https://doi.org/10.2352/ISSN.2470-1173.2018.12.IQSP-342

Development of a Perceptually Calibrated Objective Metric for Auto White Balance

Elaine W. Jin [a], Yixuan Wang [b], Wentao Liu [c] a. Nvidia Corporation, Santa Clara, California, USA; b. Apple, Cupertino, California, USA; c. University of Waterloo, Waterloo, Canada.

Abstract

This study aims at developing an image quality metric for camera auto white balance (AWB), with a transform to just noticeable differences (JNDs) of quality in pictorial scenes. In this study, a simulation pipeline was developed for a Nikon D40 DSLR camera, from raw capture to rendered image for display. Seven real-world scenes were used in the study, representing capture conditions in outdoor daylight, indoor fluorescent lighting, and indoor incandescent lighting conditions. Two psychophysical experiments were performed, and 38 observers participated in the study. In study one, method of adjustment was used to explore the color aims for individual scenes. In study two, a softcopy quality ruler method was used to refine the color aims and define the quality falloff functions. A quartic function was used to fit the results from the softcopy ruler study, forming the proposed objective metric for camera auto white balance.

Introduction

The impact of digital photography has been growing due to the ubiquitous presence of the smartphones. The cameras equipped on these smartphones have seen steady improvement in quality, and today the quality level has reached and even surpassed that of the compact digital still cameras. To continue the improvement on camera image quality, it is essential to develop objective metrics for image quality that can correlate with human perception.

Image quality attributes can be divided into two major groups, artifactual attributes and preferential attributes [1]. Artifactual attributes refer to defects introduced by imaging systems, including attributes such as noise, blur, and color shading. For these attributes the ideal aim positions are well defined, i.e. free of the defects. Preferential attributes refer to attributes that are preferential in nature, including all color/tone rendering related attributes such as color saturation, contrast, exposure, and white balance. It is challenging to define the ideal aim positions for such attributes because the aim positions are likely observer and scene dependent,

Auto white balance (AWB) is a camera auto function that detects the scene illumination and compensates for it in image rendering. In an ideal scenario, the camera image processing pipeline can perfectly detect the scene illumination, and rendering a neutral patch to be perfectly neutral in the given output color space (e.g., the sRGB color space) [2]. In such a scenario the measurement of the quality of the camera AWB function becomes somewhat straightforward. For example, in the Imatest Master software tool the white balance error is defined as the average CIELAB chroma of patches 20 – 23 on an X-Rite ColorChecker Classic target [3] [4].

There are complications to this ideal scenario. It has been reported that the preferred color/tone reproduction in photographic images are different from the colorimetric reproduction [5-6]. For example, people prefer a blue-sky image to have much higher purity than the real sky blue. Today's smartphone cameras also adjust the white balance rendering aims according to the scene illumination. Fig. 1 shows the images of an X-Rite ColorChecker Classic target captured by two smartphone cameras under two lighting conditions. The two daylight images looked very similar to each other, and the neutral patches are close to neutral. In comparison, the images captured under tungsten light are rendered warmer, with the neutral patches moving towards the red/yellow direction. Furthermore, it is interesting to notice that these two images looked very different from each other, with one having much higher chroma than the other. These observations inevitably lead to the following questions: what are the preferred white balance aims for images captured under different lighting conditions? What is the quality falloff when the rendering positions deviate from the preferred color aims? The purpose of this study is to answer these two questions.

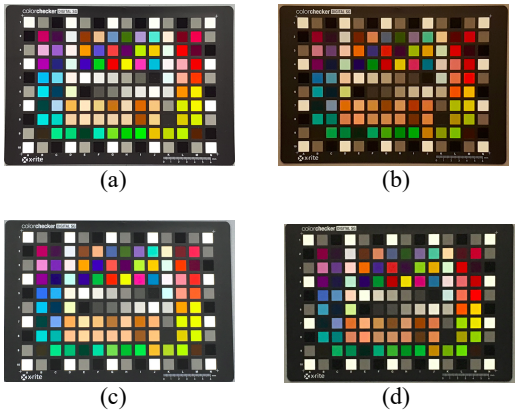


Figure 1. Images of X-Rite ColorChecker Classic captured by two smartphone cameras under two light sources. Top row: smartphone #1. Bottom row: smartphone #2. Left column: daylight light; Right column: tungsten light.

