

Evaluation of DSC (Digital Still Camera) Scene Analysis Error Metrics - Part 1

Jack Holm
Hewlett-Packard Labs
Palo Alto, CA, USA

Ingeborg Tastl
Sony US Research Labs
San Jose, CA, USA

Steven Hordley
University of East Anglia
Norwich, United Kingdom

Abstract

A fundamental problem in digital photography is the estimation of scene colorimetry from raw DSC image data. Currently, a standard is under development in this area (ISO 17321-2).[1] In the development of this standard, few subjective experiments have been carried out until now relating to the estimation of scene colorimetry from non-colorimetric raw data, where no assumptions are made concerning the spectral radiance correlation statistics of the scene. Furthermore, there is not much information available concerning whether it is appropriate to use some assumption about spectral radiance correlation statistics when the statistics of the actual natural scene are unknown.

This paper presents the first part of a study involving psychophysical tests to answer the following questions that are essential for the specification of a scene analysis color space, and for the specification of methods for transforming raw DSC data into scene colorimetric data.

1. What is the most appropriate error metric to be used for the determination of transformations from raw DSC data to scene colorimetry estimates, when no assumptions are made concerning the scene spectral radiance correlation statistics? The crucial point is to find the error metric that corresponds best with human perception.
2. How does this new error metric compare to existing criteria, and how do the existing criteria compare to each other when used to determine transformations based on specified spectral radiance correlation statistics assumptions?
3. Given that optimal error metrics are used to determine transformations, how do several spectral radiance correlation statistics assumptions compare to each other and to the maximum ignorance case when applied to natural scenes where the actual statistics are unknown?

Several of the transformation methods specified in ISO 17321, and other methods that are extensions of the 17321 methods were applied to raw data from two DSCs with different spectral sensitivity characteristics. These DSCs were used to capture images of a variety of natural scenes, and the resulting images were processed using the different characterization transforms based on different error metrics. Critical visual evaluation of the final images by expert observers was used to eliminate the obviously poor characterizations. Psychophysical experiments were conducted to differentiate the performance of the remaining candidates.

Introduction

An emerging paradigm in digital photography is the separation of the color reproduction process into scene analysis and color rendering. This paradigm offers increased flexibility and utility, because rendering algorithms can be generic to a variety of digital cameras, and even to film scene capture. Furthermore, the conceptual understanding of pictorial color image processing is also improved by explicit identification of the steps required to obtain a good photograph. From a practical point of view, an image captured with a DSC is first transformed into a scene-referred representation, which to a large extent is independent of the capturing device/system. The scene-referred representation can be seen as an interface between scene analysis and color rendering.

Over the last decade, CIE based colorimetry has been successfully employed to introduce device independent components into digital color imaging workflows. Some of these components can be successfully extended to digital photography. However, as digital color imaging expands into a broad range of applications, a number of areas requiring further investigation become apparent. With pictorial imaging of natural scenes, specific areas require further study. Those areas include the modeling of the perceived appearance of natural scenes, and maintaining preferred reproduction appearance on a variety of output media. The subject of this paper is a comparison of different approaches to perform a colorimetric scene analysis under the constraint of capturing devices that do not fulfill the Luther condition.

Background

Reasons Why Typical Digital Cameras Are Not Colorimetric

If one takes an engineering approach in comparing the human visual system (HVS) to a DSC, it becomes apparent that there is a significant design constraint difference. The HVS uses a lens that is not corrected for chromatic aberration. When designing a three-channel sampled image capture system that uses such a lens, a good scheme is to have two finely sampled and relatively broad color channels close to the wavelength of optimum focus, and a third sparsely sampled channel further away from the wavelength of best focus. Such a scheme allows the best capturing of scene detail using the simple lens, while maintaining reasonable sensitivity. With achromatic lenses, as are used in DSCs, signal to noise considerations dominate and relatively narrow, equally spaced and independent spectral sensitivity curves are best.

To illustrate this situation it is possible to calculate the ISO speed of two hypothetical digital cameras. The first uses narrow and independent (but realistic) RGB spectral sensitivities (see figure 2), and the second uses HVS cone sensitivities. The RGB sensitivities will result in an ISO primary noise speed[2] approximately 5 times that resulting from the cone sensitivities.

Decades of experience in color photography and television have shown that acceptable color reproduction for typical three-dimensional scenes can be achieved using non-colorimetric RGB sensitivity curves. Minor color errors can be traded off for large speed gains. Because of this situation, it is unlikely that digital photography camera manufacturers will sacrifice speed to achieve the slight improvements in color analysis provided by colorimetric capture. However, if a good error metric for non-colorimetric capture can be found, it is likely that perceptual color analysis can be improved with little or no speed tradeoff.

