Image Quality Evaluation of HDR Displays

Zhengnan Ye[1], Jueqin Qiu[1], Haisong Xu[1], M. Ronnier Luo[2], Stephen Westland[2]

[1] State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China

[2] School of Design, University of Leeds, Leeds LS2 9JT, UK

Abstract

We conducted psychophysical experiments to investigate the impacting factors on the image quality of HDR displays. The result indicated that the OLED display has advantages over IPS and VA LCDs due to its lower minimum luminance level and less pixel interaction, thus being an appropriate choice to display HDR contents.

Keywords

high dynamic range (HDR); display; image quality; psychophysical experiment.

1. Introduction

Demand of accurate and pleasant image reproduction for displays has increased in recent years. Being limited by their small dynamic range, however, common displays can hardly reproduce the actual luminance of real scenes precisely. In response to the call of displaying contents with higher dynamic range and better luminance accuracy, high dynamic range (HDR) displays were developed. Differing from the traditional standard dynamic range (SDR) displays, HDR displays generally have higher peak brightness and lower minimum luminance level, thus providing a wider dynamic range to reproduce more details of images or videos. To reproduce HDR contents, both HDR signal source and HDR display device are necessary. [1] The former provides the scene's real luminance information when capturing, and the latter uses a device independent electro-optical transfer function (EOTF), namely perceptual quantizer (PQ), to convert electric source signal to optical output signal, producing the same luminance as recorded in the HDR contents. Hanhart *et al.* [2] compared HDR displays with SDR displays and pointed out that the former have obvious advantages over those with lower dynamic range. Thereby it is desiderated to investigate the performance comparisons among different HDR displays as well as the external and internal factors that affect the display quality. In this study, the psychophysical experiments were carried out to evaluate the image quality of three HDR displays with different light-emitting mechanism and panel technology, and further to discuss how the image attributes and viewing conditions impact the overall preference of the observers.

2. Experiments

Three displays adopted in this study are all 65-inch HDR TVs, one of which is with organic light-emitting diode (OLED) technology and the other two are liquid crystal displays (LCDs) equipped with VA panel and IPS panel, respectively. Henceforth the three displays are denoted with the abbreviations of OLED TV, LCD TV A (VA panel) and LCD TV B (IPS panel), respectively. The detailed peak and minimum luminance values of the three displays are listed in Table 1, in which it is worth noting that the peak luminance was measured on a small square pattern (10% area of screen) located in the central regions of the displays rather than on a full-screen pattern. To protect its panel, the OLED display is designed to automatically decline the luminance when presenting large areas of bright contents, so the level of peak luminance for OLED display would be somewhat lower if the measurement is conducted on a full-screen pattern. Figure 1 plots the color gamuts of the three test displays along with sRGB and DCI-P3 [3]. The area ratios of OLED TV, LCD TV A and LCD TV B are 126.6%, 126.2% and 125.6%, respectively, with sRGB as reference, and 100.8%, 100.5% and 100.0%, respectively, with DCI-P3 as reference, in CIE1976 u'v' diagram.

Table 1. The peak and minimum luminance levels of the 3 Displays

	OLED TV	LCD TV A	LCD TV B
Peak luminance (cd/m^2)	661.7	1133.5	377.2
Minimum luminance (cd/m^2)	0.0001	0.0029	0.0096

