https://doi.org/10.2352/ISSN.2470-1173.2017.12.IQSP-250

A methodology for perceptual image quality assessment of smartphone cameras – color quality

Susan Farnand[1], Young Jang[2], Lark Kwon Choi[2], and Chuck Han[2];
[1]Rochester Institute of Technology, Program of Color Science, Rochester, NY
[2]Qualcomm, San Diego, CA

Abstract

This research project was undertaken to develop a procedure for evaluating perceptually-based pictorial image quality for smartphone camera captures. Tone quality, color quality, and sharpness and noise were evaluated in separate experiments. In each test, observers scaled overall quality and then the individual image quality characteristic of test images from a variety of smartphone cameras relative to an anchor image. Results were reported in 2016 on the individual image quality characteristics relative to overall quality as well as on the development of objective measurements that correlate with the visual ratings for sharpness and noise. (Farnand, et al., 2016) In this work, the visual ratings for color quality were assessed relative to objective measurements with the results from this analysis indicating that high correlations between the two can be achieved. The perceptual results correlated best with colorimetric information taken from the test images, rather than images of test charts that were captured under lab conditions, which did not necessarily contain colors representative of the colors important to the scenes. Results also indicated that contrast information was needed in addition to the colorimetric information in order to achieve high correlation between subjective and objective information for all scenes. Further, results for a beach scene suggest that sand may serve as a useful memory color for predicting device capture perceptual performance.

Introduction

Having a methodology for assessing smartphone camera image quality is advantageous for both those who design and develop the cameras as well as those who use them. Camera engineers need to quickly and reliably assess the impact of the system decisions they make. Smartphone customers who are armed with a quantitative understanding of the image quality can use this information to make informed decisions between products. The VICTOR (Visually Integrative Camera Test and Open Report) project was initiated to establish such a methodology. The objective of this research project was to develop a procedure for evaluating perceptually-based pictorial image quality for smartphone camera captures, including metrics for tone and color quality, sharpness, and noise.

In past work, it has been shown that proper tone reproduction is crucial to achieving well-rated perceptual color quality (Frey and Farnand, 2010). It has also been shown that measurements of the colors in the images themselves provides a better assessment of perceived quality than colors in a test chart that are not significant in the scenes. This work confirms these two findings, indicating that contrast information as well as measurements of key scene colors are needed to consistently achieve high correlations between subjective and objective color quality results.

Experimental Methodology

The experiment was conducted in three segments: one each for evaluating tone quality, color quality, and sharpness and noise quality. Each segment comprised six or seven test scenes photographed with 20 smartphones. Seven scenes covering a variety of content were used in the color quality evaluation. The scenes included skin tones, sky, grass, wood, sand, brick, and food (green peppers), Figure 1. Four of these scenes were captured indoor and three outdoor. The dominant colors in each scene ranged from being highly chromatic to being near neutral. In past print-related studies, it has been determined that accurate reproduction of key colors in a scene, as opposed to reproducing a large number of colors that may or may not be related to the given scene (such as an IT8 target), provides the best perceived match to the original. (Farnand, 2003) With consumer photographs, where the original is generally not present, preferred reproduction (which may or may not be accurate) relates to the observer's memory of the original scene. Colors that are important in observers' memories typically include skin tones, foliage, and sky. These colors were carefully developed for the Macbeth ColorChecker capture test chart. (McCamy, 1976) For this reason, test scenes that included these colors, along with food and natural materials such as wood and brick, were chosen for this study.