Classes of Scene Analysis

If one assumes non-colorimetric capture, it is desirable to classify “scenes” into three general types for analysis:

Class A analysis - known and fixed colorant behavior and illumination

An example of Class A “scene” analysis is the scanning of an image on a known photographic material. With Class A analysis it is generally possible to characterize DSCs to produce “scene” colorimetry, as long as there are at least as many camera channels as colorants, and the spectral sensitivities of the color channels of the camera are well suited to analyzing the amounts of each colorant present. Class A analysis is typically supported using targets designed according to ISO 126413.

Because of colorant behavior at different concentrations, colorant interactions, surface reflections, and illumination and capture optical and geometric considerations, it is sometimes desirable with Class A analysis to use a DSC characterization employing a 3D CLUT.

Class B analysis - statistically expected colorant behavior and limited and fixed illumination behavior

This is a limited generalization of the Class A approach. The assumption is made that all the spectral radiance distributions present in the scene can be decomposed into x basis functions. Frequently this decomposition is divided into two parts: first assuming a known illumination spectral power distribution (SPD), and second assuming the scene spectral reflectance distributions can be decomposed into x basis functions. Separation of the illumination SPD allows the same basis functions to be used for all sources, assuming the spectral reflectance distributions present in scenes of interest can be decomposed into the x basis functions.

Class A and B analyses are equivalent when the x basis functions completely describe the scene, and the DSC has at least x color channels which are capable of effectively determining the amounts of each basis function present. In practice, further statistical assumptions are typically made about the different basis functions to allow a DSC with less than x color channels to estimate the amounts of the x basis functions. Also, the number of basis functions used may be limited so that they only represent the scene spectral radiance or reflectance distributions (referred to as the scene spectral correlation statistics) to a limited degree of accuracy. Consequently, color errors in Class B analysis generally arise from two sources: the camera does not have sufficient color channel capability to accurately analyze the x basis functions, and/or the x basis functions do not have the capability to represent every spectral distribution present in the scene.

Typically when Class B analysis is employed, six to eight basis functions are used to represent all possible scenes. It is important to note that some spectral distributions will not be reproducible using these basis functions. The six to eight basis functions typically chosen tend to be smooth, because smooth spectral distributions are most common in nature. When these basis functions are used, the results of the color analysis cannot go outside the gamut of colors that can be represented by the basis functions. For example, spectral colors in a scene will never be mapped into spectral colors in the analysis. The analysis contains a bias that is represented by the basis functions.

Class C analysis - unknown and variable colorant behavior and unknown and variable illumination behavior

This is the most general case, and is sometimes referred to as the “maximum ignorance” case.[4-7] There is disagreement in the color science community about whether it is desirable to assume maximum ignorance for scenes. Measurements of surface reflectance distributions consistently support decomposition into at most a dozen basis functions, yet this approach limits the analysis gamut. Scenes are comprised of more than surface reflectances, the behavior of surface reflectances in scenes is different from that of patches due to specular reflection components and three-dimensional illumination and capture geometry effects. Furthermore colors can be found in some scenes which do not arise from reflectance distributions. There are also questions about the perceptual effects of different types of errors when using Class C analysis.

In the debate about the relative merits of Class B and Class C analysis of natural scenes, the fundamental question is: Which analysis method estimates the HVS perception of scene colors most reliably? This question goes beyond CIE colorimetry. There are known differences between cone responses calculated using CIE color matching functions and those in the literature, but there is also overwhelming evidence that CIE color matching works from decades of experience in a number of industries. However, virtually all this experience is based on the matching of two-dimensional patches of spectrally smooth colorants using a controlled illumination geometry. What happens when some stimuli are not spectrally smooth, and a group of stimuli are taken together in the context of an image? One goal of this study is to determine whether Class B or Class C analysis is most appropriate for natural scenes, through subjective evaluation of images of natural scenes produced using each analysis method.

Reasons Why Conventional Error Metrics May Not Predict the Objectionability of Color Errors for Pictorial Scenes

If one abandons the goal of achieving perfect colorimetric scene analysis, the question arises how to analyze scenes to produce color errors with minimal objectionability. There may be several answers to this question, but a first approach is to base the color characterization on minimizing one of the ΔE metrics that take advantage of perceptually uniform color spaces. This approach has been shown to work well when applied to the capture of flat art using Class A analysis, and Class B analysis where the errors introduced by the limited number of basis functions, and the corresponding analyses of these basis functions by the DSC, are insignificant.

