Assessment of the Current Light-Fade End-Point Metrics Used in the Determination of Print Life: Part II

David J. Oldfield and John Paul Twist
Eastman Kodak Company
Rochester, New York

Abstract

In Part I of this study, presented at IS&T's 19th Non-Impact Printing Conference, we tested the validity of current light-fade end-points for predicting the print-life of consumer digital output prints. It was demonstrated that current light-fade end-points generally under-predict the print life of home consumer prints, as perceived by typical end-users. The objective of P~~p~~art II of this study is to go into more detail about the first-phase conclusions, as well as to assess~~ing~~ possible alternative metrics. This study investigates how well end-points derived from colorimetric parameters correlate to end-user perception. End-points are evaluated based on both the 1976 CIELab and CIE-2000 color difference formulae. Color regions of interest, such as skin tones and neutrals, are looked at independently to improve the correlation of colorimetry-based metrics to the psychophysical results.

Introduction

The image permanence of photographic prints containing the cherished memories of a family are expected to last a lifetime. Companies whose products provide these photographic prints desire a means of predicting the print life of a typical print population. Historically, this has been accomplished by performing accelerated print-life testing. The focus of this study will be on the exposure to radiant energy as a stress; this is commonly called light-fade testing. Currently, print life is determined by looking at Δ-status A densities from an initial density of 1.0. Illustrative end-points based on percent density loss from an initial 1.0 density have been established.[1,2] In Part I of this study, the current end-points were assessed and found to be inadequate. The current end-points did not correlate well with psychophysical results and were generally under predicting print lives. This paper will provide additional insight~~s~~ into the reasons why densitometric end-points fail to properly predict print life of consumer prints. Also, the psychophysical data will be compared with various colorimetric parameters to identify potentially useful relationships. Traditionally, colorimetric color-difference measurements have been aimed at correlating small color changes in color patches, not complex scenes. Also, they have been designed to look for "just noticeable differences" as opposed to "just unacceptable" changes. With this in mind, every effort to enhance signal to noise has been pursued. This was accomplished by isolating color regions of global interest (skin tones, neutrals, memory colors, saturated colors, hair color) to see if they correlate better to psychophysical data.

Methods

Densitometric-Based Metrics

In Part I of this study, an interval scale was generated from psychophysical data where a 1.0 value was designated as a "just noticeable" difference, and a 3.0 value was designated a "just unacceptable" difference between the original scene and a degraded scene. The interval scale was derived from a psychophysical study using a category-sort methodology. The categories were:

(1) Completely acceptable
(2) Moderately acceptable
(3) Marginally acceptable
(4) Marginally unacceptable
(5) Moderately unacceptable
(6) Completely unacceptable

The interval scale was generated from the judgments of 40 subjects for 16 scenes;~~,~~ each scene contained 26 levels of degradation for a total of 416 stimuli for each system. A total of 8 separate studies were conducted to test the robustness of current end-point metrics. This included dye-based and pigment-based inkjet systems, thermal dye transfer, electrophotographic, and conventional silver-halide systems. The inkjet systems included both porous and swellable media types. Table 1 shows the maximum Torgerson value reached by each system.

Several systems never reached a Torgerson value of 3.0. This is because the original experimental design was based on percent density loss. All of the systems had, at minimum, 50% density loss in at least one color record. This would allow all the existing end-points to be fully assessed.

Table 1. Maximum Torgerson value by system.

System #	System	Maximum Torgerson Value
1	Dye-Based Inkjet on Porous Media (1)	2.06
2	Dye-Based Inkjet on Porous Media (2)	>>3.00
3	Dye-Based Inkjet on Swellable Media	2.60
4	Pigment-Based Inkjet on Porous Media	>>3.00
5	Retail Thermal Dye Transfer System	2.16
6	Home Thermal Dye Transfer System	>>3.00
7	Electrophotographic System	>>3.00
8	Conventional Silver-Halide System	1.31

