The Effect of Image Content on Color Difference Perceptibility

Susan P. Farnand
Eastman Kodak Company, Rochester, New York

Abstract

Considerable work has been accomplished regarding the perceptibility of color differences for simple images. Less is known regarding color difference perceptibility when complex images are involved. In this research, psychophysical experimentation was conducted to determine if image content, in terms of memory color or image component size, is a significant factor in color difference perceptibility. A portrait, a nature scene, an image containing strictly man-made objects, and four mosaic images composed of different size patches were examined. The experimental results suggest that image component size affected the perceptibility of color differences while the presence of memory colors had no conclusive effect.

Introduction

Possessing a lucid understanding of color difference perceptibility or the minimum color change needed between an original and a shifted copy for the average observer to see a difference between them, is of utmost importance for such industries as those producing automotive paints or textiles. These industries have pressing quality control questions regarding how much colors can change from batch to batch, dye lot to dye lot, before that difference will become evident to their customers. Consequently, past color difference perceptibility research typically incorporated simple color patches, representative of such things as paint chips or textile samples, as the original and shifted copies. The case for research involving pictorial images is somewhat different. With pictorial images, such as photographic prints, the question that has most concerned researchers is color difference acceptability or how much the color can shift between original and copy before the average observer finds the change unacceptable. The crucial question for the photographic industry was how much the color could change between the original scene or print to the photographic print or reprint before customers were no longer willing to pay for it. Research into this question revealed that observers were much less tolerant of color shifts in prints such as portraits or natural scenes that contain colors for which they have some mental record or idea of what "looks right".[1,2] It seems reasonable that the presence or absence of "memory colors", especially skin tones, affects the acceptability of color differences in complex images. It would be interesting to know whether this memory color effect on color difference acceptability extends to the case of color difference perceptibility. It would also be interesting to know if other aspects of image content effect color difference perceptibility. Perhaps merely the presence of recognizable objects to provide visual cues is enough to affect color difference perceptibility. Perhaps something altogether different such as image complexity is relevant to the understanding of color differences in complex images. The relative degree of complexity of an image may play an important role if a significant relationship exists between it and simultaneous contrast. Experimental evidence in fact suggests that the degree of induced color is proportional to the size of the inducing field.[3] To examine the effect of scene content on color difference perceptibility, psychophysical experimentation involving several image types including mosaic images comprised of different sized tiles, was conducted.

Technical Approach

To investigate the effects of image content on color difference perceptibility, a pass-fail experimental procedure having an reference pair for comparison was conducted with images having varied scene content. The observers participating in the experiment were asked to evaluate whether the color difference of each of an array of sample image pairs appeared to have a larger or smaller color difference than a reference image pair. Seven different scenes were examined in the psychophysical experimentation; a portrait, a nature scene, a scene containing man-made objects, and four mosaic images. The portrait, "Tim", depicted a young boy. For the nature scene, or "Trees", a fall scene that was comprised of similar colors to those in the portrait scene was used. Sewing thread of colors similar to those contained in the portrait and nature scenes was used for the image composed of man-made items, "Yarn". The four remaining images were mosaics; one comprised on half-inch squares, "Half", one quarter-inch squares, "Quarter", one eighth-inch squares, "Eighth", and one sixteenth-inch squares, "Tiny". The mosaic images were generated from colors contained in the portrait, nature scene, and thread scene. The mosaics, the portrait, the "Yarn" image, and the simple patch used as the reference to which all the sample images were compared, all had the same integrated L*, a*, and b* values. Two types of variation were evaluated in the experiment; shifts in the +a* direction in CIELAB color space and shifts in the -b* direction. For each type of variation, six sample pairs of each image, three pairs having smaller and three having larger color differences relative to the reference pair, were used in the experiment. With six levels of ΔE^*ab for each of seven images in each of two directions in CIELAB color space, a total of 84 comparisons were made by each observer.

The experimental results were compiled and SAS statistical software was used to perform probit analysis in each of the examined CIELAB spatial dimensions. In probit analysis, frequency of rejection responses are transformed into probability units or "probits". In this experiment, the frequency with which a particular sample pair was deemed to have a greater color difference than the reference pair served as the experimental response. The relationship between the frequencies of "greater than" responses as a function of increasing color difference is assumed to approximate a cumulative normal distribution. The chi-square goodness of fit test is used to examine the validity of this assumption. The 50% probability or threshold point represents the stimulus that is visually equivalent to the reference. SAS also provides upper and lower 95% fiducial limits around this point that essentially establish a 95% confidence interval. If these intervals overlap substantially, then there is no apparent impact of image content on color difference perceptibility for the images evaluated in the experiment. Conversely, if the intervals do not overlap at all, an effect of image content on color difference perceptibility is reasonably assumed. If the intervals overlap somewhat, some degree of uncertainty remains regarding the impact of image content.

