DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-219

Effect of Capture Illumination on Preferred White Point for Camera Automatic White Balance

Ben Bodner, Yixuan Wang, Susan Farnand
Rochester Institute of Technology, Munsell Color Science Laboratory
Rochester, New York

Abstract

Cell phone camera manufacturers use a variety of illuminant estimation methods to automatically white balance images, which often produce very different results. The goal of this study is to estimate aim white point as a function of the light source used during image capture. The results of this study may support a standard aim point for automatic white balance algorithms. Images with eight different scene light sources and five scene contents were captured and displayed using color correction and color management techniques. A Method of Adjustment experiment was conducted to determine the average preferred white-point and variation in preference between observers for images captured under different light sources and with different scene contents. This research provides information about how the preferred white-point for balancing an image changes dependent on the actual scene light source and the scene content.

1 Introduction

The human visual system is incredibly adept at identifying changes in illumination and adapting to maintain the appearance of objects. [1] This ability is called color constancy and is an important step in modeling accurate color appearance. [2] In photography, mimicking the adaptation of the visual system through a process called white balancing is key to producing acceptable images. To white balance an image, the three color channels of the camera response are each scaled to account for the color of the light source. Virtually all consumer digital cameras feature automatic white balance, which uses image information to estimate the illumination and balance the image accordingly.

Automatic white balance can be accomplished through many different methods, from simple averaging to more complex statistical or learning based algorithms. [3] [4] Because of the importance of automatic white balance in manufacturing digital cameras, a standard for evaluating the quality of white balance is being developed by those in the cell phone camera industry. Cameras from various manufacturers often produce very different results when photographing the same scene. The ability of an algorithm to accurately estimate illumination from an image can easily be determined, but images are rarely balanced precisely on the light source in the scene.

There are numerous factors at work in the human visual system that explain why the goal of automatic white balance should not be based solely on accurate illuminant estimation. Color constancy is often imperfect, with the visual system only partially adapting to scene illumination. This incomplete adaptation is a benefit, as the color and intensity of illumination can provide important information about the environment. [2] Further, the human visual system doesn't adapt the same way to emissive displays as it does to a hardcopy print or to a real scene, and changing the contents of a displayed image can impact the degree of adaptation without changing the color of illumination [5], so predicting the appearance of a digital image displayed on a screen is not straightforward. It has also been demonstrated that for many types of scene content, perceptually accurate color reproduction is not the best way to create aesthetically pleasing images. For instance, people typically prefer images with bluer skies and more saturated skin tones over colorimetrically accurate renditions. [6]

Figure 1. *Identical scenes captured with two popular cell phone cameras produce different color balance.*

This study discussed here is part of an effort to establish perceptually-based goals for automatic white balancing, dependent on scene illumination and scene contents. An image set was captured in accordance with the requests of the Cell Phone Image Quality group (CPIQ), a collection of companies involved in cell phone manufacturing or related imaging software. The image set contained five scenes each captured under eight different light sources, including tungsten, fluorescent, LED, and natural outdoor lighting. A series of three experiments were proposed, each with a unique purpose. The first was a method of adjustment test to find general trends in preferred appearance throughout the image set. The second is a paired comparison test to help narrow down these trends to a specific white balance aim for each illuminant/scene combination. The third would use a quality ruler to evaluate the quality falloff as white balance deviated from the aim.

This paper covers the decisions made in designing and capturing the image set, along with the results of the first experiment, in which observers manually adjusted the white balance of each image in a set so that its color looked best, in their opinion. White balance was adjusted through a graphical user interface with sliders that manipulated the white balance illuminant in CIEL*a*b* perceptual space. The range of preferences found in this experiment were used to create a paired comparison experiment to establish white balance aims.

2 Related Work

A variety of white balance algorithms exist that have different methods of estimating scene illumination and applying white balance. In this experiment, observers manually selected the color of the illumination used to balance images, so mathematical illuminant estimation does not need to be considered. To apply white balance, many published algorithms directly scale camera RGB values based on the calculated RGB response to the estimated illuminant [3] [4]. Because this study is designed around collecting information on preferred color appearance, a white balance calculation that was perceptually-based seemed more appropriate than one that existed only in camera signal space. Processing images in perceptual space provided for an intuitive user interface that allowed observers to manipulate the balance illuminant and allowed the observer data to be easily analyzed.

