Comparison of the accuracy of different white balancing options as quantified by their color constancy

J A Stephen Viggiano, jasv@acolyte-color.com
Acolyte Color Research, West Henrietta, NY, USA
http://www.acolyte-color.com

ABSTRACT

Six different methods for white-balancing digital images were compared in terms of their ability to produce white-balanced colors close to those viewed under a specific viewing illuminant. The six methods were: native camera RGB, XYZ, CAM02, ITU Rec BT.709 RGB, sharpened camera RGB, and illuminant-dependent. 4096 different sets of camera sensitivities were synthesized; 170 objects were evaluated under a canonical viewing illuminant (D65) and six additional taking illuminants (A, D50, D75, F2, F7, and F11). Each white balancing method was exercised in turn, and the mean and 90th percentile ΔE^*_{ab} were determined.

We found that illuminant-dependent characterization produced the best results, sharpened camera RGB and native camera RGB were next best, XYZ and CAM02 were often not far behind, and balancing in the -709 primaries was significantly worse. We recommend that, whenever the illuminant is identified, the illuminant-dependent technique be employed because of its superior performance.

Keywords: White balance, color constancy, photography, minimal knowledge.

Cite this paper: Viggiano, J A Stephen, Comparison of the accuracy of different white balancing algorithms as quantified by their color constancy, *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V: Proceedings of the SPIE, volume 5301.* Bellingham, WA: SPIE: The International Society for Optical Engineering, 2004, in press.

Comparison of the accuracy of different white balancing options as quantified by their color constancy

J A Stephen Viggiano, JASV@acolyte-color.com
Acolyte Color Research, 35 Mickens Bend, West Henrietta, NY USA 14586

ABSTRACT

Six different methods for white-balancing digital images were compared in terms of their ability to produce white-balanced colors close to those viewed under a specific viewing illuminant. The six methods were: native camera RGB, XYZ, CAM02, ITU Rec BT.709 RGB, sharpened camera RGB, and illuminant-dependent. 4096 different sets of camera sensitivities were synthesized; 170 objects were evaluated under a canonical viewing illuminant (D65) and six additional taking illuminants (A, D50, D75, F2, F7, and F11). Each white balancing method was exercised in turn, and the mean and 90th percentile ΔE^*_{ab} were determined.

We found that illuminant-dependent characterization produced the best results, sharpened camera RGB and native camera RGB were next best, XYZ and CAM02 were often not far behind, and balancing in the -709 primaries was significantly worse. We recommend that, whenever the illuminant is identified, the illuminant-dependent technique be employed because of its superior performance.

Keywords: White balance, color constancy, photography, minimal knowledge.

1. INTRODUCTION

Color digital cameras must be able to capture images (either still or moving) under a wide variety of illumination conditions and internally process them so they appear without a significant color cast when displayed or printed. The aspect of this process which deals with chromatic issues is referred to as white balancing. The white balancing parameters depend primarily upon the illuminant under which the image was captured (the Taking Illuminant, which must be somehow inferred from the captured image itself), and the illuminant under which the resulting image will be presented (the Viewing Illuminant). The problem of white balancing has, based on literature published within the past few years, largely been separated into two basic steps: taking illuminant identification, [1-7] and the actual processing. The processing step is the subject of this paper.

1.1 Alternatives for white balancing

A perhaps subtle factor, but nevertheless possessing a large impact on the final result, is the space in which the scaling should be performed. Certainly, the camera's native RGB is one option. The camera's native RGB is customarily transformed into something appropriate for display, such as sRGB. [8] One may first scale the native camera RGB, then transform into sRGB (or some other device space) before storing the image.

A second alternative which suggests itself is to perform the scaling in the output space. Under this option, the native camera RGB would first be transformed into the output space, then the scaling would be performed. However, as a step (perhaps implicit) in the conversion of the camera's native RGB into the output space, the native camera RGB signals are frequently transformed into XYZ tristimulus values. This suggests a third alternative space in which the scaling may be performed: not before the conversion, as with the first alternative, or after the conversion, as in the case of the second, but in the midst of the process.

Although the problem of chromatic adaptation is different from that of white balancing, there are similarities. A fourth alternative space in which the scaling may be performed is a linear transformation of XYZ. We have selected the proposed CAT02. [9]

Finlayson, Drew, and Funt reasoned that certain linear combinations of sensor sensitivities, which they termed sharpened sensors, would approximate ideal (from a color constancy point of view) sensitivities, because their effective bandwidths would be narrower than for the unsharpened sensitivities, and the sharpened sensitivities more closely approximate Dirac delta functions which produce perfect color constancy. [10] We include a sharpened version of each sensitivity combination as a fifth alternative.