There was a 2016 study on the effect of capture illumination on preferred white point for camera automatic white balance [7]. In that study, an extensive list of illumination sources was studied, including daylight sources (5500K and 13000K), fluorescent light sources (6500K and 5000K), incandescent light sources (3000K), and LED light sources (3000K, 4100K, 5000K). Method of adjustment was used to explore the color aim positions for various illuminations. The results of the study suggested that the preferred white balance positions could vary and were scene illumination dependent.

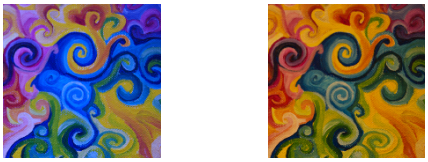
One key difference between the present study and the 2016 study is in experimental design. In the 2016 study, large variations

in responses were observed between observers. One plausible reason was that the observers in the study did not receive sufficient perceptual cues to anchor their psychophysical responses. For example, a landscape scene (shown in Fig. 2) was used as one of the test scenes in the 2016 study. the same raw input image could be rendered either as a normal daylight scene (Fig. 2(a)), with a gray patch CIELAB values as [a*, b*] = [0.598, -0.654); or as a sunset scene (Fig. 2(b)), with a gray patch CELAB values as [a*, b*] = [3.763, 5.06]. When instructed to adjust the color appearance of the test scene to an optimal quality, the observer could choose either position with high confidence, resulting in large variations among observers.

(a) (b)

Figure 2. The same raw input image was rendered to two plausible color positions, (a) daylight appearance; (b) sunset appearance.

A second possible reason for getting large variations among the observers could be the lack of memory colors in some of the test scenes. Memory colors refer to colors of familiar objects such as face, sky, and green grass. The Abstract Painting scene used in the 2016 study contained none of the memory colors (see Fig. 3). The observers could adjust the white balance position to any arbitrary colors without degrading image quality. This could be another source of large variations among the observers.



(a) (b)

Figure 3. The Abstract Painting scene rendered to two plausible color positions, (a) with blue tine; (b) with yellow tint.

In the present study we intend to avoid both sources of variations by introducing memory colors in the test scenes and informing the subjects on the capture scene illumination.

The goal of the present study is to determine the white balance color aim positions for the commonly used light sources in camera testing: outdoor daylight, indoor fluorescent light, and indoor tungsten light. Furthermore, this study aims at defining the quality falloff functions when the color positions deviate from the ideal positions.

It is worth noting that this work is part of the image quality metric development effort carried out by the IEEE P1858 CPIQ working group. This working group recognizes that camera 3As (auto exposure, auto focus, and auto white balance) are important functions for image quality, and is in the process of developing image quality metrics for these attributes.

Methods

Image Processing

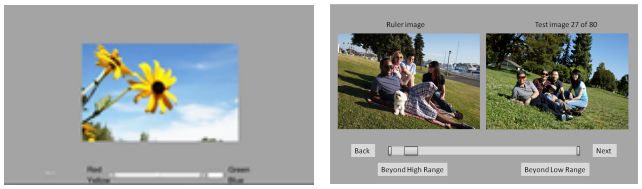
In this study test stimuli were created by adjusting color positions in CIELAB space on high quality pictorial images as well as a color target. The color target used in this study is an X-Rite ColorChecker Digital SG chart (see Fig. 1). During the capture process each pictorial scene capture was followed by a second capture of the color target. The pictorial images were evaluated by human observers, and the results compared with objective metrics of white balance derived from the color target.