A perceptual white balance calculation described by Reinhard, et al [7] consists of converting the three camera color signals to XYZ tristimulus values, and then to cone-like responses called ρ, γ, and β using the CAT02 transform. [8]

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = M_{camera} \left(\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{raw} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{\substack{Camera \\ offset}} \right) \quad (1)$$

$$\begin{bmatrix} \rho \\ \gamma \\ \beta \end{bmatrix} = M_{CAT02} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2)$$

Once in CAT02 space, a von Kries adaptation transformation is applied using the ργβ of the scene light source and display white point.

$$\begin{bmatrix} \rho \\ \gamma \\ \beta \end{bmatrix}_{balanced} = \begin{bmatrix} \frac{\rho_{Display}}{\rho_{Scene}} & 0 & 0 \\ 0 & \frac{\gamma_{Display}}{\gamma_{Scene}} & 0 \\ 0 & 0 & \frac{\beta_{Display}}{\beta_{Scene}} \end{bmatrix} \begin{bmatrix} \rho \\ \gamma \\ \beta \end{bmatrix} \quad (3)$$

The resulting balanced ργβ values should approximate the scene's appearance under the display's illumination, achieving color constancy. The balanced ργβ can then be transformed back to XYZ tristimulus, and then to RGB display drive signals.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = M^{-1}{}_{camera} M^{-1}{}_{CAT02} \begin{bmatrix} \rho \\ \gamma \\ \beta \end{bmatrix} \quad (4)$$

Creation of a transform from camera signals to tristimulus values requires a camera characterization using color patches with known reflectance and knowledge of the scene illumination. A transform from tristimulus values to display drive signals requires a similar display characterization where a set of RGB triplets that sample the display gamut are measured. [9]

A framework for psychophysical tests using digital images related to perceived image quality was developed in previous work by the CPIQ. [10]

3 Proposed Method

This study will consist of three experiments. Thus far, the first experiment has been completed. An image set using 8 light sources with a variety of color temperatures and 5 scenes with different contents was captured. More images may be added in later experiments to investigate findings in the earlier experiments. First, a method of adjustment experiment will provide information on preferred white balance of each of the images in the set. Observers directly adjusted the color of the scene light source used in the adaptation calculation in Eq. (3). Second, a paired comparison experiment will be used to determine the distribution of preferences and hopefully establish a specific white balance aim for each image. Third, a quality ruler experiment will compare the perception of image quality degradation due to poor white balance with previously established quality metrics used in the softcopy quality ruler experiment. [10]

3.1 Creating an Image Set

The image set used in this experiment was expected to meet certain requirements set by the CPIQ. Scenes needed to be lit with natural and artificial light sources, both covering a range of color temperatures. Images were needed with and without people, with certain memory colors including common foods, green plants, and blue sky, and without any memory colors. The image set also had to be small enough that the psychophysical experiments could be completed by an observer in a reasonable amount of time. This meant including multiple features in some images. To satisfy these requirements, five scenes were decided upon,

- flowers,
- oranges,
- an abstract painting with no memory colors,
- an person with medium skin tone,
- an person with light skin tone and a blue denim shirt,

and eight light sources were chosen,

- 2 Outdoor: Direct Daylight (5500K) and Shade (13000K),
- 2 Fluorescent lights (5000K and 6000K),
- 1 Tungsten light (3000K),
- 3 LED lights (3000K, 4100K, and 5000K).

The spectra of the light sources were measured using a spectroradiometer and their colorimetry were calculated. The color of the light sources are shown in CIEL*a*b* space in Fig. 2, with the display white point as the reference white. Scenes with mixed lighting were discussed as an important area of study, but were left for future experiments. One extra outdoor scene was captured in the experiment to include blue sky in the image set.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-219

Another factor to consider was the background used in each scene. Most digital images captured by consumers would include a complex background that may provide clues about the lighting conditions of the scene. At the very least, it's usually obvious whether an image was captured indoors or outdoors. However, changing the contents and color of the background might also have a significant effect on preferred white balance without regard to the illumination. It was decided that some of the scenes should have backgrounds that indicated their location, while others should have the same background for all light sources. Specifically, the two scenes with people used different backgrounds while two of the scenes without people used an off-white sheet as a background in all images. The abstract painting did not have a background. Images could then be grouped into scenes with a plain background and a complex background to see if noticeable trends existed between the two.