The final alternative we consider is to perform the white balancing and color conversion simultaneously. Finlayson, *et al.,* have argued that the overwhelming majority of taking illuminants may be represented by a proxy of a small number of alternatives. [2] One may compute an illuminant-dependent transformation from native camera RGB to the output space for each of these alternative taking illuminants. Using first one of the existing techniques to identify the taking illuminant, the appropriate transformation may be selected and then applied to the native camera RGB.

The five alternative methods for which we evaluate the accuracy of white balancing are:

1. Native Camera RGB
2. Primaries based on the ITU Rec BT.709 Phosphor set (those used in sRGB);
3. XYZ Tristimulus Values;
4. CAT02 RGB;
5. Spectrally Sharpened Sensors; and
6. Illuminant-dependent Characterization.

1.2 Similar Work

In their recent paper, Woolfe, *et al.,* examined a question similar to that posed by this paper. They used a single camera sensitivity set, whose sensitivity spectra coincided with the CIE Color Matching Functions for the 2 degree observer. [11] This removed any doubt surrounding the conversion between camera RGB and XYZ. However, it utilizes a set of sensitivity spectra which are rarely used in practice for a variety of reasons (*e.g.,* difficulty in fabrication, noise amplification caused by aggressive matrixing). In our study, we trade a small amount of uncertainty surrounding the RGB to XYZ transformation for sensitivity spectra which more closely approximate those used in actual cameras. We also exercise a large number of sensitivity spectrum combinations. Therefore, we feel our work complements that addressed in the other paper.

Woolfe, *et al.,* concluded that the sRGB primaries were very unfavorable for performing white balancing. They computed an optimized set of primaries which maximized a color constancy-based error metric.

1.3 Symbols and Expressions

RGB	A matrix with three columns, representing the camera's radiometrically linear RGB output
XYZ	A matrix with three columns, representing the XYZ tristimulus values
t	As a subscript: Denotes the taking illuminant; as a superscript: Denotes matrix transposition
c	Subscript denoting the canonical illuminant
v	Subscript denoting the viewing illuminant
wb	Subscript denoting white-balanced quantities
A	Matrix for converting RGB into XYZ; $XYZ \approx RGB \cdot A$
D	Matrix with one column for each sensitivity (Device sensitivity matrix)
C	Matrix with three columns, containing the CIE Color Matching Functions, 2° observer
S	Diagonal matrix containing a spectral power distribution (illuminant)
$\mathbf{B}^t\mathbf{B}$	Square symmetric matrix containing the auto-inner product of the characterization set
diag	Diagonal operator; transforms a vector into a diagonal matrix

2. EXPERIMENTAL APPROACH

Our experimental approach is based on computation and simulation, and involves the generation of a large number of camera sensitivity spectra combinations, determining a 3x3 matrix which permits the approximation of CIE XYZ tristimulus values under a canonical illuminant, exercising the six scaling options, each under several taking illuminants, and comparing the XYZ tristimulus values thus computed to those obtained under the canonical illuminant. This comparison is a measure of color constancy. White balancing options which produce tristimulus values close to those which actually prevail under the canonical illuminant will be deemed better than those which produce tristimulus values further from those obtained under the canonical illuminant.

2.1 Taking illuminants assumed known

We assume the illuminant under which an image is captured (termed the Taking Illuminant) is known. As was pointed out earlier, much work has been done on automatic illumination identification based solely on scene content, making this a reasonable assumption. We assume that we know the actual illuminant, not just the camera RGB of a non-selective object.

2.2 Quantification of Color Constancy

Raw linear camera RGB may be computed by multiplying together the camera's spectral sensitivity, the spectral power distribution of the taking illuminant, and the reflectance spectrum of an object, then summing (or integrating) over all wavelengths. The raw camera RGB is then transformed into XYZ and white balanced using the formulae which appear in Section 2.3 below. This will result in a different set of white-balanced XYZs for each of the five methods. These XYZs are compared to the XYZ tristimulus values of the object as viewed under the canonical illuminant. The method which produces the closest match between the white balanced color and the color under the canonical illuminant may be deemed the most color constant.

This process may be repeated for a number of different objects. We refer to such a set as the "evaluation suite." The method which tends to produce the closest matches between white balanced XYZ and XYZ as computed under the canonical illuminant would be deemed the most color constant.

Metrics such as average or 90th percentile ΔE^*_{ab}, because they are small when colors are closely matched and large when they are not, may be deemed quantifications of "color inconstancy," rather than color constancy. Thus, one would look for the method which would tend to minimize such metrics.

