proposed and detailed by Day, Taplin and Berns[9] 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 PhotoResearch-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 (CIEDE2000[10]) were 1.4 and 2.2, respectively.

Figure 2. Lab setup

Software interfaces were developed in Matlab® on the extension provided by the high-level Psychophysics Toolbox®.

Twenty-four observers participated in the experiment, fourteen of whom identified themselves as experts (curators, publishers, photographers, e.g.). Their ages ranged from mid-20s to early-60s.

Ranking experiment without the original in MCSL

The setup for the experiment was almost the same as the previous one shown in Figure 2 except that the light booth and originals were absent during the experiment. The instructions were changed accordingly. Observers were asked to choose the preferred image on the display.

Fourteen observers from early-20s to early-50s completed the experiment without the presence of the original. Among them five participants were experts working in the field of artwork reproduction. Four participants were shown the originals briefly prior to the experiment, while the rest were observers in the experiment with the original conducted approximately two months earlier. Therefore, all the observers had witnessed the original artwork, but some had fresher memory of the originals' appearance.

Web-based ranking experiment without the original

A web application, including a graphical user interface and a database on the backstage, was implemented to allow the assessment of fine art reproductions in an uncontrolled environment by observers from almost anywhere with little constraint on test conditions, as long as reasonable Internet speed and a web browser were available.

The main interface of the web-based application is shown in Figure 3. During the experiment, observers were asked not only to pick their preferred image, but also to click in the area of the image that most influenced their decision.

Eighty-eight observers did the experiment, with about an equal number of male and female participants. Observers' age ranged from 20 to 65. Given that observers did the experiment under their own testing conditions, the size of the display that was used by observers ranged widely from 13" to 30". In addition, about two-thirds of the people used Macintosh® operating system to do the experiment while the rest used Microsoft Windows® platform. The vast majority of observers finished the experiment either in Firefox® or Safari®, with a few using Internet Explorer® or Google Chrome®.

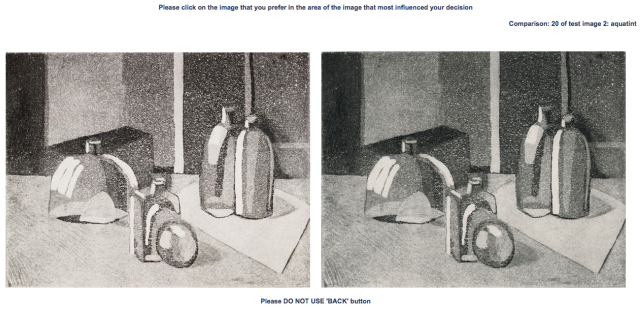

Figure 3. *Interface of web-based ranking experiment*

One of the sixteen reproductions used in the first two experiments was replaced by a reproduction that ranked well in a separate hard copy experiment.[5] The inclusion of it was to evaluate the perceived quality of its soft copy counterpart. As a result, when comparisons were made across all three experiments, the image that was replaced and the one included to replace that image were both excluded.

The software was built in Java© in Eclipse®. Tomcat® was used as the web server and Mysql® as the database. iWeb® was used to design user interfaces.

Results

Predictability of the ranking results with the original

Colorimetric accuracy was used to predict and model the ranking results, since the originals were available in the experiment, and observers were asked to choose a better representation of the originals.

When the original artwork was captured, a UTT target was included in the scene to provide insight into the complete image quality of the capturing system.[11] However, different camera settings were found to be used by several of the institutions to capture the UTT target and the artwork, thus making the parameters extracted from the UTT target hardly predictable to the image quality of the captured artwork.

As a result, to quantify the color accuracy, the hard copy prints made from the soft copy reproduction provided by participating institutions were used. The hard copy reproductions were printed in Printing Application Laboratory (PAL) in Rochester Institute of Technology (RIT). The accuracy in the color reproduction of the hard copy was thought to be highly dependent on the quality of the soft copy images. Therefore, the reflectance factors at several locations on the hard copy reproductions, and on the originals were measured, and the color differences (CIEDE2000) were calculated under CIE Illuminant D65 and used as an indicator of color accuracy.

