The Quantitative Aspects of Color Rendering for Memory Colors

Karin Töpfer and Robert Cookingham
Eastman Kodak Company
Rochester, New York

Abstract

Color reproduction is a major contributor to the overall perceived image quality of conventional and digital imaging systems. Currently, no universally accepted metric exists that quantifies the impact of color attributes on perceived image quality. Color reproduction is often assessed using a difference metric based on distance in a relevant color space (e.g. ΔE* in CIELAB) for a representative target. Such metrics only provide a qualitative indication of how color reproduction is changing, because color errors in different regions of color space have unequal impact on perceived image quality.

We report the results of an investigation of the quantitative aspects of color rendering of memory colors. In this work we have explored the responses of subjects to scenes where only the foliage, blue sky, or skin tone reproductions have been altered in a systematic way such that the image quality of the images can be determined by observers. From these measurements, a response surface can be developed that predicts the decrease in image quality as the rendition moves away from an optimal region.

Introduction

Color and tone reproduction has always been of fundamental importance in determining the image quality produced by conventional photographic systems. The advent of digital imaging only makes color and tone reproduction more critical because of the inherent flexibility in digital photography and the possibility of delivering preferred color reproduction. As imaging systems have become more diverse with both digital and conventional photographic components, the need has arisen to quantify color quality on an absolute basis. Quantifying color quality requires an absolute color quality metric to make successful predictions of system capability and performance.

The task of producing an absolute color quality metric is extremely complex given the interdependency of various color attributes in an image. The issues are further complicated by the distinction between accurate, corresponding and preferred color reproduction.[1] Image quality with respect to color and tone attributes, such as contrast, saturation[2] and the reproduction of memory colors (e.g. skin, sky and foliage[3]) are much more subject to preferences, especially near the optimum color reproduction for a system. In this respect, color quality attributes differ from other image quality attributes such as noise, streaking, contouring etc. Generally, the presence of any of the latter artifacts is unanimously perceived as a quality degradation.

However, recent advances in psychophysical techniques[4] and the availability of digital images manipulated through image simulation have made the successful development of an absolute color quality metric feasible. In this work we report the results of three investigations of the quantitative aspects of the color rendering of memory colors. We have explored the responses of subjects to scenes where only the memory color reproductions have been altered in a systematic way. From these measurements, a response surface can be developed that predicts the decrease in image quality as the rendition moves away from an optimal region in color space. This paper communicates the results of the three studies dedicated to preferred memory color reproduction.

Design of Stimuli

While early color preference studies of memory colors[3,5] were carried out creating physical masks for individual images, selective manipulations of color space or an individual color attribute can now be accurately and quickly performed using three-dimensional look-up tables (3D LUTs). Lightness, hue and chroma are widely accepted as the visually relevant dimensions of color space as embodied by the Munsell Book of Colors and several other perceptually uniform color order systems.[1] There is wide agreement that, under a specified set of scene illumination and reproduction viewing conditions, these three dimensions of color space are the most important attributes characterizing the color appearance of hardcopy and softcopy reproductions. Therefore, it is intuitive to modify the reproduction of memory colors in terms of lightness, hue and chroma. In each of the studies reported here, the specific region of color space was selectively modified while keeping all other colors constant. In terms of statistical design, lightness, hue and chroma are the factors in this experiment, while the response function is quality degradation or enhancement in terms of just noticeable differences (JNDs) compared with a reference position of known absolute image quality.

Each experiment is intended to provide a quantitative answer to three questions concerning the reproduction of color for a given memory color.

1. What are the aims for color reproduction (optimum memory color rendition) under well-defined viewing conditions?
2. What is the curvature of the response surface near the optimum, i.e. how robust is the optimum position?
3. What is the system performance, in terms of absolute quality, over a broad range of typical conditions of usage, e.g. characterization of the variability of photofinishing, as opposed to peak system capability?[6] In these cases, larger deviations from the optimum reproduction might be encountered, and the asymptotic behavior of the response surface becomes important.

