https://doi.org/10.2352/issn.2169-2629.2019.27.35

Perceptual Estimation of Diffuse White Level in HDR Images

Fu Jiang, [*] ***Mark D. Fairchild,*** [*] ***Kenichiro Masaoka*** [**]***;*** [*] ***Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, New York USA;*** [**] ***NHK Science & Technology Research Laboratories, Setagaya, Tokyo/Japan***

Abstract

Diffuse white is an important concept in all color appearance models. However, there is a lack of research directly about the perceived diffuse white in HDR images. Three experiments were conducted to explore the perceptual estimation of diffuse white in real HDR images when presented on displays. The first experiment showed that the perceptual estimation of diffuse white relied more on the image content than the clipped peak luminance levels. Experiment II used images with different neutral density filters rather than clipping. The normalized luminance levels of the perceptual estimation were similar, depending on image content but not on image absolute luminance levels. Moreover, in both experiments, no significant difference can be found between expert and naive observers. The variance across images agreed with each other between experiment I and II. However, both results demonstrated that the absolute luminance level of observers' estimation is higher than the calibrated diffuse white level. Experiment III focused on exploring the impact of measurement methodologies on the absolute luminance level of the estimated diffuse white. Results of experiment III verified that the absolute luminance level of the perceptual estimation can be manipulated by the measurement methodology but the variance across image content is stable.

Introduction

It is known that chromatic adaptation is an important mechanism of human visual perception. In most current color appearance model, from the simplest *CIELAB* to *CIECAM02*[1], or any other color appearance models [2], chromatic adaptation transforms require one to adopt the "white" as an adapting point, or anchor. The "white" here refers to diffuse white / matte white, which is the white level from a perfect diffuser. Usually for the reflective objects, the uniform neutral matte object with highest possible reflectivity is considered as diffuse white in all current color appearance models. It should be noted that the diffuse white/matte white should be separated from the highlight white, which is a highly reflective specular object and often perceived as a light source or self luminous. The bidirectional reflectance distribution function (BRDF) is used to precisely describe the material property with respect to how "matte" and how "specular" it is.

Often the peak luminance of the display is used as diffuse white in display characterization. In the classic *Berns'* and *Days'* display colorimetric methods [3–5], the peak of the display was used as diffuse white in the colormetric color difference error minimization. It is a reasonable assumption as in standard dynamic range (SDR) display peak luminance is normally close to the diffuse white when presenting real images. For many current HDR displays, a special "white" optical channel is added to achieve a higher peak luminance. The peak luminance is usually used to demonstrate the highlights, light sources, or bright high-chroma colors. For these HDR displays, it would not be a reasonable assumption that the peak luminance is the diffuse white in characterizing the display or image appearance. Moreover, in evaluating 3D color gamut volume of the display, the diffuse white is important for all color appearance models. Therefore, determining the diffuse white is a fundamental factor in calculating the 3D color gamut volume. In *Masaoka's* initial research about 3D color gamut volume of HDR displays, different luminance levels were used as "white" in the simulations [6]. In *Jiang, et al.'s* following research about building a mathematical model predicting 3D color gamut volume of HDR wide-color-gamut (WCG) displays, two assumptions about "white" level were made, constant 200 nits and 20% of the peak luminance [7]. Additionally, ITU-R made a recommendation for using around 20% of peak luminance in HLG encoding system for 1000 nits peak luminance HDR WCG displays [8]. Additonally, *Baek* published the first study about the perceptional color gamut volume [9]. In *Baek's* experiment, two different "white" levels were used in calculations and compared with experiment data for side-by-side image evaluation. This means two different "white"s were used in a single frame on one display, which is questionable. *Jiang, et al.* presented another study about the perceptual color gamut volume through a psychophysical experiment [10]. In *Jiang's* experiment, the paired images were presented one by one separately with a few seconds interval. Theoretically, different diffuse white levels could be used for the two images. However, the perceptual estimation of diffuse white did not show too much difference between the pair of the same image content and different peak luminance levels [10].