Another goal of this study is to determine whether ΔE minimization is appropriate for Class C analysis, and if not, what is appropriate.

Experimental Design

The question to be answered comes down to the selection of a preferred method or methods for DSC characterization for pictorial natural scene analysis. The candidate methods differ based on the error criterion minimized to determine the best transformation. The error metrics evaluated are listed in table 1.

RGB Color Space Error Minimization

In table 1, two items are worthy of comment. The first regards error minimization in RGB color spaces, and the associated nomenclature to distinguish these color spaces. The use of the ITU, PC and RIMM designations refers to the source of the RGB primaries on which the color spaces are based. ITU means that the color space for error minimization is based on an equi-energy transformation of the ITU-R BT.709/3[8] primary chromaticities with a CIE illuminant D65 white point. PC means that the color space for error minimization is based on monochromatic primaries with wavelengths of 450, 540, and 620 nm. and an equi-energy white point. RIMM means that the color space for error minimization is based on an equi-energy transformation of the proposed RIMM RGB primary chromaticities[9] and CIE illuminant D50 white point.

Table 1. Error Minimization Criteria for DSC Scene Analysis Characterization Transformations

Class B analysis, Macbeth Color Checker spectral reflectance correlation statistics, illumination dependent transformations.

CIE XYZ error minimization
(or any other linear transform thereof)
CIE L*u*v* ΔE minimization
CIE L*a*b* ΔE minimization
CIE ΔE94 minimization
ITU double gamma error minimization

Class C analysis, maximum ignorance, illumination independent

CIE XYZ ΔE minimization
CIE L*u*v* ΔE minimization
CIE L*a*b* ΔE minimization
CIE ΔE94 minimization

Class C analysis, maximum ignorance, illumination SPD weighted

ITU linear error minimization
ITU double gamma error minimization
PC linear error minimization
PC double gamma error minimization
RIMM linear error minimization
RIMM double gamma error minimization

The ITU and RIMM primaries were not used directly to construct a color space because they are not relative to an equi-energy white point. By convention, the color matching functions used for Class C characterizations assume an equi-energy white point. This matter will be discussed in more detail in the next section. Here it is sufficient to note that the ITU and RIMM space primaries were transformed from those of ITU-R BT.709-3 and RIMM using the method outlined in Annex B of ISO 17321 WD4.[1]

The importance of the PC wavelength primaries has been noted by Brill, et.al.[10] The specific wavelength values used in this study were obtained by determining the maxima of curves obtained by differencing logarithmic cone signals as specified by Hubel, et.al[11] and rounding to the nearest 10 nm. (see table 3). The PC primaries are already based on an equi-energy white point, so no transformation of the type used for the ITU and RIMM primaries was necessary.

Table 2. XYZ To RGB Conversion Matrices for Determining Color Matching Functions Based on Various Primaries and an Equi-Energy White Point

CIE XYZ to ITU RGB Conversion Matrix

3.08034	-1.53711	-0.54323
-0.92145	1.87593	0.04552
0.05314	-0.20418	1.15104

CIE XYZ to PC RGB Conversion Matrix

2.00177	-0.55776	-0.44401
-0.79985	1.66278	0.13707
0.00894	-0.01895	1.01001

CIE XYZ to RIMM RGB Conversion Matrix

1.29772	-0.25557	-0.04215
-0.52499	1.50808	0.01691
0.00000	0.00000	1.00000

Table 3. Maxima of Curves Generated by Differencing Logarithmic Cone Signals

Based on cone response functions determined by Smith & Pokorny[12]

Rmax = 620 nm.
Gmax = 538 nm.
Bmax = 445 nm.

Based on cone response functions determined by Stiles & Estevez[13]

Rmax = 615 nm.
Gmax = 536 nm.
Bmax = 447 nm.

In table 1, the designations “linear” and “double gamma” indicate whether the color space is linear with respect to radiance, or if the ISO RGB OECF specified in ISO 17321 WD4 was applied to make the color space more perceptually uniform. In the latter case, the choice of the nonlinear function used is somewhat arbitrary. The ISO RGB OECF was derived from the sRGB nonlinear encoding function specified in IEC 61966-2-1,[14] and is used in ISO 17321 to maximize compatibility with sRGB. It may be that a different nonlinear function would result in improved performance for error minimization. However, the number of variables that have to be dealt with in this study is difficult to manage. The first goal is to determine if a departure from linearity is beneficial. The sRGB nonlinear encoding has been shown to be reasonably perceptually uniform.[15] After the field of candidates has been narrowed, further investigations could be conducted to determine if the ISO RGB OECF is optimal for this application.