One method of evaluating the validity of the current end-points is a simple comparison of the predicted print life to the print life suggested by the VOC (voice of customer) study. A print life predicted by changes in densitometry assumes that changes in density, relative to one initial density, are representative of all perceived changes over the entire tone scale. To test this assumption, the changes in densitometry from the original tone scale to the tone scale predicted by the current end-point criteria were calculated over the entire tone scale. Also, the changes were calculated in densitometry between the original (time = 0) curves and the tone curve that exists at the time the VOC data becomes just unacceptable. These two sets of densitometric changes were compared to each other. If the current end-points were adequate, the VOC and the predicted curves would fall on top of each other. Figure 1 shows the method used to compare the tone scales at the predicted print life versus the tone scale at the VOC-determined print life. Only the curve of the color record that failed first, based on the current end-points, will be used in this discussion. Figures 2–5 compare the densitometric changes of the predicted (by current end-points) to actual (from VOC data) over selected initial densities from the original images. Ideally, the predicted end-points will have a densitometric change of 30% from an initial density of 1.0. To keep things simple, the plots in Figs. 2–5 compare the densitometric changes for initial densities of 0.6, 1.0, 1.5, and 2.0.

Densitometric Results

Although the differences between the densitometric end-points and the just unacceptable threshold determined psychophysically are quite significant, it should be noted that some of the systems never reached an end-point on the VOC scale. For example, although density losses of over 70% were measured for the silver-halide system, the mean Torgerson value of 1.31 did not come close to the psychophysical "end-point" of 3.0. This infers that larger values of density loss would likely need to occur before a rating of 3 was achieved. It should be noted that the silver-halide system demonstrated good neutral fade, which is a potential explanation for the high acceptability ratings for the faded prints seen in the study.

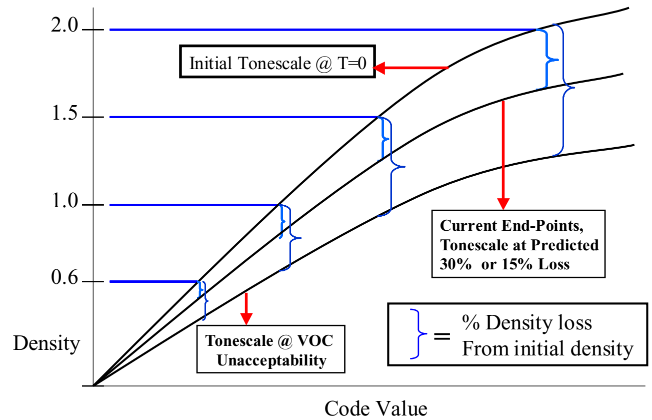

Figure 1. Plots in Figs. 2–5 are based on the Δ-density between (1) initial-density and the density from current end-points, and (2) initial density and density at judged end-points (i.e., Torgerson value = 3.0) for 4 initial densities (0.6, 1.0, 1.5, and 2.0)

The plots in Figs. 2–5 (Note: For brevity, only four systems are being shown) demonstrate how much more density loss can occur before the unacceptability threshold is surpassed. The black circles represent the Δ-densities at the point where the psychophysical data reached a Torgerson value of 3.0 (or the maximum Torgerson value if 3.0 was not reached). The gray diamonds represent the Δ-densities at the current point ANSI IT9.9 illustrative end-points reach an end-point. This shows that the current end-point criteria are underestimating print life. The different systems show varying degrees of density loss at the unacceptability threshold. The pigment-based inkjet system lost almost all of its yellow dye before a 3.0 Torgerson value was reached. The two dye-based inkjet systems on porous media required about a 50% loss in the magenta record before they approached or reached a Torgerson value of 3.0. The home thermal printer failed because of a 15% color balance difference in the neutral tone scale.

The variability among technologies also suggests that densitometric end-points are not robust enough to predict print lives across a range of technologies. Further, simply setting less stringent densitometric end-points, perhaps 50% instead of the current 30%, may improve one system's ability to correlate better to psychophysical results, but this does not improve convergence across all technologies.

To improve the relationship between psychophysical data and objective parameters, colorimetry-based metrics were investigated.