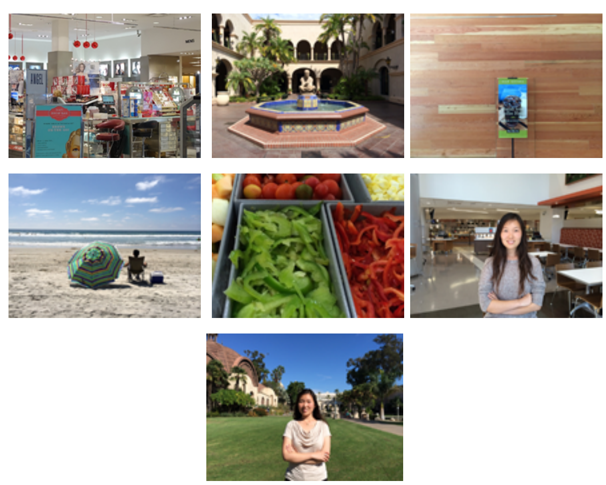

Figure 1 *Scenes used in the Color experiment, from top left – Mall, Courtyard, Wood, Beach, Peppers, Mixed Light, and Memory Colors*

The experimental setup, shown in Figure 2, was consistent across the three tests. Observers were seated in front of the HP_ZR30w display at an unconstrained viewing distance of approximately 18 inches. The room lights were extinguished for the duration of each test. Figure 2 depicts the layout of the experimental interface, which followed an anchored scaling experimental protocol (Engeldrum, 2000). The lower quality anchor image was shown on the lower left side of the display and the higher quality image on the lower right side of the display. These anchors were arbitrarily assigned the values of 30 and 75, respectively. The test images were displayed above the anchors. Observers first scaled the test images for overall quality. Then they were shown the images a second time in a different random order and asked to scale the images for the individual image quality characteristics. Twenty observers participated in each experimental segment. The observer gender, ethnicity, age, and academic program (or experience) were recorded. All observers who participated in the experiment had normal color vision and normal or corrected to normal visual acuity.

Figure 2 *Experimental setup. The room lights were turned off for the duration of the experiment.*

Results and Discussion

The experimental results reported in 2016 included those that showed high correlation (0.90 or higher) between overall image quality and color quality ratings. These results also showed that the observers in the experiment were in closest agreement in their color quality and overall quality ratings for the Wood, Peppers, and Courtyard scenes. These scenes each had one or two large areas of natural materials, which likely represent relevant memory colors. It suggests that for these types of scenes, when properly focused, color quality determines perceived overall quality.

The results reported in 2016 also indicated that images assessed as having high color quality generally are also found to have high tone quality. Of the three scenes that were included in both tests, the correlation between tone and overall quality was lower for the 'Beach' scene than the 'Courtyard' and 'Memory Color' scenes. However, the correlation for Color and Overall Quality for the Beach Umbrella scene was much higher, 0.93, than it was for tone quality, 0.82. For this, relatively low-dynamic range scene, color quality was more critical to the assessment of overall image quality than the tone quality, which may have been more difficult for observers to assess.

As part of this research, images were made with each device of the TE42 test target captured under several lighting conditions. (Altmann, 2015) None of the mean differences in lightness, hue, or chroma from the color patches in these target images correlated well with the subjective ratings of overall color quality. Past research regarding color differences between images has suggested that the evaluation of differences in colors important to the scene provide a better prediction of the perceived difference than differences between an arbitrary set of colors such as that provided by a color test chart (Farnand, 2003; Frey and Farnand, 2011).

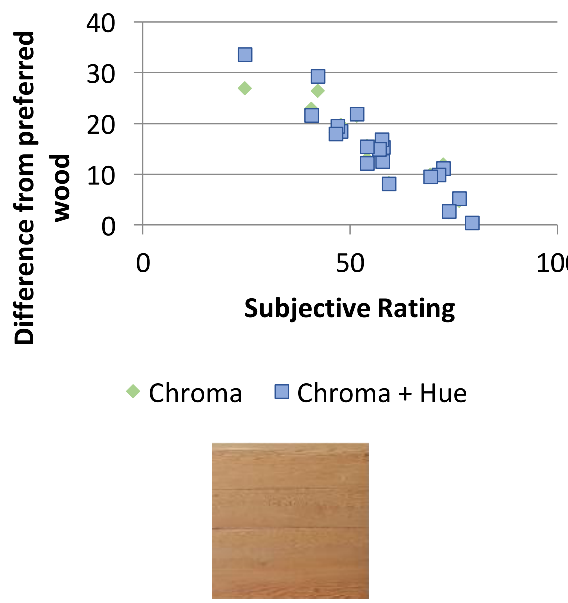

Figure 3 *Objective versus subjective results for the Wood scene using CIELAB data calculated from an area of the images, an example of which is shown below the graph.*