Williams suggested a pooled root mean square version of the separate components ΔE^* (*i.e.*, ΔL^*, ΔC^*, ΔH^*) to quantify color inconstancy. [12] While we believe this is well suited to mean-type calculations, we feel that the upper quantiles of a particular ΔE^* distribution are more visually significant. For consistency with such higher quantile calculations, we compute all statistics based on the total color error, ΔE^*.

2.3 Formulae for white balancing

The formulae for performing white balancing under the six methods are described here mathematically. We assume the camera RGBs are contained in a matrix, n x 3, so each row contains an RGB triplet, and corresponds to a particular object in the evaluation suite. (Some authors choose to place RGBs into columns. One can transpose each matrix and reverse the order of multiplication to switch from one convention to the other.)

2.3.1 White balancing in camera RGB

$$\mathbf{XYZ}_{wb} = \mathbf{RGB}_t \cdot \mathrm{diag}\,(R_{wc} / R_{wt}, G_{wc} / G_{wt}, B_{wc} / B_{wt}) \cdot \mathbf{A}_c$$

2.3.2 White balancing in ITU Rec BT.709 Primaries

$$\mathbf{XYZ}_{wb} = \mathbf{RGB}_t \cdot \mathbf{A}_c \cdot \mathbf{M}_{709} \cdot \mathrm{diag}\,(R_{709wc} / R_{709wt}, G_{709wc} / G_{709wt}, B_{709wc} / B_{709wt}) \cdot \mathbf{M}^{-1}_{709}$$

where $[R_{709wc}\ G_{709wc}\ B_{709wc}] = [X_{wc}\ Y_{wc}\ Z_{wc}]\ \mathbf{M}_{709}$ (similarly for $R_{709wt}\ G_{709wt}\ B_{709wt}$) and $\mathbf{M}_{709}$ is as specified in the Appendix.

2.3.3 White balancing in XYZ

$$\mathbf{XYZ}_{wb} = \mathbf{RGB}_t \cdot \mathbf{A}_c \cdot \mathrm{diag}\,(X_{wc} / X_{wt}, Y_{wc} / Y_{wt}, Z_{wc} / Z_{wt})$$

2.3.4 White balancing in CAT02 RGB

$$XYZ_{wb} = \mathbf{RGB}_t \cdot \mathbf{A}_c \cdot \mathbf{M}_{CAT02} \cdot \mathrm{diag}\ (R_{02wc} / R_{02wt}, G_{02wc} / G_{02wt}, B_{02wc} / B_{02wt}) \cdot \mathbf{M}^{-1}_{CAT02}$$

where $[R_{02wc}\ G_{02wc}\ B_{02wc}] = [X_{wc}\ Y_{wc}\ Z_{wc}] \cdot \mathbf{M}_{CAT02}$ (similarly for $R_{02wt}\ G_{02wt}\ B_{02wt}$) and $\mathbf{M}_{CAT02}$ is as specified in the Appendix.

2.3.5 White balancing for Spectrally Sharpened Sensors

$$\mathbf{XYZ}_{wb} = \mathbf{RGB}_t \cdot \mathbf{T} \cdot \mathrm{diag}\ (R_{Swc} / R_{Swt}, G_{Swc} / G_{Swt}, B_{Swc} / B_{Swt}) \cdot \mathbf{T}^{-1} \cdot \mathbf{A}_c$$

where $[R_{Swc}\ G_{Swc}\ B_{Swc}] = [R_{wc}\ G_{wc}\ B_{wc}] \cdot \mathbf{T}$, and similarly for $R_{Swt}\ G_{Swt}\ B_{Swt}$, and $\mathbf{T}$ is the matrix which defines the sharpening transformation.

2.3.6 Illuminant-dependent white balancing

$$\mathbf{XYZ}_{wb} = \mathbf{RGB}_t \cdot \mathbf{A}_t$$

where $\mathbf{A}_t = (\mathbf{D}^t \cdot \mathbf{S}_t \cdot \mathbf{B}^t\mathbf{B} \cdot \mathbf{S}_t \cdot \mathbf{D})^{-1} \cdot \mathbf{D}^t \cdot \mathbf{S}_t \cdot \mathbf{B}^t\mathbf{B} \cdot \mathbf{S}_v \cdot \mathbf{C}$ (this is the same formula used to compute the characterization matrix $\mathbf{A}_c$, with $\mathbf{S}_t = \mathbf{S}_v$)

3. EXPERIMENTAL METHODOLOGY

3.1 Characterization methodology

There are a variety of methods for (approximately) converting camera RGB into XYZ. We applied the Minimal Knowledge technique described in two earlier papers. [13, 14] We used the following parameters: correlation width parameter α = 50nm, ratio of mean radiance ratios at 400 and 700nm q = 0.5, and coefficient of variation $v = \sqrt{3}/3$.