In studies of white balance performance, it is important to have a calibrated color workflow. A Nikon D40 DSLR camera was used to capture both the pictorial images and the color target. An image processing pipeline was developed to process the raw capture from the camera to an output image for viewing. Camera color calibration was facilitated by the capture of the X-Rite ColorChecker Digital SG chart under the same illumination. An illumination-specific transform was derived for each scene illumination using the 140-patch colors on this color target, which converted the camera RGB values to the CIE tristimulus values. Display color calibration was performed using an analytical method developed by Day et al. [8]. An HP Z30i LCD monitor (2560 x 1600 @100 PPI) was used to display the test images in the psychophysical studies. During the display calibration, 457 color patches were displayed sequentially on the monitor. A PhotoResearch spectroradiometer PR670 was used to measure the CIE tristimulus values of the color patches. An optimization procedure was used to create a display color profile based on the monitor RGB values and the corresponding XYZ values. The display color profile includes three components: a 3x3 matrix, three 1D look-up-tables, and the flare correction offsets. The treatment in this study was to introduce color bias in image white balance. This treatment was performed in the CIELAB color space.

Figure 4. Test scenes used in the psychophysical studies. Scene names from top left to bottom right: Daylight – Grass People; Daylight – Mountain; Daylight – Sunflower; Fluorescent – Girl; Fluorescent – Flowers; Tungsten – Girl; Tungsten - Flowers.

To prepare the test images for the graphic GUI used in the two psychophysical studies, the output image from the Nikon D40 camera (3008 x 2000) was down-sampled by 2x and then center cropped to 1MP (1254 x 835) for viewing. The test images were displayed at 100% magnification during the psychophysical studies. The viewing room was dim lit, with an adapting field filled by 18% gray color.

Seven test scenes were used in the study (see Fig. 4), representing photographical contents found in typical consumer photography. The lighting conditions included direct sunlight (5500K), indoor fluorescent light (5000K), and indoor tungsten light (3000K). Memory colors in the test scenes included skin tone, sky, grass, and sunflower. In the flower scene the observer was instructed that the tablecloth in the background was near white.



(a) (b)

Figure 5. GUIs used in the psychophysical studies. (a) Method of adjustment; (b) Softcopy quality ruler.

Method of Adjustment

A Matlab tool was created to display the graphics user interface (GUI) for the study using method of adjustment (see Fig. 6(a)). The GUI consisted of a gray background and a test image of 1254 x 835, displaying at 100% magnification. There were two sliders on the GUI, allowing color adjustment along the CIELAB a* and b* axis. By moving the slider, the white balance position of the displayed pictorial image could move towards any of the red, green, yellow, blue directions.

At the beginning of each session, the subjects were shown examples of daylight, fluorescent light, and tungsten light using a SpectralLigth III light box [9]. During the session subjects were instructed to adjust the two sliders until the image looked natural to the eye. Nineteen subjects participated in the study. All subjects had normal color vision and normal or correct-to-normal visual acuity.

The objective measurement of white balance was reported using the a*, b* values of a ColorChecker mid-gray patch (H-5 on the SG chart), processed using the same image processing pipeline as that used for the pictorial images. Principal component analysis was used to fit the confidence interval for all 19 subjects and the fitting ellipses were report in Fig. 6.

The size of the ellipse represented the variation among observers, and it seemed to be both scene illuminant and scene content dependent. For daylight scenes the sunflower scene had a larger variation compared to the other two scenes. Subjects did report after the session that many color positions would seem plausible for the sunflower scene. For fluorescent and tungsten light scenes the scene with the face (left) always had a smaller variation compared to the scene with the flower (right), indicating that the variation among observers would be smaller in the presence of the skin tone. The ellipses for the tungsten light scenes were bigger than the same scene under fluorescent light, suggesting that the opinions may differ among observers regarding the color aims for images captured under tungsten light. Lastly, it was interesting to notice that all ellipses were longer along the b* axis and shorter along the a* axis, indicating that the variations among observers were asymmetric regarding a* and b* directions.

