Subjective and Objective Evaluation of Local Dimming Algorithms for HDR Images

LVYIN DUAN[1,2], **KURT DEBATTISTA**[1], **ZHICHUN LEI**[2], **AND ALAN CHALMERS**[1]

[1]Warwick Manufacturing Group, The University of Warwick, Coventry CV4 7AL, U.K.
[2]School of Microelectronics, Tianjin University, Tianjin 300072, China

Corresponding author: Lvyin Duan (Lvyin.Duan@wariwick.ac.uk)

ABSTRACT In order to enhance viewing experiences, a number of backlight local dimming (BLD) algorithms have been developed to improve the image contrast ratio and provide power efficiency for modern displays. In order to evaluate which BLD algorithm performs best for HDR images rendering on dual-panel displays, this paper presents a comprehensive subjective and objective evaluation conducted with five BLD algorithms across a number of scenes. The subjective evaluation (N = 24) required participants to rank each BLD generated image based on which they thought was the most natural looking. The objective evaluation was undertaken via the use of a novel methodology to generate the images per BLD for comparison against the ground truth High Dynamic Range (HDR) image. Resulting images were compared with the ground truth using qualitative methods namely: HDR-VDP, puPSNR, puSSIM and puVIFP. The power-saving rate of each method was also calculated. The results demonstrate a strong correlation between objective and subjective evaluation. Furthermore, results show that BLD algorithms that consider the luminance balance between backlight and LCD images perform better than straightforward BLD methods.

INDEX TERMS Local dimming algorithms, HDR displays, quality assessment, HDR image processing.

I. INTRODUCTION

Liquid Crystal Displays (LCDs) play a major role in providing improved image or video definition and colour reproduction in the flat panel display marketplace and increasingly for mobile devices. In general, LCDs require backlighting to provide the illumination source. This is located behind the LCD panel, as shown in Figure 1 [1]. However, these devices traditionally utilise a uniform backlight for the LCD panel, which leads to low local contrast ratio and high power consumption [2], [3]. To solve these problems, a number of backlight local dimming (BLD) algorithms have been proposed. These control the intensity of the backlight according to input images. BLD algorithms are particularly fundamental for HDR displays as these tend to be significantly brighter than traditional displays.

BLD algorithms are designed around the fact that not all the images displayed on the LCDs require the same amount of backlighting due to the variance in the luminance of different images. In BLD algorithms, the backlights are dimmed locally to create deeper black areas, while keeping better details in these areas. This permits high dynamic contrast
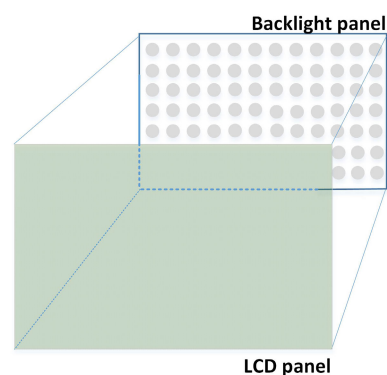
FIGURE 1. The structure of dual-panel HDR display.

ratios and higher power-saving rates in comparison with a uniform backlight setting.

HDR images are able to deliver an enhanced viewing experience to users by providing the full dynamic range that Human Visual System (HVS) can perceive at any level of adaptation. HDR images have previously been used successfully to investigate the viewing experience of displayed images [4], [5]. In this paper, five BLD algorithms are subjectively compared on an HDR display and objectively compared against a ground truth HDR image.

The primary contributions of this work are: a) The first subjective evaluation of BLD algorithms using HDR images on an HDR display. b) The first subjective evaluation (N = 24) whereby participants were asked to rank images in order of the most natural looking of five BLD algorithms; c) An objective evaluation of five BLD algorithms using HDR images, via a novel methodology, to compare BLD generated images with the ground truth using HDR-VDP, puPSNR, puSSIM and puVIFP followed by a calculation of the power-saving rate of each method; and d) An assessment of the correlation between the objective and subjective evaluation which shows, for the scenes tested, the proposed objective evaluation corresponds to the subjective results.

II. RELATED WORK

A number of BLD algorithms have been proposed which each take into consideration several factors, for example, enhancing the contrast, improving image quality and reducing display power consumption, Few of the BLD algorithms proposed so far have been specifically developed for HDR images. In this section, local dimming algorithms and evaluation methods are summarised.

