An Assessment of Reference Levels in HDR Content

Erik Reinhard

Technicolor, Rennes, France.

Jurgen Stauder

Technicolor, Rennes, France.

Michel Kerdranvat

Technicolor, Rennes, France.

Written for presentation at the

SMPTE 2018 Annual Technical Conference & Exhibition

Abstract. *The movie and broadcast industries are gaining experience with high dynamic range (HDR) video technologies, and are starting to produce HDR content at scale. This is accompanied by a learning process, particularly involving questions regarding the use of the extra dynamic range afforded by these technologies. To create further insight into the aspects of HDR content, in this paper a variety of different types of content was analysed. Manual annotation processes are used to determine the range of luminance values that, for example, diffuse white objects or white overlay graphics objects have in this content. We find that for each type of object analysed the mean luminance value is reasonably constant across different types of content, but that the spread of luminance values is remarkably large, and strongly content dependent. Our results may form a basis for understanding HDR content, and contribute toward forming opinions on defining reference levels in reports and standards, including ITU-R Report BT.2408.*

Keywords. High Dynamic Range Imaging, Movie Production, Live Broadcast, Reference Levels, Diffuse White, Graphics White.

1. Introduction

With the introduction of high dynamic range (HDR) imaging technologies into movie production as well as broadcasting, a transitionary period has begun whereby productions are taking place using conventional standard dynamic range (SDR) workflows, and new HDR workflows are developed. This inevitably involves a learning process, where broadcasters and other content producers are trying to figure out how best to create HDR content, make it as interoperable as possible with existing content and workflows, and to create content that will be appreciated by the consumer. Initial color grading experiences tended to produce HDR content with a look that was largely similar to what one would expect standard dynamic range (SDR) content to look like, albeit with an expansion of small highlights. Increasingly, however, grading of HDR content proceeds in two steps: 1) produce a grade that resembles the SDR grade 2) adjust the grade to fill the extra luminance headroom according to creative need.

The increased dynamic range offered by HDR technologies, if used well, has the potential of producing truly stunning visual experiences that will increase the sense of immersion and give consumers a veritable wow-effect. On the other hand, some may fear that content producers may over-use the capabilities of HDR technologies, in the same way that 3D stereo can be used to over-enhance the sense of depth, and in the same way that advertissements can subjectively sound much louder than movies and other shows do. For this reason an interest in how to set luminance levels is burgeoning. Reference levels, for example the luminance level at which diffuse white objects are usually reproduced, are important for setting camera parameters in live broadcast scenarios, but are also of use in the practice of color grading.

Reference levels ultimately have an impact on the look and feel of HDR content. If they are set too conservatively, i.e. too low, then the content may not have the impact that the technology is inherently capable of, and this may affect the success and widespread market adoption of HDR as a technology. If they are set too high, then content may be viewed as too bright, and this may cause sensations of discomfort, which in turn may diminish market adoption. It is therefore prudent to develop sensible best practices in the production of HDR content.

The need for a definition of reference levels was recognized, and some are now reported in ITU-R Report BT.2408[1]. They are based on initial experiences in the production of HDR content, and offer a valuable starting point. In this paper, we offer further experiments specifically addressing two reference levels, namely those of diffuse white and graphics white. These two reference levels are particularly influential on the final look and feel of content, as in essence they determine how light or dark filmed content will be, and how light or dark overlayed graphical elements will be. The experiments presented in this paper may be helpful in thinking about HDR production, so that the content that studios and broadcasters send into the world will hit that sweet spot where it is judged as impactful and immersive, without it being in any way distracting or disturbing.

In the following section, we first discuss diffuse white measurements previously published in ITU-R Report BT.2408, and add measurements on graphics white using the same protocols and datasets. Section 3 introduces a new dataset as well as its statistical features. Sections 4 and 5 then present diffuse white and graphics white measurements made using this dataset. Conclusions are drawn in the last section.

Table 1. Diffuse White Results for PQ and HLG datasets (from ITU-R Report BT.2408).