The experimental design we used is based on the central composite design for three factors, which permits the construction of a response surface including interactions between the factors while taking into account the goals of the experiment as stated above. Instead of choosing the typical 6 axial points and an inner full cube of 8 points we decided to choose 2 partial inner cubes with 4 points each. This design is illustrated in Figure 1, and it represents a fractional factorial of 7 levels in each parameter.

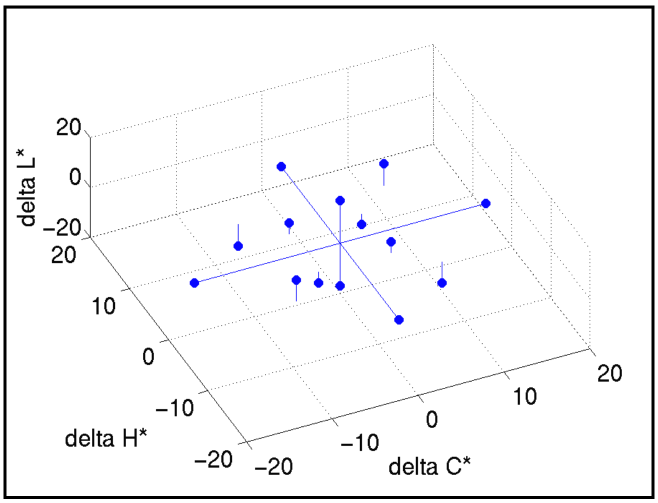

Figure 1. Typical Experimental Design

This design allows an accurate quantification of the response surface curvature near the optimum while retaining sufficient information at less favorable levels of memory color reproduction to characterize the asymptotic behavior of most reasonable color imaging systems. While we acknowledge that CIE 1976 L*a*b* (CIELAB) space is not perceptually uniform, the predictive capabilities of the CIELAB color difference equation[1] are sufficient for the purpose of designing this experiment.

Psychophysical Test Procedure

The basic psychophysical measurement method used in this study employs a softcopy quality ruler that provides a set of physical standards of reliably known quality that match test samples in all but a few carefully designed aspects of image quality. In the softcopy ruler, both the reference images that make up the ruler and the test images are displayed on high-resolution monitors. The reference images are displayed on one monitor, one reference image at a time. The quality of the reference images is varied by degrading an image due to unsharpness (i.e. blurring) the image. The test images evaluated by the judges are displayed on an identical monitor. The time required to display a given image is approximately 2 seconds per image pair, which includes the time to blank the screen, retrieve the image, and display the image. The psychophysical test-viewing environment is specifically designed for administering tests using softcopy display. Both monitors sit on a high table and are positioned so the images are a minimum distance apart and are viewed normal to the monitor faceplates. The keyboard and keypad reside on a smaller movable table at a convenient height for data entry. The viewing distance is controlled by means of a headrest. A surround lighting configuration illuminates the neutral wall behind the monitor to a luminance level of 25% of the white point luminance level of the monitor to approximate normal print viewing conditions.[7]

In the evaluation task, the judges are presented with images on both monitors with reference images displayed on one monitor and the test image on the other monitor. The judging task consists of a series of paired comparison evaluations where the judge indicates whether the right or left image is of higher overall quality. A binary search routine controls the image presentation sequence and provides a random order of presentation blocked by scene. The order of presentation of individual trial images is random within a given scene. Data collection occurs automatically as the system moves to a new test image. Judges interface with the system using a modified ALPS numeric keypad. All the judges are required to demonstrate normal visual acuity and color vision.

Development of Objective Metrics

As a result of our experiments, we obtained JNDs of image degradation or enhancement as a function of CIELAB lightness, hue and chroma modifications in the region of color space associated with a given memory color. Those modifications can be viewed as color differences from the color null position of the experiment, i.e. the color position of the softcopy quality ruler. Considerable experimental effort was devoted to optimizing the color and tone position of the null ruler images. A good starting point for the formulation of an objective metric for memory color reproduction is therefore the CIELAB ΔE^*_{ab} equation, which quantifies color differences in an almost perceptually uniform space. However, it is anticipated that JNDs of visibility of color differences do not translate into JNDs of image quality. For example, equal changes of lightness, hue and chroma in a perceptually uniform space (in terms of visibility) might affect image quality to a different degree. Furthermore, a fairly broad optimum of quality is expected because of color preferences among individuals. This is inconsistent with the functional form of the CIELAB ΔE^*_{ab} equation. In order to address these issues the objective metric was formulated according to the following principles:

