# Observer Preferences and Cultural Differences in Color Reproduction of Scenic Images

*Scot R. Fernandez and Mark D. Fairchild*
*Munsell Color Science Lab, RIT*
*Rochester, New York*

## Abstract

Observer preferences in the color reproduction of pictorial images have been a topic of debate for many years. Through a series of psychophysical experiments we are trying to better understand the differences and trends in observer preferences for pictorial images, determine if cultural biases on preference exist, and finally generate a set of preferred color reproduced images for future experimentation and evaluation. The results yielded that statistical difference between the peaks of preference of image quality may exist between cultures, but that the cultural difference observed is most likely not of practical significance for most applications. The analysis of a second experiment yielded that the intra-observer repeatability of an observer is about half of the variation between observers. Furthermore the analysis demonstrated that preferences on images with faces have a much tighter range of preference in comparison to images without faces.

## Introduction

With the recent prevalence of digital imaging, many of the constraints of traditional imaging systems have been lifted. Unfortunately, with the newfound flexibility of digital imaging, new complexities in quantifying color quality have been generated. Often minimizing some color difference metric is the standard goal in understanding the limits of color quality and color reproduction of an imaging system. A color difference metric, in its simplest form such as $\Delta E^*_{ab}$ or $\Delta E^*_{uv}$, is a Euclidean distance metric used to quantify the distance between a pair colorimetric coordinates in either CIELAB or CIELUV color space, respectively, quantifying the difference between two stimuli.[1] Theoretically, the perceived difference between two colors is uniform throughout a given color space, and one unit of difference corresponds with one unit of perceptual difference.[2] The intent of minimizing a color-difference metric or maximizing the colorimetric accuracy between an original image or scene and its reproduction through a cross-media reproduction system is known as a colorimetric reproduction objective.[3] A colorimetric objective will produce a reasonable reproduction, but further work is required to understand why it doesn't always produce the best reproduction of an image. For example, previous research efforts support the idea that observers would prefer object colors to be reproduced with greater saturation in comparison to the original, and that certain memory colors such as grass, skin, and sky are remembered with slightly different hues and with greater purity.[3] Furthermore, it is known that an observer maintains the ability to rate the quality of an image with or without the original image present.[4] Without the original image present, observers are rating the quality of an image in reference to some psychological concept of an idealized image.[5] So the goal of our color reproduction intent should sometimes be to match the psychological concept of an image, known as preferred image reproduction, rather than some arbitrary image said to be the original, which is a colorimetric image reproduction.[6]

Preferred image reproduction techniques should be viewed as an enhanced or customized version of a colorimetric objective. Thus, when evaluating preferred image reproduction, we need to move from a color-difference metric to the degree of apparent match between a reproduced image and its internal memory reference, which has been labeled as naturalness.[7] It is commonly understood that pictorial image quality has a positive correlation with naturalness, so an image of high quality is one that has a high degree of naturalness.[6,7]

## Experimental

The goal of this research is to better understand the considerations needed for preferred color reproduction of pictorial images, specifically pictorial images of unknown colorimetric origin. The three specific interests of this research are to build tolerances of observer preference in colorimetric dimensions for hard and soft-copy images, to determine if psychological biases of preference can be linked to cultural differences, and finally to create a set of "preferred" images for both hard and soft-copy image display for future experiments.

The psychophysical experiments described in this paper are a continuation of research discussed in a paper presented at the 9th Color Imaging Conference.[8]

**Experiment I - International Image Characteristic Ranked Order**

This psychophysical experiment asked observers to rank order sets of images from best to worst based on preference. Each set of images represented a ramp of a single global colorimetric manipulation to an image. The experiment was completed at four different research facilities: Chiba University (Japan), University of Derby (UK), Xerox (USA), and RIT (USA). Due to the unique nature of this experiment, each testing location was supplied a book of image sets and a user interface posted on the World Wide Web was utilized to record the observer's responses.