Difficulties with Non-Equi-Energy White Points and Class C Analysis

The second noteworthy item with respect to table 1 is the subdivision of the error minimization criteria for Class C analysis into illumination independent and illumination SPD weighted criteria. This subdivision is necessary because of the structure of CIE colorimetry and the associated error metrics. With these metrics, colors outside the spectral locus do not have clear meaning. Unfortunately, with Class C analysis, the test colors effectively are spectral colors, so it is not possible to accommodate white point changes without pushing some of the test color outside the spectral locus.

This limitation of CIE colorimetry with respect to imaging is worthy of further study. For example, say that a deep yellow object is photographed in a scene where the viewers adapted white point has the chromaticity of CIE illuminant D65. If a reproduction is produced where the viewers adapted white point has the chromaticity of CIE illuminant A, it is possible that the chromatic adaptation transform used to go from D65 to A will push the deep yellow outside the spectral locus. How should this color be represented?

When spectral colors are the test colors, any change in the adapted white will push some colors outside the spectral locus because all the test colors start out on the border. This means that for the CIE ΔE error criteria to be used for minimization, the adapted white has to be assumed to be equi-energy, and only one transformation can be determined per camera per error criterion. This approach will at least result in all the aim values for the colors being well defined. The colorimetry estimates produced by the camera may still result in colors outside the spectral locus, because the cameras are not colorimetric. This situation is particularly difficult when negative CIE XYZ values are produced.

Several methods were implemented to deal with negative XYZ values for the DSC estimated colorimetry. The first was to extend the functions for converting CIE XYZ values to CIE L*u*v* and L*a*b* values into the negative region. The second was to clip all negative XYZ values to zero. The third was to constrain the DSC characterization transformation determined to not produce any negative XYZ values for the spectral colors. It turned out that the characterization transformations produced using these methods were not too different when the Macbeth Color Checker patches were used for Class B analysis. It was found that none of the ΔE criteria produced viable characterization transformations for Class C analysis.

Because of these complications, a spectral weighting method is specified for Class C analysis characterizations in ISO 17321 WD4. This method is not a chromatic adaptation transform, but a weighting of each spectral color based on the amount of illumination spectral power at that wavelength. This weighting reduces the influence of spectral colors that are present in smaller quantities in the illumination source. In the extreme, if no power is present at a particular wavelength for a particular illumination source, no attempt is made to minimize analysis error at that wavelength.

Even though the spectral weighting method is different from a chromatic adaptation transform, it can still produce aim colors outside the spectral locus because of the re-normalization of the spectral responses required to determine white point preserving characterization transforms as specified in ISO 17321 WD4. However, with the RGB color space candidates, negative values are well defined and do not present a problem. The spectral weighting values are used as opposed to performing a chromatic adaptation transform because the validity of applying a chromatic adaptation transform to spectral colors is uncertain, and because it seems more advantageous to make full use of the illumination SPD as opposed to only its chromaticity. In any case, the purpose of the psychophysical experiments described below is to determine the validity of whatever method is used.

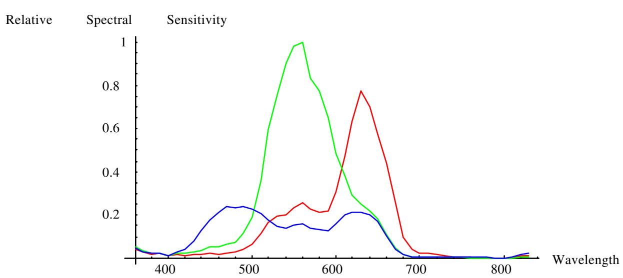

Figure 1: Camera P spectral sensitivities

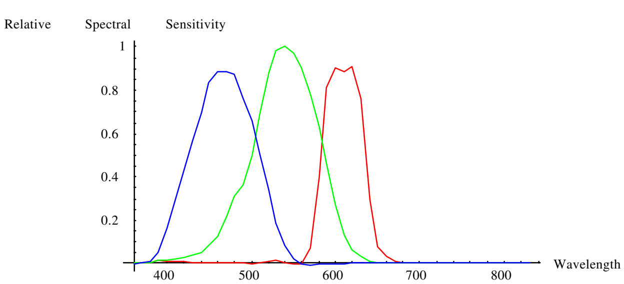

Figure 2: Camera R spectral sensitivities

Experimental Procedure

The DSC characterization methods listed in table 1 were evaluated by capturing a number of natural scenes using two DSCs with substantially different spectral sensitivity curves. These cameras are at the extremes of a broad range of RGB sensitivities as found in current DSCs (see figures 1 and 2). Non-RGB DSCs (such as CMYG) were not considered because with non-RGB cameras an unfortunate tradeoff exists between ISO speed and color saturation.[16] With non-RGB cameras (and to some extent even with RGB cameras with broad sensitivities), the question of how to analyze colors is strongly impacted by the speed tradeoff. With these cameras, optimal color reproduction may be sacrificed to increase speed.