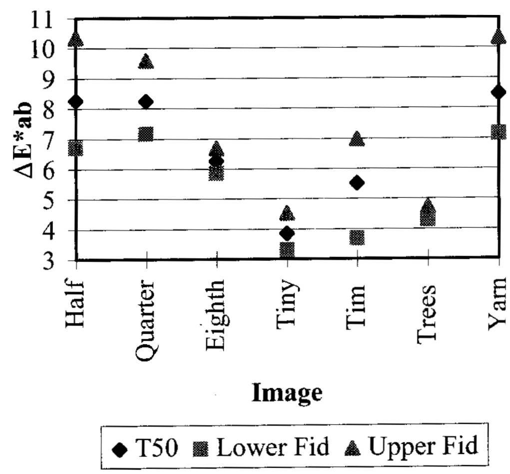

Figure 1. The 50% threshold points and related fiducial limits for the a-shifted images with all observers considered. The reference pair had a shift of 6.0ΔE*ab*

Results and Discussion

The graph in Figure 1 shows the thresholds and limits for the a*-shifted images. Three statistically different groups of images resulted. The "Half", "Quarter", and "Yarn" images constitute the group having the highest difference thresholds, about 8 ΔE*ab. The "Eighth" image, which alone makes up the second group, has a difference threshold of about 6 ΔE*ab, similar to the 6.0 ΔE*ab shift in the reference pair. The "Tiny" and "Trees" images, the final group, have difference thresholds of around 4 ΔE*ab. The interval for the "Tim" image overlaps those of the two latter groups and so is not easily separated from either of these groups. Since statistically significant groups of images occurred, it appears likely that there is some effect of image content on color difference perceptibility. The images having recognizable objects did not consistently appear to be different from those that did not. Images comprised of different sized patches, however, did appear to be different from one another with thresholds tending to decrease as the size of the patches decreased. It is interesting that the "Half" and "Quarter" images had higher thresholds than the "Eighth" and "Tiny" images since many observers commented that they would tend to concentrate on particular patches that often were similar in color to the reference patches when viewing the "Half" and "Quarter" images. These particular patches were mid-tones which individually tended to shift more than the reference to compensate for the light and dark patches that experienced smaller individual shifts. Such observers should have found the "Half" and "Quarter" images to have lower thresholds than the "Eighth" and "Tiny" images. While a few of these observers may have individually had lower thresholds for the larger patch images, the overall result, as well as the result for any subgroup of observers, was that the larger patch mosaics had the higher thresholds, for the patches included in this experiment. At some point as patch size increases, the threshold must shift back toward that found for a simple patch, given that the sample patch color does not differ greatly from the reference patch color.

The conclusions for the pictorial images are not as clear. A statistically significant difference between the "Yarn" and the "Tim" and "Trees" images is apparent. A specific reason for this difference is not as evident. The "Yarn" image does not have memory colors as the "Tim" and, to a lesser extent, the "Trees" images do. However, it seems more likely that the difference was a result of the areas of each image on which the observers tended to concentrate. Many of the observers stated that they concentrated on the skin or hair areas of the portrait and the dried grass area of the "Trees" image. It is possible that the colors in these regions shifted more than the reference patch to compensate for colors in other areas which were shifting less than the reference patch. Whatever the reason, a statistically significant difference between the "Yarn" and the "Tim" and "Trees" images does exist, suggesting that image content may indeed impact color difference perceptibility.

The graph in Figure 2 shows the thresholds and limits for the b*-shifted images. This graph indicates, once again, that there are three groups of images that are statistically different from one another, although there are a few differences from the a* shifted case. The "Half", "Quarter", and "Yarn" images still make up the group of images having the highest difference thresholds, about 13ΔE*ab. The "Trees" image, in this case, however, makes up the second group, with a difference threshold of over 11 ΔE*ab. The "Tim" image this time joins the "Tiny" and "Eighth" images in the final group, which have difference thresholds of between 9 and 10 ΔE*ab. The reference image pair for the b*-shifted image set had a ΔE*ab of about 10.4. Although there were some changes in which images are statistically different from other images, the same basic result holds; larger patches yielded larger thresholds.

The difference in the pictorial images that occurred for the a*-shifted image set is also evident for the b*-shifted set. Again, the "Yarn" is statistically significant different from the "Tim" and "Trees" images. In this case, however, the "Tim" and "Trees" images are also significantly different with the "Tim" image having the lower threshold. One reason for the low threshold for the b*-shifted "Tim" image may be found in the results for skilled observers. For this image, the difference threshold for the skilled observers is substantially lower than the threshold value for the naive observers. Skilled observers are likely more conditioned to looking at the skin tone areas in a portrait image. It may be possible that skilled observers could more objectively analyze the entire image when the little boy was turning red than when the little boy was turning blue.