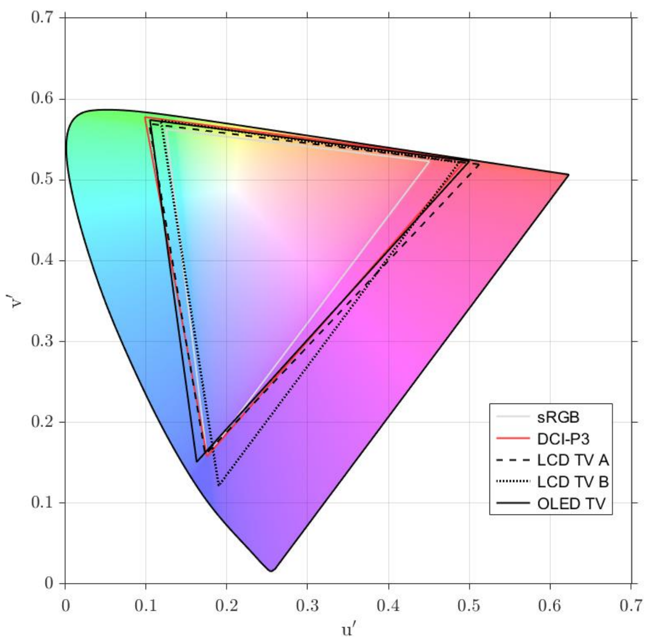

Figure 1. The color gamuts of the three test displays as well as sRGB and DCI-P3 for comparisons at CIE1976 u'v' diagram

To investigate how the viewing condition would influence the image quality and the observers' preference, 4 sessions of visual experiments were conducted with different combinations of viewing angles (front view and 45° side view) and lighting conditions (<5 lx dark and 200 lx ambient lighting). Figure 2 illustrates the experimental setup for the different sessions. The HDR contents were generated and controlled by a PC, and the signal was transferred by a HDR signal generator, then conveyed to the test displays by a splitter. The viewing distances for the three displays were all set to 240 cm, i.e. the three times of the display's short edge length, according to the IEC standard. [4, 5] Meanwhile, in the side view sessions of the experiments, the three displays were rotated counterclockwise by 45° around their rotation axis, as demonstrated by the dashed ones in Figure 2. At the distance of 240 cm, the viewing angles of the front view and side view conditions are 38.0 degrees and 27.3 degrees for the observers, respectively. A total of 6 sets of fluorescent tubes with the correlated color temperature of about 5800K provided an overhead lighting, which resulted in a 200 lx ambient lighting at the centers of the displays' screens.

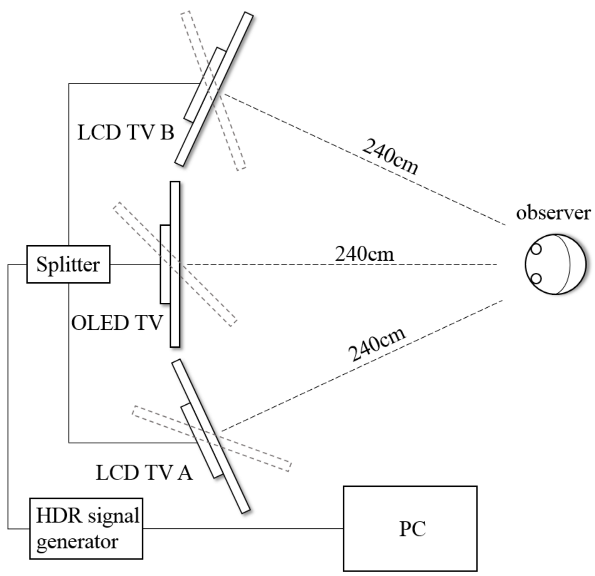

Figure 2. The experimental setup (top view)

A panel of 30 observers, including 6 females and 24 males, with normal color vision aging from 21 to 28 years old, participated in the visual experiments via the psychophysical method of category judgement. For every assessment, a test image was presented simultaneously on the three displays. The observer was asked to compare the displayed contents on the three screens and record two categorical scores for each display: one for a certain image attribute and another for the test image's overall preference, thus for every test image, 6 scores were collected for each observer. The duration for one session was approximately 25 minutes, including a 2-minute dark or ambient lighting adaptation, depending upon the test session being under dark or 200 lx condition, at the beginning of the experiment. The 4 sessions of experiments for every observer were separated into 2 parts to avoid visual fatigue.

In the experiments, 7 types of image attribute as well as the overall preference were evaluated, in which the attributes are briefly described as follows:

Peak brightness: the perceptual brightness of the image's brightest region.

Blackness: the perceptual darkness of the image's black region.

Colorfulness: the saturation of the colored regions of the image.

Gradation: the details of the highlights and dark region.

Contrast: the integral contrast of the image.

Reality: the image's reduction comparing to a real scene.

Artifacts: the image impairments that degrade image quality, e.g. halos and light leakage.