After capture, the necessary steps were applied to the captured images in order to obtain scene colorimetry estimates using the characterization methods mentioned above. Samples for subjective evaluation were then produced and evaluated.

Image Processing

The image processing steps were as follows. The first two steps were not necessary with camera R because this camera outputs linear 12-bit dark-current subtracted sensor data directly.

1. Linearize raw image data to 12-bits using an inverse OECF LUT calculated from a focal plane OECF determined according to ISO 14541.[17]
2. Subtract dark current based on the average linearized 12-bit digital code value for the optical black CCD pixels.
3. Subtract a DC camera veiling flare value equal to 3% of the mean focal plane digital code value for each channel.
4. Apply channel multipliers appropriate for the camera spectral sensitivities and scene illumination spectral power distribution.
5. Clip the channels so the maximum digital code value for each channel is the same as the maximum value for the channel with the smallest channel multiplier.
6. Demosaic the CFA image data using bilinear interpolation.
7. Apply the 3x3 color matrix determined for the specified characterization method.
8. Manually scale and clip the digital code values for each image to produce a satisfactory overall image lightness level.
9. Apply a LUT to convert to sRGB with the CRT black point luminance taken to be 1% of the white point luminance as specified in PIMA 7667 WD1.

Experimental Setup

For the initial subjective experiments, the images were viewed on a CRT display set up in an e-sRGB viewing environment as specified in PIMA 7667 WD1. The measured values for the viewing environment and display are listed in table 4. Figure 3 is a plot of the measured display L* compared to the ideal e-sRGB display L* values.

Table 4. Measured display characteristics

Cosine-corrected ambient illuminance level measured in the plane of the display faceplate - 61 lux.

Display luminance measured from the observer position with the display turned off - 1.06 cd/m²

Display luminance and chromaticity as a function of input digital value:

Digital Value	Luminance(Y)	*x*	*y*
0	1.25	0.3212	0.3492
12	1.46	0.3325	0.3518
25	2.01	0.3368	0.3508
38	2.83	0.3358	0.3481
51	4.41	0.3346	0.3464
64	6.64	0.3330	0.3444
76	8.98	0.3294	0.3439
89	12.6	0.3243	0.3405
102	16.2	0.3214	0.3403
115	20.5	0.3208	0.3355
128	24.6	0.3188	0.3352
140	30.8	0.3182	0.3333
153	39.1	0.3172	0.3325
166	47	0.3168	0.3331
179	56.3	0.3160	0.3320
191	65.2	0.3147	0.3310
204	74.2	0.3140	0.3305
217	81.1	0.3128	0.3282
230	93.7	0.3136	0.3287
243	111	0.3147	0.3265
255	123	0.3138	0.3271

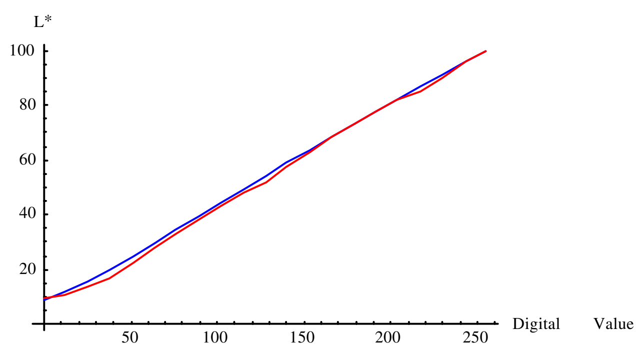

Figure 3. Measured and ideal L as a function of digital value*

A Matlab interface was constructed for conducting the experiments. This interface presents pairs of images to the subject, who is then forced to choose one image. The instructions given to the observer were as follows: Which of the two displayed images has a higher color accuracy (the colors are closer to the colors in a real scene)? The analysis of the paired comparison data was done in the same way as previous experiments in the ISO TC42 committee.[18-21] These previous experiments served as the basis for development of ISO 12232.

