Preferred Color Spaces for White Balancing

Feng Xiao[*a], Joyce E. Farrell[b], Jeffrey M. DiCarlo[c] and Brian A. Wandell[a,b,c]
[a]Department of Psychology, Stanford University
[b]ImagEval Inc Menlo Park, CA 94025
[c]Department of Electrical Engineering, Stanford University

ABSTRACT

When rendering photographs, it is important to preserve the gray tones despite variations in the ambient illumination. When the illuminant is known, white balancing that preserves gray tones can be performed in many different color spaces; the choice of color space influences the renderings of other colors. In this behavioral study, we ask whether users have a preference for the color space where white balancing is performed. Subjects compared images using a white balancing transformation that preserved gray tones, but the transformation was applied in one of the four different color spaces: XYZ, Bradford, a camera sensor RGB and the sharpened RGB color space. We used six scenes types (four portraits, fruit, and toys) acquired under three calibrated illumination environments (fluorescent, tungsten, and flash). For all subjects, transformations applied in XYZ and sharpened RGB were preferred to those applied in Bradford and device color space.

Keywords: white balance, color spaces, color preference, psychophysics.

1. INTRODUCTION

The digital imaging color-processing pipeline is designed to reproduce a visually pleasing and reasonably accurate color image. Several processing steps in the pipeline are intended to parallel processing in the human visual pathways.[1] For example, image sensor signals are frequently transformed to a calibrated color space in order to make the pipeline data approximately consistent with the human encoding of wavelength. The use of calibrated color spaces is one way in which the digital imaging pipeline is coordinated with human vision.

White balancing is a second way that the color-processing pipeline parallels human visual processing. A century of experiments have shown that the visual pathways change their color sensitivity as ambient lighting conditions vary. For example, in the presence of a powerful long-wavelength ambient illumination, the sensitivity of the long-wavelength cones is reduced.[2-4] The reproduction and original are often viewed under different ambient lighting, so that visual sensitivity is quite different between the acquisition and rendering contexts.[5] White balancing algorithms attempt to account for these changes in human visual sensitivity under different ambient illuminant conditions.

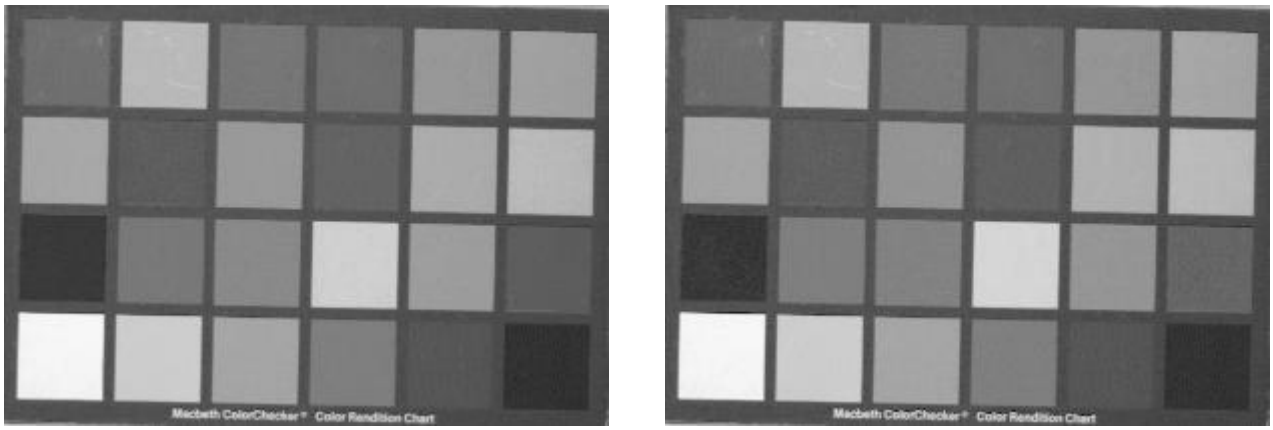
Some form of white balancing is implemented in most color processing pipelines.[1,6,7] Our understanding of human visual white balancing is incomplete,[8] and there are no quantitative models that predict accurately which implementations are most successful. In this paper we describe behavioral experiments to evaluate white balancing options for digital photography. Specifically, (a) we introduce psychophysical methods for measuring subjects' preferences between white balancing options and (b) we measure user preferences for color balancing operations performed in several different color spaces. We focus our evaluation on white balancing methods that are appropriate for the color-processing pipeline of digital cameras in devices with limited computational power.