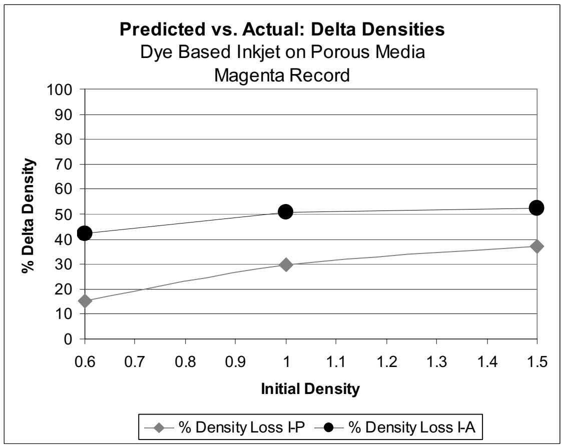

Figure 2. Magenta record of a dye-based inkjet system on porous media.

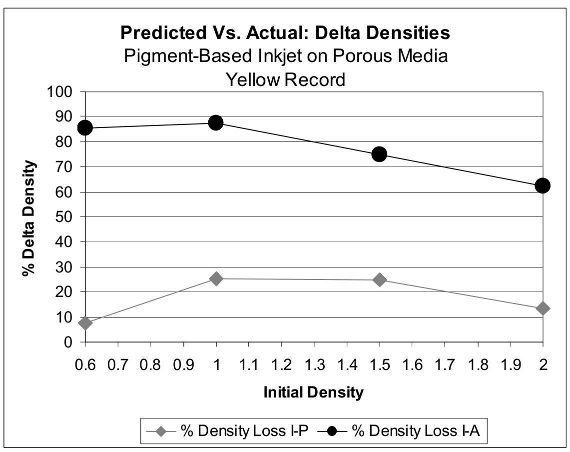

Figure 3. Yellow record of a pigment-based inkjet system on porous media.

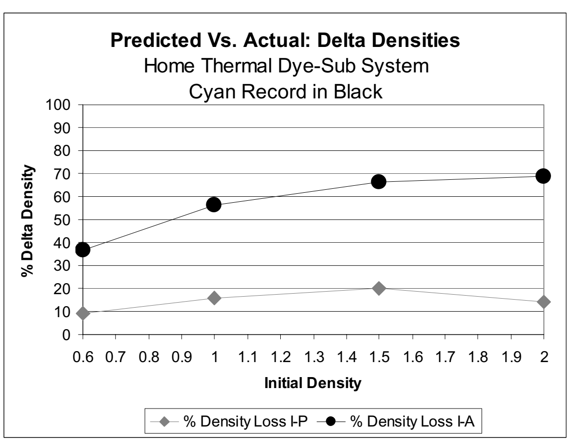

Figure 4. Cyan record of a home thermal dye transfer system.

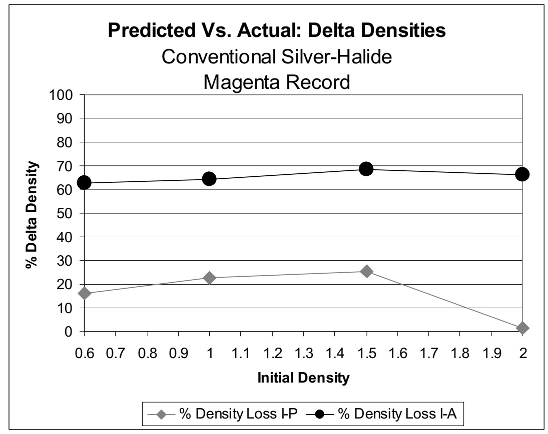

Figure 5. Magenta record of a conventional silver-halide system.