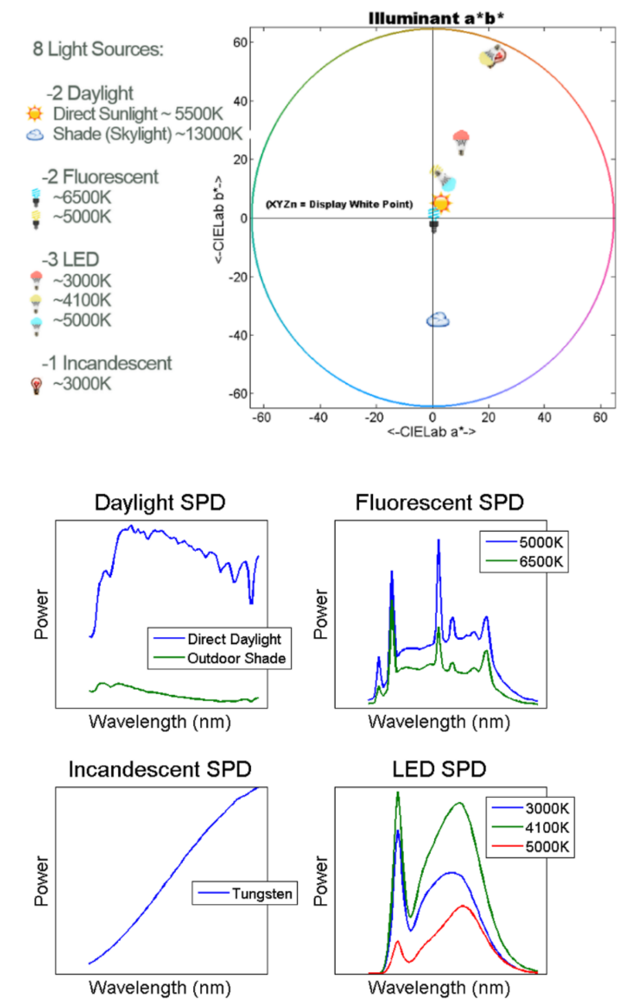

Figure 2. *a*b* (above) and spectral power distribution (below) of the 8 light sources.*

When capturing images, aperture and ISO were fixed to maintain consistent depth of field and image noise throughout the set. Exposure time was adjusted to maintain consistent exposure and avoid clipping throughout the set. The ratio of red, green, and blue camera responses to a white patch was dramatically different between the warmest and coolest light sources, and to avoid clipping both highlights and shadows in all three channels, the overall exposure in terms of lightness was not completely consistent. For instance, the images using tungsten illumination had a much higher red response than green, while the outdoor

Figure 3. *The six scenes used in the image set. Each scene was captured under 8 different light sources, except the outdoor landscape (top-center), which was only captured once.*

shade images had a much lower red response. Attempting to match exposure during capture would either cause the tungsten images to clip highlights in the red channel, or outdoor shade images to clip shadows in the red channel. However, exposure was successfully adjusted in post-processing without any clipping and quantization artifacts.

3.2 Image Processing Pipeline

Images were captured with a Nikon D40 DSLR camera in raw format. The raw files were converted to linear 16-bit images using Adobe DNG Converter. The Nikon D40 was characterized using a standard Color Checker SG target to produce a transform from camera raw signals to XYZ tristimulus values. A different characterization was necessary for each light source to produce acceptably accurate results.

Table 1: Delta-E2000 statistics for the camera characterizations of each light source using the 140 patch Color Checker SG.

	ΔE_{2000} Mean	ΔE_{2000} 90th Percentile	ΔE_{2000} Max
LED 5000K	1.3	2.4	6.0
LED 4100K	1.5	3.1	6.8
LED 3000K	1.3	2.3	5.2
Direct Daylight	1.6	3.2	6.6
Outdoor Shade	1.8	3.5	5.9
Fluorescent 5000K	1.5	2.8	5.3
Fluorescent 6500K	1.7	3.5	5.6
Incandescent	2.3	4.5	6.9