3.2 Viewing, canonical, and taking illuminants

In a typical consumer workflow, captured images are viewed on a computer monitor, either as an end unto itself, or before subsequent processing/retouching, or before printing. It makes sense, then to base the viewing conditions on those which prevail for a computer monitor. Of interest here is the monitor's white point. In the United States, this is often that of CIE Illuminant D65. Therefore, we selected D65 as the viewing illuminant. This is the illuminant under which the objects in the evaluation suite are assumed to have been viewed.

Although it may be possible to have the viewing illuminant, under which the objects in the evaluation suite are evaluated, be different from the canonical illuminant, under which the scaling is assumed to be optimized, doing so introduces an additional step in the process (an additional scaling). This, in turn, injects uncertainty, and, therefore, is better avoided. Accordingly, we selected the canonical illuminant to be the same as the viewing illuminant, viz, D65.

For the taking illuminants, a representative sample of six were selected, with correlated color temperatures ranging from 2853 to 7500 Kelvins. This spans a variety of lighting conditions, from near-dawn/sunset to midday north light. Of the six, three were fluorescent. Specifically, the taking illuminants were: A, D50, D75, F2, F7, and F11. Note that the CIE x, y chromaticity coordinates of F7 are nearly identical to those of D65, and F7's Color Rendering Index is 90. If white balancing is entirely dependent upon x, y chromaticity coordinates, then all balancing options should work about equally well. If, on the other hand, white balancing is more involved than this, some techniques may be expected to work better than others.

All illuminants used in this investigation were normalized so their Y tristimulus values were identical.

3.3 Evaluation Methodology

If a white balancing technique performs well, the colors predicted by the technique, in conjunction with the provided characterization, should match those under the viewing illuminant. In other words, regardless of the taking illuminant, good white balancing should produce results which are as indistinguishable as possible from the objects in the scene itself, viewed under the viewing illuminant. In order to formulate an evaluation methodology, it is necessary to specify a suite of objects whose colors are to be compared, the metric used to compare them, and which statistic (or statistics) of the various metrics to serve as an overall measure or summarization.

For the evaluation suite, we selected the 170 objects of Vrhel and Trussel. [15] This suite contains spectra of natural and manufactured objects, and contains several samples of human skin of various complexions.

The metric we used is the CIE 1976 L*, a*, b* color difference, ΔE^*_{ab} (hereinafter simply ΔE^*). For each combination of camera, viewing illuminant, and white balancing method, 170 ΔE^*s were generated. They were summarized for each of these combinations of experimental factors as both the arithmetic mean, or average, and as the 90th percentile.

3.4 Camera sensitivities

All camera sensitivities were simulated on the spectral range of 400, 410, …, 700 nanometers. Each sensitivity spectrum was a Gaussian form, parameterized solely by the location of its peak wavelength and its full width at half height. Sixteen Red sensitivity spectra were generated as a factorial cross of peak sensitivity locations of 595, 600, 605, and 610 nanometers and full-width at half heights of 55, 60, 65, and 70 nanometers. These sixteen Red sensitivities were crossed factorially with sixteen Green sensitivities (peak sensitivities at 535, 540, 545, and 550 nanometers, and full-width at half heights of 50, 55, 60, and 65 nanometers) and sixteen Blue sensitivities (peak sensitivities at 440, 445, 450, and 455 nanometers and full-widths at half heights of 40, 45, 50, and 55 nanometers). A total of 4096 camera sensitivities were thus generated and exercised.

3.5 Sharpening transform methodology

Finlayson, *et al.,* provide a variety of techniques for computing the sharpening transform matrix **T**. Although they point out that the different techniques they describe often produce quite similar results, there are some differences. [10] Therefore, we disclose the specifics of the methodology we applied in this study.

We used their “Data-Based Sharpening,” in which the transform is optimized for a specific ensemble of reflectance spectra, together with a taking illuminant, canonical illuminant, and camera sensitivity set. The ensemble of reflectance spectra we used were those of a Macbeth Color Checker, and the taking illuminant we used was the average of a selected subset of the taking illuminants considered in this study. We first used all taking illuminants, only to find that the prominent spectral lines in F2 and F11 tended to swamp the results. Accordingly, the taking illuminant used in computing the sharpening transform was the average of illuminants A, D50, D75, and F7 (all were scaled to possess unit Y tristimulus values). The canonical illuminant was D65.