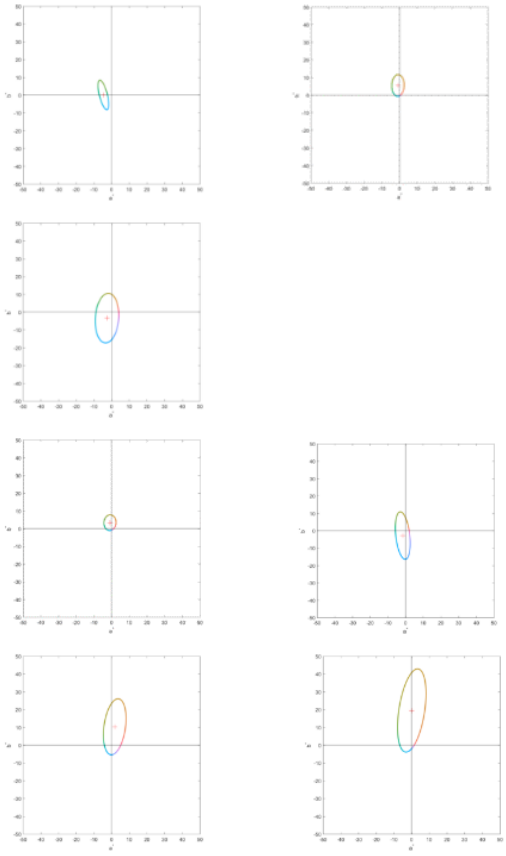
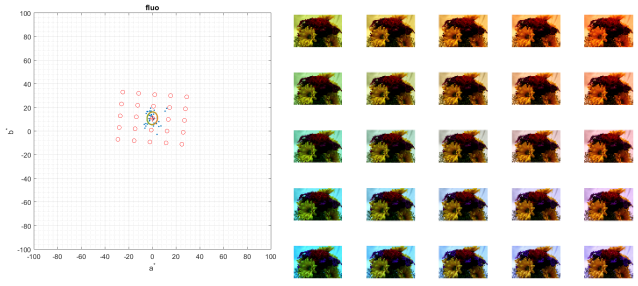


Figure 6. Test results from the method of adjustment study. From top left to bottom right: Daylight – Grass People; Daylight – Mountain; Daylight – Sunflower; Fluorescent – Girl; Fluorescent – Flowers; Tungsten – Girl; Tungsten - Flowers. All plots used the same scales on the x-axis and y-axis.



(a) (b)

Figure 7. 5x5 sampling for creating test images used in the softcopy quality ruler study.

The test stimuli used in the softcopy quality ruler experiment were created based on the results from the method of adjustment study. Fig. 7 shows one example image set. A 5x5 sampling grid was created in the CIELAB a*b* space (Fig. 7 (a)). The spacing and orientation of this grid were determined by the size and orientation of the ellipse. A set of 25 test stimuli were created using the 25 points on the grid (Fig. 7(b)), representing white balance positions covering a large range around the center point in the CIELAB space. The CIELAB a*b* values were also created from the gray patch on the ColorChecker SG chart to represent the target metric values. The sampling grid was unique for each of the seven test scenes.

Softcopy Quality Ruler Study

A softcopy quality ruler method, as depicted in ISO 20462 Part 3, was used in the second psychophysical study [10-13]. In the softcopy quality ruler method, two images were displayed side-by-side on a monitor, a ruler image that varied in sharpness, and a test image that varied in white balance (see Fig. 5 (b)). The subjects were instructed to adjust the sharpness (and hence quality) of the ruler image (left image) to match the quality of the test image (right image). The result of the match was recorded as a calibrated value of the ruler image in display on the Standard Quality Scale (SQS), as defined in ISO 20462 Part 3. Because the ruler images are calibrated using sharpness, the viewing distance must be specified and controlled. In this experiment, the viewing distance was set at 864 mm.

Figure 8. Ruler images used in the softcopy quality ruler study.

Care was taken to match the scene contents of the ruler images and the test images for easy comparison. Fig. 8 shows the 5 ruler scenes used in this study. The two bottom scenes were used in testing for both the fluorescent light and the tungsten light scenes.

Same as in the first experiment, the subjects were shown examples of daylight, fluorescent light, and tungsten light using a SpectralLight III light box at the beginning of the session. This was to ensure that there was no ambiguity in understanding the capture conditions.

Nineteen subjects participated in the study. All observers had normal color vision and normal or correct-to-normal visual acuity. Each subject performed quality matching for a total of 175 test images, including 7 scenes and 25 treatment levels. The test was divided into two sessions to avoid viewer fatigue.

Results

The SQS responses from all 19 subjects were averaged to produce the final SQS values for the 25 color positions for each scene. Fig. 9 shows the SQS results in the CIELAB a* b* space.