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-219

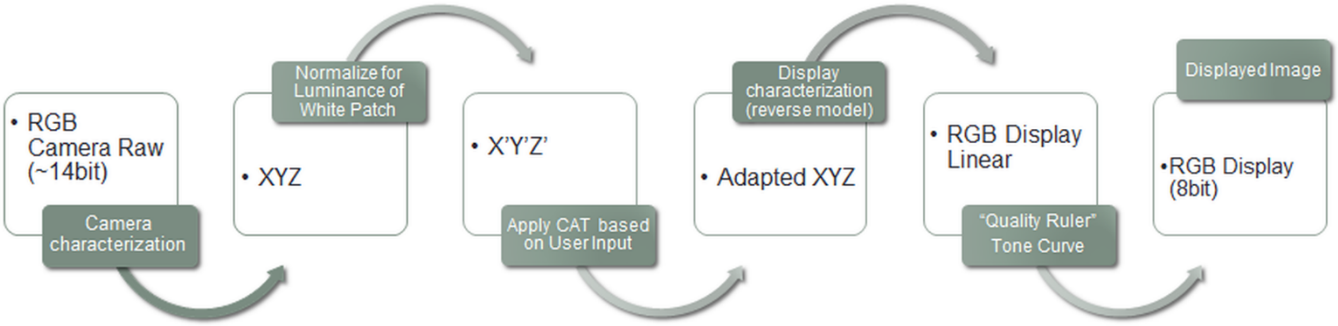

Figure 4. *The image processing pipeline. Observers adjusted the chromatic adaptation transform in step 3 using a GUI.*

To avoid clipping, some images had to be captured with a different exposure level than others, as explained in Sec. 3.1. An image of a Color Checker was captured for each illuminant, so tristimulus images could be normalized using the luminance of a white patch. Images were normalized so that a white patch would have a CIEL*a*b* lightness of 65 when displayed, leaving ample room for displaying specular highlights. All L*a*b* calculations were made using the display white point as reference white. Images were processed using 16-bit values, preventing any noticeable quantization artifacts due to scaling.

Normalized tristimulus images were transformed to cone-like responses ργβ using the CAT02 matrix, and white balance was applied using the equations from Reinhard discussed in Sec. 2. To create the von Kries matrix used in Eq. (3), the color of the scene illumination was manually selected by observers in CIEL*a*b* space, then transformed to ργβ. White balanced images were converted back to tristimulus values and then to display drive values.

The display was characterized using the method proposed by Day, et al, [9] creating a matrix transform from tristimulus values to scalar RGBs and lookup tables from scalar RGBs to RGB code values. The characterization was based on measurement of 453 RGB triplets distributed throughout the displays' gamut. Throughout the processing pipeline, the only changes made to the original scene tristimulus values are the luminance normalization and chromatic adaptation. If tone reproduction is left unaltered, the resulting displayed images are low in contrast compared to images one would expect from a consumer digital camera. To remedy this, the LUTs produced by the display characterization were modified by averaging them with the tone curve used in the image processing pipeline for the softcopy quality ruler protocol [10], boosting contrast enough to produce acceptable appearance.

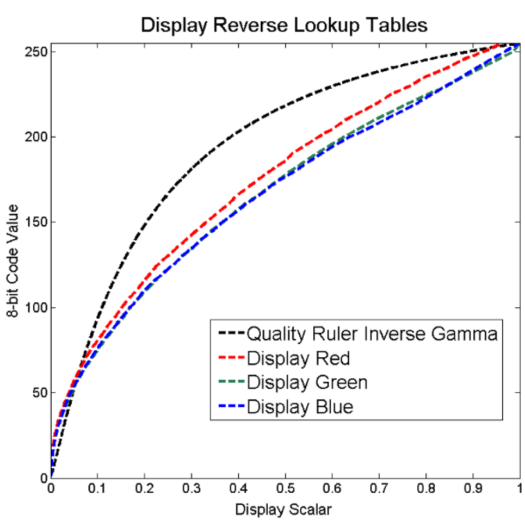

Figure 5. *The display characterization LUTs and the Quality Ruler tone curve. An average of the two was used to boost contrast while accounting for the display's gamma.*

3.3 Experimental Setup

A graphical user interface was created using Matlab. Two sliders below a displayed image controlled the a* and b* coordinates of the scene light source used for white balance. The scene light source had a fixed L* of 100. The reference white of the L*a*b* space was the white point of the display. L*a*b* values from the user interface were converted to XYZ tristimulus and then to ργβ using the CAT02 transform. The ργβ cone-like responses are used in the white balance calculation shown in Eq. (3). The a* slider was labeled with "+Green" and "+Red" on either end and the b* slider was labeled with "+Blue" and "+Yellow". Observers were asked to act as if they had taken the images on their cell phone and planned to share them online, but first to adjust the red-green balance and yellow-blue balance of each image until they felt it looked its best. The sliders were controlled using a mouse.