1. The CIELAB ΔE^*_{ab} equation was expressed in terms of lightness, hue and chroma differences.
2. Variable weighting factors, w_c and w_h, were introduced for chroma and hue, which, including the remaining weight for lightness, w_l, sum to unity.
3. In order to improve the robustness of the metric (applicability to systems with a very different distribution of lightness, hue and chroma errors compared with the one that was studied here), a mean over several patches of a color set was used.

These requirements are met by a modified CIELAB ΔE^*_{ab} equation of the form

$$O = \frac{1}{n}\sum_{i=1}^{n}\sqrt{w_l \Delta L_i^2 + w_c \Delta C_i^2 + w_h \Delta H_i^2} \quad (1)$$

where O is the objective metric, n is the number of color patches included, i is the index of the patches and ΔL, ΔC, ΔH are CIELAB lightness, chroma and hue differences between the optimum position as determined in the experiment and a particular test level. In some cases interaction terms between two of the three attributes had to be included.

While ΔE_{mod} is the appropriate objective metric for memory color reproduction, the functional form of this expression still does not address the expected behavior near the optimum color reproduction position. It is anticipated that JNDs of visibility of color differences do not translate into JNDs of image quality, and that small changes in the objective metric, O, cause virtually no degradation in quality because of a wide distribution of color preferences. This issue is addressed by a psychophysical "transform" between JNDs of quality degradation or enhancement, JND, and the objective metric, O:[4]

$$JND = JND_o - \left(\frac{O}{\Delta O_\infty} - \frac{R_t}{\Delta O_\infty^2} \ln\left(\frac{R_t + \Delta O_\infty \cdot O}{R_t}\right) \right) \quad (2)$$

where O is the objective metric defined in Equation 1 and ΔO_∞, R_t and JND_o are regression parameters.

Results and Discussion

Using the functional relationships defined in Equations 1 and 2, a non-linear regression produces the predictive curve shown in Figure 2 for skin tones. This is a mean curve for all ethnic groups evaluating scenes containing persons of different skin types. The solid line is the best-fit prediction and the dotted curves show the 95% confidence intervals. The performance of two practical imaging systems representing the current state of conventional and digital imaging are shown as the • and the * symbols.

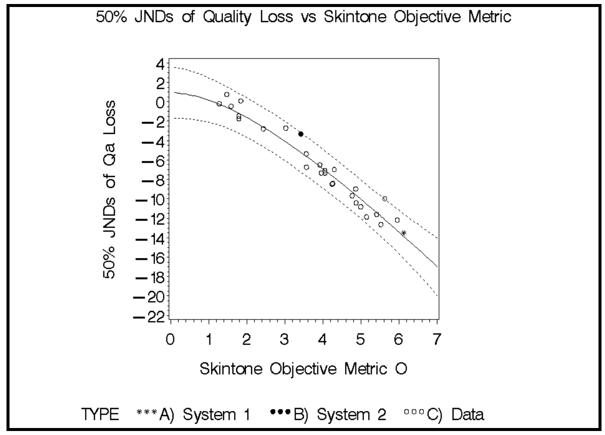

Figure 2

Extrapolated values of the objective metric predict an area of color space where improved skin tone reproduction is possible. This color and tone position would then be an enhancement relative to the reference image color and tone position. Although it is generally risky to extrapolate beyond the boundaries of an experimental design, in this case the modified central composite design completely encompasses the region of color space where the predicted optimum position occurs. The curve shown in Figure 2 represents the manner in which the image quality decreases away from the optimum in any direction, and it can be thought of as a radius from the optimum of a response surface. Image simulations with the memory colors reproduced at the optimum color and tone position demonstrate improved color reproduction thus confirming the model.