	PQ Dataset			HLG Dataset		
	R	G	B	R	G	B
cd/m^2	231.8 (76.6 – 665.6	244.2 (80.6 – 703.4)	193.3 (58.9 – 594.1)	222.1 (134.7-373.5)	204.3 (123.6-345.4)	231.3 (141-386.7)
%HLG	77.1	78.1	73.5	76.6 (8.3)	75.0 (8.4)	77.4 (8.2)
%PQ	59.5 (11.3)	60.0 (11.3)	57.6 (12.0)	59.0	58.1	59.4

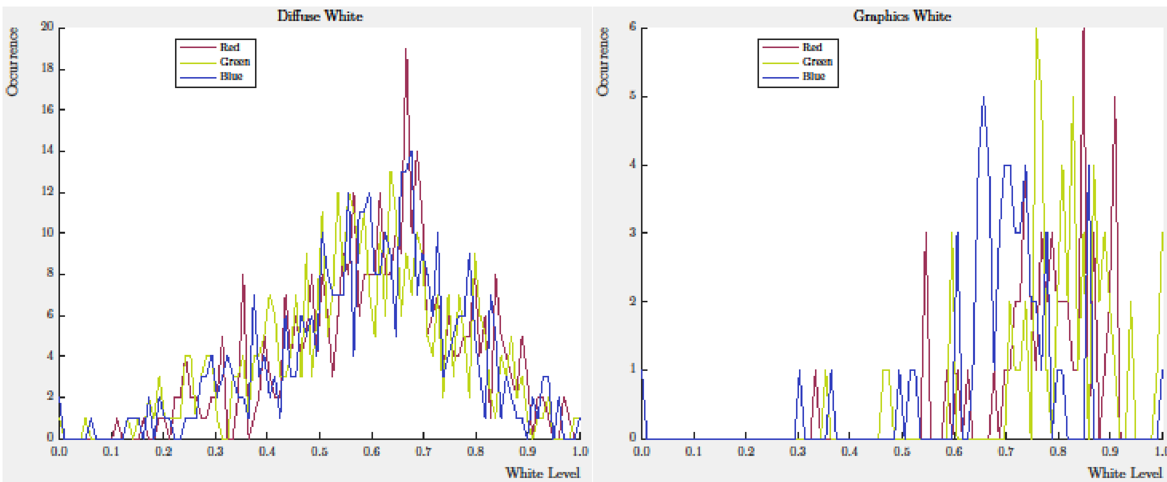

Figure 1. HLG encoded sports event. Left: the distribution of diffuse white (from ITU-R Report BT.2408). Right: graphics white.

2. Diffuse White and Graphics White

ITU-R Report BT.2408[1] presents measurements on diffuse white elements in an HLG[2] encoded set of frames from a live sports broadcast, as well as a PQ[2,3,4] encoded image dataset (a dataset which is documented in ITU-R Report BT.2245[5]). In this section, we briefly summarize the discussion on diffuse white elements of ITU-R Report BT.2408 and add a new set of -measurements on graphics white, obtained from the same HLG encoded sports broadcast.

Diffuse white elements in HLG encoded live broadcast content ("Dodgers Game") were analysed by taking one frame every five seconds, and manually clicking in each frame on patches that represent directly illuminated diffuse white elements, without being over-exposed. The frames were displayed on a Sony BVM 300 professional grading monitor. The total number of analysed frames was 152, and the number of diffuse white pixels identified in this manner is 378. The content was a base-ball game, interspersed with commercials, and containing scenes from a game played in daylight and a game played at night under artificial illumination.

Table 2. Analysis of graphics white in HLG encoded live broadcast content. The %HLG column was measured, while the remaining columns were derived from these measurements (152 frames, 60 points analysed).