To predict the ranking results, simple linear regression was performed for each image. The z-score converted from the ranking data was used as the response, and mean color difference was used as a predictor. The fit for *Aquatint* is shown in Figure 4.

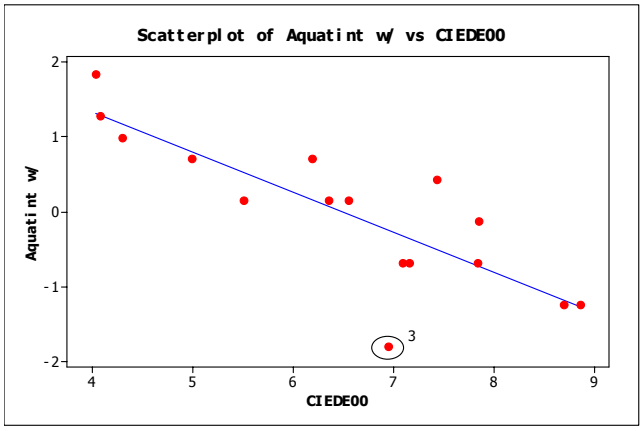

Figure 4. *Z-scores obtained for the Aquatint w/ the original present vs. color difference*

In Figure 4, the y-axis is the z-score of the ranking data. The higher the z-score, the better the image was ranked. The x-axis is the mean color difference. The regression line fits most data points well, indicating that the color difference on the corresponding hard copy served a reasonably good predictor (the p-value for the color difference predictor was almost zero, and the R^2_{adj} was 66.3%). The negative slope of the regression line meant that the higher the color difference, the worse the reproduction was ranked.

However, the 3rd reproduction seemed to be an outlier in Figure 3. It was ranked the worst while its color difference was not very large. The 3rd reproduction of *Aquatint* had a much lighter appearance, which might result from an over-boost of lightness. To verify this, the lightness histograms of the 3rd reproduction and the mean of the rest are shown in Figure 5.

In Figure 5, the lightness distribution of the 3rd reproduction (bottom) was much more skewed to the right, in agreement with its much lighter appearance. The lightness difference between the hard copy reproduction and the original, ΔL^*, was therefore included to improve the model. For *Aquatint*, more variance could be explained by having ΔL^* in addition to color difference in the model, as the R^2_{adj} increased from 66.3% to 84.8%. Both color difference and ΔL^* were found to be significant (p-value: 0.001 for ΔL^*).

For all test images, the color difference predictor was found to be highly significant in predicting the experimental results w/ the original, indicating that the criterion used by observers to rank the images with the original was colorimetric accuracy.

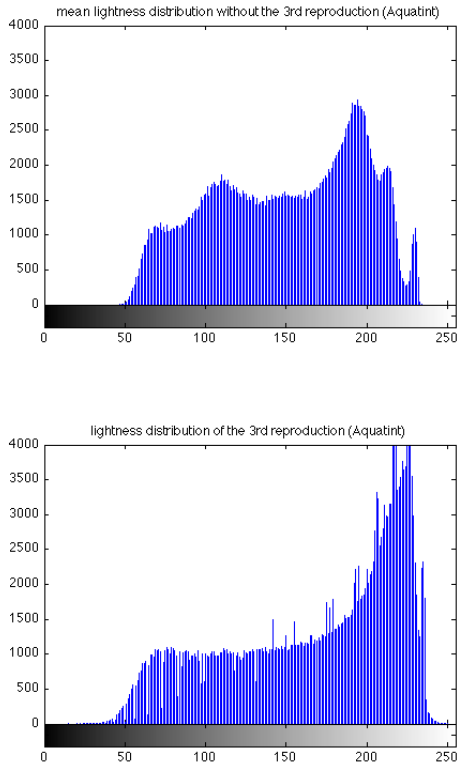

Figure 5. *Lightness distribution for Aquatint*

Predictability of the ranking results without the original

Both ranking experiments w/ and w/o the original were carried out on the same characterized display. The predictability of the ranking results w/o the original by colorimetric difference was also examined by simple linear regression model. For all the images except *Mountain*, the color difference predictor was found to be insignificant and the R^2_{adj} very low, indicating an overall lack of correlation between the ranking results obtained w/o the original and the colorimetric accuracy of reproductions.