Three experiments are included in this paper. The first and the second experiment focused on the perceptual evaluation of diffuse white level and the third experiment showed the impact of measuring methodologies on the absolute luminance level of the estimated diffuse white when HDR images were presented on displays. Experiment I was partially reported in [10]. It should be noted that our experiments are different from the research about lightness perception in HDR images. *Allred's* research focused on mapping HDR to a series of lightness, where the highest patch was always perceived as white [11]. This paper explores diffuse white in HDR images when presented in displays and comparing that with the calibrated diffuse white in HDR images.

Experiment Design

Experiments I and II adopted the same configuration as shown in Figure 1, a SONY 30 inch Trimaster 4K display with two black plates blocking any ambient lights. This SONY display could reach a stable 1000 nits peak luminance. However, the SONY display could not reach full screen 1000 nits due to the Average Picture Level (APL) limitation. In our experiments, the APLs of the images are low enough that the display could reach a stable 1000 nits peak luminance. A LG 55-inch OLED TV was used in experiment III, Figure 4. In all experiments, the observers were sitting in a dark room with curtains or stands blocking the flare.

Figure 1: Configuration for experiment I and II, two black matte stands were used to block any possible flare onto the display.

Figure 2: 6 images used for the three experiments.

Figure 3: Example of image#1 with five level gray patches on the center with black padding.

Experiment I

Images for Experiment

Figure 2 showed the images used in experiment I. All the 6 images were taken with a well-calibrated HDR camera. During capture, a color checker was used to calibrate the exposure. The exposure was following the ITU-R suggestion that the diffuse white should be calibrated to 200 nits for 1000 peak luminance. All 6 original images would have a 1000 nits peak luminance when shown on the SONY display. The peak luminance of these 6 images was clipped to the additional three different experimental levels: 500 nits, 300 nits, and 200 nits. Therefore, there were 24 images for evaluation. A five-level methodology was used in this experiment. In each image five gray patches were added at the center. Figure 3 illustrates an example of the five levels. The five level patches were padded with black, with width of 20% of the patch size. The absolute luminances of the five levels were: 176 nits, 285 nits, 439 nits, 640 nits, and 879 nits.

Observers

In total, 21 observers participated in the experiment, including 10 expert observers and 11 naive observers. The expert observers were identified by their background knowledge of a perfect diffuser/lambertian surface in radiometry. They were graduate students majoring in imaging science or color science, and faculty members in related area.

Observer's Task

Each observer was asked to use a keyboard indicating their estimation of the diffuse white. They were asked to pick two consecutive levels, where their estimation lay in between (or pick the same level twice if they believe their estimations were very close this level). All observers were instructed regarding the difference between diffuse white and specular white/highlight white before the experiment.

Experiment II

Images for Experiment II

The same 6 original images, Figure 2, were used in experiment II. For each original image, three different neutral density filters were used for scaling the whole image luminance level. The transmittances of the three filters are: *0.7*, *0.5*, and *0.35*. Including the original one, each image generated 4 images, in total 24 images for observation. As discussed before, these images were calibrated with 200 nits diffuse white level when shown on the display. Therefore, for different filters, the calibrated diffuse white would be 200 nits, 140 nits, 100 nits and 70 nits when shown on display.

Observers

20 observers participated in experiment II, 10 expert observers and 10 naive observers. 10 expert observers were graduate students / faculty members from imaging science or color science.

Observer's Task

The same five-level methodology was used in experiment II. The same five levels were used. As in experiment I, observers were asked to pick two levels out of five. The five levels were also scaled by the filters. All observers were instructed to the difference between diffuse white and specular white/highlight white before the experiment.

Experiment III

Figure 4: Configuration for experiment III.

Images

The aim of experiment III is evaluating the impact of the measuring methodology. Therefore, two different methodologies were adopted. In addition to the five-level method in experiment I and II, a single patch with method of adjustment was also used in this experiment. Figure 4 shows the configuration of this experiment. Due to the average picture level limitation of the TV, only 25% of the whole TV was used for presenting the image, and the remaining area was set to black. Also the example in Figure 4 is an example of the single patch method, where one adjustable patch was added on the center of the image. Observers were sitting in a dark room during the experiment. Observers were seated about 120 cm from the TV, during the experiment.