	HLG Dataset		
	R	G	B
cd/m^2	631.1 (495.0 – 806.2)	631.5 (498.8 – 802.6)	633.0 (498.6 – 805.2)
%HLG	92.9 (3.8)	92.9 (3.7)	93.0 (3.7)
%PQ	70.2	70.2	70.2

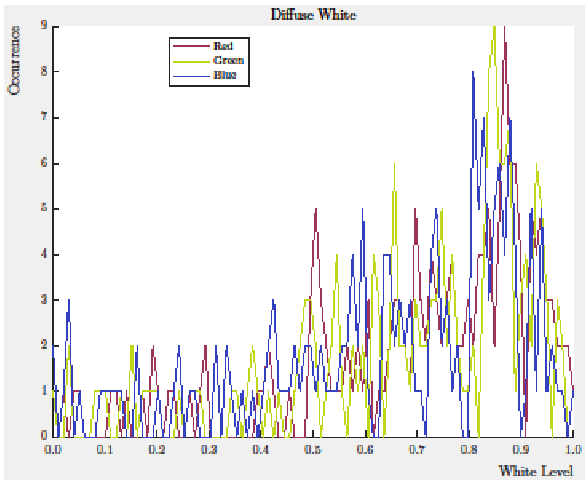

Figure 2. Distribution of diffuse white in a PQ encoded dataset of (from ITU-R Report BT.2408).

The pixels identified in this manner represent values as %HLG (i.e. as normalized HLG code word values). The mean and standard deviation are given in the HLG dataset section of Table 1. These numbers were subsequently converted to cd/m^2 and to %PQ (normalized PQ code word values; see Table 1), assuming a display peak luminance of 1000 cd/m^2. Finally, Figure 1 (left) shows a histogram of the distribution of diffuse white levels for each of the red, green and blue channels, with the horizontal axis indicating values in %HLG.

Graphics white was analysed in the same manner, by providing button clicks in the same set of images. As not all images contained inserted graphics, within the 152 frames analysed, only 60 patches with graphics white were identified. The distribution of values is shown in Figure 1 (right), while the statistics derived from this distribution are shown in Table 2. The %HLG column contains measured values, while the remaining columns were derived from these measurements through calculation.

Using the same method, an additional dataset of 54 PQ encoded EXR images containing diffuse white patches was analysed. The images are graded for a 1000 cd/m^2 display device (a Sony BVM 300). The linear EXR images were first PQ encoded. A total of 169 white patches

Table 3. Features of each group of frames.

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1920x1080	√	√			√		√	√	√	√			√	√	√
1280x720			√	√		√					√	√			
Images	16	10	37	35	18	17	15	136	123	63	38	49	332	114	46
Diff. White	45	25	118	50	60	52	21	88	173	84	117	115	360	6	23
Gr. White	7	-	8	11	10	7	1	37	13	2	13	7	-	2	-
Live sport outdoors					√	√	√	√							
Live sport indoors									√	√	√	√			
Talk show	√	√	√	√											
Movie													√		
Nature														√	
Standup															√

were identified. Their distribution is shown in Figure 2 and the derived statistics are shown in Table 1 (PQ dataset section). In this table, the %PQ column was measured from the pixels that were selected, whereas the columns indicated with cd/m^2 and %HLG were calculated from the %PQ column.

First, diffuse white elements are analysed. For the HLG-based live broadcast content, the observed mean diffuse white level is 75% HLG, which is the same as the recommmended reference level given in ITU-R Report BT.2408. However, the corresponding observed standard deviation was about 8.3% HLG, which - for an assumed 1000 cd/m^2 signal - translates to a range from around 123 to 345 cd/m^2 (i.e. $\mu \pm 1\ \sigma$). This suggests that the diffuse white level in live broadcast content varies significantly.

These results are broadly replicated with the additional data set of 54 PQ encoded EXR images. Here, the mean diffuse white level was determined to be around 60 %PQ, which is close to 58 %PQ as recommended in ITU-R Report BT.2408. The observed standard deviation is 11% (in %PQ), however, which translates to a range between around 80 and 700 cd/m^2 ($\mu \pm 1\ \sigma$). The variability of diffuse white in this dataset is therefore signficant, and it is even larger than measured in the HLG-produced live broadcast content.