1.1 White balancing principles

A complete white balancing algorithm comprises two parts. First, the system uses the image data and statistical base rate information to estimate the current ambient lighting conditions. Second, using this estimate, the system converts the acquired sensor data to a format that is appropriate for the targeted rendering environment.

[*] Correspondence: Email: xiao@white.stanford.edu; Telephone: (650) 725-1255.

Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications IV, Morley M. Blouke, Nitin Sampat, Ricardo J. Motta, Editors, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5017 (2003) © 2003 SPIE-IS&T · 0277-786X/03/$15.00



(a) (b)

Figure 1. Macbeth Color Charts white-balanced in CIE (a) XYZ color space and (b) sensor RGB color space.

Many algorithms have been developed for the first part of white balancing: estimating the ambient illumination. Barnard's doctoral thesis [9] provides a good review of these algorithms. The second part of the algorithm, however, is often implemented using a simple computation: the acquired data are linearly transformed from sensor values into a new coordinate frame.[10-12] In many cases, particularly for devices with limited computational power (PDA or cell phone cameras), the transformation comprises a fixed general linear transformation followed by a variable diagonal transformation that scales the resulting coordinates. The three parameters in the diagonal transformation are set based on the estimated ambient illuminant. We refer to this architecture as the generalized diagonal transformation (GDT) method.[13] We study the GDT method for two reasons: First, this method parallels the well-known von Kries coefficient law of human visual system.[3] Second, many illuminant estimate applications derive only the RGB values of the white point (hence the name white balancing). For these methods, a diagonal matrix that preserves the white point can be easily derived from these RGB values.

1.2 White balancing using generalized diagonal transformations

The generalized diagonal transformation white-balancing model can be expressed in equation form as:

$$RGB_{out} = F_{out} \times D \times F_{in} \times RGB_{in} \tag{1}$$

RGBin: 3 x 1 linear input (camera) RGB values
F_{in}: 3 x 3 matrix (fixed) transforming RGB to an intermediate color space
D: 3 x 3 diagonal matrix whose values vary with illuminant
F_{out}: 3 x 3 matrix (fixed) transforming intermediate color space value to display color space value
RGB_{out}: 3 x 1 linear output RGB values

Each fixed matrix, F_{in}, transforms the input (camera) RGB space to an intermediate color space independent of illumination change. The diagonal matrix D is chosen based on the specific image data to guarantee that the estimated white point will be rendered as white. Another fixed matrix, F_{out}, maps the scaled values into linear output RGB values. We describe the linear steps here; additional steps may be needed to correct for the non-linearity (gamma-correction) of most output devices (such as the industrial standard sRGB monitor).[14]

Table 1: CIELAB error for the 24 Macbeth Color Charts in the two images in Figure 1.

6.1145	6.8777	8.4035	4.8370	3.0919	7.4780
13.5337	11.8102	11.1386	4.5647	9.7461	12.3994
13.3608	8.6204	17.4100	9.4425	9.0038	9.2261
0.5312	0.7179	0.8895	0.9768	1.5068	5.8923

When the illumination estimate is accurate, the GDT calculation guarantees that the scene white point is mapped into the display white point. The GDT calculation does not guaranteed the accuracy of other colors. In the example shown in Figure 1 and Table 1, achromatic targets white balanced in different color spaces have color differences at the noise level; other color targets have much higher color differences.