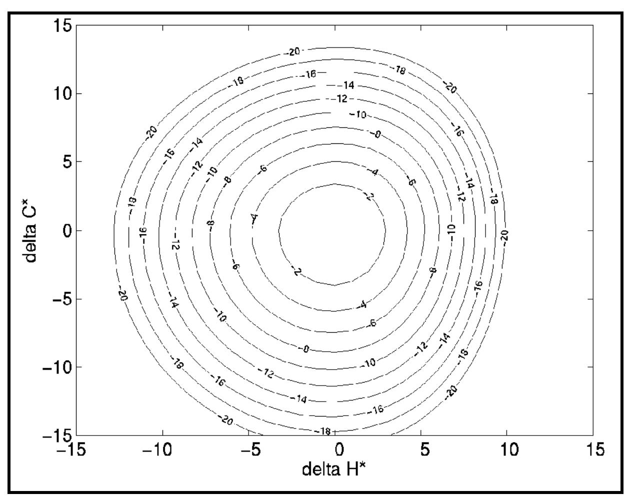

Figure 3. Contours, measured relative to the optimum in JNDs of absolute quality, as a function of CIELAB ΔC^*_{ab} and ΔH^*_{ab} distance from the optimum reproduction of Caucasian skin.

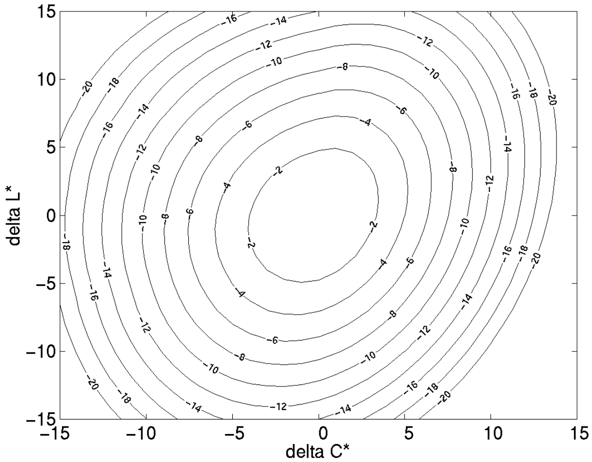

Figure 4. Contours, measured relative to the optimum in JNDs of absolute quality, as a function of CIELAB ΔL and ΔC^*_{ab} distance from the optimum reproduction of Caucasian skin.*

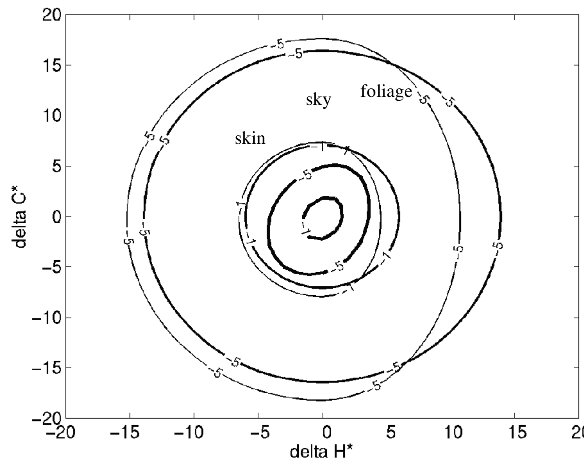

Figure 5. JND contours for skin (thick line), foliage (medium line) and sky (thin line) as a function of CIELAB ΔC and ΔH* around the optimum reproduction. The -1 and –5 JND contours are shown.*

Because the method generates a response surface in lightness, chroma and hue, sections through this surface reveal important features regarding optimal color reproduction for memory colors. In Figures 3 and 4, sectional views of the response surface in the region of color space representative of Caucasian skin tones are shown. Figure 3 shows the hue versus chroma cross section for Caucasian skin. The elliptical shape of the contours suggests that hue deviations from the optimum position are more detrimental to color quality than chroma shifts, confirming that different weights for lightness, hue and chroma are required in order to predict image quality. The assymmetric shape of the contour indicates a rapid decrease in quality for positive ΔH^*_{ab} (yellow) renditions.