Colorimetry-Based Metrics

In an attempt to improve the relationship between psychophysical data and objective parameters, colorimetry-based metrics were investigated. Others have previously tried to correlate colorimetry-based metrics to the current densitometric end-point.[3,4] However, as shown in Part I, the current densitometric end-points are inadequate. In this paper, colorimetry-based metrics will be compared with the results of the psychophysical study data. It should be emphasized that colorimetry-based color difference formulae are based on color differences of flat field patches not complex images. Also, these metrics tend to focus on just noticeable differences as opposed to how much color difference is tolerable

Both the 1976 CIELab color difference formula and the CIEDE2000 formula were investigated. The 1976 CIE color difference ΔE_{ab} is shown in Eq. (1) and the ΔE_{00} is shown in Eq. (2).[5,6] Note that ΔE, alone, is simply the magnitude of color change. The parameters L*, a*, b*, C', and H' give perceptual direction to the changes in color. CIEL*a*b* provides metrics based on a "perceptually uniform" color space. L* refers to the lightness scale, a* refers to the red-green axis, and b* to the yellow-blue axis. CIEDE2000 was developed to improve the correlation of measured color differences to perceived color differences. This was accomplished by (1) adding weighting factors for lightness, chroma, and hue, (2) adding a correlation factor based on the arithmetic means of both chroma and hue, and (3) applying a scaling factor to CIELab a* to improve predictability of neutrals.

$$\Delta E_{ab} = ((\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2)^{1/2} \quad (1)$$

$$\Delta E_{2000} = \{(\Delta L^*/K_L S_L)^2 + (\Delta C'/K_C S_C)^2 + (\Delta H'/K_H S_H)^2 + R_T \,.\, [(\Delta C'/K_C S_C) \,.\, (\Delta H'/K_H S_H)]\}^{1/2} \quad (2)$$

Colorimetry was calculated for both colorant tone scales and the "real color target" (RCT). The RCT contains 160 color patches drawn from specific color regions within the scenes used in the study. The colors included in the RCT stress memory colors (flesh tones, sky, foliage, fruit, neutrals, etc.). The purpose of this target is to enhance any signal from the colorimetric parameters. See Figs. 6—8 to see the respective color regions included on the RCT. Plots of the saturated colors, neutrals, red, and brown colors are not included.

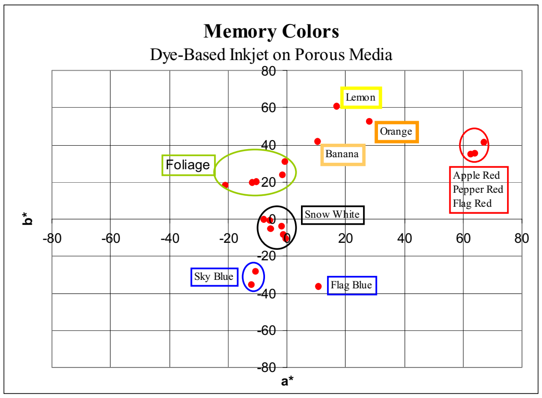

Figure 6. Memory colors.

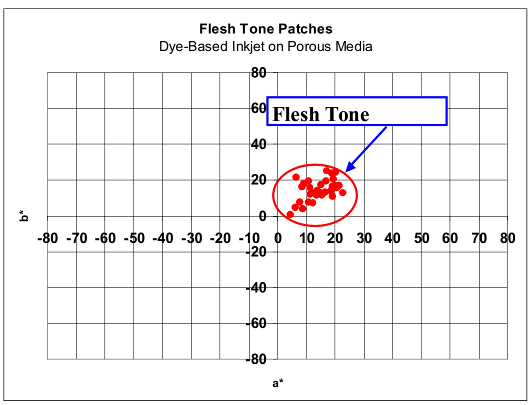

Figure 7. Flesh tones.

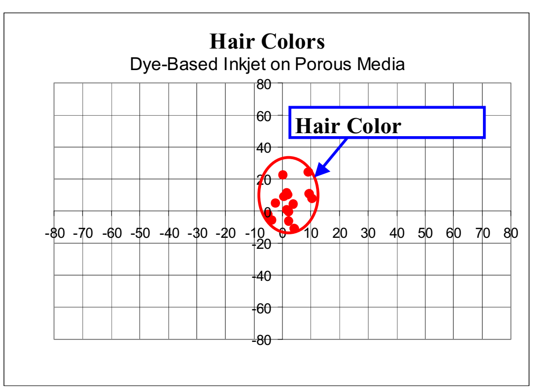

Figure 8. Hair colors.