Table 5. Class C Analysis Error Metric Subjective Results

Camera P Scene z-scores

Scene	ITUlin	ITUdg	PClin	PCdg	RIMMlin	RIMMdg
Mt. Moran	-1.84	1.6	-0.75	1.64	-1.86	1.21
Carousel & Flowers	-0.26	0.28	0.28	-0.56	0.69	-0.43
Chris & Cotton Candy	0.72	0.18	-0.54	-0.72	0.36	0.00

Camera R Scene z-scores

Scene	ITUlin	ITUdg	PClin	PCdg	RIMMlin	RIMMdg
Roses	0.42	0.64	1.66	1.35	0.37	-4.44
Grapes	-0.76	-1.16	1.79	1.44	-1.61	0.3
Meat	-2.19	-2.36	2.64	2.35	0.08	-0.52
Citrus Aisle	-1.52	-1.6	-0.61	1.73	0.43	1.57
Straw-berries	-5.53	-3.41	0.65	3.24	4.64	2.74
Outdoor scene	-0.13	-0.72	0.00	-0.29	1.27	-0.13

Combined Results

Metric	ITUlin	ITUdg	PClin	PCdg	RIMMlin	RIMMdg
Mean z-score	-1.23	-0.73	0.57	1.13	0.49	0.03
Max z-score	0.72	1.6	2.64	3.24	4.64	2.74
Min z-score	-5.53	-3.41	-0.75	-0.72	-1.86	-4.44
Overall Rank	6	5	2	1	3	4

Experiment 2 - Comparison of different error metrics for Class B analysis based on Macbeth Color Checker correlation statistics

The results obtained so far in the subjective experiments involving the error metrics for Class B analysis based on Macbeth Color Checker patches are presented in table 6.

Results

The following characterization methods were eliminated as candidates from the subjective experiments related to Class C analysis because of their obviously poor performance:

1. CIE XYZ error minimization fails if the illumination is very different from equi-energy.

2. CIE L*a*b* and zero-clipped L*a*b* ΔE minimization fails if the illumination is very different from equi-energy.
3. CIE L*a*b* ΔE minimization, with the characterization matrix constrained to only produce positive XYZ values, fails in general.
4. CIE L*u*v* ΔE and CIE ΔE94 minimization fails in general.

In all of the above cases, a characterization method was not considered to have failed unless it was relatively indisputable that the results produced were unacceptable.

Experiment 1 - Comparison of different error metrics for Class C analysis

The results obtained so far in the subjective experiments involving the remaining error metrics for Class C analysis are presented in table 5.

Table 6. Class B Analysis Error Metric Subjective Results

Camera P Scene z-scores

Scene	XYZ error	ITUdg error	L*a*b* ΔE	L*u*v* ΔE	ΔE94
Mt. Moran	-1.84	1.6	-0.75	1.64	-1.86
Carousel & Flowers	0.73	0.28	0.28	-0.56	0.69
Chris & Cotton Candy	0.9	-1.71	0.36	0.63	-0.18

Camera R Scene z-scores

Scene	XYZ error	ITUdg error	L*a*b* ΔE	L*u*v* ΔE	ΔE94
Roses	0.57	1.32	-1.23	0.23	-0.89
Grapes	-1.03	-0.15	0.28	0.75	0.15
Meat	-0.28	-0.17	0.73	-0.56	0.28
Citrus Aisle	-1.16	-1.74	0.72	1.44	0.74
Straw-berries	-2.21	-3.41	2.28	1.87	1.47
Outdoor scene	-0.11	1.17	-0.42	-0.2	-0.44

Combined Results

Metric	XYZ error	ITUdg error	L*a*b* ΔE	L*u*v* ΔE	ΔE94
Mean z-score	-0.49	-0.31	0.25	0.58	0.00
Max z-score	0.9	1.6	2.28	1.87	1.47
Min z-score	-2.21	-3.41	-1.23	-0.56	-1.86
Overall Rank	5	4	2	1	3

Experiment 3 - Comparison of Class B and Class C analysis using the same error metric

The results obtained so far in the subjective experiments comparing Class B and Class C analysis with both using the ITU based error metrics are presented in table 7.

Table 7. Class B and Class C analysis comparison using ITU based error metrics

Camera P Scene z-scores

Scene	Class B ITUlin	Class B ITUdg	Class C ITUlin	Class C ITUdg
Mt. Moran	-2.49	-0.28	0.61	2.16
Carousel & Flowers	-0.14	-0.3	0.29	0.15
Chris & Cotton Candy	1.01	0.44	0.52	-1.97

Camera R Scene z-scores

Scene	Class B ITUlin	Class B ITUdg	Class C ITUlin	Class C ITUdg
Roses	0.42	0.64	1.66	1.35
Grapes	0.00	0.13	0.00	-0.13
Meat	1.68	0.44	-0.95	-1.17
Citrus Aisle	-0.02	0.15	0.00	-0.13
Straw-berries	2.3	0.34	-2	-0.64
Outdoor scene	0.3	0.44	-0.3	-0.44

Combined Results

Metric	Class B ITUlin	Class B ITUdg	Class C ITUlin	Class C ITUdg
Mean z-score	0.34	0.22	-0.02	-0.09
Max z-score	2.3	0.64	1.66	2.16
Min z-score	-2.49	-0.3	-2	-1.97

The z-score results presented in tables 5, 6 and 7 are based on 9 subjects for all scenes except the "Chris & Cotton Candy" scene, which had 7 subjects. The 95% confidence limits for 9 test persons is 0.65, and for 7 test persons is 0.74.