In this experiment, only images#1, #2, and #4 were used. For the five-level method, the absolute luminance levels of the five patches are 195.9 nits, 255.2 nits, 335.8 nits, 423.3 nits, and 573.9 nits. For the single patch method, 10 levels were provided: 150 nits, 175.8 nits, 205.8 nits, 240.8 nits, 281.9 nits, 330.6 nits, 386.8 nits, 452.5 nits, 529.0 nits, and 598.6 nits.

Observers

Only expert observers participated in this experiment. Results from experiment I and II showed not much difference between expert observers and naive observers. Therefore, graduate students / faculty members from color science were recruited for this experiment.

Observer's Task

All observers participated in both methods. For the five-level method, the observers were asked to pick two levels out of the five the same as in experiment I and II. For the single-patch method, observers could adjust the level closest to their estimations of the diffuse white.

Result and Analysis

All the results and analysis of the three experiments are presented in this section. In experiment I and II, the average result and a comparison between expert observers and naive observers are analyzed. In experiment III, comparison between different methods is demonstrated.

Result of Experiment I

Since it is well known that the perceived lightness is more linear to the log scale of the absolute luminance level. All the plots and analyses will be on a log scale of the absolute luminance level. This corresponds with the span of the five levels, which are almost linear in log scale. Figure 5 plots the overall average result of experiment I. For each image, there are four corresponding data points of different peak luminances, 1000 nits, 500 nits, 300 nits and 200 nits as shown in legend. The five solid lines represent the five provided patch levels. Each data point is plotted with plus-and-minus one standard error bar.

Firstly, it can be seen that the variance across the six images is bigger than the variance across 4 different peak luminances. For images#1, #2, #3, and #6, the data of all four different peak luminances are within each other's one standard error. For image#4, result of 1000 nits peak luminance is higher than that of 300 nits and 200 nits peak luminances. For image#4, there is not an obvious white object in the image, which means there is no white reference when observers made estimations. Therefore, a bit larger variance of this image than the others is reasonable. For image#5, the estimated diffuse white of 200 nits peak luminance is lower than that of 500 nits and 300 nits peak luminances. In image#5, the white sweater includes the white objects with a gradient luminance changing. The white object is used by the observers as reference when the observers made their estimations. The texture/detail of that is affected when the image was clipped into different peak luminance levels. When used as a reference, the white object losing texture would have an impact on the estimation. This assumption would need more experiment to verify.

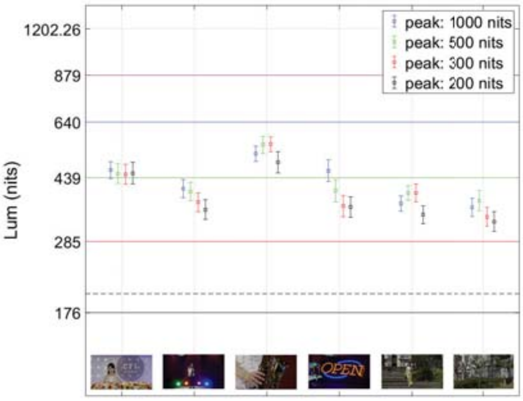

Figure 5: Result of experiment I. Vertical axis is the absolute luminance in log scale. Gray, red, green, blue and magenta solid lines represent the five gray patch levels. Dashed line represents the calibrated 200 nits diffuse white level.

The 21 total observers were categorized into expert observers and naive observers. Comparison between the two groups is made to determine if any expertise would affect the result. Figure 6 plots the result of 10 expert observers on the left and that of 11 naive observers on the right. Firstly, results of both group showed agreement with the overall group result that the variance across the 6 images is larger than the variance of 4 different peak luminance levels of the same image content. Therefore, this conclusion, stability of diffuse white across different peak lumiances, is valid regardless of expertise. Moreover, the trend across images is similar between the expert observers and the naive observers, as well as that in Figure 5. However, the expert observers showed slightly smaller variance over different peak luminance levels in images#5, and #6 than naive observers. The variance across different peak luminance levels in image#4 is larger of the expert observers than that of the naive observers.Therefore, no significant difference can be found between the expert observers and the naive observers. This suggests that expertise of the "diffuse white" would not necessary keep the result more consistent.