The extremely low predictability was not beyond expectation. First of all, the originals were excluded intentionally in the experiment, thus making it impossible for direct comparisons. Secondly, despite the fact that some observers were asked to view the images right before the experiment, their results were found to be not significantly different from those who attended the experiment but experienced the originals weeks before. Therefore, color memory was either unreliable or irrelevant when critical judgments on color were required in the experiment.

The only exception was *Mountain,* for which the predictor of color difference was found to be significant. Further analysis was made as shown in Figure 6. The color difference of one reproduction was far greater than the rest, thus making that soft copy least preferable even when the original was absent. In Figure 6, the 5th reproduction was an influential point. Were it removed, the color difference predictor would be insignificant (p-value: 0.109), and the regression line would become almost horizontal, indicating little predictability of the ranking data w/o the original by the color difference measured on the hard copies.

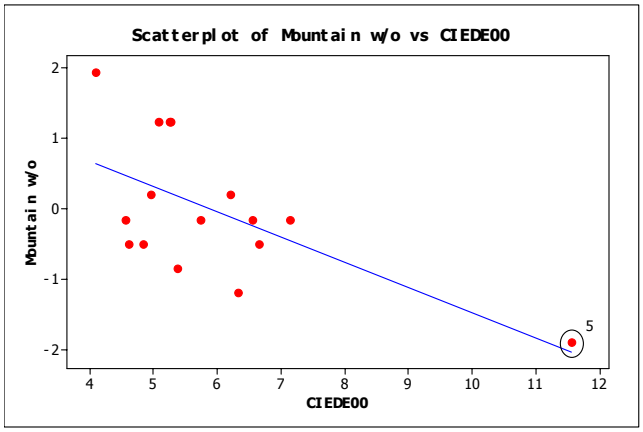

Figure 6. *Mountain w/ the original vs. color difference*

In the web-based experiment, the color accuracy in reproducing images on observers' own displays could hardly be guaranteed. Therefore, the predictability of data from the web-based experiment by color differences was not tested.

Comparisons of ranking results across all three experiments

The ranking results from all three experiments were compared and contrasted. Given the ordinal nature of the ranking data, the Spearman rho correlation coefficients[12] were calculated across the three experiments for each image. For all six images, the Spearman rho correlations between the web-based ranking results and those obtained w/o the original were highly significant. The z-scores converted from the ranking data in all three experiments for *Firelight* are plotted in Figure 7 as an example.

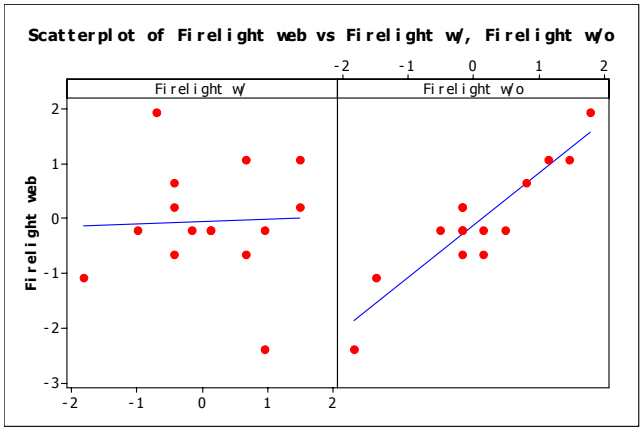

Figure 7. *Firelight web vs. w/ and w/o the original*

In Figure 7, almost all the data points aligned along the regression line in the plot on the right, indicating the high correlation between the ranks of images in the web-based experiment and that w/o the original. On the other hand, in the plot

on the left, the data points were scattered, indicating a lack of correlation for *Firelight* between the ranking results in the web experiment and that w/ the original.