Initial Conclusions

Transformations based on the CIE XYZ error and all three CIE ΔE criteria tested were found not to be acceptable for Class C analysis. A new error metric is needed for determining Class C characterization transformations. The proposed new RGB based metrics seem to perform reasonably well, with the PC double gamma criterion scoring highest. Unfortunately, this error metric has not yet been tested for use in Class B analysis.

Of the Class B analysis error criteria tested, L*u*v* ΔE minimization scored highest, closely followed by L*a*b* ΔE minimization. The z-score differences for this experiment were not as large as with the Class C analysis experiment. Nevertheless, it looks probable that CIE XYZ error minimization produces poorer transformations than CIE L*u*v* and L*a*b* ΔE minimization. This is significant in view of the widespread use of XYZ error minimization. ITU double gamma error minimization also did not perform as well as the L*u*v* and L*a*b* ΔE minimizations, but in experiment 1 it performed poorly compared to the PC double gamma error minimization for Class C analysis so no conclusions can be drawn regarding the use of RGB error minimization for Class B analysis.

The z-score differences between Class B and Class C analysis were well below the level of statistical significance for the combined results, but significant differences were observable for some specific scenes. This is consistent with the idea that the performance of Class B analysis depends on how accurately the spectral radiance correlation statistics assumptions match the actual scene spectral radiance correlation statistics. If the match is good, one would expect Class B analysis to outperform Class C analysis. The opposite would be expected if the statistical assumptions were not a good match to the scene. Unfortunately, the design of the experiment reported here is not optimal because of the choice of the ITU based RGB error minimization criteria.

In all cases, the results obtained were strongly scene dependent. In choosing the best error minimization criteria it is necessary to evaluate a wide range of scenes, and to recognize that no single criterion will give the best results for all scenes.

Continued Investigations

Experiment 1 as outlined here will be continued. More observers will be added to the CRT display based experiments, and a comparable set of experiments may be conducted using print samples. After extensive exposure to the samples, the authors have the impression that it is easier to differentiate between print samples than CRT display samples, so it may be possible to evaluate smaller differences and/or use smaller numbers of observers. There is also the question of whether the results of the experiments will be the same for print samples as for CRT display samples.

Experiments 2 and 3 will not be continued as is. Initial results indicate that the ITU primaries are the worst of the three candidate sets on which to base error minimization. Experiments 2 and 3 will therefore be continued with the PC primary based error minimization substituted.

The authors hope to also continue this work with three new experiments as follows:

Experiment 4 - Comparison of different Class B analysis spectral correlation statistics assumptions using the error metric preferred for the Macbeth Color Checker statistics. Candidate statistics include those of the patches used for the CIE Color Rendering Index,[22] and those of the SOCS database.[23]

Experiment 5 - Comparison of the best Class B analysis method to the best Class C analysis method.

Experiment 6 - Investigation of the effects of more sophisticated color rendering algorithms applied to scene-referred images. The question to be answered with this experiment is whether the additional color rendering step affects the choice of analysis method.