Result of Experiment II

Firstly, it should be noted that in this experiment II ND filters scaled whole images. Therefore, the image content did not change but the luminance of the image changed with different filters. Figure 7 showed the overall average result of experiment II of the 20 observers. The vertical axis is log scale of the normalized luminance level, which was normalized by the peak luminance. Again,

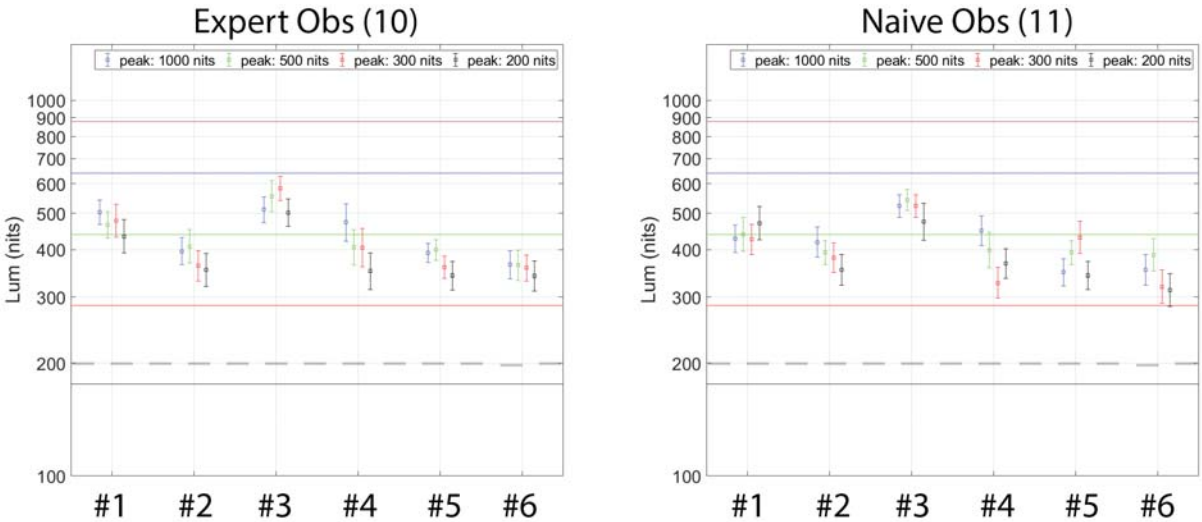

Figure 6: Comparison between 10 expert observers' result (left) and 11 naive observers' result (right) of experiment I.

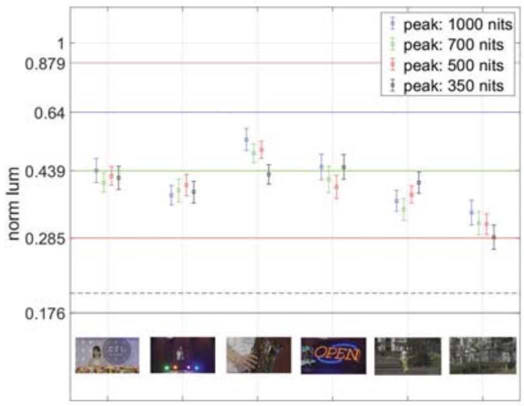

Figure 7: Result of experiment II. Vertical axis is the log scale of normalized luminance level. Gray, red, green, blue and magenta solid lines represent the five gray patch levels. Dash line represents the calibrated diffuse white level, 20% of the peak luminance.

the five solid lines represent the five patch levels. Each data point is the mean of the group chosen diffuse white levels and error bar is one standard error. It can found that variance across images is bigger than the standard error within each image. This agrees with the conclusion from experiment I. Also the general trend of chosen diffuse white across images agrees with that from experiment I, highest in image#3, followed by images#1, #2, #4, then #5, and #6. This again supports that the estimated diffuse white is more relatively an image-based choice. The standard errors of all different ND filters of all images are very close. This standard error is very close to that from experiment I as well.