The Spearman rho correlation coefficients between web-based results and w/ the original, and those between w/ and w/o the original were almost always insignificant at 0.05 confidence level except for *Photo*. For *Photo*, the p-values for the correlation coefficients were near the borderline of 0.05 (p-value: 0.06 for correlation between web and w/, and 0.04 for correlation between w/ and w/o for *Photo*).

Analysis of the latent structure in the test images

Based on the rankings of images by observers, the latent structure within the test images could be understood. Clustering analysis was used, and the result for the experiment w/ the original is shown in Figure 8. This figure shows that *Daisy* was ranked more closely to *Aquatint* and *Photo,* both of which were of near-neutral appearance.

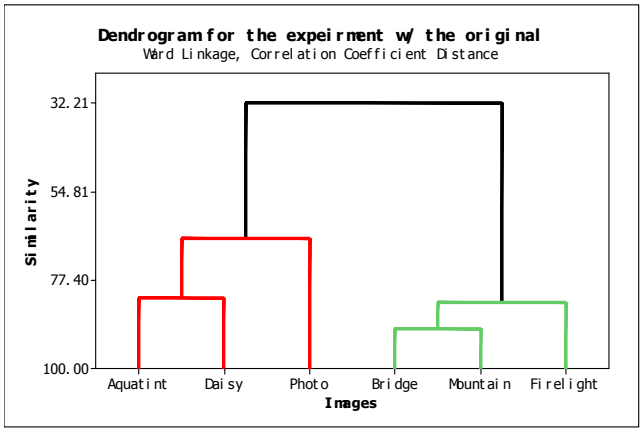

Figure 8. *Dendrogram for the experiment with the original in MCSL*

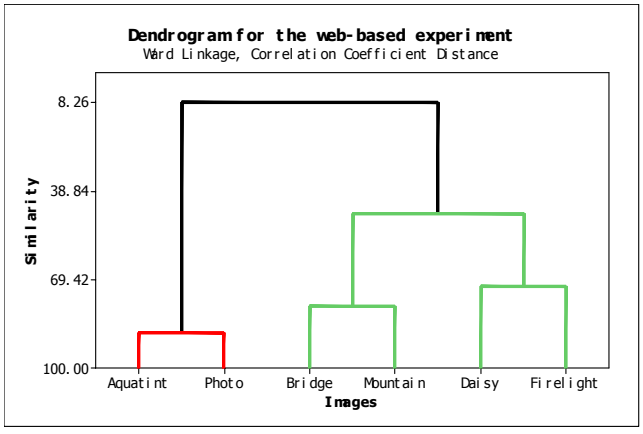

Figure 9. *Dendrogram for the web-based experiment*

The dendrogram for the web-based experiment is shown in Figure 9. In this figure, *Aquatint* and *Photo* were ranked in one group and the rest were clustered in the other group. The grouping of images for the experiment w/o the original in MCSL was the same as that in Figure 9.

By comparing Figure 8 and 9, the main difference was how *Daisy* (Figure 17 in the Appendix) was ranked w/ and w/o the presence of the original. Daisy was ranked closer to neutral color images when the original was available (Figure 8), but more similarly to colorful images when the original was absent (Figure 9). It might result from the shift in criterion employed by observers to judge perceived image quality. When the original was present, observers could be reminded constantly about the whitish color of the petals on the original in the light booth. However, when the original was absent, reproductions of Daisy with more chromatic yellowish petals generally became more appreciated.