There are an infinite number of possible choices for fixed matrices F_{in} and F_{out}. Each pair corresponds to a different choice of an intermediate color space. In this study we examine four pairs that correspond to meaningful intermediate color spaces:

- 1931 CIEXYZ color space (Figure 2c)

- Bradford RGB color space (Figure 2d): Bradford RGB color space [15] is widely used in color appearance models (CIECAM97s, CIECAM02) for modeling illumination changes. Color-appearance models have been used to guide digital photography where the goal is often to create pleasing color reproduction of the original scenes. This type of reproduction is often called "preferred color reproduction" [16] or "color-preference reproduction".[5]

- Sensor RGB color space (Figure 2a): Simplicity is the main advantage this color space. The adjustment of analog gains of each channel in camera sensors can be seen as hardware implementation of the scaling operation.

- Sharpened sensor RGB color space (Figure 2b): Sensor sharpening [17] may be attractive for this application; system responses based on narrowband sensors follow the diagonal transformation, unlike systems based on broadband sensors. Sensor sharpening therefore appears to be a sensible transformation before application of a diagonal transformation.

The question we address in this paper can now be put simply: For white balancing using a generalized diagonal transformation architecture, which fixed linear transformation do people prefer? We have developed a behavioral testing methodology to answer this question. This methodology can be applied to many different questions about the user preferences in image rendering.

2. METHODS

2.1 Camera calibration

The experimental images were obtained using a scientific-grade color CCD camera (QImaging Regita 1300, 1024 by 1280 Bayer pattern at 10 bits/pixel).

2.2 Scene setup

The experimental images were obtained in a studio containing three common types of controlled light sources: fluorescent office lighting with measured correlated color temperature (CCT) 3035K, tungsten with CCT 2992K and flash with CCT 5626K (flash is very close to CIE daylight at CCT 5626K). The experimental scenes included four portraits: two Caucasian (male and female) and two Asian (male and female) individuals. We also used a still-life scene containing fruit and a scene containing colorful plastic toys. The fruit scene contains familiar objects; the toy scene contains unfamiliar objects with very saturated colors. Samples of the images are shown in Figure 3. Scene white points were measured using a halon surface placed in the scene under each of the illuminants (Table 2). A total of eighteen test images were captured (6 scenes under 3 illuminants).

Table 2: RGB of scene white points

RGB \ Illumination	Fluorescent	Tungsten	Flash
Red	0.7777	1.0000	0.7691
Green	1.0000	0.7374	1.0000
Blue	0.4143	0.3369	0.7432