Second, graphics white was measured in the HLG-based live broadcast. Graphical elements included logos, score panels as well as various elements displayed during commercial breaks. The mean graphics level was measured to be 93 % HLG, which translates to 632 cd/m^2 under the assumption of a 1000 cd/m^2 signal. The standard deviation is low at 3.7% (%HLG), so that the range is between 400 and 800 cd/m^2 ($\mu \pm 1\ \sigma$). It appears that in the content analysed, graphics white is routinely placed significantly above diffuse white.

3. Reference Levels in SDR and Corresponding HDR Content

Here, another workflow of HDR content production is considered, whereby both SDR and corresponding HDR content exist. A popular and viable route to obtaining HDR content is to

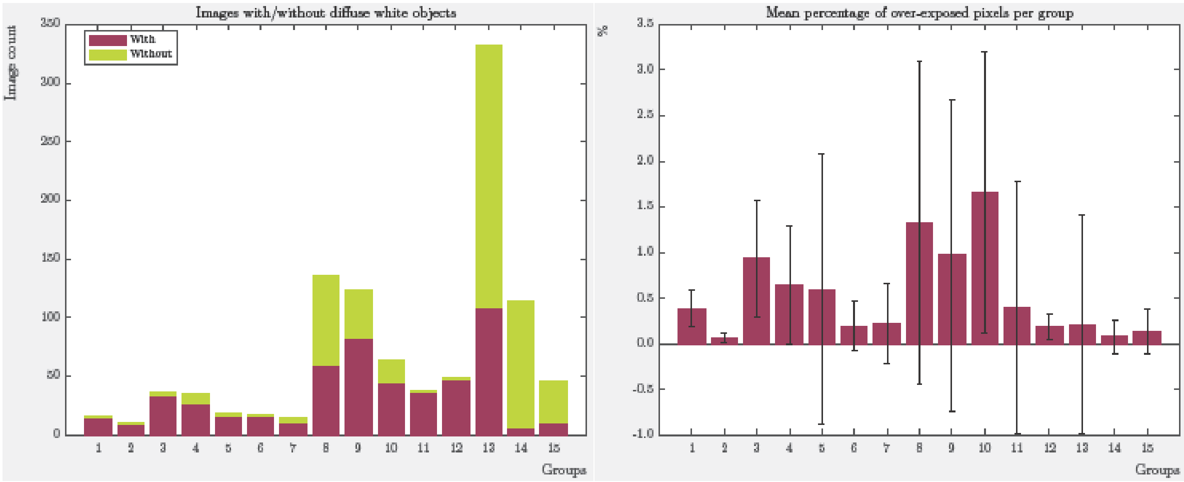

Figure 3. Left: the number of images in each group with and without diffuse white objects being present. Right: the mean percentage of pixels (and standard deviation) considered over-exposed in the SDR dataset. A pixel is over-exposed if its pixel value is 254 or 255.