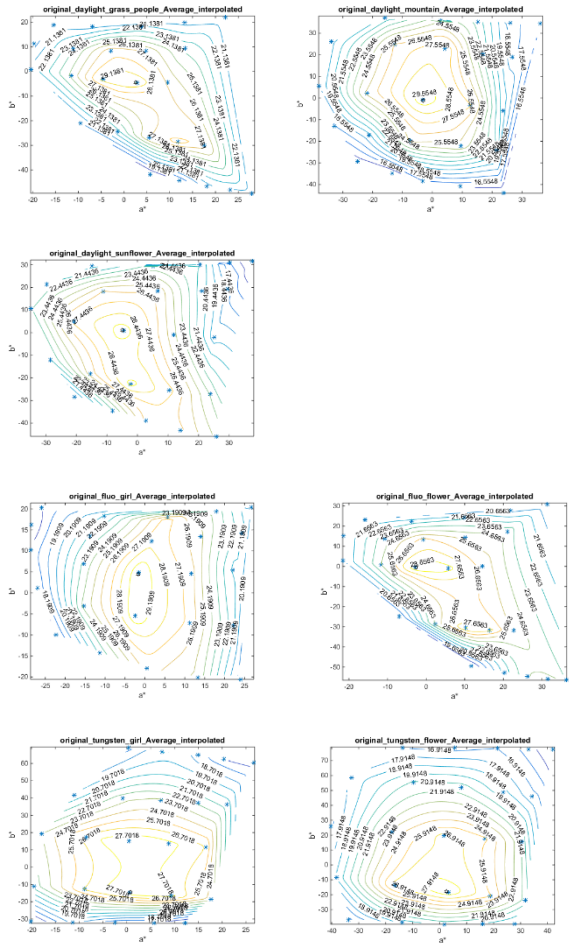


Figure 9. Test results of the softcopy quality ruler study, following the same order as in Fig. 6. The blue symbols are the data from the experiment. The contours were SQS values interpolated based on the measurement data. The x and y axis are the CIELAB a and b* axis.*

It can be seen from Fig. 9 that the SQS results have complex contour shapes. A quartic function was selected to fit the SQS values as a function of the target a* b* values. Eq. (1) shows the quartic model used to fit the data.

$$SQS = p1x^4 + p2\,y^4 + p3x^3\,y + p4x^2\,y^2 + p5x\,y^3 + p6\,x^3 + p7\,y^3 + p8x^2\,y + p9x\,y^2 + p10x^2 + p11\,y^2 + p12x\,y + p13x + p14y + p15 \quad (1)$$

$$QL = MaxSQS - SQS \quad (2)$$

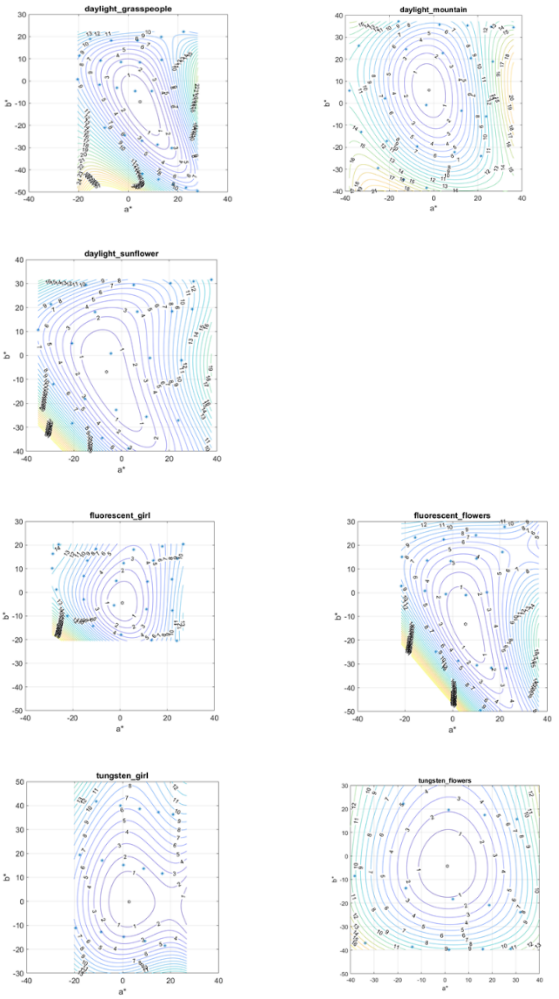


Figure 10. Modeling QL results of the softcopy quality ruler study, following the same scene order as in Fig. 5. All plots used the same scales on the x-axis as well as the y-axis. The x and y axis are the CIELAB a and b* axis. The blue symbols are the data from the experiment. The interval between contours is 1 JND.*