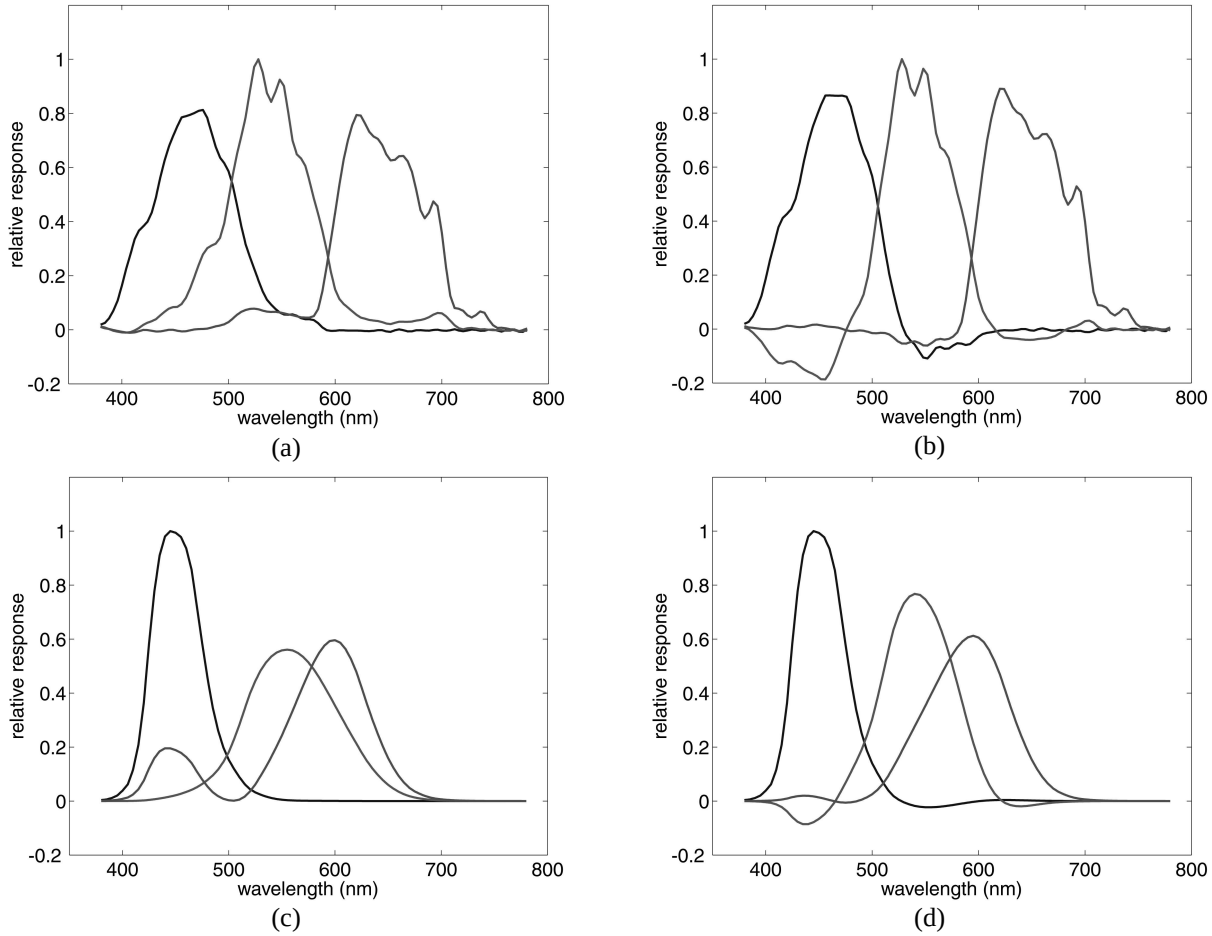


Figure 2. Spectral responsivity curves of (a) the camera sensor (b) the sharpened sensor (c) CIE-XYZ and (d) Bradford RGB.

2.3 Stimulus preparation

Each of the eighteen captured images (six scenes by three lights) was rendered into four different display pictures by white balancing in four different intermediate color spaces: 1931 CIEXYZ, Bradford, sensor RGB and sharpened sensor RGB. So there are totally 72 display pictures. The appendix provides detailed information about how to derive the pair of fixed matrixes, F_{in} and F_{out}, corresponding to each intermediate color space. Once we define F_{in} and F_{out}, the diagonal matrix D for each of the sixty-four pictures can be derived as:

$$\begin{bmatrix} \frac{R_{display}}{R_{scene}} & 0 & 0 \\ 0 & \frac{G_{display}}{G_{scene}} & 0 \\ 0 & 0 & \frac{B_{display}}{B_{scene}} \end{bmatrix} \tag{2}$$

Where $RGB_{display}$ and RGB_{scene} are the RGB values of the display and scene white point in intermediate color space respectively. This choice of diagonal parameters guarantees that the scene white point is mapped into the display white point.

Figure 3. Six scenes used in this study

The output device is a calibrated CRT display, which is very close to the specification of the industrial standard sRGB display. Therefore, the linear RGB_{out} values from Equation (1) were normalized and transformed to display values according to the sRGB specifications.[14]

2.4 Behavioral methods

The experiments were controlled in Matlab, using extensions of the Psychophysics Toolbox.[18,19] Six observers (four females and two males) with normal color vision participated in this study. Their ages range from 22 to 50 years old. All subjects were familiar with digital images and three had previously participated in psychophysical experiments. Subjects sat approximately 50 cm from the surface of the display. Each image subtended a viewing angle of 17° x 13.4°. In every trial, two images were displayed side by side with a 0.5° horizontal gap separating the images. The background was uniformly gray. The room light was off and the maximum luminance of the monitor was 80 cd/m^2.