The test contents involved in the experiments included HDR format static images and video. Except for Gradation, each attribute contained 5 individual test images to be evaluated. For Gradation, there were 6 test images, 3 of which were with good highlight details and the other 3 with dark details. Besides the static images, an HDR video with the duration of about 1 minute was also employed to be watched and then evaluated by the observers. To exclude the influence by the presenting order of the test contents, all the images were displayed randomly in the experiments, thus the sequence of the image attributes to be assessed was also random.

In the visual evaluations based on the category judgement method, 9 categorical grades on the scale of subjective feelings were employed for the image attributes and overall preference assessment, according to ITU-T P.910 standard [6], of which a higher grade represents a better performance and a more preferred level. Particularly, the attribute Artifacts has only 5 grades on the scale, of which a higher grade corresponds to a less perceived level of image impairments, and especially the grade 5 means no perceived image impairments.

All the scores typed by the observers were stored in a laptop. In the psychophysical experiments, a total of 26280 raw visual scores were collected, i.e. 3 displays × 4 viewing conditions × 30 observers × (36 test images × 2 evaluations for the attribute and overall preference respectively + 1 test video for overall preference). Then, Z-score was adopted in the data processing so that the categorical scores could be transformed to scale values to remove the bias and differences among the observers, by which the image attributes and the preferences of observers could be clearly quantified and further analyzed.

3. Results and Discussion

The validity of the collected visual data was firstly tested. The coefficient of variation (CV) of the observations was used to describe the level of variability among the judgements made by different observers. The inter-observer accuracy in terms of CV values for various attributes and viewing conditions are shown in Table 2, which maintain a fairly low level through the test, indicating a high consistency among the observers during the whole experiments. It can be found that the CV values of Gradation and Artifacts are slightly higher than those of other attributes, which is mainly due to the extra test image and less judgement categorical grades, respectively.

Then, the scale values of overall preference for the test static images and video with respect to various viewing conditions were calculated, as shown in Table 3. It can be obviously seen that the OLED TV is more preferred than the other two displays, and this tendency is very robust in all situations, no matter with static images or video for front view or side view mode in dark or 200 lx ambient lighting condition. Due to the intrinsic property of VA panel structure, the LCD TV A has a relatively small effective visual angle range, thus its performance severely declines in the side view condition, which is a detrimental factor for indoor displaying. The very similar results under dark and 200 lx ambient lighting conditions also imply that the lighting level has little effect on the HDR display.

Table 2. Inter-observer accuracy in terms of CV values for the tested attributes in all the 4 sessions

	Peak brightness	Blackness	Colorfulness	Gradation	Contrast	Reality	Artifacts	Image Preference	Video Preference
Dark, front view	13.35	12.41	12.71	13.04	12.18	13.33	18.84	13.89	10.20
200 lx, front view	11.87	10.71	12.23	12.22	10.38	12.47	15.41	12.09	9.91
Dark, side view	12.83	12.86	11.66	16.29	13.83	15.19	13.71	14.62	11.06
200 lx, side view	12.25	12.69	12.49	16.22	12.90	14.59	16.54	14.03	11.36

Table 3. The scale values of overall preference for the test static images and video

	OLED TV		LCD TV A		LCD TV B	
	Image	Video	Image	Video	Image	Video
Dark, front view	3.03	3.22	2.20	2.29	2.29	2.12
200 lx, front view	3.09	3.24	2.26	2.35	2.25	2.04
Dark, side view	3.17	3.27	1.62	1.28	2.46	2.37
200 lx, side view	3.22	3.22	1.81	1.39	2.46	2.32

Figure 3 plots the scale values of the 7 tested attributes under the 4 viewing conditions. The trends of the performances for individual attributes are rather similar to that of the overall preference. By comparing the results in Figure 3, it is clear that the ambient lighting condition has merely little affect to the attribute scale values, while the viewing angle largely influences the observers' perception. It can be seen that, for the VA LCD TV A in the side view condition with comparison to the front view mode, the Contrast, Artifacts and Peak brightness performances decline, and the Colorfulness and Blackness perceptions drop more severely, but the change of viewing angle just slightly depresses the scale values of Reality and Gradation. These defects found in the side view condition contributed a lot to the relatively poor preference performance for the VA LCD TV A . However, the IPS LCD TV B shows a steady performance in all conditions, in relative to the VA LCD TV A. From the comparisons between Figure 3 (a) and (c), or Figure 3 (b) and (d), it can be seen that the superiority of OLED TV over the other two LCD TVs for most image attributes, especially for Peak brightness and Contrast, becomes more evident. It is interesting to note that the scale values of Peak brightness for OLED TV under side view mode are even higher than those under front view condition, given the fact that the actual luminance of the screen does not increase. Besides, the t-test results also support that the attribute scale values as well as overall preference of the OLED display have a significant difference in comparison with those of the other two LCDs.