In this experiment, 20 observers can be divided into two groups: 10 expert observers and 10 naive observers. Figure 8 showed the comparison between the two groups. Comparison showed similarity between the expert observers and the naive observers in general trend, and that variance across 6 images is larger than that within the images of 4 different ND filters. For different image contents, the expert observers showed even larger variance across different ND filters on image#5 than the naive observers, and slightly smaller on the rest images. The mean estimated diffuse white levels of the expert observers are also very close to that of the naive observers for all images.

Result of Experiment III

Figure 9 plots the results of 9 expert observers. The blue line and the blue markers indicate the result of the five-level method, the red line and the red markers indicate the result of the single patch method. All markers were also plotted with one standard error. The vertical axis is the absolute luminance in log scale. Both the mean and the standard error were calculated in log scale. Firstly, it can be found that the results of the two methods showed similar trends across images, that the result of image#1 are close to that of image#2 and significantly higher than that of image#4. However, in terms of the absolute luminance level, the result of the five-level method is higher than that of the single patch method. The absolute luminance level is around 350 nits for images#1 and #2, and 300 nits for image#4. For the single patch method, for all three images the absolute luminance level is around 70 nits lower than that from the five-level method. The result showed that the absolute luminance level of the estimation depends highly on the measurement method.

The diffuse-white level of all the three images is calibrated to 195 nits when presented on this TV. That was 200 nits for experiment I. They are very close. So it is reasonable to compare the results of this five-level method with the result from experiment I directly. It can be found that the mean value in this experiment is slightly lower than that from experiment I. This could be caused by the difference between the provided absolute luminance levels in the two experiments. In experiment I, the brightest patch is much brighter than that in this five-level methodology. The simultaneous contrast could impact the overall perceived lightness of the five levels if different five levels were provided. Also the standard error of this experiment in Figure 9 is much smaller than that from experiment I (Figure 5). This is more due to the fact that

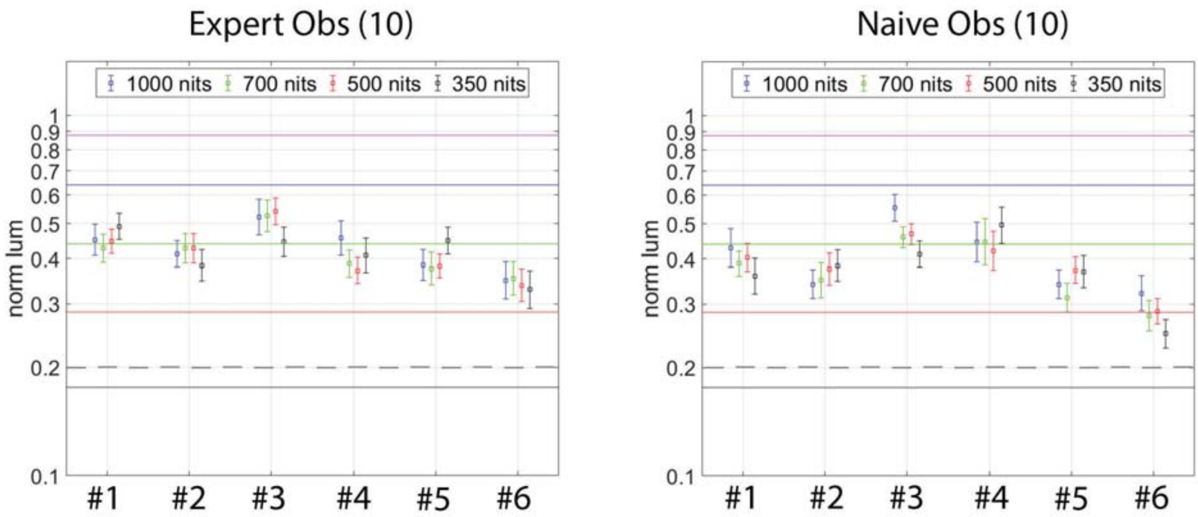

Figure 8: Comparison between 10 expert observers' result (left) and 10 naive observers' result (right) of experiment II.