Each of the eighteen camera images had four different color balanced versions. In each experimental trial, two of these four versions were chosen randomly and presented side by side on the calibrated CRT monitor. The subjects were instructed to indicate which image they preferred. Each pair was shown 16 times in the whole session and the order of presentation was randomized across subjects. Each subject made to 1728 preference judgments. Subjects were given as much time as they needed for each trial; subjects were permitted to take breaks as they needed as well. Most subjects finished the session in 80 minutes.

3. RESULTS

3.1 Analysis methods

We analyzed the data using two methods: one-dimensional scaling and rank-order analysis. One-dimensional scaling uses the preference judgments between stimulus pairs to generate estimates of image quality on an interval scale. This method requires many judgments and we performed analysis based on the pooled data across the subjects. Rank-order analysis derives an ordinal relationship between stimuli. With this method, it is possible to determine whether the

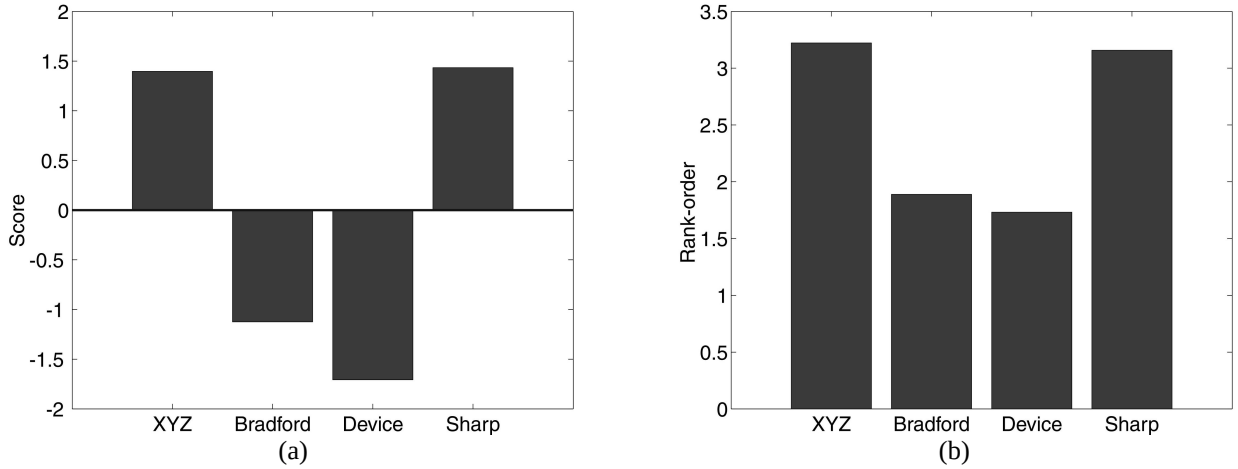


Figure 4. The preference is data are analyzed using (a) one-dimensional scaling and (b) rank-order methods.

preference ordering across stimuli is consistent across subjects.

- One-dimensional scaling method

We assume that the four pictures in each group can be positioned on a one dimensional quality line. We estimate the distance between the two pictures in units of the standard deviation with the inverse cumulative-normal function (z-score):

$$d_{i,j} = \sqrt{2}z\left(\frac{C_{i,j}}{C_{i,j} + C_{j,i}}\right) \tag{3}$$

Where $C_{i,j}$ is the total number of times that picture i was chosen over sample j for all subjects.