The inputs of the three test displays were from a same HDR signal generator, thus this subjective diversity must be caused by their different physical parameters. Compared with the relatively high minimum luminance level of the LCDs (0.0029 cd/m^2 for VA LCD TV A and 0.0096 cd/m^2 for IPS LCD TV B), the OLED TV has a perfect minimum luminance level, which is as low as 0.0001 cd/m^2, thus its perceived overall contrast and dynamic range are rather high. This superiority of the OLED display can be greatly enlarged when evaluating the attributes of Blackness and Contrast. The LCDs use LED backlights to illuminate the panel, which would cause light leakage and pixel crosstalk. These defects of the LCDs lead to a local contrast degradation, while the self-illuminated OLED panel avoids these problems. The performance of Peak brightness also shows the superiority of the OLED display: a higher local contrast and less image impairments make the perceptual luminance of highlight regions being brighter than its actual level. Though the VA LCD was physically brighter, it was visually dimmer than the OLED display because of the degradation of local contrast caused by the halos and glare on its screen (the proper luminance for displays to avoid glare is 600 cd/m^2) [7]. Although the areas of color gamut of the three displays are very near, as can be seen in Figure 1, the visual results of Colorfulness show a big difference among them, which may also be caused by the different perceptions of visual contrast of the test displays.

It was noticed that the dark regions of VA LCD TV A were brighter than those of the other two displays, resulting in its poor performances of Contrast and Blackness. For the VA LCD, when the displayed images contain both bright and dark contents, the luminance of backlight has to be set to a high level to provide enough details on the bright region. In addition, because of the leakage, the dark pixels, especially those on the edges of bright and dark regions, would also be lit and so has higher luminance values than what they actually should be. On the other hand, for those test images in the Gradation group, containing large areas of highlight contents, the OLED display was found to be inferior to the VA LCD mainly due to its power consumption limit, which caused a decline of integral luminance and the loss of bright details in the presented images.

Based on the achievements from this study, the multiple image attributes for a single test image will be visually evaluated so that the relationships of image attributes and their contributions to the overall preference would be deeply explored in the future. Furthermore, the statistical method like factor analysis could be applied to revealing the key factors that impact the overall image quality of HDR displays.