Figure 4, depicting the lightness – chroma plane, shows a significant chroma/lightness interaction as elliptical contours with the major axis of highest quality reproduction approximately parallel to the saturation axis.

Although we surprisingly found approximately the same weights for lightness, hue and chroma for all three memory colors investigated, the sensitivity to deviations from the optimum reproduction differed substantially. This behavior is illustrated in Figures 5 and 6. Figure 5 shows the -1 and –5 JND contours for all three memory colors in the hue – chroma plane, while Figure 6 contains a similar plot for the lightness – chroma plane. The distance between the -1 and –5 JND contours is significantly smaller for skin tones compared with the other two memory colors illustrating the high sensitivity of observers towards sub-optimal skin reproduction.

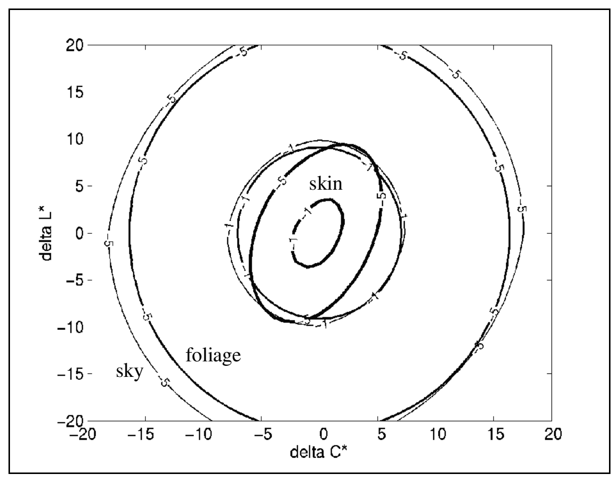

Figure 6. JND contours for skin (thick line), foliage (medium line) and sky (thin line) as a function of CIELAB ΔL and ΔC* around the optimum reproduction. The -1 and –5 JND contours are shown.*

Differences in the shape and the orientation of the contours can also be observed. In the case of the reproduction of foliage, the contours of overall quality occur symmetrically about an optimal hue. For the blue sky memory color, the contours are not symmetric with respect to hue, but indicate that quality decreases rapidly in the direction of magenta especially at higher chroma positions.

Conclusion

The results reported here are the first in a series of quantitative studies of the attributes contributing to the perception of image quality in color images. We anticipate these investigations will lead to the development of an integrated metric to predict the perceived image quality of an image. In its final form, the metric will be based on perceived color and tone scale quality of images generated by a color imaging system in terms of an absolute scale and evaluated physically using measurements of reproductions of representative color test targets.

References

1. R. W. G. Hunt, "The Reproduction of Colour", 4th Edition, Fountain Press, 1987.
2. S. N. Yendrikhovskij, "Color Reproduction and Naturalness Constraint", Thesis, Technische Universiteit Eindhoven, 1998.
3. R. W. G. Hunt, I. T. Pitt, L. M. Winter, "The Preferred Reproduction of Blue Sky, Green Grass and Caucasian Skin in Colour Photography", *J. Photogr. Science* **22**, 144, 1974.
4. B. W. Keelan, "Characterization and Prediction of Image Quality", *Proceedings of IS&T PICS 2000 Conference*, p. 197, 2000.
5. J. E. Pinney, L. E. DeMarsh, *J. Photogr. Science* **11**, 249, 1963.
6. R. B. Wheeler, "Use of System Image Quality Models to Improve Product Design", *Proceedings of IS&T PICS 2000* Conference, p. 204, 2000.
7. E. J. Giorgianni, T. E. Madden, "Digital Color Management", Addison-Wesley, 1997.

Biography

Karin Töpfer received her Masters degree in Physics from Dresden University of Technology in 1983 and a Ph.D. in Photophysics from Dresden University of Technology in 1988. Since 1993 she has worked at Eastman Kodak Company, first in the U.K. and later in Rochester, NY. Her work has primarily focused on image quality, including sources of grain in conventional silver halide systems, image quality modeling and psychophysics. She is a member of IS&T and a Fellow of the Royal Photographic Society.