Thumbnail representations of the image set utilized in this experiment are in Figure 1.

To create the sets of manipulated image, the images were adjusted along eight different CIELAB dimensions. The colorimetric dimensions chosen were a logical extension of experience from adjusting manipulating images, and later correlated to the analysis of previous research.[8] Four of the dimensions affected color balance (additive shifts of a* and b*); the other four manipulations were lightness (a gamma adjustment of L*), contrast (a sigmoid adjustment to L*, with an threshold at 50.0 L*), Chroma (multiplicative adjustment to $C_{ab}$* at a constant $h_{ab}$), and Hue rotation ($h_{ab}$ rotation at a constant $C_{ab}$*). The direct and indirect dimensions of adjustment are two of the color balance dimensions that manipulated the image along the 45• axes of the a* and b* coordinate system.

The eight manipulations were applied to the eleven-image set to generate eighty-eight sheets of randomly ordered six-image sheets that varied around the nominal image. Each sheet demonstrated the effect of a single adjustment applied globally, and consisted of three steps above and below the original image. The increments were clearly perceivable, but not objectionably large. The increments used to generate the image sets are in Table 1.

The sheets were printed on a Fujix Pictrography 3000, at a resolution of 300 dots per inch. The printing system was characterized using a 10x10x10 LUT, and a tetrahedral interpolation technique. The printer's forward characterization was utilized to convert the RGB images into CIELAB space, were all manipulations were done and then the inverse characterization was utilized to convert the CIELAB images back to RGB. This workflow of starting in the printer's gamut minimized gamut issues. A pictorial representation of a print sheet from the experiment is presented in Figure 2. This sheet represents an example of an adjustment of lightness. In addition to the placement of the manipulated image sets being randomized within each sheet, the order of image and applied manipulation were randomized throughout the entire book of image sets.

The observers of each sub-population were then asked to rank each sheet images from best to worst based on preference utilizing an online user interface that recorded the entire set of response files to the Center of Imaging Science at RIT. The sub-population statistics are presented in Table 2, and in total seventy-seven observers participated.

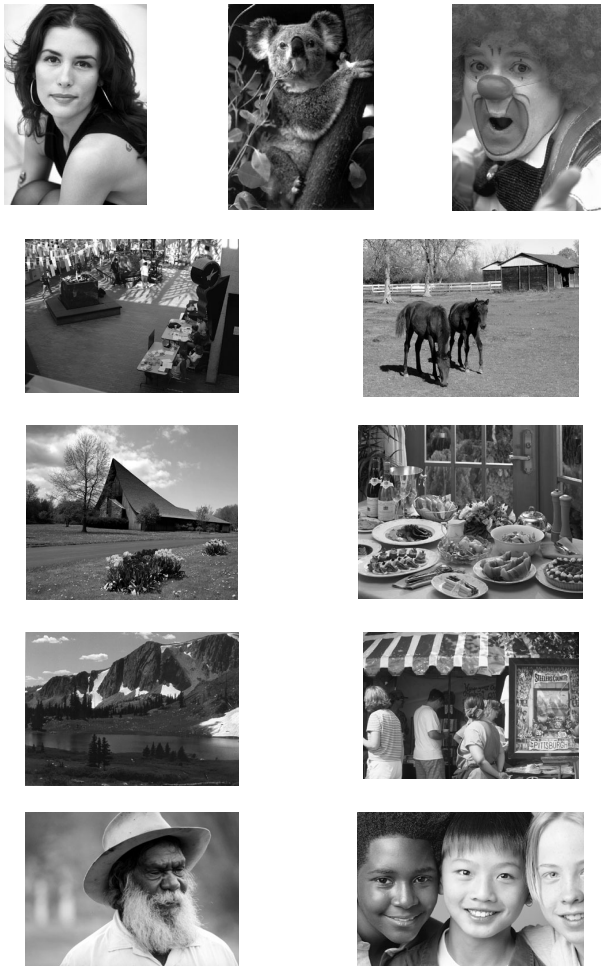

*Figure 1. Image set for Experiment I & II– (From left to right, top to bottom) 1. Model, 2. Koala, 3. Clown, 4. Indoor Scene, 5. Horses, 6. Church, 7. Dinner, 8. Mountains, 9. Art-fair, 10. Bearded Man 11. Harmony.*