Images were shown in random order. The total range of the sliders was fixed to 110 a* units and 210 b* units, but the end points were randomized. The initial slider position for each image was the center point, but the value corresponding to center of the

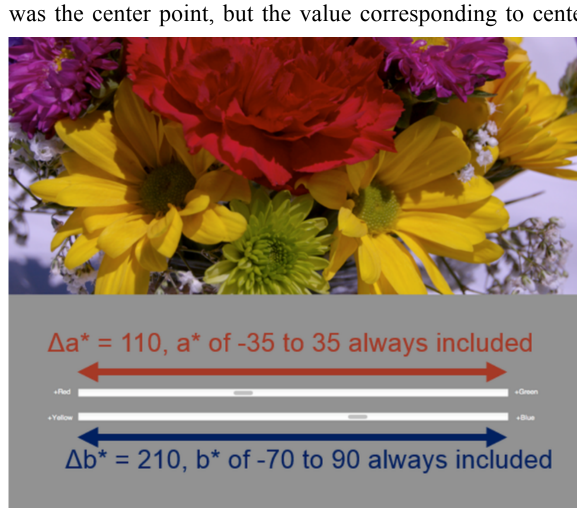

Figure 6. *The sliders used in the GUI. The range on the sliders is fixed but the values that slider represents is randomized.*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-219

slider and the initial image appearance were random. The randomization of the endpoints was set to always include a* values of -35 to 35 and b* values of -70 to 90 as options. The authors felt the required range covered the limits of reasonable appearance in the image set. Because the slider range was fixed, the amount of change produced by identical slider increments was consistent across all images. The change produced by the minimum increment was not noticeable.

An HPZR30w display, with 30-inch diagonal and 2560x1600 resolution, was used to conduct the experiment. The display background was set to a mid-gray at half the maximum drive level, with correlated color temperature 6600K and luminance 75 cd/m^2. The experiment was conducted in a dark room, with a gray wall behind the display lit to have similar color to the display background. The immediate surround on the back wall had a correlated color temperature of 5300K, with 80 cd/m^2 luminance just above the top of the display and 55 cd/m^2 luminance to the sides.

Images were scaled with a downsampling algorithm used in the softcopy quality ruler protocol and cropped to have a height of 835 pixels. Most images had a width of 1254 pixels, except the images of the abstract painting, which were only 851 pixels wide due to the dimensions of the painting. Images were displayed in the center of screen.

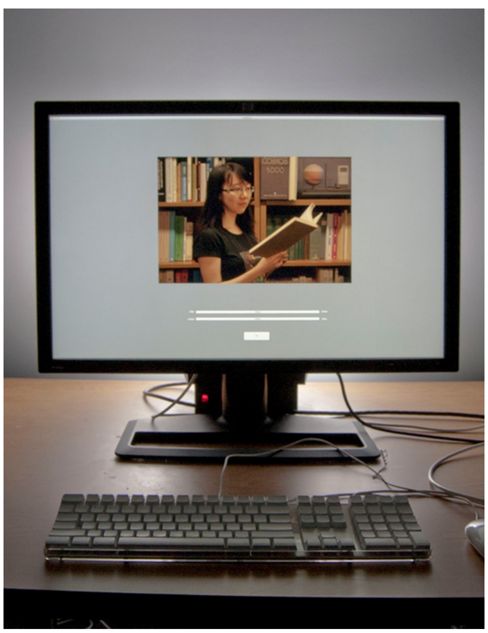

Figure 7. *The experimental setup. The room is dark except for the rear wall, which is lit with a light positioned behind the display.*

4 Experimental Results

Twenty-one observers participated in the method of adjustment experiment. Their results were evaluated by calculating the appearance of a color checker mid-gray patch using their selected scene light sources for white balance. The raw camera signals of the gray patch under the selected illumination were processed using the pipeline described in Sec. 3.2 and the L*a*b* coordinates of the displayed gray patch were used for analysis. An image that is left uncorrected, with a white balance where the scene illumination is identical to the display white point, would result in a gray patch with a* and b* equal to the actual scene light source. An image that is completely corrected, with a white balance where the scene illumination is identical to the actual scene light source, would result in a gray patch with a* and b* equal to zero. The results are discussed in terms of preferred appearance, referring to the calculated a* and b* of the displayed gray patch using the selected white balance for each image.

The shifts in preferred appearance between images was much larger in the b* direction than the a* direction. This was expected, as the color of different light sources vary mostly in b*, and shifting appearance in a* quickly produces images that look unnatural, especially in the negative (green) direction. Generally, a more reddish tone was preferred, as all but three images had an average preferred appearance with positive a*, and two of the negative a* images were of the abstract painting with no memory colors.