Colorimetric Results

Colorimetry data was tabulated for both the CIELab 1976 metrics and the CIEDE2000 metrics. The data is tabulated including both the color tone-scale data and RCT memory color data. The tone-scale data consisted of seven color sub-categories. The data shown represent the parameter values found at a Torgerson value of 3.0 (or the maximum value if 3.0 is not achieved) relative to the initial tone-scale value of 1.0: (A value of 1.0 was chosen to adhere to current conventions, however, all densities will be used to define statistical end-points)

(1) Red
(2) Green
(3) Blue
(4) Cyan
(5) Magenta
(6) Yellow
(7) Neutral (t)
(8) Grand Average

The RCT was also sub-categorized into specific color regions:

(1) Skin Tones
(2) Hair Color
(3) Saturated Colors
(4) Non-Global Memory Ceolors
(5) Neutrals (p) – p for neutrals seen in the pictures
(6) Grand Average of all RCT Colors

Skin tones, hair color, and neutrals were viewed as global memory colors meaning they are very common in photospace and warranted their own data grouping. An argument can be made that sky blue, foliage, and snow should also be global but, for this study, they were grouped together in the non-global memory colors sub-category.

Table 2 shows the mean 1976 ΔE_{ab} value for each of the sub-categories shown above for each of the eight systems in the study. The Δ-values are determined by subtracting the time zero CIELab metric value from the t @ Torgerson = 3.0 (or maximum value if 3.0 is not achieved). Note that some systems, especially the home thermal dye transfer system, had color records that did not achieve a tone-scale density of 1.0. In these cases, the CIELab metrics were determined from the maximum density achieved. Table 3 shows the same data using ΔE_{00} (ΔE 2000). —The CIE ΔE_{00} metrics have performed significantly better in psychophysical testing versus the CIELab 1976 metrics, generally getting better agreement with the perceived color changes of color patch judgments.[6] The "K" weighting factors were set at unity. Optimizing the weighting factors may improve results and provide improved statistical fitting of the data. Although all of the colorimetric parameters, including ΔL^*, Δa^*, Δb^*, ΔC^*, Δh^*, and all the equivalent CIE-2000, have been determined;, only ΔE will be discussed in the interest of brevity. A rigorous statistical analysis of all factors will need to be completed to identify any possible relationships that exist. When all of the color categories are rank ordered those categories most sensitive to change will have smaller values. Comparing the mean ΔE_{ab} for each of the ~~8~~eight systems versus~~.~~ the various color parameters will give some insight to the relationship between which colors are changing in the pictorial scenes and the sensitivity to those changes. The most conspicuous observation in the RCT data is that the subjects appear to be most sensitive to changes in neutral color appearance. This is because at the judged end~~-~~-point, the mean ΔE_{ab} value for the neutrals (p) is very small, relative to the other sub-categories. The neutral will be affected by changes in any color record and, as such, should reflect color biases occurring within a system. Therefore, the ΔE_{ab} for the neutrals is a good indicator of the sensitivity of the judges to image degradation. The mean ΔE_{ab} for the pictorial neutrals ranged from about 6–~~-~~18 for the 8 systems in the study, with an overall mean of about 10 ΔE_{ab}. The maximum ΔE_{ab} occurred in a system that had a large blue shift, suggesting the judges are perhaps less critical of neutrals shifting toward blue. Although skin tones may demonstrate higher sensitivities in some systems, they fail to do so in all systems. The RCT does suggest that changes in the neutrals, skin tones, and saturated colors could be significant independent of system. The cyan value is relatively small compared with the other tone-scale data. This is likely the~~a~~ result of the systems failing because of strong signals from the magenta and yellow records. Averaging ΔE values for the tone-scale data can be misleading. Relationships can be observed from color changes within a single system but quickly dissolve as we apply those relationships across systems. Inspection of the ΔE_{ab} for the weakest colorant in each system at the point of judged failure provides an interesting contrast to the value of the ΔE_{ab} for the pictorial neutrals. The ΔE_{ab} for the first colors to fail, using current end-points, range in values from 27 to 83~~.~~ (discounting system 7, which had very low pure cyan tone scale). These values are much larger than those of the pictorial neutrals. Also, 5 of the systems failed first in the magenta record;~~,~~ however, the ΔEab values for the magenta record range from 27 to 62. Finally, system 4's value for ΔEab in the yellow record is about 83. This suggests that end-points based solely on the first color to fail may not be the solution.