**Experiment II – Image Characteristic Adjustment**

In this psychophysical experiment, observers used a graphical user interface to manipulate a set of images until the images best matched their perception of the best possible color reproduction of the image. In order to incorporate all of the objectives in this phase of research the experiment was done using two different interfaces. For this phase of research the colorimetric dimensions of manipulation and the image set were the same as in Experiment I.

**Table 1. Adjustment Ranges and Increment Values for Experiment I.**

| | Starting Value | Ending Value | Increment |
|---|---|---|---|
| **Gamma adjustment** | 0.55 | | 0.15 |
| **Sigmoidal adjustment** | 0.55 | 1.30 | 0.20 |
| **Chroma adjustment** | 0.75 | 1.55 | 0.11 |
| **Hue Angle adjustment** | -0.07 | 1.30 | 0.035 |
| **a* adjustment** | -7.50 | 0.11 | 3.00 |
| **b* adjustment** | -7.50 | 7.50 | 3.00 |
| **Direct adjustment** | -7.50 | 7.50 | 3.00 |
| | -7.50 | 7.50 | 3.00 |
| **Indirect adjustment** | -7.50 | 7.50 | 3.00 |
| | 7.50 | -7.50 | -3.00 |

**Table 2. Breakdown of Observer Population for Each Cultural Subpopulation of Experiment I**

| Ethnic Background | Chinese | European | American | Asian | American | Japanese |
|---|---|---|---|---|---|---|
| **Testing Location** | Derby | Derby | RIT | RIT | XEROX | Chiba |
| **Number of Female Observers** | 2 | 2 | 6 | 2 | 2 | 3 |
| **Number of Male Observers** | 8 | 8 | 12 | 5 | 3 | 20 |
| **Age Range of the Observers** | 23–43 | 22–39 | 17–39 | 28–31 | 29–44 | 21–31 |

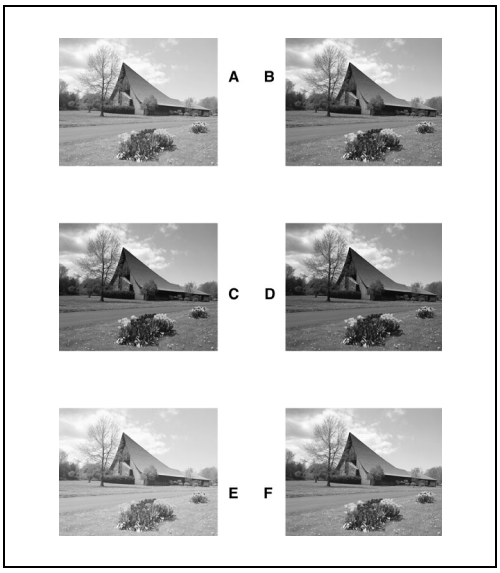


*Figure 2. Sample sheet manipulated images from Experiment I.*

This experiment was conducted on a 22" Apple Cinema Display, and each observation was made in a darkened environment.

The first graphical user interface (GUI) in this phase randomized the image order and allowed the users complete freedom to manipulate an image along all colorimetric dimensions. This allowed each observer to make adjustments in any order they choose and also allowed them the ability to return to any of the previous dimensions as many times as needed until they obtained their desired image. This user interface was utilized for the repeatability aspect of this phase of research; therefore three different observer populations were required. The first intent was to evaluate a large population for just one observation, the second was a medium size population with multiple observations, in this case five observations were made by each, and finally a small population with many observations.