When considering all forty-one images and twenty-one observers as independent observations, a strong positive correlation exists between the actual scene light source b* and preferred appearance b*. The Pearson correlation coefficient between uncorrected b* and preferred b* was 0.40, well above the critical value of .07 needed for 95 percent confidence with 861 observations. In other words, preferred appearance was directly related to the color of the scene light source, despite observers having no knowledge of the scene light source in most cases. This trend can be seen in Fig. 8, where the arrows point from the uncorrected appearance to preferred appearance, averaged across all scenes. It looks as if all scenes are corrected towards some central point, but their degree of correction depends on the scene light source's b*. Interestingly, this central point is not the origin in a*b* space, perhaps because people do not fully adapt to emissive displays, making the use of the display white point as reference white inaccurate. [5] This trend of appearance skewing towards the scene light source is similar to the incomplete adaptation of the human visual system. The uncorrected appearance of images, which indicates their scene illumination, should not have influenced observer decisions because the initial image appearance was randomized across a range much larger than the range of illumination color.

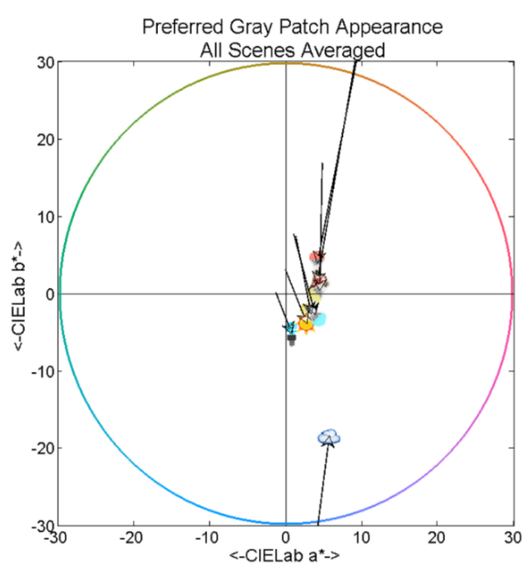

Figure 8. *Preferred appearance, indicated by the icons, seems to directly depend on the color of the actual scene illumination, indicated by the starting point of the arrow.*

DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-219

The observers had no knowledge of what the uncorrected appearance of each image was, but it seems their decisions were somehow influenced by the scene light source. Most images did not have backgrounds that indicated what the scene illumination was. Whatever the cause of this trend, it seems that white balance should be incomplete to some degree, similar to chromatic adaptation in the visual system.

The results also showed that scene contents have a large impact on preferred appearance, even when the scene illumination is consistent. It was difficult to identify any specific relationships between certain types of scene contents and preferred appearance. One notable trend was that images with people had a much more yellow/orange appearance on average than those without any people, except for one light source (outdoor shade). This could be explained by a preference for skin tones that look more tan. [6] Based on this, a white balance algorithm that included input from facial recognition could improve perceived image quality by producing a more yellowish appearance if people were present in the image. Hopefully, further experiments will reveal more trends like this for other types of content, but a larger, more diverse image set may be required to identify specific relationships between content and preferred white balance.

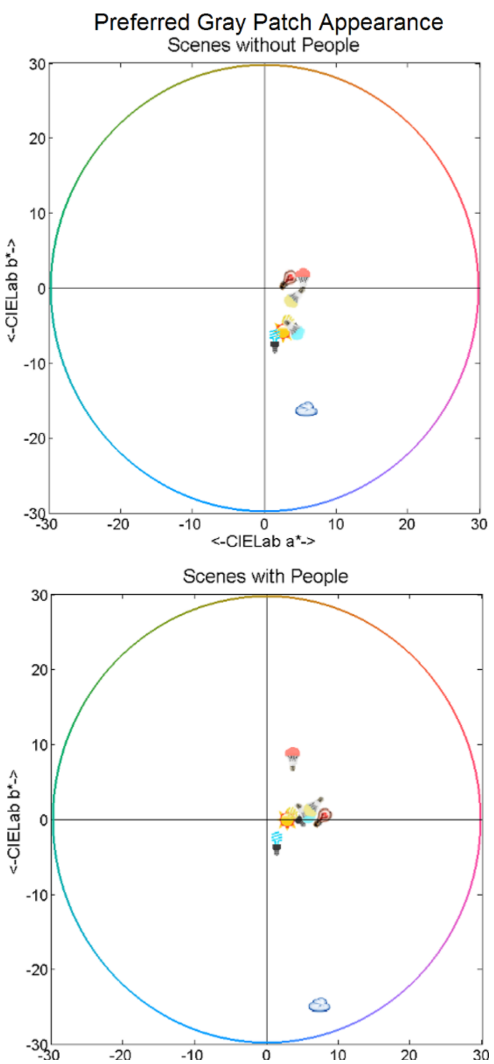

Figure 9. *Average preferred appearance shifts towards yellow when people are included in a scene.*