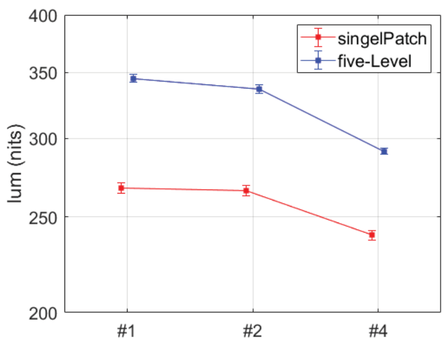

Figure 9: Comparison between result of single-patch adjustment methodology and the five-level methodology for the image#1, #2, and #4.

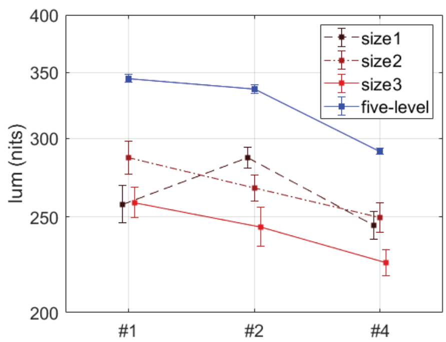

Figure 10: Comparison of result of single-patch adjustment with three different patch sizes. *size 1* is twice of that in five-level as in Figure 3, *size 2* is four times of that, and *size 3* is six times of that.

the range of the five levels in experiment III is much smaller than that from experiment I. Therefore, the standard error is smaller as the chosen levels were closer to each other.

Additionally, for the single patch method there are three different sizes of the patch, which are twice of that in the five-level method (*size 1*), four times of that (*size 2*) and six times of that (*size 3*). The comparison between the three sizes and the five-level method is plotted in Figure 10. Firstly, for all three different sizes, all the mean data are lower than that of the five-level method. Within the same image, more variance across different sizes is found for image#2 than image#1 and #4. This is an interesting result. With the patch size increasing, the patch is covering more and more the white shirt of the lady in image#2 (see Figure 11) while not covering any white reference in image#1 or image#4. According to the feedback from the observers, their priority in making estimations/decisions are using a white object as reference. Therefore, covering the white object in the scene would result in a larger variance in image#2. This is also an attempt of exploring the area-lightness effect in complex images from *Gilchrist* [12].

Discussion & Conclusions

Observers' estimations of the diffuse white on the HDR images were explored in three experiments. In experiment I, only the highlights/bright high chroma colors were clipped while the most areas were kept the same. In experiment II, each image was scaled by the ND filters. The ND filter scaled the whole image. In both experiments results, the mean data varied more across the image than different peak luminances or ND filters. Also the general trend across the six images agrees with each other in the two experiments. R^2 between the mean of experiment I and experiment II is 0.79. This validates that the observers' estimation is more image-based than peak luminance or scales. Moreover, the observers were categorized into two groups: expert observers and naive observers. Almost no difference be found by between the two groups in both experiments. However, the expert observers' result showed slightly smaller standard error than that of the naive observers. Expert observers' estimations were slightly higher than naive observers.

The result from experiment I showed that clipping the peak luminance, keeping most matte intact, has very limited impact on observers' estimation of diffuse white level. This tells that

Figure 11: Preview of original image#2 (top left), with *size 1* patch (top right), with *size 2* patch (bottom left), and *size 3* patch (bottom right).

observers' estimation depends more on the matte objects, which corresponds with feedback from observers about their strategies. Results from experiment II showed that observers can adapt completely even 0.3 ND filter, which has 60 nits calibrated diffuse white level. In experiment III, the comparison between different methods demonstrated that the absolute level of observers' estimation changes with measurement methods. The result of the single patch method is almost 70 nits lower than that from the five-level method, and it is closer to the calibrated diffuse white level 200 nits. The big difference between the single patch method and the five-level method can be explained by the simultaneous contrast effect [13]. Very bright patch in the five-level method makes the perceived lightness of the rest patches much lower compared with the single patch method. So, if the probable measurement methodology, size location and texture of the patch, is used, there is a high possibility that the estimation can be even closer to the 200 nits. For choosing a diffuse white level in image appearance calculation/color gamut volume, a constant level, the calibrated diffuse white level if the image/video has, can be used, even recommended as a standard for industry practical usage. Due to the peak luminance limitation of the displays, in experiment I and II, the maximum peak luminance is 1000 nits and the peak luminance is around 700 nits. As for the displays having a higher peak luminance, the conclusions would need a verification experiment.