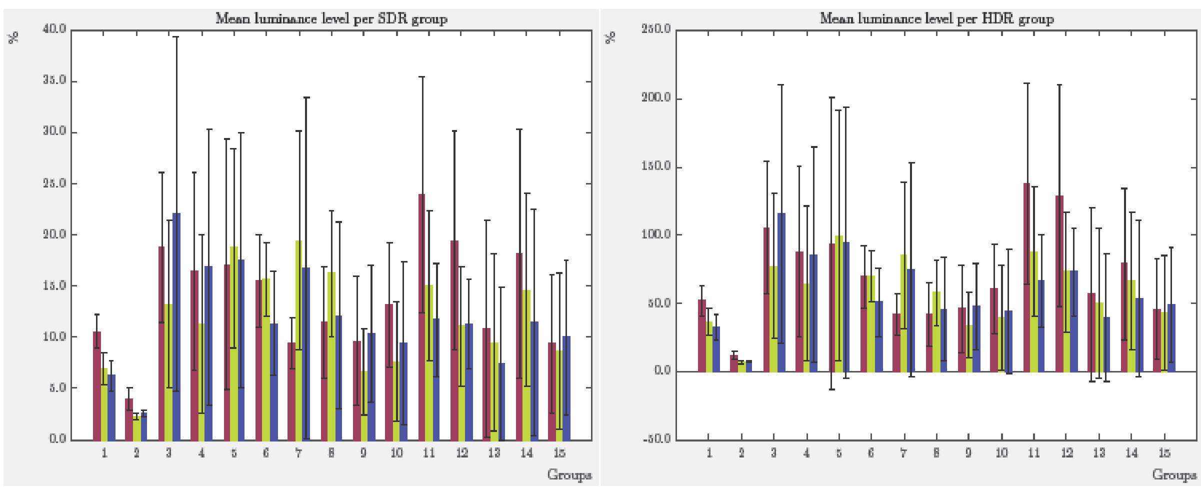

Figure 4. The image means and standard deviations for the SDR and the HDR groups.

apply an inverse tone mapping operator to SDR content[6]. Graded SDR content, for example, can be inverse tone-mapped semi-automatically to produce corresponding graded HDR content, leaving the colorist with at most a small number of corrections to apply. Graded SDR content can so be converted to colorist-approved HDR content. A total of 15 groups of different SDR source materials, comprising a total of 1049 images where converted to HDR by a colorist in this manner. It should be emphasized that although the colorist used an inverse tone mapper, he had the option of making additional corrections, should this be deemed necessary. This gives us a large dataset of SDR frames, with a corresponding set of HDR frames that are amenable to analysis. The SDR frames are graded to 100 cd/m^2, while the HDR frames are graded to 1000 cd/m^2. Thus, for analysis the images are linearized and scaled to 100 and 1000

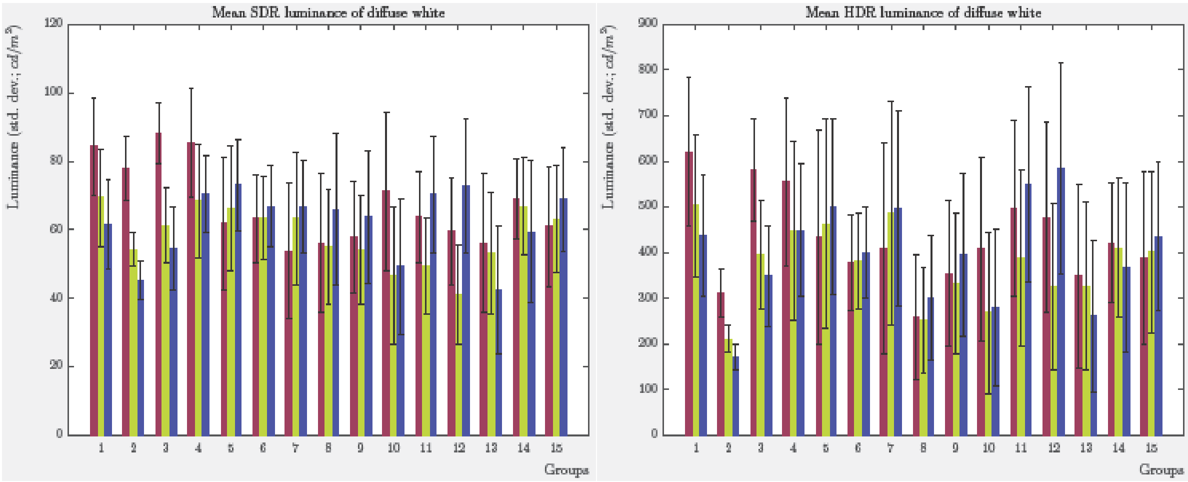

Figure 5. Means and standard deviations of diffuse white objects per SDR image groups (left) and HDR image groups (right).

respectively, using ITU-R BT.709[7] colorimetry. Table 3 lists the characteristics of the content, also showing the number of identified diffuse white and graphics white elements (1337 and 118 elements total, respectively). Measurements were obtained by manually identifying pixels that depict diffuse white objects that were not over-exposed, while still receiving direct illumination. In this case, the SDR source material was used to perform the annotation, and the images were displayed on a standard computer monitor (a Dell U2413). The approach is otherwise identical to the one described in Section 2.