Areas (Colors)-of-interest clicked by observers in the web-based experiment

In order to learn more about the ranking decisions made by observers, in the web-based experiments, observers were asked specifically to click not only on the preferred image but also in the area that contributed most to their pick between each pair of images. From the mouse clicks, regions-of-interest were identified. User clicks were overlapped on *Bridge* in Figure 10 as an example. Since the reproduction that was ranked best received more clicks than the rest, it was of most interest in understanding which part of the image was most contributive to the top rank.

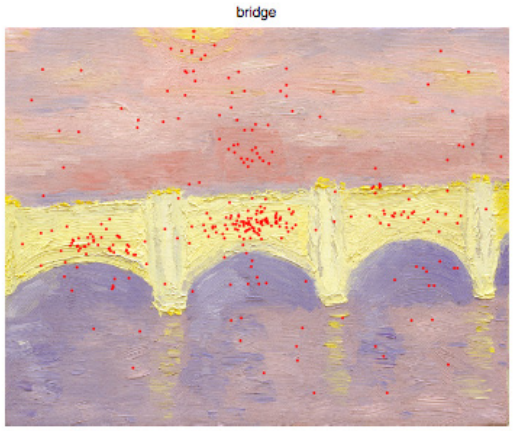

Figure 10. *User clicks overlaid on Bridge*

For *Bridge*, the 4th, 8th, 9th and 12th reproduction were ranked the best among sixteen reproductions and the 4th one is shown in Figure 10. The mouse clicks by observers were indicated by the red dots. The pattern of the red dots looked too clustered to have occurred by chance. Besides, the attention of observers was overwhelmingly drawn to the objects in the image, bridge, e.g. in Figure 10. Moreover, the central part of the objects was highlighted much more by the red dots than the boundaries of objects. However, if the edges of objects shared similar colors with the areas that were clicked by observers, accurate color reproductions of the edges would be of similar importance.

To investigate observers' interest in the edges of objects in the painting, both the areas clicked by observers (the red dots in Figure 10), and the pixels that were of similar colors to those clicked by observers were examined as shown in Figure 11.

In Figure 11, the red dots represented either locations that were clicked by observers or pixels of similar colors to those clicked by observers. The horizontal boundary between the sky and

the bridge, and that between the bridge and water could be identified easily, as they were not covered in red dots. By analyzing Figure 10 and 11, the ranking decisions made by

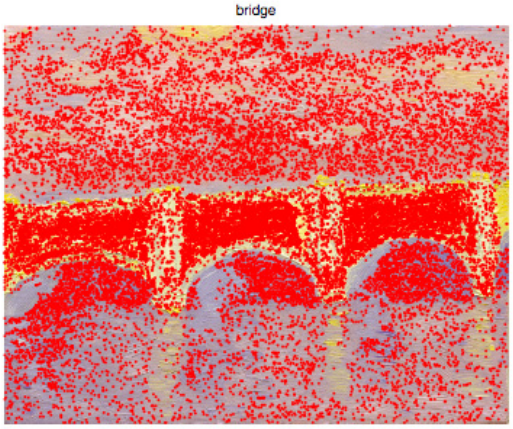

Figure 11. *Color-of-interest overlapped on Bridge*

observers were more likely to be influenced by the color reproduction of objects (especially the central part of them) in the image. On the other hand, the boundaries of objects in the image were of much less importance when preference judgments were made by observers.

Similar results were found for the other images, except for *Photo* and *Firelight*, both of which had human faces in the image. The intensity map of *Firelight* is shown in Figure 12 as an example.

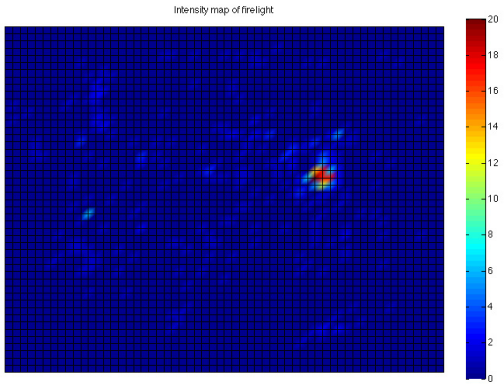

Figure 12. *Intensity map of Firelight*

In Figure 12, the color bar represented the intensity of user clicks on the image. The warmer the color, the more clicks were made by observers in that region. Given that the user clicks on *Firelight* were highly concentrated, simply overlapping the mouse clicks on the image might underestimate observers' clicks in overcrowded areas. Intensity map was therefore plotted in Figure 12 to more accurately reflect the distribution of the clicks by observers.