When comparing the data between the ΔE_{ab} and ΔE_{00}, the first observation is the smaller values in the ΔE_{00} because, in part, it is more closely related to lightness, hue, and chroma. The RCT neutrals (p), skin tones, and saturated colors maintain higher than normal sensitivity. There is still significant variability between systems. The overall system mean ΔE_{00} ranged from about 10 to 20. A range of 6–13 is even tighter when the mean ΔE_{00} for the neutrals (p) is inspected. If we again at the value of the ΔE_{00} for the first color record to fail based on current end-point criteria, the values range from 3 to almost 40. This is an improvement over ΔE_{ab} but still too variable to establish robust end-points over all the systems. The range of data was 10–38 ΔE00 for the 5 systems that failed in the magenta record. The change in the other color sub-categories is contributing to the variability seen in the magenta records.

Table 2.

Mean ΔE_{ab} 1976									
Attribute	**System 1**	**System 2**	**System 3**	**System 4**	**System 5**	**System 6**	**System 7**	**System 8**	**System Mean**
Neutrals (p)	10.94	15.55	6.11	18.02	7.33	9.31	7.23	8.67	**10.39**
Cyan	18.65	11.81	3.86	9.49	**23.17**	**4.39**	11.29	1.55	**10.53**
Saturated	22.63	26.44	14.99	36.81	15.73	17.85	14.22	21.12	**21.22**
Skin T~~t~~ones	18.50	26.49	12.09	45.52	13.85	9.02	16.71	30.26	**21.55**
Green	33.12	20.49	7.98	58.62	21.11	7.20	16.14	11.45	**22.01**
Memory	23.31	25.16	18.99	42.41	15.64	21.66	12.91	19.23	**22.41**
Neutral (t)	23.08	30.60	9.11	43.22	18.92	30.38	10.98	14.10	**22.55**
Browns/Reds	20.56	25.45	22.20	51.26	18.16	16.61	13.10	19.86	**23.40**
Blue	29.16	36.11	30.32	11.78	24.98	11.24	26.82	31.67	**25.26**
Yellow	25.82	18.74	11.73	**82.63**	13.54	6.96	38.70	6.05	**25.52**
Hair Color	24.90	27.88	23.04	60.06	20.16	35.88	12.74	12.93	**27.20**
Magenta	**31.46**	**52.30**	**26.66**	10.24	17.67	3.67	**47.16**	**61.62**	**31.35**
Red	28.17	43.64	14.94	53.69	27.77	28.13	21.24	52.19	**33.72**
Grand Avge.	23.87	27.74	15.54	40.29	18.31	15.56	19.17	22.36	**22.86**

Table 3.

Mean ΔE 2000									
Attribute	**System 1**	**System 2**	**System 3**	**System 4**	**System 5**	**System 6**	**System 7**	**System 8**	**System Mean**
Cyan	8.43	5.45	1.98	3.70	**10.01**	**2.58**	5.13	1.04	**4.79**
Neutrals (p)	8.56	13.08	5.55	9.06	7.41	7.16	6.12	8.64	**8.20**
Yellow	7.28	5.35	3.16	**34.66**	3.26	2.52	12.62	1.46	**8.79**
Green	12.67	8.94	3.16	31.88	9.51	3.84	8.10	4.97	**10.38**
Saturated	15.21	17.98	9.60	17.59	10.41	11.55	8.92	14.38	**13.20**
Memory	16.04	18.46	12.43	24.00	12.09	15.79	8.91	13.34	**15.13**
Blue	16.83	23.17	16.83	2.10	17.33	9.70	15.82	23.97	**15.72**
Magenta	**14.94**	**25.56**	**10.88**	3.50	8.58	2.19	**22.08**	**38.33**	**15.76**
Skin Ttones	12.54	23.33	10.26	26.53	10.58	6.57	12.31	26.42	**16.07**
Browns/Reds	15.33	21.39	15.66	26.42	15.98	13.49	10.80	17.88	**17.12**
Red	12.88	27.34	8.45	27.35	15.59	13.50	13.17	35.09	**19.17**
Neutral (t)	24.19	28.62	7.68	23.23	21.57	28.91	10.54	14.39	**19.89**
Hair Color	20.18	22.96	19.10	29.16	17.08	26.32	13.36	13.08	**20.15**
Grand Avge.	14.24	18.59	9.59	19.94	12.26	11.08	11.38	16.38	**14.18**