Inter-observer variation for each image was very high, which was expected in an experiment based on opinions of what is or is not aesthetically pleasing. The ellipses in Fig. 11 show the standard deviation along the first and second principal components of variation. The ellipses may be surprisingly large, with a typical standard deviation being 10 to 25 ΔE_{ab} along the first principal component. However the four images in Fig. 10 represent the edges of a standard deviation ellipse and all have reasonable appearance. The amount of variation likely means that preference should not be used to set any specific aims for white balance. Increasing the number of observers may reduce the amount of variation, but future experiments will probably need to ask a more specific question of observers than their preference.

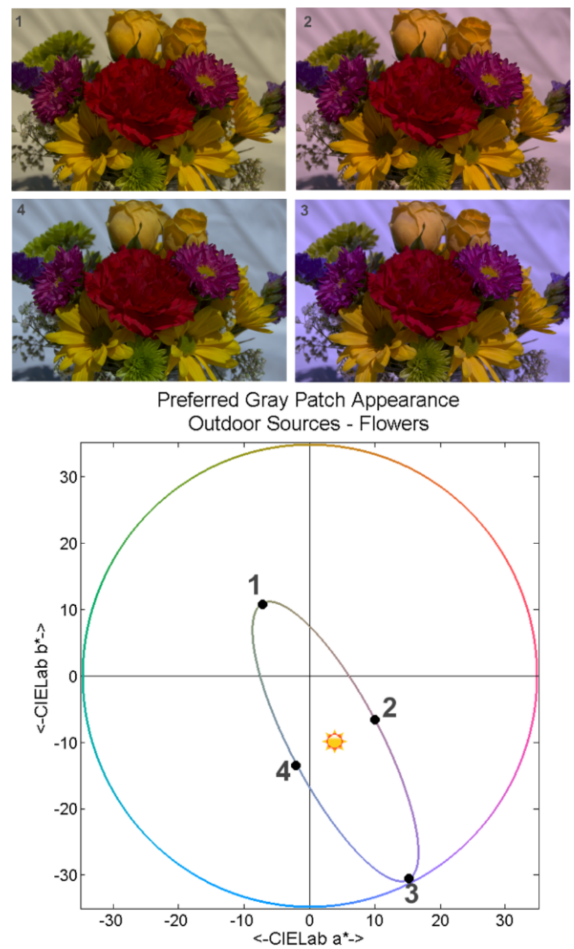

Figure 10. *The standard deviation of preferred appearance for flowers under outdoor lighting (bottom)). Images produced using preferred appearance one standard deviation away from the mean in each of the four principal component directions for flowers under direct daylight (top).*

In most cases, the inter-observer variation is quite directional and has a somewhat consistent shape. In a few cases, such as the abstract painting, the variation seems random. Usually, the combination of standard deviation ellipses across all light sources is similar in shape to the Planckian locus of natural light sources, as shown in Fig. 11. This visual assessment was confirmed by statistical analysis of the properties of the standard deviation ellipses. Significant correlation existed between b* and ellipse angle, with a negative coefficient of 0.57, just over the critical value of 0.41 for 95 percent confidence with 21 observations. This

DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-219

indicates a tendency for the variation ellipses to rotate in a clockwise direction as b* of the mean increased, just as the Planckian locus curves to the right as b* increases. The total amount of variation was inconsistent between scenes and light sources, without many noticeable trends.

Images with people did tend to have lower variation than those without, which fits with previous research that suggests variation in preferred skin tone appearance is quite small compared to other types of content. [11] The lower variation could also be due, in part, to the scenes with people having complex backgrounds while the other scenes did not. The backgrounds could provide some expectation of the color of scene illumination or the variety of colors in the background could limit the white balance options that produce reasonable appearance.

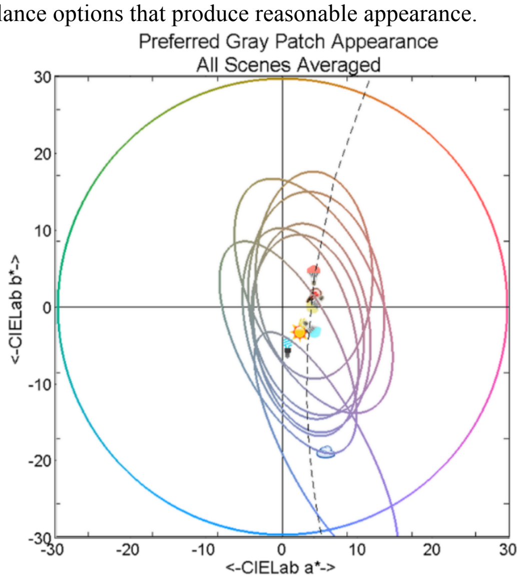

Figure 11. *Standard deviation ellipses for all light sources averaged across all scenes. The general shape is similar to the Planckian locus (dotted line).*