A final preference position (score) for picture i is assigned as the mean distance between that picture and the three other pictures in its group:[20] [21]

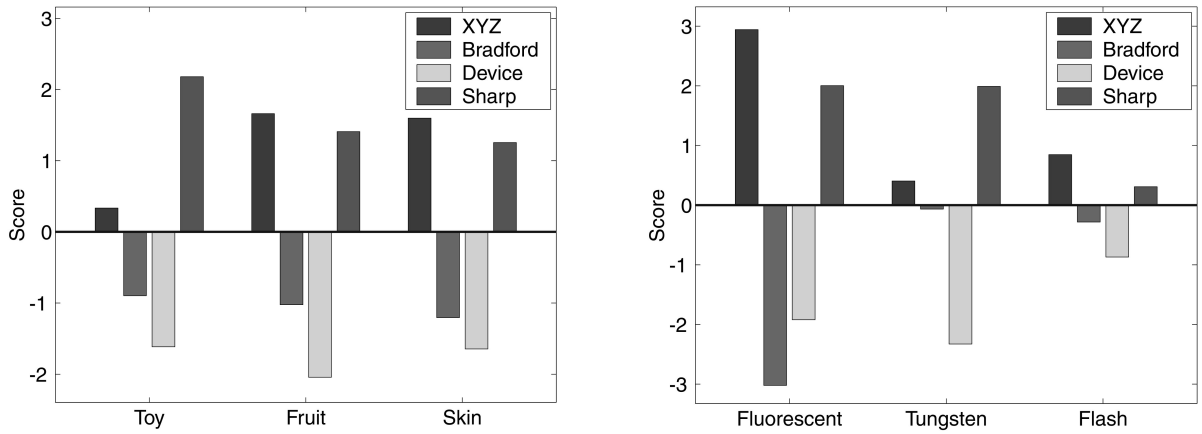
$$q_i = \frac{\sum_{j=1, j\neq i}^{N} d_{i,j}}{N} \tag{4}$$

- Rank-order analysis method

In rank-order analysis, the quality of each picture in each group is assigned a number from 1 to 4 based on the total number of times that it was chosen over other three pictures. To be consistent with the one-dimensional scaling analysis, we define 4 as the highest quality and 1 to be lowest quality. If there is a tie between pictures, both receive a score that is the mean of their original scores. Rank-order analysis was applied to each subject and we analyzed the consistency across subjects using the Kendall coefficient of concordance method.[22]

3.2 Overall color space preference

The one-dimensional scaling method and the rank-order methods lead to very similar conclusions (correlation coefficient of 0.9975): among all subjects white balancing in XYZ and sharpened RGB were preferred to Bradford and sensor RGB color spaces. The preference scores for XYZ and sharpened RGB were two standard deviations higher than the other spaces. The difference between XYZ and sharpened RGB color spaces were not significant, Bradford color space holds a slight edge against sensor RGB color space.



(a) (b)

Figure 5. Color space preference for (a) different types of scenes and (b) different illuminants.

3.3 Interaction between color space preference and scene type or illuminants

Figure 5 shows the preferred color space for white balancing grouped either according to scene or illuminant type. The score is consistent across both scenes and illuminants. The XYZ and sharpened RGB color spaces are preferred; the XYZ color space was most often preferred for images that include familiar objects (fruit or people) or fluorescent and daylight illumination.

3.4 Consistency across subjects

The degree of agreement for the eighteen groups is shown in Table 3. The Kendall coefficient of concordance ranges from 0.4667 to 0.9444 (1 for perfect agreement and 0 for no-agreement) for 15 out of these 18 image groups. The other three groups (two portraits and the toy) have a much lower agreement among subjects.

Table 3: Kendall coefficient of concordance across subjects

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Coeff	0.8	0.6	0.94	0.94	0.9	0.6	0.6	0.47	0.63	0.47	0.48	0.18	0.3	0.67	0.47	0.68	0.52	0.06

4. DISCUSSION

For this study, we found white balancing in XYZ and sharpened RGB color space was most often preferred to white balancing in Bradford or camera RGB color space. Such a clear preference is somewhat surprising when we consider the small difference in the spectral curves when comparing sensor RGB and sharpened sensor RGB color space (Figure 3 a and b): Such small differences result in large preference differences.