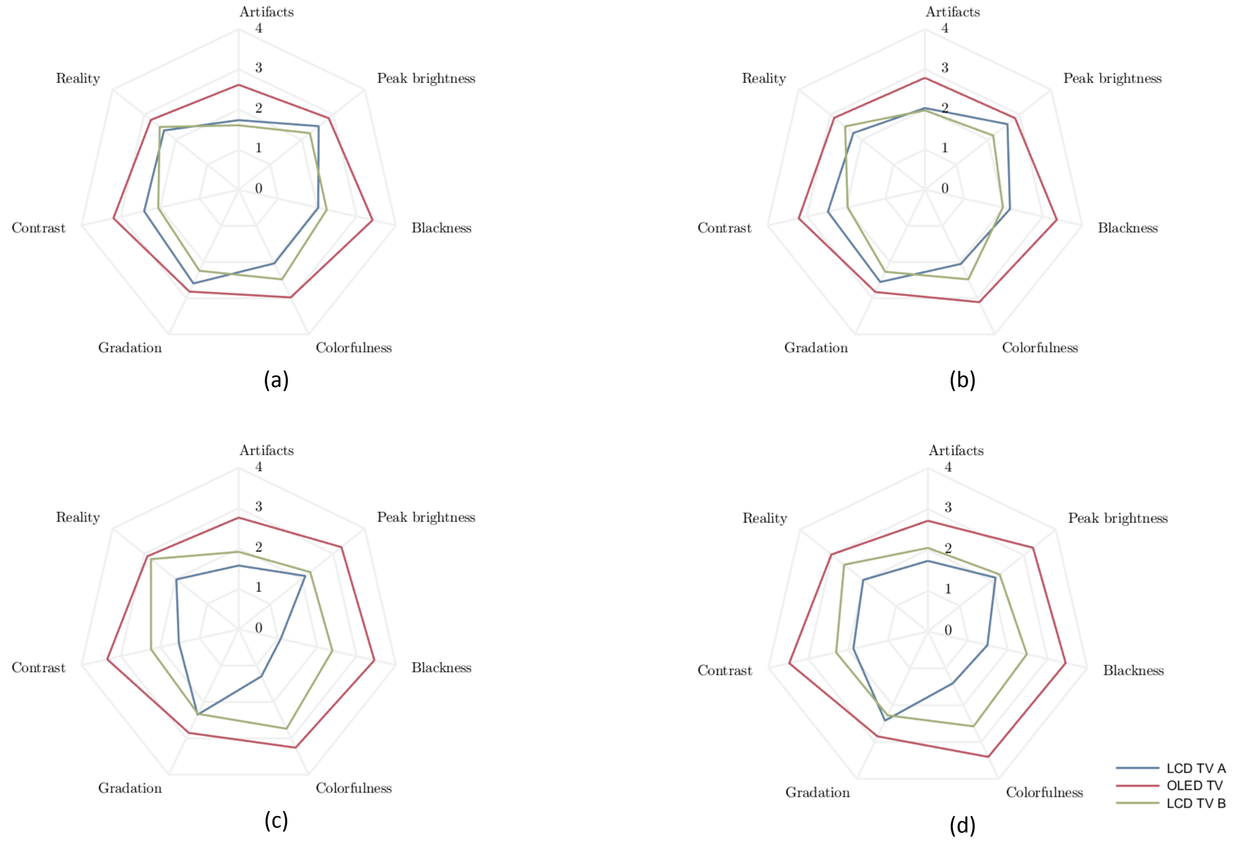

Figure 3. The scale values of individual image attributes under the viewing conditions of: (a) dark, front view, (b) 200 lx, front view, (c) dark, side view, and (d) 200 lx, side view

4. Conclusion

Using 3 large size HDR displays with different light-emitting mechanism and panel technology under various viewing conditions, four sessions of psychophysical experiments were performed to visually evaluate the image quality of HDR displays. The subjective results indicate that the lower minimum luminance level and the absence of image impairments could effectively improve the image quality of the OLED HDR display. Herewith, it is verified that the OLED display is most suitable for presenting HDR images as well as videos among the tested display devices due to its high local contrast, excellent black level, superior color saturation, wide viewing angle, and almost invisible image impairments.

5. References

[1] S. Miller, M. Nezamabadi, and S. Daly, “Perceptual signal coding for more efficient usage of bit codes,” *SMPTE Motion Imaging J.*, vol. 122, no. 4, pp. 52–59, 2013.

[2] P. Hanhart, P. Korshunov, T. Ebrahimi, Y. Thomas, and H. Hoffmann, “Subjective quality evaluation of high dynamic range video and display for future TV,” *SMPTE Motion Imaging J.*, vol. 124, no. 4, pp. 1–6, 2015.

[3] “EG 432-1:2010 - SMPTE engineering guideline - digital source processing #x2014; Color processing for D-cinema,” *SMPTE EG 432-12010*, pp. 1–81, 2010.

[4] IEC 61966-4, “Colour measurement and management in multimedia systems and equipment.” International Electrotechnical Commission, 2000.

[5] IEC 62341-6-1, “Ed. 1: Organic light emtting diode (OLED) displays.” International Electrotechnical Commission, 2007.

[6] ITU-T Recommendation P.910, “Subjective video quality assessment methods for multimedia applications,” *Int. Telecommun. Union, Geneva*, pp. 1–42, 2009.

[7] J. Fang, H. Xu, W. Lv, and M. R. Luo, “Proper luminance of HDR TV system,” *SID 2016 Symposium Digest of Technical Papers*, vol. 47, no. 1, pp. 806–808, 2016.