The second user interface used the same colorimetric dimensions of manipulation, however the user was limited to adjust along one dimension at a time. The user was allowed to adjust the single dimension as many time as he or she needed but were limited to only adjust the dimension that was presented. Lightness, contrast, chroma, and hue rotation were presented one time for each image and color balance was done twice, first individually as a* or b* and then as a* and b*. Once the observer adjusted each dimension to the best possible color reproduction of the image along the one dimension the observer was asked to rate the overall color quality of the image, using the same scale mention earlier. This user interface was only used to evaluate one population size, a large population for a single observation.

The original images were converted from RGB digital counts to CIELAB values using the forward printer characterization from Experiment I. This was done to increase the amount of correlation between experiments. In order to invert the adjusted image from CIELAB values back to RGB values, the inversion of a characterization incorporates the use of a 3x3 matrix with three linearly interpolated one-dimensional look-up tables. The major design decision for this phase of research was how to calculate the adjusted images. The primary concern was to determine which order the colorimetric manipulations should be applied to an image, and furthermore how to preserve the ability to be able to undo the application of any

manipulation in any order. The solution was to always recalculate the adjusted image from the original image file, and to build the colorimetric manipulations into one function so that the adjusted image is always calculated in the same manner allowing the observer the ability to reasonably predict the resultant image from one manipulation to another. The order that the colorimetric functions were integrated is as follows: lightness, color balance, contrast, and then chroma and hue rotation. The observer population consisted of students, faculty, and staff. Table 3 presents the breakdown of the observer population.

**Table 3. Breakdown of Observer population for each sub-population of Experiment II**

Experiment II. Version I

| | Number of Observers | Number of Trials | Percent Male | Age Range |
|---|---|---|---|---|
| **Data Set A** | 31 | 1 | 68 | 22–71 |
| **Data Set B** | 10 | 5 | 90 | 22–37 |
| **Data Set C** | 1 | 15 | 100 | 25 |

Experiment II. Version II

| | Number of Observers | Number of Trials | Percent Male | Age Range |
|---|---|---|---|---|
| **Data Set D** | 30 | 1 | 70 | 22–60 |

## Results and Discussion

### Experiment I - International Image Characteristic Ranked Order

The analysis of this experiment was done in two steps. Both steps of the analysis compared four sub-populations Americans, Chinese, Europeans, and Japanese against the entire population of the experiment to determine if a difference in preference existed. The first evaluation implemented Thurstone's Law of Comparative Judgments to develop scales of preference for each adjustment dimension. This analysis combined the results of all the images, and compared the composite results for the entire image set for each dimension. Sample results of this analysis are seen in Figure 3. The two plots in Figure 3 are of the same data. The first plot allows one to visualize the shape and distribution of preference for each sub-population in relation to each other. The second plot allows one to understand the error associated with the interval presented for each sub-population. The composite results of this analysis are summarized in Table 4. For the dimension depicted in Figure 3, it is obvious that the Japanese group has a shifted preference for a lighter image.

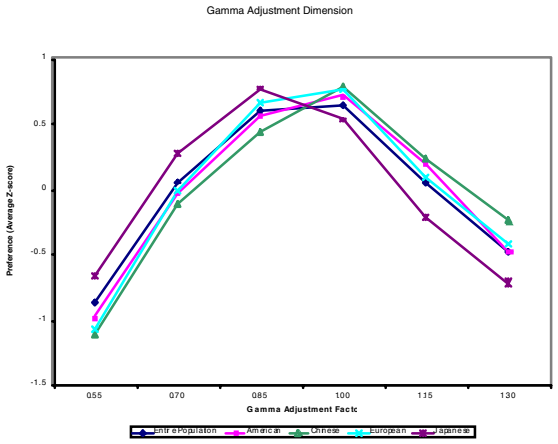


*Figure 3a. Results for the Gamma Adjustment Dimension for the Thurstone's Analysis*