We further emphasize that the sharpening transform is specific to the set of sensitivity spectra used in a particular camera. Sensitivity spectra which are relatively broadband will tend to have more aggressive sharpening, and those relatively narrow in bandwidth will tend to have weaker sharpening. Hence, we have no single sharpening transform to report.

3.6 Computational environment

The experiment was conducted on a Linux workstation equipped with an AMD Athlon processor running at 1.8 GHz. The Octave language was used to produce the simulated sensitivity spectra, perform the characterization under the canonical/viewing illuminant, exercise the various white balancing methods, and compute the evaluation figures of merit. The results were further summarized (in order to produce Tables 4 and 5 below) using a script written in PERL.

4. EXPERIMENTAL RESULTS

The results for three camera sensitivity combinations appear in Tables 1 through 3. Certainly, some of the sensitivity combinations produce smaller characterization errors (average and 90th percentile ΔE^* under D65, which requires no white balancing) than others; it is our belief that readers would find results for sensitivities which produced smaller characterization errors more interesting than others. Accordingly, we include the results for two such sensitivities. Table 1 contains the results for Red peak at 600nm, bandwidth of 70nm; Green peak at 545nm, bandwidth of 65nm; and Blue peak at 450nm, bandwidth of 50nm. This produced characterization errors of 0.83 (average) and 2.02 (90th percentile).

	Red	Green	Blue
Peak Sensitivities, nm	600	545	450
FWHH Bandwidths, nm	70	65	50

Results under Canonical Illuminant (D65): Average: 0.83; 90th Percentile: 2.02

Taking Illuminant		**Camera RGB**	**XYZ**	**CAT02**	**sRGB**	**Sharp**	**Illum. Dep.**
A	Avg:	2.74	6.56	4.16	22.55	2.32	1.93
	90%:	6.43	10.30	9.20	41.52	4.66	3.85
D50	Avg:	1.17	2.38	1.51	1.96	1.13	0.97
	90%:	2.99	3.88	3.25	3.82	2.16	2.02
D75	Avg:	0.78	1.51	0.90	0.96	0.94	0.86
	90%:	1.71	3.19	1.81	2.26	2.13	2.02
F2	Avg:	4.59	3.82	4.87	8.22	3.70	1.90
	90%:	8.94	6.94	8.72	13.94	7.09	4.35
F7	Avg:	1.55	2.96	2.97	2.97	1.49	1.06
	90%:	3.03	4.53	4.54	4.54	2.92	2.57
F11	Avg:	3.54	5.94	6.70	9.84	2.69	2.29
	90%:	8.52	9.29	13.63	16.18	6.54	4.79

Table 1. Average and 90th Percentile ΔE* for a camera with low characterization error.

In Table 2 are the results for the Prime Wavelengths, which are reportedly near-optimal locations for peak sensitivities. [16] The bandwidths are the maximum for each channel.

	Red	Green	Blue
Peak Sensitivities, nm	605	540	450
FWHH Bandwidths, nm	70	65	55

Results under Canonical Illuminant (D65): Average: 1.04; 90th Percentile: 1.88

Taking Illuminant		**Camera RGB**	**XYZ**	**CAT02**	**sRGB**	**Sharp**	**Illum. Dep.**
A	Avg:	2.90	7.47	4.64	25.74	2.24	1.86
	90%:	6.26	11.33	8.92	48.71	4.43	3.70
D50	Avg:	1.34	2.75	1.75	2.57	1.19	1.06
	90%:	2.81	4.43	3.12	4.98	2.09	2.00
D75	Avg:	1.00	1.59	1.07	1.15	1.20	1.11
	90%:	1.73	3.02	1.62	2.13	2.00	2.01
F2	Avg:	4.72	3.59	4.87	8.18	3.77	1.56
	90%:	8.85	6.77	8.54	14.19	7.01	3.77
F7	Avg:	1.71	2.90	2.90	2.91	1.67	1.15
	90%:	2.89	4.19	4.22	4.24	2.80	2.28
F11	Avg:	3.97	6.28	6.91	10.78	2.92	2.26
	90%:	9.35	9.78	12.78	16.69	7.61	4.73

Table 2. Average and 90th Percentile ΔE* for Prime Wavelength Peak Sensitivities, Broad Bandwidths

In Table 3, the results for a camera whose peak sensitivities occur at the Prime Wavelengths, as in Table 2, but with the narrowest bandwidths considered in this study. Broader bandwidths tend to result in smaller characterization errors, but require more aggressive matrixing, which amplifies noise. Therefore, sensitivities which suffer greater characterization error, but require less aggressive matrixing, may be preferable. This prompted this camera's inclusion in the limited number of results we could present in this paper.

