A method for assessing the color stability requirements for a printing device

Mary Nielsen, Eyal Shelef, Norman Burningham, Shlomo Harush, Eyal Duzy; Hewlett-Packard Corporation (Israel and USA)
(eyal.shelef@hp.com, Mary.nielsen@hp.com, Shlomo.harush@hp.com, eyal.duzy@hp.com)

Abstract

"Consistent color" is a core requirement for any printing device. It is a very important and sensitive requirement that cannot be ignored, if we wish to meet customer needs. The written color stability specifications used in the industry (ISO, SWOP, etc..) are based on ink parameters such as density and dot area of individual ink patches. These measurements are empirically derived process control parameters and for calibrating printing machines, but it is not clear how these parameters relate to image color acceptability. In addition, we know of no quantitative data that would identify the differences between different print use models.

This paper describes a method to asses the color variation values that are required by any printing device to meet market requirements for a particular use model. The method will be described through a description of a test done to asses the acceptability of color stability variations in commercial offset printing.

Test description

Within the printing industry there are two use models that have different requirements for stability. The first refers to a customer who has printed some material and later wants to print more of the same basic document. Can the press calibrate with sufficient accuracy to get to an acceptable match between the two runs? A second use model refers to the differences in color caused by tone scale variations during a long run. Maintaining a successfully low variability within a run is a process stability issue. We knew of no quantitative data that would identify the differences between these two use models.

This test is based on the perception of image quality of color variations as evaluated by 66 observers (in this test the observers were experienced professionals in the field of commercial offset printing). It defines an objective metric for color variations that relates well with the visual quality evaluation of the observers.

In this test acceptability statistics for two different models of use were obtained. Five test images incorporating a range of image content were created with controlled color variations.

The images used in the experiment were selected to represent a range of commonly imaged documents (the images are detailed in Appendix A).

Twenty-one variations of the color tone scale were created of each image in addition to a nominal print at aim calibration set up. These variations were selected to sample the range of the functioning of a normal hp-Indigo 3000 press. In addition, several different standard sets of color patches were printed at the same imaging conditions of the visually evaluated samples. Data from selected subsets of color patch measurements were found to correlate with the acceptability for the tone scale variations as a function of image content. [1],[2]

Experimental Plan

a. Determine the acceptability of color variations for images that would be viewed side-by-side.

b. Determine the acceptability of color variations for images that would represent differences in the length of a print run.

c. Define an objective metric that correlates well with the perception of image quality for the variations in parts a and b above.

Evaluation Methodology

The Method of Constant Stimuli should be chosen for the analysis. In the procedure each test variation was compared one at a time to the nominal image created with aim tone scale. The observers were asked to make two decisions for each comparison pair:

1. Is the color difference acceptable if the two images were to be placed side-by-side in a publication?

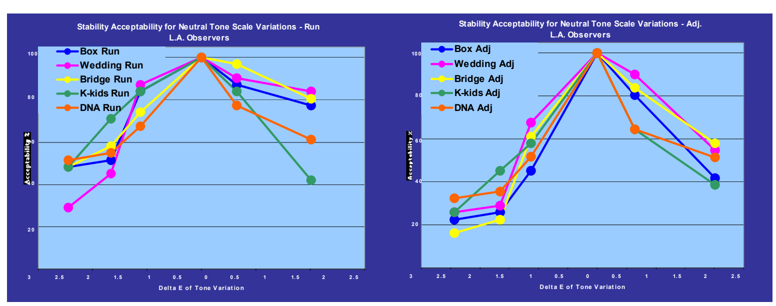
2. Is the color difference acceptable if it represented the maximum variation seen within the length of a print run?

Thus for each of the five image contents there were 42 data points, or a total of 110 judgments per observer. This number of judgments was selected so that the observer could complete the task in one hour or less.

The Results

Acceptability of color differences as a function of use model

Selected examples of the percent acceptability of the color differences in the test images are presented in the following graphs for both use models:



A reasonable goal for a high quality system would be an acceptability of 90%. If we use this limit we can immediately see in the above example that the curves fall below that goal very quickly as color error increases. The judgments for the "adjacent" use model are generally more critical that those for the "within run" model as would be expected.

An ANOVA analysis was done to detect the main factors that influence the acceptability of the color variations. In this test the first, second and third order interactions were detected [3].