To help characterize the dataset, the number of images that contain diffuse white objects is shown in Figure 3 (left). Only images are shown that have one or more diffuse white objects selected in the experiment. Movie frames (group 13, drawn from different movies) as well as nature shows (group 14) tend to have relatively fewer diffuse white objects than the other groups included in the experiment. The mean percentage and standard deviation of over-exposed pixels were measured in the SDR source materials, as shown in Figure 3 (right). For most frames, especially those coming from movies and nature shows, the percentage of over-exposed pixels is very low, while for some sports shows more over-exposed pixels seem to be inevitable.

Further, the mean luminance in each image was calculated. Figure 4 (left) shows the mean per group (and standard deviation) for the SDR dataset, and Figure 4 (right) shows the results for the corresponding HDR dataset. For the entire SDR dataset, the mean luminance is (13.9, 11.8, 11.8) for each of the RGB channels, while this increases to (70.6, 59.4, 58.7) for the HDR dataset. Figure 4 shows, however, a large variability within and between groups, as indicated by the standard deviations.

4. Diffuse White in SDR and Corresponding HDR Content

In the following, the SDR and corresponding HDR datasets are analysed for diffuse white objects. Figure 5 shows the mean and standard deviation of the measured diffuse white objects

Table 4. Luminance means (standard deviations) of diffuse white and graphics white objects measured over all groups.

	SDR			HDR		
	R	G	B	R	G	B
Diffuse white	64.2 (20.5)	54.8 (17.5)	58.2 (20.9)	414.2 (203.9)	350.6 (180.9)	377.1 (205.2)
Graphics white	78.8 (15.0)	79.3 (14.3)	81.0 (13.2)	567.7 (231.5)	572.7 (231.7)	579.5 (215.1)

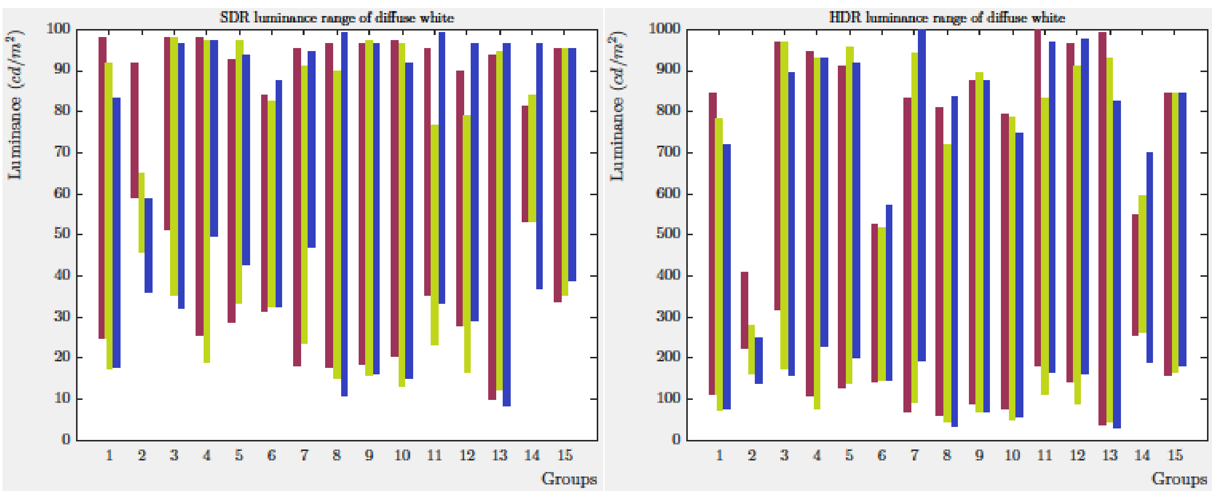

Figure 6. Luminance range of diffuse white objects per SDR image group (left) and HDR image group (right).

per group and per RGB channel for the SDR dataset (left) and the corresponding HDR dataset (right). The means and standard deviations aggregated over all groups are shown in Table 4.