A second mathematical conversion is needed to express the white balance errors in the quality loss space. The mathematical equation for this conversion is show in Eq. (2). Use of the quality loss space would allow the results from white balance to be combined with results from other image quality attributes for prediction of an overall image quality using the multivariate formulism [14].

The model fitting process is as following:

(1) A quartic function, i.e. Eq. (1), was used to fit the mean SQS values in each test scene as a function of the target a*b* values. Table 1 shows the parameters used for individual test scenes;
(2) The maximum SQS value was derived for each test scene based on the quartic model, also shown in Table 1;
(3) A QL function was calculated from Eq. (2), results shown in Fig. 10.

Table 1. Parameters used in the SQS quartic model fitting, and the derived Max SQS value for converting SQS to QL values.

Model Fitting Parameters	Daylight-Grasspeople	Daylight-Mountain	Daylight-Sunflower	Fluorescent-Girl	Fluorescent-Flowers	Tungsten-Girl	Tungsten-Flowers
p1	4.30524E-06	4.68349E-06	8.20377E-07	9.53214E-06	4.65095E-06	1.06442E-05	-8.65255E-08
p2	-2.4752E-06	-4.41778E-07	-3.20494E-06	-1.58085E-06	-4.03236E-06	-6.4835E-07	-2.52362E-07
p3	1.85874E-05	2.21687E-06	-1.90869E-06	-6.27231E-06	-1.04169E-06	-2.47294E-06	-5.64879E-07
p4	2.64524E-05	6.27516E-06	-2.67061E-06	-1.22866E-06	-5.75255E-06	-1.60596E-06	1.4641E-06
p5	1.03412E-05	2.44742E-06	-1.46526E-06	-1.6479E-05	-1.00311E-05	2.58577E-06	-2.08627E-07
p6	-0.000333848	-4.85936E-05	3.48521E-05	-6.44073E-05	2.20721E-07	0.000129298	-9.44539E-06
p7	-8.99289E-05	9.94819E-06	-5.14708E-05	0.000340915	2.03605E-05	0.000111538	5.3948E-05
p8	0.000261758	-4.83943E-05	0.000171184	0.000351532	0.000671449	-1.87023E-05	-1.41141E-05
p9	0.000507166	0.000107986	0.000331959	0.000552646	0.000495102	-0.000111135	9.56828E-06
p10	-0.024890675	-0.018491006	-0.010658138	-0.023816928	-0.018161854	-0.014092945	-0.006867797
p11	-0.01230918	-0.006767424	-0.000903276	-0.011079798	-0.006025495	-0.007276888	-0.0042381
p12	-0.015697172	-0.006139414	-0.000582775	0.003460064	-0.006881368	0.000181348	0.000246717
p13	0.089194224	-0.04512049	-0.17492333	0.051322782	0.102773551	0.073356955	0.015120296
p14	-0.104289982	0.070606074	-0.051212174	-0.118665408	-0.093657563	-0.00232567	-0.040295814
p15	29.30305418	30.82317105	28.87388957	30.10681166	28.89018847	30.06442457	29.42557927
Max SQS	29.93110121	31.08617245	29.68234616	30.40770812	29.73230004	30.16299433	29.52360452

Some observations can be made based on the quality falloff contours shown in Fig. 10.

(1) The neutral position, i.e. a*b* = [0, 0], is within 1 JND of quality loss for all test scenes. This suggests that the color aim positions under all 3 test illuminants are near neutral.
(2) The presence of skin tone tends to make the quality falloff faster compared to those without skin tones.
(3) Quality falloff is slower along the b* axis compared to the a* axis;
(4) Under tungsten lighting quality falloff is slower along the +b* direction compared to –b* direction, indicating an asymmetric quality falloff behavior

Conclusions and Discussions

Two subjective studies were performed to explore the observer perception of image quality variations due to changes in white balance position. In the first study, method of adjustment was used to explore the color aim position. In the second study two, softcopy quality ruler was used to refine the color aim position and creating quality falloff contours around the color aim position. A quartic model was used to fit the results obtained in the softcopy quality ruler experiment. Model fitting parameters were generated for all 7 test scenes.