In Figure 12, most of the areas were in blue, thus having mouse clicks less than 6 according to the color bar. By comparing *Firelight* (Figure 15 in the Appendix) and its intensity map in Figure 12, user clicks were highly clustered on the cheek of the lady's face (the reddish areas in Figure 12). The number of clicks on the lady's cheek was about three or four times of that in other regions on the painting. Given the far higher priority of the lady's face, whether the color of human skin was reproduced pleasingly to observers would be determinant to the ranking of the soft copy reproduction. Similar concentration of user clicks was found on the face of the lady in *Photo*.

Discussion

The ranking data obtained w/ the original was found to be predictable by the mean color difference measured at selected areas on the hard copy reproductions and on the originals. It indicated the transferability of color accuracy from soft copy to hard copy reproductions. However, the color difference metric became insignificant when predicting the ranking results for the experiment w/o the original. The difference might result from the change in criterion of observers when making evaluations of the image quality w/ and w/o the original.

If the soft copy were to be viewed together with the original, color accuracy rather than image preference should be of higher priority. Furthermore, from the web-based experiment, more attention was drawn to the central part instead of the boundaries of the objects in the image. Dominant colors of those areas-of-interest could be extracted so that color patches could be made and adjusted instead of making visual editing on the whole image. An example is shown in Figure 13 for *Bridge*.

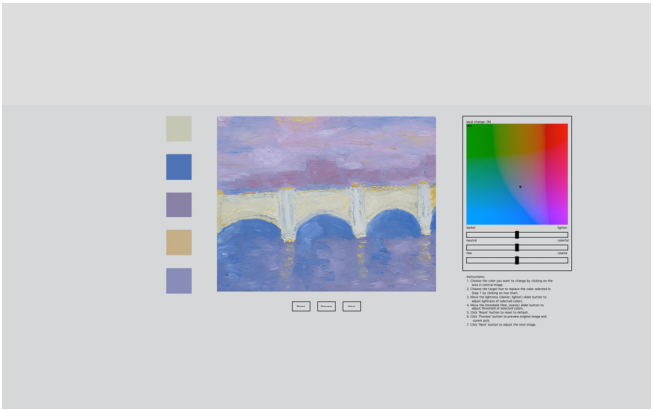

Figure 13. *Extraction of dominant colors for adjustment for Bridge*

In Figure 13, color patches of dominant colors in *Bridge* were shown to the left of the image, and the adjustment tools were shown to the right. Dominant colors could be either handpicked by observers or extracted automatically from the salient objects in the image. By addressing the areas that were selected more often, not only could the reproduction be made within a shorter amount of time, but it was also more likely to be appreciated.

However, one assumption that was untested in this solution was that observers were looking at the same locations when making ranking decisions w/ and w/o the originals. To verify this, observers could be asked to identify the regions that contributed most to their ranking decisions with the presence of the originals. Or the pattern of user clicks obtained in the web-based experiment could be compared with the selected areas on the hard copy reproductions used to measure the color difference. A good

correspondence would be indicative of similar regions used by observers to judge images invariant to the presence of the originals.

On the other hand, if the soft copy were to be viewed without the original, like when people viewed artwork images on museum websites, the soft copy that was more colormetrically accurate might not always be appealing. For example, the 14th soft copy reproduction was ranked best overall when the originals were available. However, it was ranked worst generally in the experiment w/o the original and in the web-based experiment. The huge contrast might not be representative of all the other reproductions, but the seemingly implausible result and the lack of correlation between the rankings w/ and w/o the originals should not be overlooked. Separate workflows might be needed when creating soft copy reproductions depending on the presence of the original and the audience.