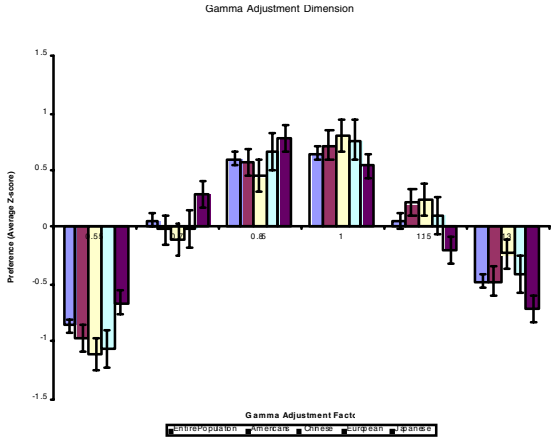


*Figure 3b. Results for the Gamma Adjustment Dimension for the Thurstone's Analysis*

**Table 4. Summarized Results of Thurstone's Analysis**

| Adjustment Dimension | Comments |
|---|---|
| Gamma | Japanese group has a shifted preference towards a lighter image in comparison to all other sub-groups |
| Sigmoid | Chinese group demonstrate a shifted preference to more contrast in comparison to the Americans and Japanese |
| Chroma | The Eastern Hemisphere has a shifted preference to more chroma compared to the Americans |
| Hue Rotation | Hue Rotation demonstrated little peak preference for any sub-group |
| a* | Japanese demonstrate a preference towards redder or warmer images than Americans |
| b* | Chinese group demonstrate a shift towards bluer or cooler images |
| a*b* Direct | No Particular Trends |
| a*b* Indirect | No Particular Trends |

**Table 5. Results of Student-T Mean Statistical Difference Evaluation**

| | American | Chinese | European | Japanese |
|---|---|---|---|---|
| **Entire** | 2 | 2 | 0 | 3 |
| **American** | | 3 | 1 | 3 |
| **Chinese** | | | 0 | 2 |
| **European** | | | | 2 |

| | American | Chinese | European | Japanese |
|---|---|---|---|---|
| **Entire** | Direct, Indirect | Sigmoid, b* | | Gamma, a*, Indirect |
| **American** | | Hue, b*, Indrect | Direct | Gamma, a*, Indirect |
| **Chinese** | | | | Gamma, b* |
| **European** | | | | Gamma, indirect |

**Table 6. Observer Inter- & Intra-Variability in Making Preferred Images**

| | Inter-Observer Variability | | | |
|---|---|---|---|---|
| | MCDM – based on $\Delta E^*_{94}$ | | | |
| | **Minimum** | **Maximum** | **Mean** | **St. Dev** |
| **Data Set A** | 2.38 | 17.70 | 7.36 | 3.58 |
| **Data Set D** | 2.44 | 20.89 | 8.23 | 4.16 |

| | Intra-Observer Variability | | | |
|---|---|---|---|---|
| | MCDM – based on $\Delta E^*_{94}$ | | | |
| | **Minimum** | **Maximum** | **Mean** | **St. Dev** |
| **Data Set B** | 1.04 | 12.37 | 4.51 | 2.53 |
| **Data Set C** | 2.50 | 11.04 | 6.04 | 2.48 |

The second evaluation of this experiment calculated the peek response of each of the eighty-eight sheets within the experiment and utilized the student-t distribution and an alpha value of 5% to calculate if statistical difference existed between the mean responses of each dimension between cultures. The results of this analysis are presented in Table 5. This first table simply outlines how many dimensions of the eight tested differed for each pair of cultures tested. The second table specifically lists which dimensions differed.