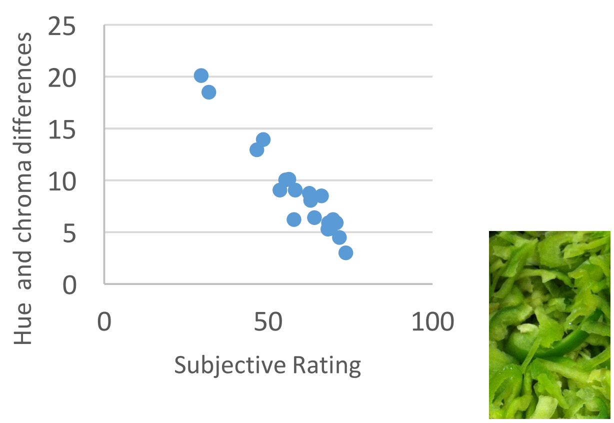

Figure 4 *Objective versus subjective results for the Peppers scene*

Two scenes that demonstrate the impact of evaluating relevant colors are the Peppers and Wood scenes. The subjective performance for the Wood scene could be reasonably well predicted (correlation coefficient of -0.93) with a weighted combination of the chroma and hue differences from the preferred color wood in this experiment. (The CIELAB values were determined for selected relevant areas of the images, shown to the right in Figures 3&4. The RGB to L*a*b* calculations were performed in Matlab.) As seen in Figure 3, chroma played a key role in an effective objective measure. Similar to the Wood scene, in the Peppers scene, Figure 4, a weighted sum of the hue and chroma differences could be used to predict the subjective ratings. For Peppers, however, hue played a larger role than chroma, though both were needed to achieve an objective measure that well predicts the subjective ratings. For the scene, observers preferred their peppers to be green rather than yellow (or gray), though peppers that went too far toward cyan were objectionable as well.

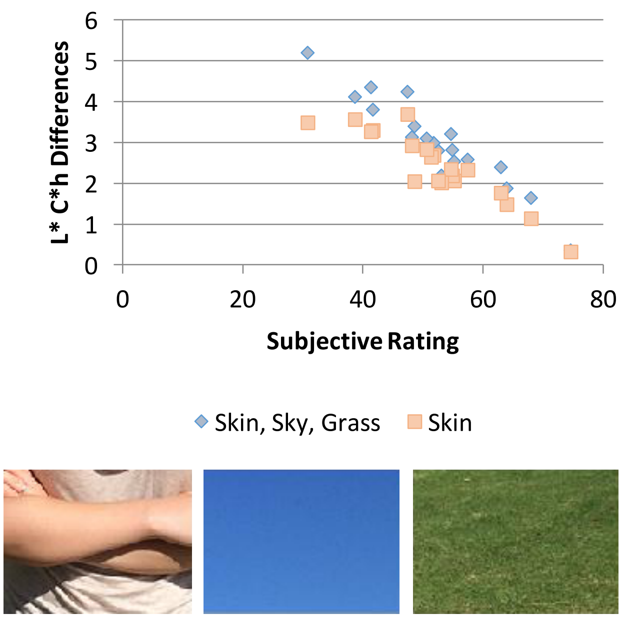

Figure 5 *Objective versus subjective ratings for the Memory Colors scene, where objective results were calculated based on combined skin (weighted highest), sky, and grass (areas shown above) and with skin alone. Areas of the images used in the CIELAB calculations shown below the graph.*