References

1. ISO 17321 WD4, Graphic Technology and Photography - Colour characterisation of digital still cameras (DSCs) using colour targets and spectral illumination, 1 October 1999. http://www.pima.net/standards/iso/tc42/WG20/WG20_POW.htm
2. ISO 12232:1998(E), Photography - Electronic still picture cameras - Determination of ISO speed.
3. ISO 12641:1997(E), Graphic Technology - Prepress digital data exchange - Colour targets for input scanner calibration.
4. G.D. Finlayson and M.S. Drew, "The Maximum Ignorance Assumption with Positivity," Proceedings, IS&T/SID Fourth Color Imaging Conference, p. 202-204, 1996.
5. P.M. Hubel, J. Holm, G.D. Finlayson, and M.S. Drew, "Matrix Calculations for Digital Photography," Proceedings, IS&T/SID Fifth Color Imaging Conference, p. 105-109, 1997.
6. G.D. Finlayson and M.S. Drew, "White-Point preserving color correction," Proceedings, IS&T/SID Fifth Color Imaging Conference, p. 258-261, 1997.
7. G.D. Finlayson and M.S. Drew, "White-Point Preservation Enforces Positivity," Proceedings, IS&T/SID Sixth Color Imaging Conference, p. 47-52, 1998.
8. ITU-R BT.709-3:1998, Parameter values for the HDTV standards for production and international programme exchange
9. K.E. Spaulding, G.J. Woolfe, and E.J. Giorgianni, "Reference Input/Output Medium Metric RGB Color Encodings (RIMM/ROMM RGB)," Proceedings, IS&T PICS Conference, p. 155-163, 2000.
10. M.H. Brill, G.D. Finlayson, P.M. Hubel, and W.A Thornton, "Prime colors and color imaging," Proceedings, IS&T/SID Sixth Color Imaging Conference, p. 33-42, 1998.
11. P.M. Hubel, J. Holm, and G. Finlayson, "Illuminant estimation and colour correction," Colour Imaging: Vision and Technology, ed. L.W. MacDonald and M.R. Luo, p. 73-95, Wiley, 1999.
12. V. C. Smith and J. Pokorny, "Spectral sensitivity of colorblind observers and the cone pigments," Vision Research 12 2059-2071, 1972.
13. G. Wyszecki and W.S. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formulae, second edition, Wiley, 1982.

14. IEC 61966-2-1:1999, Multimedia systems and equipment - Colour measurement and management - Default RGB colour space - sRGB
15. PIMA 7667 WD1, Photography - Electronic still picture imaging - Extended sRGB color encoding - e-sRGB http://www.pima.net/standards/it10/IT10_POW.htm#7667
16. R.L. Baer and J. Holm, "A Model for Calculating the Potential ISO Speeds of Digital Still Cameras based upon CCD Characteristics," Proceedings, IS&T PICS Conference, p. 35-38, 1999.
17. ISO 14524: 1999, Photography - Electronic still picture cameras - Methods for measuring opto-electronic conversion functions (OECFs)
18. J. Holm and S. Süsstrunk, "An EIQ - Subjective Image Quality Correlation Study," Proceedings, IS&T's 47th Annual Conference/ICPS, p. 634-640, 1994.
19. G.A. Gescheider, Psychophysics - Method, Theory, and Application, second edition, Lawrence Erlbaum Associates, 1985.
20. J.C. Bartleson and F. Grum, ed. Optical Radiation Measurements: Vol. 5, Visual Measurements, Academic Press, 1984.
21. W.S. Torgerson, Theory and Method of Scaling, Wiley, 1958.
22. CIE Technical Report 13.3:1995, Method of Measuring and Specifying Colour Rendering Properties of Light Sources
23. J. Tajima, "A Huge Spectral Characteristics Database and its Application to Color Imaging Device Design," Proceedings IS&T/SID Sixth Color Imaging Conference, p86-89, 1998.

Biography

Jack Holm is an imaging and color scientist at the Hewlett-Packard Laboratories. Prior to joining HP he was an independent technical consultant, and prior to that served on the faculty at the Rochester Institute of Technology. He is a long-standing participant in standards activities relating to digital photography and color management, and is the convener of ISO TC42 WG20, a joint working group of the ISO photography and graphic technology committees concerning digital still camera color characterization. He is project leader for ISO 17321-2, ISO 14524, ISO 7187, ISO 5800, ISO 2240, and ISO 6, and is the ISO TC42 liaison to both the CIE and the ICC. He is active in the IS&T and last year was co-chair of this conference. Mr. Holm holds a B.S. in Physics from Texas A&M University and an M.S. in Imaging Science from RIT.

Ingeborg Tastl is a digital imaging scientist at Sony's US research laboratories, working in the fields of digital photography and color science in general. Before joining Sony in 1998, she worked as a Postdoc at ENST in Paris and as a Postdoc at the Vienna University of Technology. She got her M.S. degree and her Ph.D. degree in computer science from the Vienna University of Technology. She is a member of IS&T and actively involved in the development of standards in the field of Digital Photography.

Steven Hordley obtained his BSc in Mathematics from the University of Manchester (Manchester, England) in 1992. He then obtained an MSc degree in Image Processing from Cranfield Institute of Technology (Cranfield, England) in 1996. Between 1996 and 1999 he studied at the University of York (York, England) and the University of Derby (Derby, England) for a PhD in Colour Science. In September 1999 he was appointed a Senior Research Fellow within the School of Information Systems at the University of East Anglia (Norwich, England). His research interests are in the fields of colour imaging and physics-based vision, and he has particular expertise in the areas of colour constancy, colour correction, and mathematical modelling.