A. BLD ALGORITHMS

The most straightforward BLD algorithms use local image characteristics to determine the backlight value. Funamoto *et al.* [6] proposed the use of maximum and average intensity of a given segment. The maximum algorithm sets the intensity of each LED to the maximum pixel value of the corresponding segment image. This method can lead to high power consumption and may be sensitive to noise. The mean method tends to produce excessively dim backlighting and can lead to significant clipping artefacts. To overcome such limitations more complex methods have been proposed. Cho and Kwon [7] used a correction term to adjust the average pixel intensity and considered the local difference between the maximum and average luminance. In addition, they also used a new method to reduce the clipping artefacts of LCD images displayed on the LCD panel by increasing the backlight luminance from the average luminance. A similar method developed by Zhang *et al.* [8] also computed a correction term as the ratio of the difference of maximum and average luminance. Other methods, such as that introduced by Nam [9]; use both local and global brightness in order to find a better trade-off between enhancing local contrast and preserving the overall appearance of the LCD images, and a roll-off method was used to keep better image details in the high-level grey areas. The BLD algorithm developed by Kim *et al.* [2] is based on a decision rule. This is used to search the optimal dimming value by comparing the light-leakage measure and the clipping measure to keep the light-leakage and clipping lower. Other BLDs were developed to preserve the image quality. For example, Cho *et al.* [10] used an image metric to obtain the intensity of the backlight and refined these values by considering both local block lighting and the lighting

from neighbouring blocks. Kang and Kim [11] considered the pixel distribution of an image using multiple histograms to improve the image quality. Similar methods include Nadernejad *et al.* [12] and Chen *et al.* [13]. Lin *et al.* [14] also used a histogram-based method to compute the cumulative distribution function (from a global histogram) and used its inverse curve to map a weighted mean of the maximum and average pixel values of each backlight segment to the resulting backlight values. Shu *et al.* [15] took the local dimming of LED backlight LC displays as an optimisation problem and obtained a higher visual quality with less power consumed. Zhang *et al.* [16] also proposed one optimal method to maintain a balance between compensated image quality and power saving. Cha *et al.* [17] presented an efficient optimised BLD method for edge-lit lighting-emitting diode backlight to reduce image quality fluctuation. Another category of backlight modulation methods, such as those proposed by Albrecht *et al.* [18] and Hong *et al.* [19], are based on a point spread function (PSF) to exploit the knowledge of light diffusion and model how light diffuses from a source. There have also been other approaches, such as those introduced by Burini *et al.* [20] and Mantel *et al.* [21], which focus primarily on achieving a trade-off between clipping and leakage. Forchhammer and Mantel [22] extended the method proposed by Mantel *et al.* [21] further to multiple viewers talking into clipping and leakage as well as reflections of the ambient light.

Although there have been many BLD algorithms developed for enhancing image quality and saving power, these methods mostly target LDR images. In order to render HDR images on dual-panel displays, Seetzen *et al.* [23] developed a method to solve this problem by splitting HDR images into two layers using square root of the image luminance channel. To assess the impact of HDR image rendering on both subjective and objective scores, Zerman *et al.* [24] developed a method for HDR image rendering for the SIM2 HDR47 display [25].

B. BLD ALGORITHMS EVALUATION

While a number of BLD algorithms have been proposed over two decades, few evaluations have compared the characteristics of the different BLD algorithms.

Error metric methods are objective and used to evaluate image quality based on theoretical models. For example, Kang and Kim [11] used Peak Signal to Noise Ratio (PSNR) to evaluate the quality of LCD images. Besides PSNR, Burini *et al.* [26] also used Mean Square Error (MSE) and labPSNR which comes from PSNR to assess the impact of colour distortion of LCD images. To determine which characteristics of local backlight displays influence quality assessment, Mantel *et al.* [27] conducted subjective and objective evaluations to investigate which aspects, such as clipping and leakage, are relevant for perceptual quality. Mantel *et al.* [28] extended their work to investigate the impact of ambient light and peak white levels on the perceived quality of videos displayed using BLD methods. However, these approaches
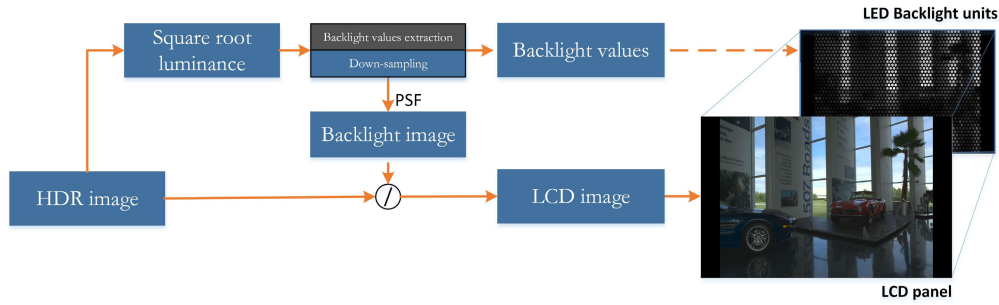
FIGURE 2. Local dimming algorithm baseline method for HDR images.