		Red	Green	Blue
Peak Sensitivities, nm		605	540	450
FWHH Bandwidths, nm		55	50	40

Results under Canonical Illuminant (D65): Average: 1.69; 90th Percentile: 4.00

Taking Illuminant		**Camera RGB**	**XYZ**	**CAT02**	**sRGB**	**Sharp**	**Illum. Dep.**
A	Avg:	2.39	6.16	4.94	20.33	1.92	1.63
	90%:	5.71	11.61	12.17	33.40	3.77	3.58
D50	Avg:	1.77	2.45	2.08	2.50	1.64	1.61
	90%:	4.04	4.68	5.10	5.10	4.13	3.90
D75	Avg:	1.68	2.29	1.70	1.67	1.77	1.75
	90%:	4.04	4.45	3.61	3.95	3.75	3.79
F2	Avg:	3.98	3.10	4.13	5.78	3.47	2.22
	90%:	7.30	6.59	7.85	11.66	6.36	5.05
F7	Avg:	2.20	3.38	3.38	3.38	2.19	1.91
	90%:	4.27	5.36	5.38	5.38	4.35	4.18
F11	Avg:	3.65	8.55	9.43	12.04	3.05	2.73
	90%:	9.05	11.89	15.67	18.29	7.67	5.49

Table 3. Average and 90th Percentile ΔE* for Prime Wavelength Peak Sensitivities, Narrow Bandwidths

The overall results, for all 4096 camera simulators, appear in Tables 4 and 5. Table 4 details the number of times a particular white balancing method performed better than its peers, both in terms of average and 90th percentile ΔE*. Table 5 contains the average figures of merit (average and 90th percentile ΔE*) produced by each method, by taking illuminant.

Taking Illuminant		Camera RGB	XYZ	CAT02	sRGB	Sharp	Illum. Dep.
A	Avg:	51	0	0	0	19	4026
	90%:	114	0	0	0	242	3740
D50	Avg:	177	2	0	8	411	3498
	90%:	100	8	0	11	810	3167
D75	Avg:	656	6	2200	828	57	349
	90%:	666	1	3098	214	42	75
F2	Avg:	0	0	0	0	0	4096
	90%:	0	0	0	0	1	4095
F7	Avg:	451	0	0	0	65	3850
	90%:	1714	10	7	7	1	4095
F11	Avg:	0	0	0	0	27	4069
	90%:	0	0	0	0	0	4096
All Illums	Avg:	1335	8	2200	836	597	19618
	90%:	2594	19	3105	232	2063	17142

Table 4. Number of times each method produced the “best” results, based on Average and 90th Percentile ΔE*.

Taking Illuminant		Camera RGB	XYZ	CAT02	sRGB	Sharp	Illum. Dep.
A	Avg:	2.96	8.29	6.39	24.06	2.55	2.12
	90%:	6.12	13.55	12.26	44.99	4.96	4.20
D50	Avg:	2.03	3.35	2.74	2.97	1.95	1.86
	90%:	4.14	5.97	5.36	5.39	3.87	3.68
D75	Avg:	1.86	2.21	1.82	1.93	1.94	1.92
	90%:	3.63	4.25	3.39	3.82	3.78	3.79
F2	Avg:	4.13	5.88	6.98	8.75	3.52	2.20
	90%:	7.56	9.68	10.81	14.25	6.57	4.53
F7	Avg:	2.17	3.55	3.56	3.56	2.18	1.95
	90%:	4.01	5.84	5.86	5.86	4.05	3.98
F11	Avg:	3.69	7.31	8.03	10.73	3.10	2.64
	90%:	8.36	11.26	14.59	17.36	7.29	5.31
All Illums	Avg:	2.81	5.10	4.92	8.67	2.54	2.11
	90%:	5.64	8.43	8.71	15.28	7.63	4.25

Table 5. Average and 90th Percentile ΔE*s averaged across all 4096 cameras.

5. DISCUSSION

Overall, the illuminant-dependent characterization produced the lowest average ΔE*, which implies the highest degree of color constancy. The average ΔE* for white balancing using sharpened sensors, which was second lowest, was 2.54, which is 20 percent higher than the 2.11 obtained using the illuminant-dependent characterization. Further, the balancing in native camera RGB produced averages which were 33 percent higher than the illuminant-dependent technique. The illuminant-dependent method produced the best results, based on average ΔE*, an overwhelming 80 percent of the time over all illuminants (96 percent of the time for fluorescent illuminants and 64 percent of the time for non-fluorescent illuminants). Based on this, one may argue that images captured under fluorescent illuminants tend to have a greater need for illuminant-dependent white balancing than those captured under incandescent lighting or normal daylight.