The combination of these two analysis techniques is important. The Thurstone's analysis allowed us to understand the shape of the response interval from each cultural group for each adjustment dimension. This information identified any trends in the cultural biases for example the Japanese trend noted above and in the chroma dimension it appears that a difference does exists despite the fact that chroma did not test positive as a significant difference in the second evaluation. The advantage to the second evaluation is that it's a quantitative test of statistical difference and clearly defines were statistical difference exists between the most preferred response for each sheet. However this analysis can present no statistical difference between two groups of peak responses while the previous analysis demonstrates significant differences in the preference curves, such as chroma. From this analysis it is clear that there are statistically significant cultural differences, however it appears that they might not be that important in most practical applications. Finally when the Thurstone's analysis was repeated for each individual image/manipulation pair the shapes of each dimension preference curve across the set of images were very consistent, further diminishing the idea that huge differences between cultures exist.

**Experiment II – Image Characteristic Adjustment**

The first analysis of this experiment was to understand the variability between observers (inter-observer) and also to understand the repeatability within an observer (intra-observer) to make a preferred image. The results of this evaluation are presented in Table 6, and the statistics are based on the individual results of each image in comparison to its mean image.

The mean color difference from the means, where the mean represents the optimal image of a given population, was calculated using a pixel-by-pixel color difference calculation. It is interesting to note that the variability within an observer is about half of the variability between observers.

The next colorimetric evaluation was to determine how close the average image of each population was to the starting image. Table 7 represents this data, and this validates that the starting images were likely inside the circle of observer variability. This was important to this research because the goal is to better understand an enhancement of a colorimetric objective. If the starting point of manipulations was too far away from the endpoint then the manipulation would be correcting a flaw in the characterization not allowing us insight to preferred color reproduction.

**Table 7. Color Difference Between the Optimal Image and Starting Image**

| | Difference between Original Image and Mean Image | | | |
|---|---|---|---|---|
| | Color difference based on $\Delta E^*_{94}$ | | | |
| | Minimum | Maximum | Mean | St. Dev |
| Data Set A | 2.35 | 7.57 | 5.05 | 1.64 |
| Data Set B | 2.23 | 10.32 | 6.60 | 2.57 |
| Data Set C | 4.15 | 10.24 | 7.12 | 2.07 |
| Data Set D | 2.88 | 8.70 | 4.55 | 1.78 |

Unfortunately the MCDM analysis presented above does not allow one to visualize the observer variability. Therefore the final evaluation of inter- and intra-observer variability was to make actual print sets to demonstrate the variability. To better understand which image was the most variable or least, the image set was rank ordered by standard deviation. This demonstrated that the images with the smallest standard deviation of color difference from the mean image were all images with people in them. In Data Set A (31 obs. – Ver. I), B (10 obs. – Ver I), and D (30 obs. – Ver II), the four primary face images were all in the top six for each experiment. These images are Model, Man, Clown, and Harmony. Data Set C is based only on one person so the subtle deviations were noticed. For the print sets made the least variable image chosen was Model and the most variable image chosen was Mountains.

**Crossover Analysis of the Experiments I & II**

The final analysis of this research was to generate sets of preferred images from each of the previous experiments and compare the results. The first obstacle was to decide how to compute the mean image, either by averaging the end adjustment points or by regenerating each optimal image and then averaging the images. To aid in the decision, the mean pixel-by-pixel color difference was calculated between the two techniques of calculating a preferred image, utilizing Data Set A from the Adjustment Experiment. The results revealed that the difference between the two different techniques is negligible; therefore the decision was to calculate the preferred images based on the average of the adjustments rather than the average of images. This decision was made for computation ease and because averaging the adjustments is more similar to how the peaks were generated from Experiment I.

The results of this analysis demonstrate that there is little difference between the four techniques utilized to generate preferred images. This is visually confirmed by the image sets generated, which were printed on a Fujix Pictrography 3000 using the characterization technique and considerations from the hard copy experiment.

## Conclusions

Observer preference incorporated into current image reproduction techniques should be viewed as an enhanced or customized version of a colorimetric reproduction objective. The images in this research were not a complete colorimetric reproduction of the original scenes because there were no colorimetric measurements of the original scenes to compare with the reproduced images. However, the idea of a need for a customized reproduction objective is still the underlying theme of this research.