The study results suggest that the aim color positions for white balance seems to be close the neutral position for all light sources used in this study. Furthermore, the quality falloff function seems to be light source dependent and can vary with scene content.

Even though we attempted at reducing variations among the observers by providing additional perceptual cues in scene object color and scene capture illumination, there were still significant variations among the observers and among test scenes. As a result, it is challenging to identify illumination specific color aims and quality falloff functions. In future studies we intend to verify the results from this study by adding more test scenes and more observers.

The quartic model used in this study has many parameters and it may be too complex a model to use in describing the AWB quality falloffs. Further studies may consider using other functions such as 2D skewed normal function.

References

[1] B. Keelan, Handbook of Image Quality: Characterization and Prediction, New York: Marcel Decker, pp. 6 – 9, 2002.

[2] https://www.w3.org/Graphics/Color/sRGB.html

[3] http://www.imatest.com/

[4] http://xritephoto.com/colorchecker-targets

[5] R.W.G. Hunt, The reproduction of Colour, 5th ed. England: Fountain Press, pp. 222 – 240, 1995.

[6] B.W. Keelan, R. Jenkin, and E.W. Jin, "Quality versus color saturation and noise", Proc. SPIE 8299, 829914, 2012.

[7] B. Bodner, Y. Wang, and S. Farnand, "Effect of capture illumination on preferred white point for camera automatic white balance," IS&T International Symposium on Electronic Imaging 2016 IQSP-219.1, 2016.

[8] 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.

[9] https://www.xrite.com/-/media/xrite/files/manuals_and_userguides/g/gmb_spl3_manual_en.pdf

[10] B.W. Keelan and H. Urabe, "ISO 20462, A psychophysical image quality measurement standard," in Image Quality and System Performance, edited by Yoichi Miyake and D. René Rasmussen, Proceedings of SPIE-IS&T Electronic Imaging, Vol. 5294, pp. 181 - 189, 2004.

[11] ISO 20462-3:2012 Photography -- Psychophysical experimental methods for estimating image quality -- Part 3: Quality ruler method.

[12] E.W. Jin, B.W. Keelan, J. Chen, J.B. Phillips and Y. Chen, "Softcopy quality ruler method: Implementation and validation," Proc. SPIE 7242, 724206, 2009.

[13] E.W. Jin and B.W. Keelan, "Slider-Adjusted Softcopy Ruler for Calibrated Image Quality Assessment", Journal of Electronic Imaging, 19(1), 011009 (Jan-Mar 2010), 2010.

[14] B.W. Keelan, "Predicting Multivariate Image Quality from Individual Perceptual Attributes", Proc. IS&T's PICS 2002 Conference, Portland, Oregon, Society for Imaging Science and Technology, Springfield, Virginia, 82–87, 2002.

Author Biography

Elaine Jin received the B.S. and Ph.D. degrees in Optical Engineering from Zhejiang University and her Ph.D. degree in Psychology from the University of Chicago. She has worked in the imaging industry for 15+ years, including employment at Eastman Kodak Company, Aptina Imaging, Intel Corporation, and Google. Currently she is an image quality manager at Nvidia. She served as the IQSP conference chair at EI18. She is a major contributor to the IEEE P1858 CPIQ Standard.

Yixuan Wang received the Master's degree in color science at Rochester Institute of Technology. Her research involves multi-spectral imaging, image processing and image quality. She is now working in Apple as a camera image quality engineer. Her research interests include image quality, human vision and perception, and camera 3A.

Wentao Liu received the B.E. degree and the M.E. degree from Tsinghua University, Beijing, China in 2011 and 2014, respectively. He is currently a Ph.D. candidate with the Electrical & Computer Engineering Department, University of Waterloo, Canada. His research interests include perceptual quality assessment of images and videos, image aesthetics, and quality of experience.