To explore the generality of this conclusion, we need to consider the several issues:

- Non-colorimetric property of camera sensor

Because the camera is non-colorimetric, there is residual error transforming the camera RGB into colorimetric color spaces. The amount of error depends on the optimization method used; more colorimetric sensors would minimize this error. Obtaining images using a hyper-spectral camera would overcome this problem and might also be a useful tool for evaluating a wide range of non-colorimetric camera sensors.

- Limited number of scenes and illuminants

This study investigated a small subset of possible scenes and illuminants. Even with this small set, we find that preferences depend on both the image content. We need to explore how well these results generalize to other types of scenes.

- Perfect white balancing might not be preferred

In this study, we focused on color balancing methods that map the scene white point into the display white point. However, incomplete color adaptation would map the scene white point differently.[5] An imperfect white point might be preferred to the GDT color balancing.

- Other post-processing

The generalized diagonal transformation scheme is a very simplified model that is suitable for devices with limited computational power. For devices with more computational power, more complicated post-processing might be involved and similar psychophysical procedures can be performed for these devices.

5. CONCLUSION

We presented a generalized diagonal transformation method for white balancing color images. This method is useful in devices with limited computational power. We ask which color space is preferred when performing a diagonal white balancing correction. To answer this question, we introduced psychophysical methods for measuring subjects' preferences and for comparing white balancing operations performed in four different color spaces. For the specific camera and six scenes, we found that white balancing in XYZ and sharpened RGB color spaces were preferred by a significant margin to white balancing in Bradford and device color space. The significant preference differences we observe are surprising because the spectral difference between, say, the sharpened space and sensor RGB spaces is rather small (see Figure 2ab). It would be useful to understand how such large preference differences can arise despite only modest differences in the white balancing color space.

ACKNOWLEDGEMENTS

We thank Hagit Hel-or, Xuemei Zhang, Ulrich Tobias Barnhoefer, Peter Catrysse for their help and comments on the paper. We also thank numerous people for their willing to be pictured. This work was supported by the Media-X program and Programmable Digital Camera project at Stanford University.

6. REFERENCES

1. B. A. Wandell, A. El Gamal, and B. Girod, Common Principles of Image Acquisition Systems and Biological Vision. Proc. of the IEEE, **90**, 5-17, 2002.
2. H. von Helmholtz, *Treatise on Physiological Optics*, Optical Society of America, Rochester, NY, 1911.
3. J. von Kries, *Chromatic adaptation*,In *D.L. MacAdam*, (ed.) *Sources of Color Science*, The MIT Press, Cambridge, MA, 120-127,1970.
4. B. A. Wandell, *Foundations of Vision*, Sinauer Associates, 1995.
5. M. D. Fairchild, *Color Appearance Models*, Addison-Wesley Longman, Inc, 1997.
6. Hsien-Che Lee and Robert M. Goodwin, Colors as Seen by Humans and Machines. Recent Progress in Color Science, 18-22, 1997.
7. Janne Laine and Hannu Saarelma, Illumination-Based Color Balance Adjustments. *Eighth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications*, 202-206, 2000.
8. J. M. Kraft and D. H. Brainard, Mechanisms of color constancy under nearly natural viewing. Proc. Natl. Acad. Sci. USA, **96**, 307-312, 1999.
9. Kobus Barnard, PhD thesis, *Practical Colour Constancy*, Simon Fraser University, 1999.
10. B.A. Wandell and J.E. Farrell, Water into wine: Converting scanner RGB to tristimulus XYZ. *SPIE: Device-Independent Color Imaging and Imaging Systems Integration*, 1909, 92-101, 1993.
11. P. M. Hubel, J. Holm, G. D. Finlayson, and M. S. Drew, Matrix Calculations for Digital Photograpphy. *Fifth IS&T/SID Color Imaging Conference*, 105-111, 1997.
12. Jack Holm, Adjusting for the Scene Adopted White. *IS&T's PICS Conference*, 158-162, 1999.
13. G. D. Finlayson, M. S. Drew, and B. V. Funt, Color constancy: generalized diagonal transforms suffice. J. Opt. Soc. Am. A, **11**, 3011-3019, 1994.