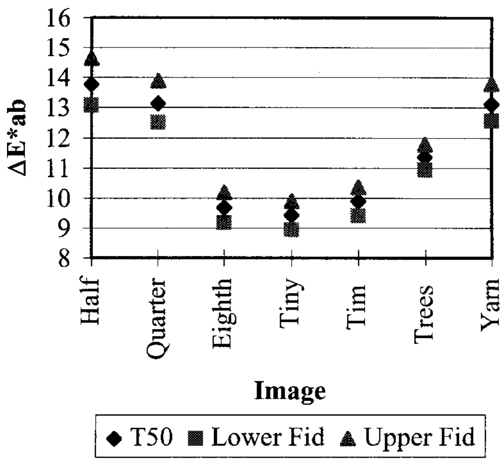

Figure 2. The 50% threshold points and related fiducial limits for the b-shifted images. The reference image pair for the b*-shifted image set had a ΔE^*ab of about 10-1/2.*

The b*-shifted "Tim" image was one of the few images for which observer experience mattered. Of the experimental factors evaluated, including image order and observer experience, gender, and age, image order had the most significant impact. In this experiment, half of the observers evaluated the a*-shifted images first, progressing up through a fixed random order while the other half looked at the b*-shifted images first, counting down through the same random order. For the a*-shifted images, the observers who viewed the a*-shifted set first had considerably noisier results than the observers who viewed the b*-shifted image set first. As a consequence, despite larger shifts in CIELAB space, the b*-shifted image set produced less noisy results than the a*-shifted image set, as seen by the generally tighter fiducial limits set on the thresholds of the b*-shifted images, Figure 2. Also, the observers who viewed the a*-shifted images first had higher thresholds for the a*-shifted set than the observers who viewed the b*-shifted image set first. For the b*-shifted images, the observers viewing the a*-shifted images first had lower thresholds than the b* first observers. These results indicate that observers became more sensitive to color differences in the sample pairs or harsher in their judgments as the experiment progressed. This shift in sensitivity led to noisy results for some of the a*-shifted images. A short trial run allowing observers to establish viewing criteria prior to the actual experiment would likely have been helpful for limiting experimental noise.

One other interesting result was that observer age appeared to have an effect for both "Eighth" images as well as the b*-shifted "Half" and "Tiny" images. The younger observers had a lower threshold for the b*-shifted "Half" image and older observers tended to have lower thresholds for the other three images. Older observers, who generally have lower visual acuity, may be integrating the colors represented by the individual patches to a larger degree than younger observers.

Conclusion

Psychophysical experimentation was conducted to investigate the impact of scene content on the perceptibility of color differences. The analytical results suggested that scene content in pictorial images had a significant effect, although the reasons are not clear. The "Yarn" image, which contained no memory colors, had a significantly higher threshold than either of the other two pictorial images, indicating that the average observer had a higher tolerance to color shifts in this image than for the portrait or nature scene. However, the difference between the "Yarn" image and the other pictorials may have been the result of observers concentrating on specific areas of the "Tim" and "Trees" images that were shifting more rapidly and, as a result, more distinctly showing the color shift than any areas of the "Yarn" image. The presence of skin tones may also impact color difference experimental results because, when skin tones were present in an image, observers, especially skilled observers, tend to focus on those areas of the image. It seems then that to get a true idea of the effects of memory colors we need to understand if it matters whether a skin color appears in an image as a man-made object such as thread, a natural object such as dried grass, or as skin itself.

The experimental results did indicate that the size of the image elements has an impact on color difference perceptibility. The difference threshold increased for larger image elements. It seems likely that the eye may effectively integrate the colors of very small patches where larger patches would appear more clearly as separate entities. Future experimentation conducted to explore the question of image content effects on color difference perceptibility may focus on this issue of image complexity or size of image elements. Such a focus would require added attention to observer viewing distance and visual acuity. Whatever the focus of further experimentation, a short trial run to allow the observers to establish evaluation criteria should be conducted to reduce experimental noise.

Acknowledgment

The author wishes to thank Dr. Mark Fairchild, who guided this work, Mr. David McVay, who printed the test images, and all of the observers for their time and patience.

References

1. C. J. Bartleson, "Some Observations on the Reproduction of Flesh Colors", *Photo. Sci. and Eng.* **3** 114-117 (1959).
2. R. W. G. Hunt, The Specifications of Colour Appearance. II. Effects of Changes in Viewing Conditions. *Color Res. Appl.*, **2** 109-120 (1977).
3. J. S. Kinney, "Factors Affecting Induced Color", *Vision Research* **2** 503-525 (1962).

published previously in the IS&T 1996 Color Imaging Conference Proceedings, page 101