Conclusion

The current ANSI IT9.9 illustrative end-points, together with the WIR Version 3.0 are inadequate. This was illustrated by the plots of tone-scale Δ-densities of the current end-points versus the VOC data end-points. The differences between these VOC and current end-points were as high as 80% Δ-density. The current end points also failed to predict accurately across all technologies. In general, current end-points were under-predicting print lives.

Colorimetery metrics ΔE_{ab} and ΔE_{00} were calculated for many color sub-categories for all eight output systems. Although end-points could be defined for any given system, no simple solution was found to be robust over all eight systems. Judges appeared to be very sensitive to changes in the neutrals seen in the pictorial images. This was less true for skin tones and saturated colors, however, they were both still important in majority of the systems. The $\Delta E_{ab,00}$ of the color records that first failed were highly variable and large relative to the pictorial neutral data.

If a relationship is to be found between Torgerson value and colorimetric parameters, extensive multivariate statistics will need to be applied to the data. This will be the next phase of this study. Also, the evaluation of color differences in complex scenes may require the use of color appearance modeling and perhaps other novel means such as Mahalanobis distance.[7-9] The assessment of fade in complex scenes is not trivial and state-of-the-art analysis tools for complex scenes is just beginning to find it's way into standards committees.[10]

Acknowledgments

Special thanks to Doug Bugner and Joe LaBarca for their continued support of this project

References

1. ISO TC/sc N4720 (ISO/CD 18909.2), *Imaging Materials – Processed PhotographicColor Films and Paper Prints – Methods for measuring Image Stability*, (2001).
2. H. Wilhelm, *Proc. IS&T 12th Int. Symp. Photofinishing Technology*, p. 34, Springfield VAa. (2002).
3. D. E. Bugner, C. McWilliams, "Correlating Changes in Densitometry and Colorimetry in the Context of the Light-Induced Fade of Inkjet Photographic Prints," *Proc. NIP-18*, p. 753, (2002).
4. R. Hoffmann, E. Baumann, R. Hagen, "Densitometry versus Colorimetry for Permanence Investigations," *Proc.NIP-17*, p. 209, (2001).
5. CIE Publication No. 152, *Colorimetry, 2nd edition*, Commission Internationale de l'Eclairage, Vienna Austria, (1986).
6. M. R. Luo, G. Cui, B. Rigg, " The Development of the CIE 2000 Color Difference Formula," *Color Res. Appl.*, 26, p. 340, (2001).
7. C. Li, M. R. Luo, G. Cui, *Color Differences Evaluation using Color Appearance Models*, Proc. IS&T's CIC-11, p. 127, 2003
8. C. Sano, T. Sang, M. R. Luo, *Colour Differences for Complex Images*, Proc. IS&T's CIC-11, p.121, 2003
9. F. H. Imai, N. Tsumura, Y. Miyake, *Perceptual Color Difference Metric for Complex Images Based on*

Mahalanobis Distance, J. Electronic Imaging, 10, p. 385, 2001
10. CIE Technical Committee 8-02, *Colour Difference Evaluation for Images*.

Biography

David Oldfield is employed as a senior development engineer at Eastman Kodak Company in the's Inkjet Systems and Materials Division. In his 20 years at Kodak, he has worked with color negative, thermal dye transfer, and inkjet technologies. Dave's principal areas of interest are image quality, color quality, and test development. He has a B.S. in Imaging Science from the Rochester Institute of Technology.