14. IEC, Multimedia systems and equipment - Colour measurement and management Part 2-1: Colour management - Default RGB colour space - sRGB. 61966-2-1, 1999.
15. K. M. Lam, PhD thesis, *Metamerism and colour constancy*, University of Bradford, 1985.
16. R.W.G. Hunt, *The Reproduction of Color*, Fountain Press, 1995.
17. G. D. Finlayson, M. S. Drew, and B. V. Funt, Spectral Sharpening: Sensor Transformations for Improved Color. J. Opt. Soc. Am. A, **11**, 1553-1563, 1994.
18. D.H. Brainard, The Psychophysics Toolbox. Spatial Vision, **10**, 433-436, 1997.
19. D.G. Pelli, The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, **10**, 437-442, 1997.
20. F. Mosteller, Remarks on the method of paired comparisons I: The least squares solution assuming equal standard deviations and equal corrections. Psychometrica, **16**, 3-9, 1951.
21. D.A. Silverstein and J.E. Farrell, Quantifying perceptual image quality?*The Society for Imaging Science and Technology, Image processing quality and capture*, 1998.
22. S. Siegel, *Nonparametric statistics for the behavioral sciences*, McGraw-Hill, New York, 1956.

7. APPENDIX

This appendix describes how to derive the pair of fixed matrixes, F_{in} and F_{out}, for each of the four intermediate color spaces: 1931 CIEXYZ, Bradford RGB, sensor RGB and sharpened sensor RGB. The ISO working draft 17321-2 describes three main ways to resolve the optimal transformation matrix from one color space to another color space. In this study, we used a method similar to the second method in this ISO working draft. Specifically, we chose the Macbeth Color Charts (MCCs) as the characterization data patches and D65 as the characterization illuminant.

As an example, for the CIEXYZ color space, matrix F_{in} was chosen as the optimal matrix that minimizes the XYZ error when transforming camera RGB values of the MCCs under D65 ($RGB_{macbeth}$) to their corresponding XYZ values $XYZ_{macbeth}$. That is:

$$F_{in} = XYZ_{macbeth} * \text{Pseudoinverse}\ (RGB_{macbeth}).$$

Matrix F_{out} is the matrix from XYZ to linear sRGB values.

Similar procedures were carried for other three intermediate color spaces (Table 4). One additional step was needed to find the optimal matrix from sensor RGB color space to sharpened sensor RGB color space: We implemented the algorithm described in Finlayson's paper [17] and the three targeted narrow wavelength bands were chosen to be 10nm wide and centered on the peak wavelength of each channel's sensitivity.

Table 4: The pair of matrix for each intermediate color space.

Color Space \ Matrix	F_{in}			F_{out}		
XYZ	0.7002	0.5883	0.0222	3.2406	-1.5372	-0.4986
	0.2450	1.2027	-0.2345	-0.9689	1.8758	0.0415
	0.0989	-0.3736	1.7808	0.0557	-0.2040	1.0570
Bradford	0.6760	0.9073	-0.3300	2.5382	-1.2933	-0.0403
	-0.1018	1.6058	-0.3531	-0.1457	1.1165	-0.0223
	0.1123	-0.4441	1.8504	-0.0422	-0.0716	1.0225
Camera RGB	1	0	0	1.8430	0.2440	-0.4555
	0	1	0	-0.2146	1.6705	-0.3875
	0	0	1	0.0936	-0.6075	1.9313
Sharpened Camera RGB	0.9911	-0.1287	0.0353	1.9044	0.4508	-0.3826
	-0.0909	0.9420	-0.3231	-0.0532	1.8018	0.1995
	0.0092	-0.1686	0.9856	0.0491	-0.3057	1.8576