were used on LDR images; the evaluation of different BLD algorithms using HDR images has remained an open issue.

In order to evaluate HDR images, several image quality metrics have been developed and adapted to the HDR domain. These include puPSNR [29], puSSIM [30] and HDR-VDP-2 [31], which have been adopted for comparing HDR images for a number of applications [32], [33] [35]. These metrics are capable of addressing a wide range luminance and are used widely.

Power consumption is one significant aspect for which BLD algorithms were designed. To measure to what extent different BLD algorithms contribute to power consumption and contrast ratio, the statistical measure method of power consumption and contrast ratio were introduced [7], [14], [34]. However, the problem is that it is difficult to compare these BLD algorithms using their claimed results because of the difference of the hardware design, image content, and even measuring methods.

In this paper, the same settings as mentioned above were conducted using a SIM2 HDR47 display.

III. BACKGROUND

This section introduces the local dimming algorithms and reference images used in this paper for the subjective and objective evaluation.

A. BASELINE METHOD OF BLD ALGORITHMS FOR HDR IMAGES

Local dimming algorithms are mostly used on a dual–panel display which has a high-resolution panel to maintain the image details and a low-resolution backlight panel to control the contrast ratio.

Figure 2 shows the baseline method for dual-panel HDR display systems. This method, proposed by Seetzen *et al.* [23], splits HDR images into two layers according to the structure of the dual-panel HDR display. One layer is for backlight values and the other one is for LCD images. In this method, the first step is to compute the square root of the HDR image luminance. Secondly, the backlight values are computed by down-sampling the square root luminance according to the resolution of the LED array. Thirdly, the Point Spread Function (PSF) of the LED is approximated with a Gaussian filter. Then, the backlight values are

convolved with PSF and the resulting signal is upscaled to the full resolution backlight diffusion image. Subsequently, the original HDR image is divided by the backlight diffusion image in the previous step to compute the LCD image. Finally, the backlight values is transmitted to the backlight panel to drive the LEDs, and the LCD images are displayed on the LCD panel.

BLD algorithms seek to calculate the backlight values and obtain the LCD image. In this paper, several backlight value extraction methods will be used instead of the down-sampling method shown in Figure 2.

B. METHODS

In general, BLD methods for LDR images include two steps, one is backlight values extraction, the other step is LCD image compensation. The compensated image L_C computed by Eq.(1).

$$L_C = \left(\frac{BL_{full}}{BL_{HDR}} \right)^{\frac{1}{\gamma}} \times L_T \qquad (1)$$

where L_T is the luminance of the original LDR image. BL_{full} and BL_{HDR} denote the intensities of conventional (full-on) backlight and the updated backlight, respectively. γ is the gamma correction coefficient [14]. The compensated image is the LCD image to be displayed on LCD panel. However, considering the data format of HDR images and the baseline method of rendering HDR images mentioned in this paper, the compensation step used for LDR images is not suitable for HDR images anymore. The LCD image is determined by the quotient of the original HDR image and the backlight diffusion image, as shown in Figure 2. For the selected BLD methods the backlight values extraction is not affected by the compensated image. Furthermore, when considering that most BLD algorithms were designed according to the statistical characteristics of displayed images and power consumption, and few methods were developed for HDR images, five BLD methods were selected. These include four methods which are commonly used and discussed in related work, and one which has been developed for HDR images and has been shown to perform better than the built-in method (based on Seetzen *et al.*'s method [36]) of the SIM2 HDR display [25].

A brief description chronologically of each is provided below:

1) AVERAGE METHOD (AVG METHOD)

The Avg method was proposed by Funamoto *et al.* [6]. In this method, the original images are divided into several image segments according to the number of backlight units. The backlight values are calculated using Eq.(2).

$$BL = Avg_i \qquad (2)$$

where Avg_i is the average value of each image segment i.