When evaluating changes in colors taken directly from the images, the results for the color assessment include interesting findings relative to the traditional memory colors of skin, sky, and foliage. They showed that, for cameras to rank high over a range of scenes, they must reproduce skin tones well. One of the scenes in the test set, labeled Memory Colors, included the standard memory colors of skin, sky, and grass. For this scene, observer ratings of color quality are well-predicted using just the lightness and hue of the skin content (correlation coefficient of -0.91), Figure 5. Adding in characteristics of the sky and grass areas improved the correlation between the objective and subjective results (correlation coefficient of -0.95), but were not necessary to achieve a reasonable idea of the relative performance of most of the devices for this scene. The addition of sky and grass provided objective values that were more in line with the subjective ratings for the cameras with the lowest rating and one with good skin tones but for which the grass and sky were much lighter than optimal.

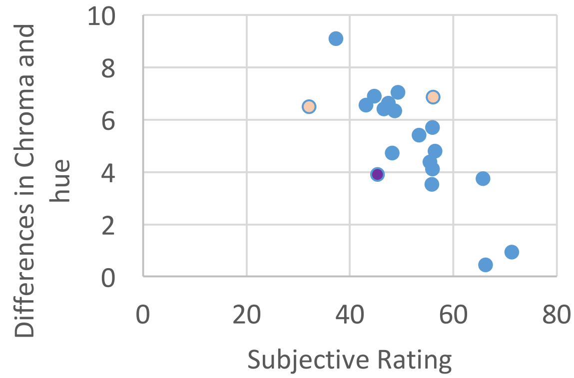

Figure 6 *Objective versus subjective ratings for the Mixed Light scene. Data points for the scenes shown in Figure 7 are shaded orange. Face area used in calculating CIELAB values shown in Figure 7*

Figure 7 *Renditions of the Mixed Light scene that were subjectively rated differently but had similar objective ratings based on chroma and hue differences, illustrating that it may be necessary for tone to be included as well.*

A second scene in the set that included skin tones, referred to as Mixed Light, was composed of a woman under indoor, mixed lighting conditions. For this scene, the difference in chroma from the optimum of the skin tones in the woman's face correlated well with the subjective results for seventeen of the twenty devices. For two of the remaining devices, contrast also had to be considered. The lowest rated device had essentially the same average colorimetric values as several of the higher rated devices. However, the distribution of these values was quite different, Figure 7. The device rated lowest had a higher percentage of light values and fewer mid-tones than the higher rated devices. These results suggest that, for skin tones to be reliably well-reproduced the values throughout the tone curve must be well-reproduced, not just the average lightness levels.

While adding information on the sky and grass colors improved the performance of the objective measures in predicting subjective ratings for the Memory Colors scene, it was found that adding colorimetric information for a white area to the analysis of the Mixed Light scene did not improve the prediction performance. The lightness, a*, b*, and chroma were uncorrelated with the subjective ratings, correlation coefficients of -0.1 to 0.1. These results suggest that skin tones are more important to overall perceived color reproduction performance than the appearance of a background white.

In contrast to the results for the Mixed Light scene, for which colorimetric data for a white area were not helpful in predicting subjective results, the experimental results for the Beach scene suggest that the relatively light, relatively neutral color of sand may serve as an important, 'whitish' memory color for image and color quality evaluation. While the sand color is relatively neutral, it was the devices that retained the chromatic information that rated higher for this scene then others. The sand needed to look like sand and not concrete. This was also true, to an extent, for the Mixed Light scene. Chroma was the driving color characteristic for this scene. It was the smartphone cameras that retained a higher level of chroma for the skin tones that were scaled higher than those for which the skin tones were driven toward neutral.