Additionally, a significantly high correlation was found between the ranking results in the web-based experiment and that in the experiment w/o the original. Given the high correlation and the absence of originals in both experiments, a similar criterion, image preference, must be shared by observers when the images were evaluated. Moreover, the preference judgments by observers were less likely to be affected by display settings, lighting conditions, and so forth, given the wide range of displays used by observers participating in the web-based experiment. One direct application would be evaluations of image quality (based on preference) online when a characterized display or a controlled environment was not readily available. By having observers perform the test through the Internet, similar results could be expected. Nevertheless, a relatively large number of observers would be helpful to eliminate the biases introduced by various testing conditions.

In the web-based experiment, the regions-of-interest were identified by having user clicks not only on the preferred image but also in the area that most influenced their decisions. While the user clicks might indicate not only the single pixel that was selected, but also the areas around it, given enough participants (about 100) to the experiment, it was believed that the distribution of the user-clicks was already a reasonably good representation of how spread out or concentrated the regions-of-interest were. The scattering of user clicks was not randomly distributed. Instead, they were object-oriented, that is, user clicks were on the objects in the image. It was reasonable because preference decisions could hardly be made until observers first figured out what the objects were and how the objects should look based on their experience. In addition, most of the clicks were clustered in the central part of the objects. Since the color and texture of the objects were usually more coherent close to the center, it was much easier for observers to click near the center without leaning really close to the display and fixating on a single pixel in the image. Similarly, the boundaries of objects in the image were the least examined by observers. For one thing, it was more ambiguous about what the color of the boundaries should be to be pleasing. For another, much more effort would be needed to click on a thin edge.

Attention-guided[13] or saliency-based[14] models were of special interest in the field of image segmentation and object recognition in order to detect and extract important image features and details. It would be of interest then to learn whether such models could be used to predict the areas-of-interest determined by user clicks in the web-based experiment. Customizations to the models might be needed, because the areas that were most contributive to the pairwise comparisons might not necessarily be the most conspicuous parts in the image. In future work, a model predicting the regions that not only drew most attention but also best differentiated images could be built based on the available visual attention models and by tuning the user-click data from the web-based experiment.

Conclusion

A series of experiments were conducted to investigate the image quality of fine art reproductions shown on the display under different testing conditions. Reproductions of six images from sixteen institutions were collected and tested. Color accuracy and image preference were used respectively by observers to rank soft copy reproductions w/ and w/o the originals in the experiment. A strong correlation was found between the ranking results in the web-based experiment and that w/o the original, indicating a lack of impact of testing conditions on the preference judgments on images. The user-clicks in the web-based experiment were object-based, and the boundaries of objects in the image received the least attention.

The use of art stimuli has been popular especially in eye tracking studies.[15, 16] In the future, an eye-tracking experiment will be conducted. By comparing the data in the current study and the eye-tracking experiment, topics on the spatial and temporal distributions of attention when observers view a piece of artwork could be investigated more thoroughly. Questions such as whether the areas that were fixated for a long time were actually used by observers to make preference decisions might be answered in a more rigorous manner.

Acknowledgements

The work was supported by The Andrew W. Mellon Foundation.

References

[1] David Saunders. High-quality Imaging at the National Gallery: Origins, Implementation and Applications. Computers and the Humanities 31: 1998

[2] Alejandro Jaimes and Fred Mintzer. Improving Quality of Digital Images of Art in Museum Collections. IS&T Image Processing, Image Quality, Image Capture, Systems Conference (PICS), April 2001

[3] Roy S. Berns. Color-Accurate Image Archives Using Spectral Imaging. Scientific Examination of Art: Modern Techniques in Conservation and Analysis, National Academies Press (2005)

[4] Yonghui Zhao and Roy S. Berns. Image-Based Spectral Reflectance Reconstruction Using the Matrix R Method. Color Research and Application, Volume 32, Number 5, October 2007

[5] Susan Farnand, Jun Jiang, Franziska Frey. Comparing Hardcopy and Softcopy Results In the Study of the Impact of the Workflow on Perceived Reproductoin Quality of Fine Art Images. SPIE-IS&T/ Vol. 7867

[6] Franziska Frey and Susan Farnand. Benchmarking Art Image Interchange Cycles Final Report 2011.