The first experiment, *International Image Characteristic Rank Order*, was an experiment designed to determine if cultural biases on the perception of image quality exists, and also to better understand in colorimetric dimensions observer preferences of hard copy images. The results of this experiment identified that cultural biases may exist between peak preferences while rating image quality, however the analysis also cites that it is probably not practical to account for these differences. The analysis also demonstrated that despite possible difference between the peak responses due to cultural difference, the shape of the preference responses were maintained uniformly across cultures, further diminishing any distinct difference between cultures. This experiment also generated a series of preference curves, which provided insight into how preferences change according to various subject matter, capture modes and overall quality of an image. This analysis demonstrated that images in which people are the primary focus of the image maintains tighter preferences, and that images of higher quality tend to have steeper peaks in preference in comparison to images of lower quality. Generally, the quality of the image is more likely to be directly linked to the quality of capture technique utilized to create the image, so better image capture also appears to generate more defined preference responses. Finally, the results also demonstrated that of the manipulation dimensions, hue rotation had the most ambiguous peaks, meaning that as a global manipulation tool hue rotation is difficult judge and does not produce a clear preference peek or curve. Furthermore each of the other tools provided did demonstrate clear preference peaks.

The second experiment, *Image Characteristic Adjustment*, allowed us to better understand inter- and intra-observer variability while generating "preferred" images. This experiment concluded that the variability within an observer is about half of the variability between observers. The evaluation of this experiment also validated that the image set utilized within this experimentation was at a good starting point to account for differences in preference rather than a flaw in characterization techniques.

The final evaluation of this research was a cross comparison between Experiments I and II, the comparison

was made by generating "preferred" image sets from the data collected from each experiment. The exercise demonstrated good consistency between experiments, leading us to believe that the information gathered in one experiment can be pieced together and directly compared to the results of the other experiment. Also based on the generation of preferred image sets, it became most apparent that the most "preferred" image is the one based on the average of individual preferred images.

## Acknowledgements

The authors wish to thank Dr. K. Braun of Xerox Corporation for providing images and helpful comments. The Xerox Corporation and the NYSTAR CAT CEIS funded this research.

## References

1. R. S. Berns, Billmeyer and Saltzman's Principle of Color Technology. NY: John Wiley & Sons; 2000.
2. M. D. Fairchild, Color Appearance Models. Reading, MA: Addison-Wesley; 1997.
3. R.W.G Hunt, I. T. Pitt, L. M. Winter. *J. of Photographic Sci.* **22**, 144 – 149(1974).
4. R.W.G. Hunt, The Reproduction of Colour, 5th. Fountain Press: UK, 1995.
5. T.J.W.M. Janssen and F.J.J Blommaert. Predicting the Usefulness and Naturalness of Color Reproduction. *J. Img. Sci. Technol.* **44**(2), 93-104(2000).
6. S. N. Yendrikhovskij, F.J.J. Blommaert, H. de Ridder. Color Reproduction and the Naturalness Constraint *Col Res Appl.,* **24**, 52-67(1999).
7. H. de Ridder, Naturalness and Image Quality. *JIST,* **40**(6), 487 – 493(1996).
8. S. R. Fernandez, M. D. Fairchild, Preferred Color Reproduction of Images with unknown Colorimetry. *Proc. 9th IS&T/SID Color Imaging Conf.*, 274-279 (2001).

## Biography

Scot R. Fernandez received his B.S. degree in Imaging and Photographic Technology in 1999 and M.S. degree in Color Science in 2002 from the Rochester Institute of Technology. This work on image quality and preferred color reproduction was part of his thesis requirement for the M.S. degree. In September of 2001, he began working towards a second postgraduate degree in Imaging Science at Rochester Institute of Technology. He is a member of the IS&T and ISCC.