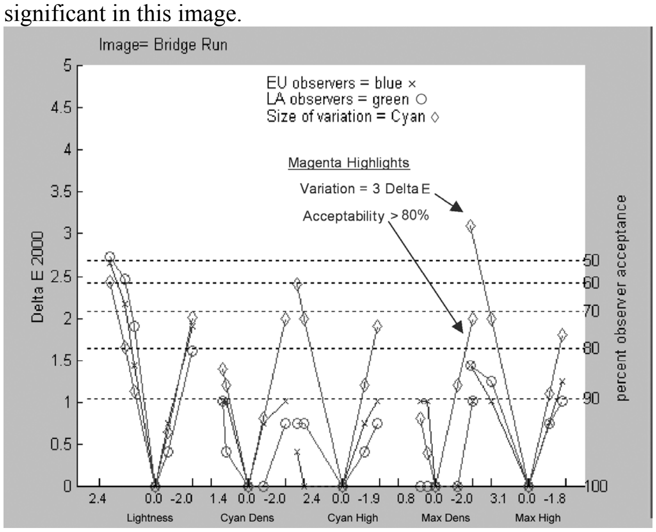
Define an objective metric that correlates well with the perception of image quality

One objective of this experiment is to create a metric that predicts the image color difference acceptability using measurements of color patches. This metric uses the mean ΔE's of patches printed with each image variation.

Actual Size of Intended change for each image variation and acceptability level

There are two noticeable trends in the data, the relative sensitivity to the parameter changed and the sensitivity to the direction of the change within each parameter. A complicating factor is that the step sizes of the changes introduced in the images were not uniform. The three kinds of parameters changed in the images are lightness, ink density and "highlight" tone curve changes.

An imbalance is a variation where a change is much more unacceptable in one direction than in the other. The following Figure shows the size of the change along with the observer acceptability of the change. The lightness (first cluster of points) changes are an example where the acceptability is proportional to the size of the actual change. The actual change in the lightest images was larger than the magnitude of the darkest change and the observer acceptability data tracks this. Other change size imbalances do not track with the observer acceptability. In this Figure the change in the magenta highlights positive direction is much larger than the change in the negative direction as seen by the cyan plot of the fifth cluster of points. In this case, the blue and green observer acceptances show that this large change was not significant in this image.



Results of a metric modeling the observer data

Several different measurement charts of color patches were printed along with the images to provide an objective measurement of the print variations (the ECI chart and the IT8-7/3 chart) and an image specific set of the 64 most common colors from each image. Image dependent optimized set of patches

Each test image has different color content resulting in different observer acceptability patterns. This results in a unique set of patches for each image for the optimum correlation to the ΔE color change. A set of "best" patches to predict the acceptability for each image was developed. A correlation factor based on the Matlab *corrcoef* function was used to compare the patch set ΔE to the acceptability data. This function provides a correlation that is independent of the scale of the data sets since the ΔE and acceptability data are in completely different kinds of units. Several combinations of patches were developed as patches were added, deleted and duplicated to maximize the correlation factor for each image.

The patch sets were created by two separate methods. First, an intuitive method relied on the image content and the exit interview notes of the observer sessions to develop logical patch sets. Second, a "blind" analysis correlated the ΔE of each patch to the image acceptability and found the best X patches, where X was allowed to vary. Patch sets were combined, weighted and mixed together until a good correlation was obtained for each image.

A Practical Metric for any Image

The next step is to determine the best correlation factor for the averaged results of the images. Although one set of patches was the target, the gray scale images proved to respond differently than the color images and had to be represented by a different set of patches. The two final patch sets for the metrics were based, first, on the average of the acceptability of the 4 color images and, second, on the grayscale image patches.

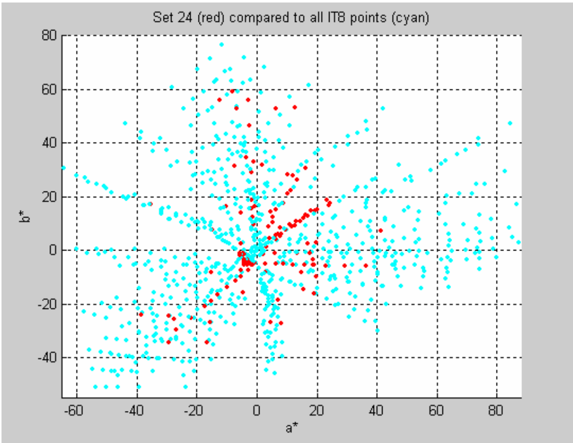
Resulting correlation and ΔE of the patches – compared to traditional method(s)

The following table shows the mean ΔE 2000 required for the optimized set of data patches for 80% and 90% acceptability of each image. For general purpose use, separate ΔE's are provided for color images and grayscale images for each use case, adjacent side by side viewing, and within a run. The ΔE 2000 that would be required of the entire set of IT8 patches is also shown for comparison with traditional methods. Traditional methods using the entire set of patches not only don't correlate as well to the observer acceptance, but results in a tighter spec, as can be seen in the following table:

Acceptability	80%		90%	
Image	ΔE IT8	ΔE Set	ΔE IT8	ΔE Set
Run Color	2.3	2.6	1.5	1.7
Run Grayscale	1.3	2.0	0.8	1.3
Adjacent Color	1.3	2.0	0.8	1.3
Adjacent Graysca	1.0	1.5	0.6	1.0

The table above shows that the ΔE spec is slightly looser (higher) when the optimized set of patches is used. Unfortunately, this still confirms that a ΔE around 2 is still required to produce acceptable color. It also shows the need for strict control of the 3 and 4 color gray balance.