In the SDR dataset, the mean diffuse white luminance is about 5 times higher than the overall mean image luminance, which can be seen as consistent with common photographic practice. In the HDR dataset, the mean diffuse white is around 6 times higher than the mean luminance, which is likely a consequence of the less restrictive nature of the medium.

The sizes of the error bars in Figure 5 indicate that the spread around the mean for diffuse white is significant, for both the SDR and HDR datasets. The full luminance range of diffuse white objects (Figure 6) shows that in extreme cases diffuse white can be placed at almost any luminance value in the available range. This is all the more remarkable, given that only well-exposed and well illuminated diffuse white objects were chosen.

5. Graphics White in SDR and Corresponding HDR Content

Graphics white elements are those that are not captured, but are inserted later. They can include logos, subtitles, score boards, etc. Figures 7 and 8 show the mean, standard deviation and range of graphics white elements for both the SDR and the HDR content. In group 7, only 1

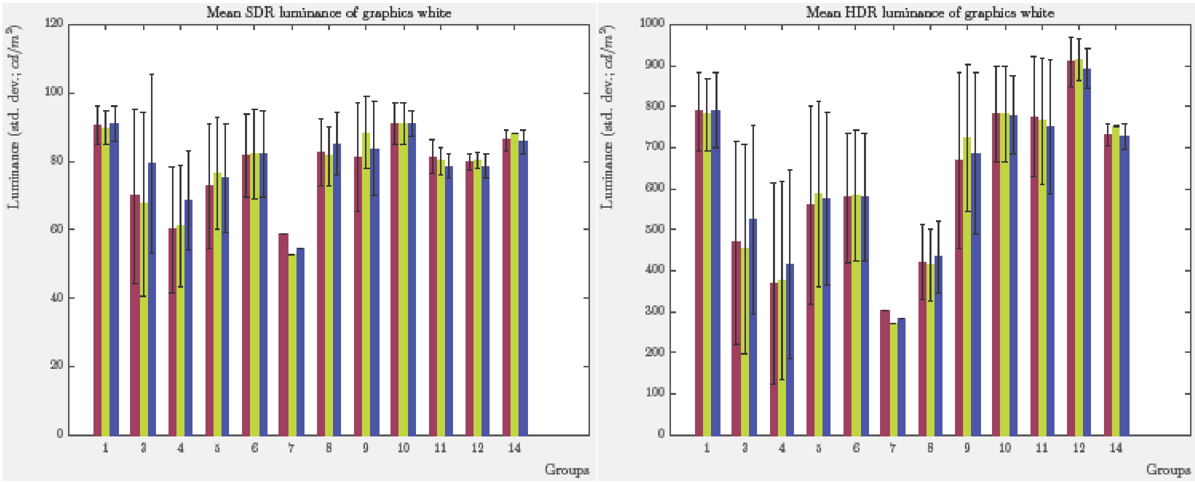

Figure 7. Means and standard deviations of graphics white objects per SDR image groups (left) and HDR image groups (right).

