Update on the INCITS W1.1 Standard for Evaluating the Colour Rendition of Printing Systems

Susan Farnand
Eastman Kodak Company
Rochester, NY USA

Frans Gaykema
Océ-Technologies B.V.
The Netherlands

William Kress
Toshiba America Business Solutions
Irvine, CA USA

Óscar Martinez
Hewlett-Packard Company
Barcelona, Spain

Ann McCarthy
Lexmark Corporation
Lexington, KY USA

Helen Shin
Xerox Corporation
Rochester, NY USA

Eric Zeise and Dmitri Gusev
NexPress Solutions
Rochester, NY USA

Introduction

The Colour Rendition *ad hoc* team is working to create a standard for evaluating colour rendition of hard copy output generated from digital input. To accomplish this, the team is considering what it means for the colour to "look right," and how to measure the ability of any given printer to properly render colour. To date, the first of these issues has been addressed. In working to define colour rendition, we started with a discussion of what is and what is not included within colour rendition. Tonal rendition was considered integral to colour rendition, while variations in colour rendition across an image where a constant code value is specified was considered as outside the realm of colour rendition, and would, rather, be under the auspices of the macro-uniformity or micro-uniformity sub-attributes. An understanding of what constituted "the printer" was also needed. It was decided that the printer would be defined as "everything that occurred after hitting the 'print' button." Finally, it was agreed that different metrics might be necessary for natural images and vector graphic images because of the adaptation effects and differences in the image content and intended use of vector graphic images. With these considerations in place, three important parameters or subattributes were identified as being necessary to describe what it means for colour to be well rendered: 1) colour quantization, defined as the ability to merge colours where needed; 2) colour scale, defined as the ability to distinguish colour where needed; and 3) colour fidelity, defined as a balance of colorimetric accuracy, in cases where a reference exists, and a pleasing overall colour appearance. These definitions and descriptions of their appearance can be found in the periodic summaries of the colour rendition work, the most recent being the 2005 IQSP Conference.[4] Colour scale is a term developed to be analogous to tone scale but which is meant to include not just changes in lightness, as might be thought of for tone scale, but changes in chroma and hue as well. This definition of colour scale should highlight the fact that tonal rendition was considered part of colour rendition.

With the definition complete, efforts turned to addressing the second question of how to measure colour rendition. This work started by looking at the subattribute of Colour Quantization. A test target was designed and generated and initial experimentation was conducted. This work has been summarized periodically as it has progressed; most recently at the 2005 IQSP Conference.[4] More recently, the focus of the Colour Rendition *ad hoc* team has been on the sub-attribute of Colour Fidelity. Following a brief review of the accomplishments in the area of Colour Quantization, progress of the Colour Fidelity work in the past year will be summarized today.

Colour Quantization

Once colour rendition was defined, the next step was to establish methodology for measuring the three subattributes of colour rendition. Initially, efforts focused on the colour quantization subattribute because agreement of a general approach for colour quantization came most easily. With colour quantization defined as "the ability to merge colours where needed", a failure in colour quantization would manifest itself in images primarily as contouring. It was clear that perceptible contours constituted a problem. Images with sufficient colour quantization would be free of the contouring that can appear in gradient areas, such as sky, walls, or faces, in pictorial images, or backgrounds, in graphic images. It seemed then that colour quantization could be adequately evaluated using colour sweeps.

A strategy of first developing a test target, followed by a method of measuring the target, was adopted because of the difficulty of having any meaningful discussion without data or, at least, an attempt at obtaining data. After much discussion and a few trial runs, it was decided that the test target should comprise colour gradients in a serpentine format. A monochrome black target having steps between consecutive patches of one printer count were printed on photographic, electrophotographic, and inkjet printers and were evaluated for contours by several members of the group. The results of these initial evaluations, which were well correlated, indicated that the visibility of contours was related to the ΔL (or ΔE) between two adjacent patches along with the density of those patches, with the contours being more apparent in the mid-density range.

Colour Fidelity

With the development of the colour quantization measurement methodology well underway, it was decided that next subattribute to be addressed would be colour fidelity. The colour fidelity subattribute of colour rendition is defined as a balance of colorimetric accuracy (hue, lightness, and chroma) in cases where a reference exists and having a pleasing, overall colour appearance, including reasonable reproduction of memory colours. Although the ability to match colours is included in colour fidelity, it is not meant to imply that matching to a target or original is always necessary or preferred. Perhaps because it includes the concept of "pleasing appearance", this attribute has been and, no doubt, will continue to be a challenge and will likely require separate test targets and metrics to characterize images with text and graphical content and others with pictorial content. While in graphical and text images a colorimetric match may often be desirable, a pictorial image with good colour fidelity should have skin tones, grass, and sky areas that look realistic. Pictorial images, in general, should look natural.

The effort to develop an evaluation methodology for colour fidelity was initiated by conducting psychophysical experimentation using a set of images printed on equipment having a wide range of image quality capability in an attempt to establish an initial scale of colour fidelity. A set of eight pictorial images that included four images containing people, two scenic images, and two images depicting building interiors was used. A custom colour chart, consisting of 64 important colours, was extracted for each image using the "Colour Table" feature included when the "Save for Web" function is selected in Adobe Photoshop®. Each custom colour chart was printed along side its corresponding image on eight different printers that spanned a range of image quality capability and included inkjet, photographic, and electrophotographic printers. An experiment was conducted in which observers were asked to scale the eight prints for each of the eight scenes in terms of colour fidelity, where colour fidelity was described as "essentially that the colours look correct and pleasing." A total of 25 observers participated in the experiment at four different locations. The experimental results were examined to determine whether there is some perceptual attribute or attributes associated with the average perception of colour fidelity that can be correlated with physical measurements of the accompanying colour charts. We first examined the patches from the custom colour chart that showed a high degree of correlation between ΔE2000 values (calculated using the highest rated image as a reference) and the observer scale values. Once the patches having ΔE2000 values that were well correlated with the observer results were determined, the L*, C*, and h values of these patches were plotted relative to the scale values. For some of the patches in each scene, these plots show reasonable trending of the colorimetric value against the observer scale values, especially for the C* values, though the correlations were never adequate to suggest that any of these parameters could be used as predictors.