The colorimetric information for the sand in the Beach scene, taken together with that of the sky, could be used to generate a reasonable relationship between the objective colorimetric data and the subjective ratings (correlation coefficient: -0.90), Figure 8. Further, the results for this individual scene correlate well with overall device performance across all scenes. The subjective ratings for the Beach scene had the highest correlation of any scene with the subjective ratings averaged over all of the scenes (correlation coefficient: 0.87). Interestingly, the objective values for the Beach scene predict the overall quality ratings for the devices averaged over all of the scenes, with a correlation coefficient of -0.83, Figure 9. This indicates that this scene, containing sand and sky, may be valuable for evaluating the general perceived color reproduction performance of smartphone devices.

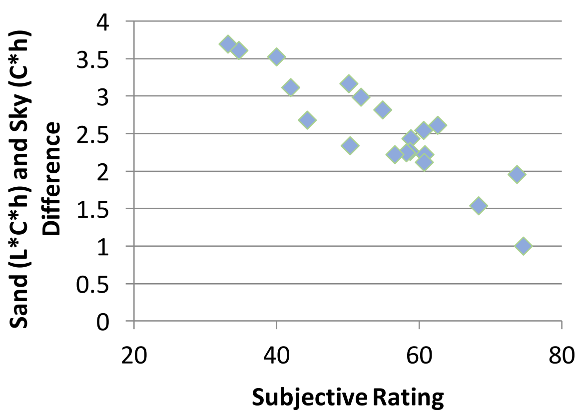

Figure 8 *Objective versus subjective results for the Beach scene*

The scene that is the least predictive of the overall color quality ratings is the Mall Scene. The subjective ratings for this scene, which is composed of primarily light colors, but no strong memory colors, do not correlate with the average ratings for the other six scenes (correlation coefficient of 0.15). The two lowest rated devices for the Mall scene were mid-rated devices, on average, for the other scenes while the three lowest rated devices for the other six scenes were rated in the mid-to-high range for the Mall scene. This may suggest that scenes containing memory colors are particularly useful in assessing color reproduction performance.

In this study, using color data from the TE42 was not effective in predicting subjective color quality. Skin tones, sky blue, and grass green colors were evaluated relative to the Memory Colors scene. None were found to correlate at a level over 0.40. The TE42 patch most similar to the wall in the Wood scene was evaluated relative to the subjective results of this scene. The results were not correlated (correlation coefficient -0.25). More promising results were found for one of the skin tone patches relative to the Mixed Light scene. The chroma differences for the Patch J1 were reasonably correlated (correlation coefficient -.80) with the subjective ratings, except for the device with the contrast issue, Figure 7. This is also the color on the TE42 chart that is most similar to sand. It was evaluated, along with the lightness differences for sky colors (Patches A4, A5, B4, & B6), relative to the subjective ratings for the Beach scene. Weak correlations were found with the ratings for the Beach scene (correlation coefficient -.64) as well as for the overall color quality ratings (correlation coefficient -.68). The highest correlation with the overall color quality ratings (correlation coefficient -.74) was found for the mid-gray G5 patch. In summary, mid-gray and sky blue lightness differences and chroma differences of a light skin tone for color patches on the TE42 target were found to most closely predict perceived color quality of pictorial images captured by smartphone cameras, though the correlation was not strong.

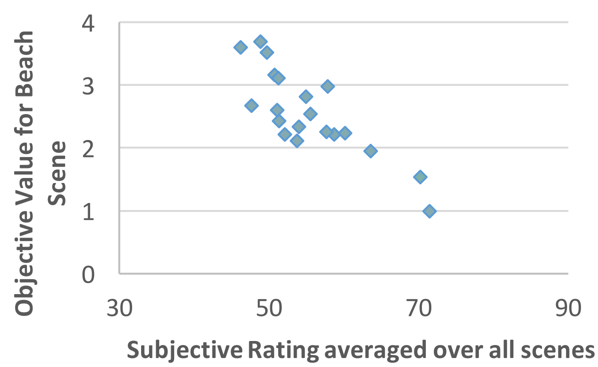

Figure 9 *Objective versus subjective results for the Beach scene relative to the overall subjective ratings*