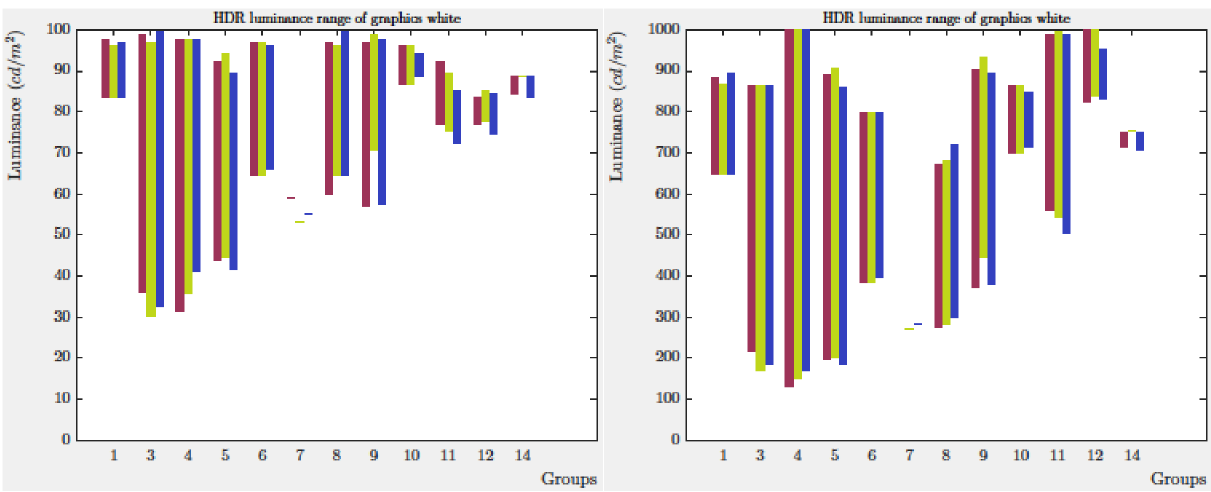

Figure 8. Luminance range of graphics white objects per SDR image group (left) and HDR image group (right).

graphics white element was identified. The aggregate values are shown in Table 4, suggesting that graphics white tends to be placed at higher luminance values than diffuse white. This is so both in SDR content and HDR content, even though the effect is stronger in the HDR content: graphics white is around 1.35 times higher in luminance in the SDR dataset, while it is 1.51 times higher than diffuse white in the HDR dataset. Interestingly, the standard deviation of SDR graphics white is lower than SDR diffuse white, but this order is reversed in the HDR dataset.

The range of values shown in Figure 8 indicates that in particular the minimum values of graphics white are relatively high. They certainly appear to be significantly higher than the minimum values of diffuse white elements, as shown in Figure 6.

6. Conclusion

In this paper, diffuse white objects and graphics white elements in a large database of SDR and corresponding HDR content was analysed. Diffuse white SDR patches were on average reproduced at a luminance of 55 cd/m^2 (as judged by reading the green value in Table 4). The corresponding mean luminance in the HDR dataset is 350 cd/m^2. This is higher than the mean HDR luminance values previously recorded on a PQ encoded set of photographics (244 cd/m^2) and a live HLG sports cast (204 cd/m^2), as discussed in Section 2. This is despite the use of the same experimental design. It suggests that in a color grading studio, at least some willingness is emerging to push diffuse white levels a bit higher. Note that these measured diffuse white levels are also higher than the reference level for diffuse white of 203 cd/m^2 which is currently recommended in ITU-R Report BT.2408.

Acknowledgements

We warmly thank Hassan Guermoud for his valuable help with the dataset used in this work. We thoroughly appreciate the discussions around HDR production practices we had with many, notably including Josh Pines, Josh Limor, Bill Redmann, Kirk Barker, Alan Stein, Marie-Jean Colaitis and David Touzé.

References

[1] International Telecommunications Union – Radiocommunication (ITU-R), Recommendation BT.2408-1, “Operational Practices in HDR Television Production,” 04/2018.

[2] International Telecommunications Union – Radiocommunication (ITU-R), Recommendation BT.2100-2, “Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange,” 07/2018.

[3] SMPTE, ST 2084:2014, “High Dynamic Range Electro-Optical Transfer of Mastering Reference Displays”

[4] S. Miller, M Nezamabadi, S Daly, “Perceptual Signal Coding for More Efficient Usage of Bit Codess”, SMPTE Motion Imaging Journal, 122 (4):52-59, May 2013

[5] International Telecommunications Union – Radiocommunication (ITU-R), Recommendation BT.2245-4, “HDTV and UHDTV Including HDR-TV Test Materials for Assessment of Picture Quality”, 04/2018

[6] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, K. Myszkowski, “High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting”, Morgan Kaufmann: Burlington, 2nd Ed., 2010

[7] International Telecommunications Union – Radiocommunication (ITU-R), Recommendation BT.709-6, “Parameter Values for the HDTV Standards for Production and International Programme Exchange”, 06/2015