One interesting bit of information determined from the data was that the standard deviation of the observer scale values was higher across the individual printers than it was over the individual scenes, suggesting more variability due to individual printer than to individual scene, which is a useful characteristic of a scene set, when trying to distinguish printer performance.

Several issues arose when examining the colour fidelity scaling experimental results, foremost of which was that, for a couple of patches for certain prints, the best and worst rate prints had nearly identical colorimetry, indicating that colour fidelity quality is driven by more than just colour change in isolation. To attempt to examine this effect further, patches that form a "tone scale," of sorts, were evaluated where possible. The results of this evaluation suggested that there were definite differences in the tone scales of the top-rated prints relative to those of the lowest-rated prints. Typically the highest-rated prints had darker dark colors and, sometimes, darker highlights as well (possibly maintaining some degree of color when the lower-rated prints had fallen off to paper-white). And the "tone scales" tended to be smoother for the higher-rated prints relative to the lower-rated prints. In follow-on experimentation, actual tone scales should be included along with the custom colour charts for each manipulated image.

It was suggested that the printers and scenes included in the initial experiment were not an adequate representation of printer image quality; that the prints from eight printers that were used might not have constituted enough images to evaluate colour fidelity quality. To remedy this situation, a follow-on experiment has been undertaken in which the number of samples of each scene is increased by making deliberate changes to the colour in the digital source files using image-manipulation software, such as Adobe Photoshop®. The first step of the experiment, which consumed an incredible amount of time, was to select some pictorial targets. To reduce the number of iterations needed, we wanted images that could eventually be used in the standard, provided that they functioned well in the experiment. We felt that, to be effective, we would need to have images that contained colours for which people have a relatively uniform idea of "correct" colour rendition. Studies indicate that images containing people would be reasonable selections, which comes as no surprise since human skin tones represent significant memory colours. However, finding images containing an abundance of skin tones that have a known pedigree, that have not be used extensively already, that would challenge the reproduction capabilities of the printers and that we have the legal right to use has been a significant challenge. To date, we have selected four images for use, Figure 1, and are looking into others. (If you know of an image that could be included, please let us know.)

We felt that our set of four images was sufficient to move ahead with the experiment. The next step was to select a set of image manipulations to be applied to the images and to custom colour charts extracted from each of the original images. The charts were again composed of 64 important colours in the images as determined by an Octree colour quantization schemes used within the GIF file format for the Paintgirl and Picnic scenes, a Median Cut colour quantization scheme for the Cashew Nuts scene, and the Perceptual palette contained in the "Colour Table" feature

included in the "Save for Web" function of Adobe Photoshop® for the Bride and Groom scene. The manipulations include incorrectly performed conversion from a wide gamut color space to sRGB, and global changes in saturation, contrast, lightness, and hue. The manipulated images and accompanying custom colour charts will be printed on a single digital silver halide printer; one similar to that used to print the image rulers for ISO 20462. Rulers using our images will also be constructed. These rulers will be used as directed in ISO 20462 to produce scale values for colour fidelity in terms of JNDs.

Next Steps

In the coming year, attention will continue to be focused on the colour fidelity subattribute. Further experimentation is planned to investigate possible relationships between perceived colour fidelity and physical attributes in images. Evaluation is also planned for the colour quantization experimentation examining contouring in the serpentine targets. Having worked through the issues associated with test targets for the colour quantization subattribute, the team will continue the process of making prints of the test targets and evaluating the results and will then consider what can be applied to the colour scale subattribute.

Conclusion

A method of quantifying the perceived image quality of the colour rendition of a printer has been defined as comprising three basic components: colour fidelity, colour scale, and colour quantization. These subattributes have been defined, related to classic colour components, and descriptions of how these attributes might be perceived in images have been developed. Development of a measurement methodology for the colour quantization subattribute has been initiated. Potential test targets have been designed and generated. Prints have been made of these targets on photographic, electrophotographic, and inkjet printers, and initial attempts of measuring contours have been performed. A colour fidelity strategy has been proposed and, again, prints have been made on photographic, electrophotographic, and inkjet printers. Further work is underway and will be reported in a formal presentation, as it is available. Also planned is a review of future work, including planning for fundamental research that will be needed to fully answer the question regarding measuring the ability of a printer to render colour.

References

1. E. K. Zeise and N. W. Burningham, Standardization of Perceptually Based Image Quality for Printing Systems (ISO/IEC JTC1 SC28 and INCITS W1.1), *Proc. NIP18,* pg. 699–702 (2002).
2. N. W. Burningham and E. N. Dalal, Status of the Development of International Standards of Image Quality, *Proc. PICS*, pg. 121–123 (1998).
3. E. N. Dalal, D. R. Rasmussen, F. Nakaya, P. A. Crean, and M. Sato, Evaluating the Overall Image Quality of Hardcopy Output, *Proc. PICS*, pg. 169–173 (1998).
4. S. P. Farnand, K. Topfer, A. McCarthy, W. Kress, O. Martinez, H. Shin, E. Zeise, and D. Gusev, Update on the INCITS W1.1 Standard for Evaluating the Colour Rendition of Printing Systems, Proc. IQSP pg. 157-162 (2005).