Based on the 90th percentile, the illuminant-dependent method produced results which could be characterized as visibly better than the next-best method considered (which was scaling in native camera RGB), whose average 90th percentile was also 33 percent higher. The sharpened sensors performed here third best, produced 90th percentile ΔE^*s which were 80 percent larger than those produced by the illuminant-dependent technique.

5.1 "Easy" and "difficult" taking illuminants

Based on the traditional method of white balancing, in native linear camera RGB, the illuminants may be classified as either "easy" or "difficult" in terms of how well the white balancing produces color constant results. Based on the 90th percentile criterion, and Stamm's rule of thumb of a ΔE^* of 6, [17] taking illuminants D50, D75, and F7 are classified as "easy," and A, F2, and F11 are classified as "difficult." It is worthwhile to note that the correlated color temperatures of all taking illuminants for which white balancing is "easy" are all fairly close to the 6504 Kelvins of the canonical/viewing illuminant. However, the correlated color temperatures of F2 (4230 K) and F11 (4000 K) were both closer to that of the canonical illuminant than Illuminant A (2853 K), yet both had higher average 90th percentile ΔE^*s for balancing performed in camera RGB. Both F2 and F11 (in particular) have extremely prominent spikes from mercury line emission and phosphor emission. F7 also exhibits spikes, but it possesses a much greater amount of continuous spectrum than either F2 or F11, and, consequently, less prominent spikes.

The only method which performed consistently well for both the "easy" and "difficult" taking illuminants used in this study was the Illuminant-Dependent Characterization. Sharpened sensors tended to perform as well as Camera RGB for the "easy" taking illuminants, and better for the "difficult" ones. The other methods performed worse than Camera RGB for the "difficult" illuminants, and thus cannot be considered favorable alternatives.

5.2 Balancing when chromaticity coordinates of taking and viewing illuminants identical

As expected, the results for taking illuminant F7, in terms of both average and 90th percentile ΔE^*, of the three methods which use the XYZ tristimulus values of the white point (XYZ, CAM02, BT.709), are essentially identical. This is because we used the actual XYZ tristimulus values of a perfect diffuser under F7, which are extremely close to those of D65, making the effective white balancing matrix essentially an identity matrix.

5.3 Poor performance of balancing in monitor phosphor space

It is significant that the balancing in the ITU Rec BT.709 primaries space produced the poorest results, with 90th percentiles approximately 3.6 times as large as those produced by the illuminant-dependent method. Most monitors use primaries which match or are quite close to these. While monitor display subsystems normally involve a non-linear electro-optical transfer function (EOTF), white balancing results will be independent of the presence of this EOTF provided it is a power-law (strict gamma correction). CRT-based display subsystems tend to have EOTFs which conform closely to a power-law relationship except in the extreme shadows. This implies that balancing in monitor RGB space will produce poor results, and people should refrain from performing all but the smallest adjustments in white balance in this space, even though it is popular to do so.

The -709 primaries probably produced poor results because they oversharpen. When oversharpening occurs, side lobes become prominent, and the transformed sensitivities bear less and less resemblance to Dirac delta functions, rather than more and more. Sharpening should be applied judiciously, using techniques such as those described in Finlayson, *et al.* [10]

6. CONCLUSIONS

White balancing in a typical monitor RGB space tends to produce large degrees of color inconstancy, and should be avoided.

White balancing using an illuminant-dependent characterization produced consistently good results. This technique is predicated on identification of the illuminant under which the image was captured, and possession of a pre-computed characterization matrix for that, or a similar, illuminant.

White balancing in native camera RGB produces good results for illuminants which are close in correlated color temperature to the viewing illuminant and do not possess large spikes, but not as good as using illuminant-dependent characterizations. White balancing in spectrally-sharpened RGB is better on average than camera RGB, but tends to produce slightly larger outlier errors.

White balancing is certainly more complex than accounting for differences in the chromaticity coordinates of the taking and viewing illuminants, because slightly better results can be obtained when other factors are accounted for. These are largely accounted for when illuminant-dependent characterizations are employed.

7. LITERATURE CITED

1. Brian Funt, Vlad Cardei, and Kobus Barnard, Learning color constancy. *Proceedings of the IST/SID 4th Color Imaging Conference,* 1996, p 58-60.

2. Graham Finlayson, Paul M. Hubel, and Steven Hordley, Color by correlation. *Proceedings of the IST/SID 5th Color Imaging Conference,* 1997, p 6-11.