References

[1] Moroney, N., Fairchild, M. D., Hunt, R. W., Li, C., Luo, M. R., and Newman, T., "The ciecam02 color appearance model," in [*Color and Imaging Conference*], **2002**(1), 23–27, Society for Imaging Science and Technology (2002).

[2] Fairchild, M. D., [*Color appearance models*], John Wiley & Sons (2013).

[3] Berns, R. S., "Methods for characterizing crt displays," *Displays* **16**(4), 173–182 (1996).

[4] Berns, R. S. et al., [*Billmeyer and Saltzman's principles of color technology*], Wiley New York (2000).

[5] Day, E. A., Taplin, L., and Berns, R. S., "Colorimetric characterization of a computer-controlled liquid crystal display," *Color Research & Application* **29**(5), 365–373 (2004).

[6] Masaoka, K., "CIELAB-Metric Color Volume for HDR Displays," 1040–1043 (11 2017).

[7] Jiang, F., D. Fairchild, M., and Masaoka, K., "Estimation of HDR WCG Display Color Gamut Volume," *Color and Imaging Conference* **2018**, 344–349 (11 2018).

[8] ITU-R, "Operational practices in hdr television production," Tech. Rep. BT.2048, International Telecommunication Union Radiocommunication, Geneva (June 2017).

[9] Baek, Y. S., Kwak, Y., Seo, S., and Lee, J., "78-4: Visual evaluation of displays 3d color gamut volume," in [*SID Symposium Digest of Technical Papers*], **49**(1), 1052–1055, Wiley Online Library (2018).

[10] Jiang, F., Masaoka, K., and Fairchild, M. D., "Subjective Verification for Color Gamut Volume of HDR WCG Display," in [*International Display Workshop*], 1026–1029, International Display Workshops Incorporated Association (2018).

[11] Allred, S. R., Radonjić, A., Gilchrist, A. L., and Brainard, D. H., "Lightness perception in high dynamic range images: Local and remote luminance effects," *Journal of vision* **12**(2), 7–7 (2012).

[12] Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., and Economou, E., "An anchoring theory of lightness perception.," *Psychological review* **106**(4), 795 (1999).

[13] Hurvich, L. M., "Color vision," (1981).

Author Biography

Fu Jiang is a third-year PhD student from Munsell Color Science Laboratory, Rochester Institute of Technology (RIT), USA. Fu received his MSc in imaging science from RIT in 2016. He works on display perception, mainly HDR display perception for his PhD project.

Mark Fairchild is Professor and Founding Head of the Integrated Sciences Academy in RIT's College of Science and Director of the Program of Color Science and Munsell Color Science Laboratory. Mark was presented with the 1995 Bartleson Award by the Colour Group (Great Britain) and the 2002 Macbeth Award by the Inter-Society Color Council for his works in color appearance and color science. He is a Fellow of the Society for Imaging Science and Technology (IS&T) and the Optical Society of America. Mark was presented with the Davies Medal by the Royal Photographic Society for contributions to photography. He received the 2008 IS&T Raymond C. Bowman award for excellence in education.

Kenichiro Masaoka is a Principal Research Engineer at NHK Science and Technology Research Laboratories, Tokyo, Japan. He received his Ph.D. in Engineering from the Tokyo Institute of Technology in 2009. He worked with Professors Mark Fairchild and Roy Berns for a six-month residency as a Visiting Scientist at the Munsell Color Science Laboratory at the Rochester Institute of Technology (RIT) in 2012. His research interests include color science and digital imaging systems. In 2017, he received SID's Special Recognition Award for his leading contributions to the research and development of a wide-color-gamut UHD-TV display system and gamut-area metrology.