Conclusions

The eventual goal for this effort was to develop a methodology that brings the individual image characteristics together into an overall metric of the perceptual quality attainable by capture systems. As a first step toward this goal, subjective data was gathered for the individual characteristics of tone, color, sharpness, and noise. The results of these experiments were reported in 2016. Initial evaluation of the color results relative to objective data for the target capture indicated that correlations were poor. In this work, the subjective ratings were compared to dominant colors in the scenes. Correlations between the visual results and objective measurements for lightness, chroma and hue differences for the individual scenes were significantly higher. The color results in aggregate indicate that the scene colors work quite well for establishing objective measures for predicting perceived camera phone performance for that scene. However, color reproduction performance for patch targets or even an array of pictorial scenes may not predict performance for a particular scene. The results for the Mall and Wood scenes indicate that devices that rank poorly overall may reproduce individual colors well. The rankings for the Mall scene do not correlate with any other scene and one of the worst performing devices for the scenes overall (only two devices were rated lower, on average) was the highest rated device for the Wood scene.

With smartphone cameras performing differently depending on scene content, even for similar lighting situations, analyzing color quality performance is complex. Certainly reproduction of colors common in smartphone captures is essential. Skin tones likely rank highest in importance in this regard. Sky, foliage, and neutrals are other color categories that have historically been important in consumer photography. This study indicates that sand may be a key color for smartphone imaging, possibly because its near neutrality may be useful for identifying white balance issues. Finally, this study reinforces previous findings that proper reproduction of tone scale is critical for high perceived color reproduction quality.

Acknowledgments

Many thanks to Qualcomm for sponsoring this work and do all of the observers who took the time to participate in the perceptual experiments.

References

Artmann, U, “Image quality assessment using the dead leaves target: experience with the latest approach and further investigations”, in proceedings of IS&T Electronic Imaging Symposium, Vol. 9396, San Francisco, CA, 2015.

Engeldrum, P. G., Psychometric Scaling: A Toolkit for Imaging Systems, Imcotek Press, Massachusetts, 2000.

Farnand, S. P., *Using E metrics for measuring color difference in hard-copy pictorial images*, *Proc. SPIE vol. 5008*, pp. 109-118, 2003.

Farnand, S. P. and Fairchild, M. D., *Designing pictorial stimuli for perceptual experiments,* Applied Optics, Vol. 53 Issue 13, pp. C72-C82, 2014.

Farnand, S. P., Jang, Y., Han, C., and Hwang, H., *A methodology for perceptual image quality assessment of smartphone cameras*, in proceedings of IS&T Electronic Imaging Symposium, Image Quality and System Performance XIII, pp. 1-5(5), San Francisco, CA, 2016.

Frey, F. S., Farnand, S. P., *Benchmarking Art Image Interchange Cycles*, Final Report for The Andrew W. Mellon Foundation, www.rit.edu/cos/colorscience/re_archived_projects.php, 2011.

Keelan, B. W. *Handbook of Image Quality: Characterization and Prediction*, Marcel Decker, NY, pp. 160-162, 2002.

McCamy, Calvin S., H. Marcus, and J. G. Davidson. "A color-rendition chart." *J. App. Photog. Eng* 2.3 (1976): 95-99.

Author Biography

Susan Farnand is a Visiting Assistant Professor in the Program of Color Science at the Rochester Institute of Technology. Her research interests include human vision and perception, color science, cultural heritage imaging and 3Dprinting. She received her BS in engineering from Cornell University, her Masters in Imaging Science and her PhD in Color Science from the Rochester Institute of Technology. She began her career at Eastman Kodak, designing and evaluating printer systems. She is Publications Vice President of the international Society of Imaging Science and Technology and serves as an Associate Editor for the Journal of Imaging Science and Technology. She participates in several Standards efforts including ISO TC 42 JWG26 Archival Imaging.