The color coordinates of the chosen set of patches in reference to the whole IT8 patches are presented. In the following graph the Flattened a* vs. b* scatter plot shows the distribution of the chosen set of patches compared with all data points in the IT8 chart. [4]



Summary of the data correlation results

This study shows the significance of light tints and a balanced neutral balance to the acceptability of images. The results show less tolerance for light tint shifts than the density changes.

The difference in image quality color requirements between Europe and U.S.A color professionals.

In order to explore differences in the responses between European (EU) and U.S.A. (LA) observers we conducted a five – factor ANOVA based on the differences in acceptance values between the two data samples. The results are summarized below with a brief discussion and comments.

Exceptions to data correlation

This test can not only provide an objective metric but can detect some exceptional image variations that the general metric does not model but are significant in the acceptability data. These changes consist mainly of small skin tone changes and color changes in office documents that have large areas of the same color.

Conclusions

A 2.6 ΔE 2000 mean of the patches defined by set 24 is required for stability within a run for an 80% acceptability of color images. This becomes 1.7 ΔE 2000 for 90% acceptability. Grayscale images are much more sensitive and require a ΔE of 2.0 for 80% and 1.3 for 90% acceptability for set 10 patches consisting of the three and four color grays of the IT8 chart. For side-by-side viewing, these specs tighten to 2.0 for color and 1.5 for grayscale images at 80% and 1.3 and 1.0 (respectively) for 90% acceptability.

Attempts to model the direction and magnitude of hue shifts of specific color regions such as skin tones did not create a better metric than ΔE 2000. It was found that images became unacceptable with skin tone hues shifting in either the magenta or green directions.

Color shifts in a critical part of an image will cause the image to be rejected. This is known as the "tent pole phenomena". Attempts to increase the sensitivity of the metric to account for these phenomena created errors in images with different content.

This example shows the information that can be derived from this kind of tests. Not only an objective, quantitative metric that correlates with human perception acceptability is found, but exceptions and main problem can be detected.

References

[1] M.R Pointer, "Measuring Colour Reproduction". The Journal of Photographic Science, Vol 34 (1986)

[2]N. Moroney, M.D. Fairchild, R.W.G. Hunt, C. Li, "The CIECAM02 Color Appearance Model", IS&T 10th Color Imaging Conference (2002)

[3] R.L. Gorsuch, "Factor Analysis", Hillsdale, Nej Jersey (1983)

[4] M. R. Luo, G. Cui, B. Rigg, "The Development of the CIE2000 Colour-Difference Formula CIEDE2000", Colour & Imaging institute, University of Derby, UK (2000)

Author Biography

Mary Nielsen has worked at Hewlett-Packard for 26 years and is a Color Imaging Architect for the Indigo Digital Press division. She received her BS in Electrical Engineering from Utah State University (1980) and her MS in Computer Science from California State University, Chico (1992). Her current work is focused on the Indigo Press Digital Front End and Raster Image Processor color management and imaging features.

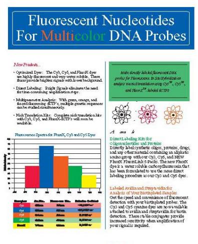
Eyal Shelef received his BSc in physics from the Tel Aviv University (2000). Since then he has worked in the Research and Development department of the Indigo division of Hewlett-Packard Company in Israel. His work focuses on color output stabilization.

Shlomo Harush received his BA in Mathematics and MSc in physics from the Hebrew university in Jerusalem (1988) . He received his MBA from Tel Aviv university (1996). He is working in the printing industry (Indigo Graphic System, Indigo and now hp). His work has focused on color and algorithms

Eyal Duzy is a Marketing Product Manger at HP Indigo Division. Eyal holds a B.Sc. degree in Biophysical Chemistry (1990) and a combined MBA / M.Sc. degree majoring in Marketing, Entrepreneurship and Applied Chemistry (1994). Prior to joining HP Indigo, Eyal held various product marketing positions in the electronics and semiconductor industry, where he was responsible for product marketing.

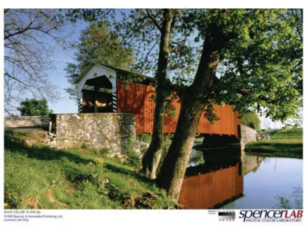
Appendix A – images used in the test

Image has well balanced Caucasian skin tones as well as important details in the highlights and shadows.



Fluorescent Nucleotides
For Multicolor DNA Probes

Represents office or marketing collateral type documents.



Contains natural memory colors of sky, foliage and grass.

Gray image.

Typical "proof" type document including a collection of objects and a variety of skin tones.