[7] Silverstein and Farrell. An Efficient Method for Paired-Comparison. J. EI 10 (02) 394-398. Jan. P. Allebach; Ed.

[8] Shigeru Sasao. Paired Comparison: A User Perspective. 24th International Forum on COCOMO and System/Software Cost Modeling

[9] 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.

[10] M.R. Luo, G. Cui, B. Rigg. The Development of the CIE 2000 Color Difference Formula: CIEDE2000. Color Research and Applicatoin. Volume 26, Number 5, Ocotober 2001.
[11] Dietmar Wueller, Hans van Dormolen, Koninklijke Bibliothek van Nederland, and Volker Jansen. Universal Test Target Technical Specification.
[12] Irwin Miller, Marylees Miller. John E. Freund's Mathematical Statistics with Applications 7th Edition. Pearson Education.
[13] Neil D. B. Bruce, Daniel P. Loach, and John K. Tsotsos. Visual Correlates of Fixation Selection: A Look at the Spatial Frequency Domain. IEEE 2007
[14] Dirk Walther, Christof Koch. Modeling Attention to Salient Proto-Objects. Neural Networks 19 (2006)
[15] Daniel J. Graham, Christoph Redies. Statistical Regularities in Art: Relations with Visual Coding and Perception. Vision Research 50 (2010)
[16] Buswell, G. T. How People Look at Pictures. Chicago: University of Chicago Press (1935)

Author Biography

Jun (Chris) Jiang is a third-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 with Prof. Franziska Frey and Susan Farnand on visual editing and cross-media color reproduction of artwork in museums.

Appendix

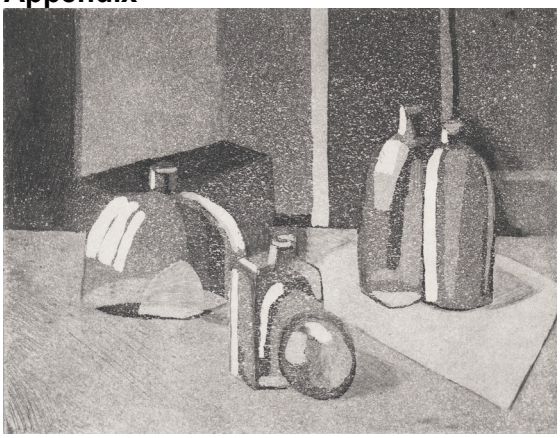

Figure 14. *Aquatint*

Figure 15. *Firelight*

Figure 16. *Bridge*

Figure 17. Daisy

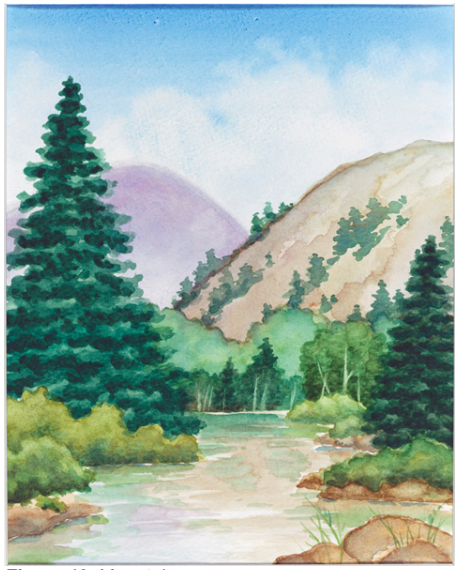

Figure 18. *Mountain*

Figure 19. *Photo*