3. Graham D Finlayson, Steven Hordley, and Paul M Hubel, Unifying color constancy. *Proceedings of the IST/SID 7th Color Imaging Conference,* 1999, p 120 - 125.

4. Shoji Tominaga, Atsushi Ishida, and Brian Wandell, Further research on the sensor correlation method for scene illuminant classification. *Proceedings of the IST/SID 8th Color Imaging Conference,* 2000, p 189 - 194.

5. Jeffrey M DiCarlo and Brian A Wandell, Illuminant estimation: beyond the bases. *Proceedings of the IST/SID 8th Color Imaging Conference,* 2000, p 97 - 101.

6. S. Tominaga and S. Okamoto, Color Temperature Estimation of Fluorescent Scene Illumination. *Final Program and Proceedings of IS&T and SID Eleventh Color Imaging Conference,* 2003, p 82-87.

7. F. Ciurea and B. Funt, A Large Image Database for Color Constancy Research. *Final Program and Proceedings of IS&T and SID Eleventh Color Imaging Conference,* 2003, p 160-164.

8. Matthew Anderson, Ricardo Motta, Srinivasan Chandrasekar, and Michael Stokes, Proposal for a Standard Default Color Space for the Internet—sRGB. *Proceedings of the IST/SID 4th Color Imaging Conference,* 1996, p.238 - 245.

9. CIE TC8-01, *A color appearance model for color management systems: CIECAM02.* Vienna: Bureau Central of the CIE, in preparation.

10. Graham D Finlayson, Mark S Drew, and Brian V Funt, Spectral sharpening: sensor transformations for improved color constancy. *Journal of the Optical Society of America A,* 1994. 11:5, p. 1553-1563.

11. G. Woolfe, K. Spaulding, and E. Giorgianni, Optimal Color Spaces for Balancing Digital Color Images. *Final Program and Proceedings of IS&T and SID Eleventh Color Imaging Conference,* 2003, p 6-70.

12. Don Williams, Color inconstancy index: a proposal and case study for consumer digital cameras. *Proceedings of the 2002 IS&T PICS Conference,* p. 181 - 184.

13. J A Stephen Viggiano, Minimal-knowledge assumptions in digital still camera characterization. I: uniform distribution, Toeplitz correlation. *Proceedings of the IST/SID 9th Color Imaging Conference,* 2001, p 332-336.

14. J A Stephen Viggiano, Minimal knowledge assumptions in digital still camera characterization. II: non-uniform distribution. *Proceedings of the 2003 IS&T PICS Conference,* p. 435-440.

15. Available online at ftp://ftp.eos.ncsu.edu/pub/spectra/objects.spectra.10.Z. Visited 2003-12-15.

16. M D Brill, G D Finlayson, P M Hubel, and W A Thornton, Prime colors and color imaging. *Proceedings of the IST/SID 6th Color Imaging Conference,* 1998, p 33 - 42.

17. Scott Stamm, An investigation of color tolerance. *1981 TAGA Proceedings,* p 156-173.

8. APPENDIX

For this study, the matrix $\mathbf{M}_{CAT02}$ transforms a row vector containing XYZ tristimulus values into another row vector containing the CAT02 RGB values. It is the transpose of the CAT02 matrix cited in the CIECAM-02 document:

$$\mathbf{M}_{CAT02} = \begin{bmatrix} 0.7328 & -0.7036 & 0.0030 \\ 0.4296 & 1.6975 & 0.0136 \\ -0.1624 & 0.0061 & 0.9834 \end{bmatrix}$$

In this study, the matrix $\mathbf{M}_{709}$ is used to transform a row of XYZ tristimulus values into a row of RGB relative luminances for phosphors which conform to those specified in ITU Rec BT.709, such that the D65 white point, 2 degree observer, transforms into a vector of three ones:

$$\mathbf{M}_{709} = \begin{bmatrix} 3.240178 & -0.969303 & 0.055644 \\ -1.537008 & 1.876083 & -0.204027 \\ -0.498489 & 0.041558 & 1.057233 \end{bmatrix}$$

Both M_{CAT02} and M_{709} transform the tristimulus values of a perfect diffuser under D65 to unit RGB, but the elements of M_{709} are larger in magnitude. This implies that the transform defined by the M_{709} matrix is sharper than that defined by M_{CAT02}, and its primaries will be tighter in the (x, y) chromaticity diagram.

9. ACKNOWLEDGEMENTS

The author thanks Dr Mitchell Rosen of RIT's Center of Imaging Science for several helpful suggestions, and Nathan Moroney of HP Labs for providing information on the CAT02 matrix.