5 Conclusions and Future Work

A method of adjustment experiment was designed and carried out as a first step to gain general information about white balance preferences. Trends in the data indicated that ideal white balance reflects the incomplete nature of chromatic adaptation in human vision. Preferred white balance had a direct dependency on the color of scene illumination even when no obvious perceptual cues indicated the scene illumination. The amount of variation between observers was quite large, but the data provides a baseline range of acceptable white balance, which can be examined in more detail in future experiments. That range, unsurprisingly, appears to be related to the color produced by natural light sources.

The graphical user interface and image set has been shared with members of the CPIQ who will conduct the experiment under similar viewing conditions with more observers. A larger pool of observers may reduce the amount of variation, or reveal important trends. The authors will continue the research with a paired comparison experiment using images white balanced based on the range of balances selected in the first experiment. The paired comparison experiment will ask the observers to select the image with the more natural color balance. Preferred appearance varies quite a lot from person to person, so the specific aims for an automatic white balance algorithm should be based on an appearance that most of the population agrees is at least reasonable. The perception of what is natural would be more consistent than the perception of what is best, and most people likely think images that look natural have a reasonable appearance.

References

[1] E. Hering, L. translator Hurvich and D. Jameson, Outlines of a Theory of the Light Sense, Cambridge: Harvard University Press, 1964, pp. 7-8.

[2] D. Jameson and L. Hurvich, "Essay Concerning Color Constancy," *Annual Review of Psychology,* no. 40, pp. 1-22, 1989.

[3] G. FInlayson, S. Hordley and P. Hubel, "Color by Correlation: A Simple, Unifying Framework for Color Constancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, no. 11, pp. 1209-1221, 2001.

[4] D. Cheng, D. K. Prasad and M. S. Brown, "Illuminant estimation for color constancy: why," *J. Opt. Soc. Am. A,* pp. 1049-1058, 2014.

[5] M. Fairchild, "Chromatic Adaptation to Image Displays," *Technical Association of the Graphic Arts,* vol. 2, pp. 807-823, 1992.

[6] R. Hunt, "Preferred Colour Reproduction," in *The Reproduction of Color*, West Sussix, John Wiley & Sons Ltd, 2004, pp. 174-175.

[7] E. Reinhard, E. Khan, A. Akyuz and G. Johnson, Color Imaging Fundamentals and Applications, A K Peters, Ltd, 2008.

[8] CIE TC 8-01, "A Colour Appearance Model For Colour Management Systems: CIECAM02," 2003.

[9] E. A. Day, L. Taplin and R. S. Berns, "Colorimetric Characterization of a Computer-Controlled Liquid Crystal Display," *Color Research and Application,* vol. 29, no. 5, pp. 365-373, 2004.

[10] E. Jin and B. Keelan, "Slider-adjusted softcopy ruler for calibrated image quality assessment," *Journal of Electronic Imaging,* vol. 19, no. 1, p. 011009, Jan-Mar 2010.

[11] B. Keelan, Handbook of Image Quality: Characterization and Prediction, New York: Marcel Decker, 2002, pp. 160-162.

Author Biography

Ben Bodner received his BS in motion picture science from the Rochester Institute of Technology (2012). He worked as an Image Scientist at Exelis, working display technology and imaging algorithm development. He is now pursuing a Masters in color science at RIT, where his research involves multi-spectral imaging, Kubelka-Munk theory, and image processing.

Yixuan Wang received her BS in pharmacy from Nankai University, China, her Masters in biomedical engineering from PUMC, Tsinghua University, China. She is now pursuing a Masters in color science at Rochester Institute of Technology. Her research involves multi-spectral imaging, image processing and image quality.

Susan Farnand received her BS in engineering from Cornell University, her Masters in Imaging Science and her PhD in Color Science from the Rochester Institute of Technology. She currently works as a Visiting Assistant Professor in the Program of Color Science at RIT. Her research interests include human vision and perception and color science. She is the

DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-219

Publications Vice President of IS&T and serves as an Associate Editor for the Journal of Imaging Science and Technology, and has served as co-chair of the IQSP conference at EI.