2) MAXIMUM METHOD (MAX METHOD)

Same as the Avg method, the Max method was also proposed by Funamoto *et al.* [6]. Its backlight values are obtained by Eq.(3).

$$BL = Max_i \qquad (3)$$

where Max_i is the maximum value of each image segment i.

3) CHO METHOD

Cho and Kwon [7] used a correction term, as shown in Eq.(4), to adjust the backlight values to take into account the local difference between the maximum and average luminance. Eq.(5) shows how to calculate backlight values.

$$correction = \frac{1}{2}\left(Diff + \frac{Diff^2}{2^n}\right) \qquad (4)$$

$$BL = Avg_i + correction \qquad (5)$$

where *Diff* is the difference between average and maximum luminance and n is n bit greyscale.

4) INVERSE OF A MAPPING FUNCTION (IMF METHOD)

Lin *et al.* [14] introduced an IMF method by inverting the cumulative distribution function (CDF) of the traditional histogram equalization with the oblique line $y = x$. The CDF is accumulated by the probability density function (PDF) from the lowest grey-level to the highest of the global histogram. Then, the zone weighting value of each image segment is defined by Eq.(6) to obtain the backlight values.

$$ZoneWeightingValue = 0.9 \times Max_i + 0.1 \times Avg_i \qquad (6)$$

where *Max* and *Avg* are the maximum and average values of each image segment.

5) ZERMAN METHOD

Zerman *et al.* [24] proposed an iterative HDR rendering algorithm. This HDR display rendering algorithm calculates the target display-referred luminance I from the input HDR image by saturating luminance values in excess of the maximum display brightness. Next, the optimal backlight target luminance map L_{opt} is determined by minimising the power consumption and maximising the fidelity to the target pixel values. Then, backlight values are obtained by the iterative procedure according to L_{opt}, and the rendered backlight on the display is obtained by convolving the values of the LED with the PSF. The LCD values of the panel are found by dividing each channel of the original image by the optimal backlight map.
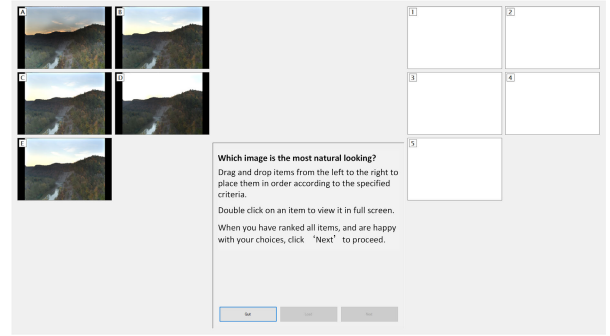

FIGURE 3. Screenshot of the evaluation software.

IV. SUBJECTIVE EVALUATION

While evaluation of BLD algorithms could, ideally, be performed via the use of objective metrics, it is yet unclear if such metrics could provide the same results as would be perceived by human participants. In order to evaluate the BLD algorithms themselves and also the performance of objective metrics on such algorithms, a subjective experiment is conducted. The experiment serves to demonstrate which methods are perceived best, and paves the way for evaluating the proposed objective evaluation via the use of traditional HDR metrics.

A. DESIGN

A ranking design was chosen for the evaluation phase as it provides a significant advantage over methods such as pairwise comparisons as a large number of samples can be viewed in relatively short times. Furthermore, such an approach has been used successfully on a number of occasions to evaluate other algorithms in imaging, for example, when comparing the quality of different HDR compression methods [35]. The motivation for ranking is to be able to distinguish between the quality of the different images that are relatively close without requiring an exhaustive full-pairwise comparison. The primary goal of the experiment was to rank and identify the order of each BLD algorithm, over 15 scenes. In particular, participants were asked to rank the presented BLD generated HDR images in order of which image they perceived to be the most natural looking on an HDR display.

15 scenes were chosen from the Fairchild database [37]. Each of the 15 scenes was represented by five images generated from the BLD algorithms introduced in Section III-B. The order of the scene presentation was randomised as was the order of the BLD methods presented. Participants were allowed to view the images as many times as they wanted but had to view them all at least once, before they ranked them.

B. MATERIALS

A graphical user interface (GUI) for the ranking-based experiment was used for displaying the 15 scenes. Figure 3 shows the custom GUI application which was specifically built for the ranking-based subjective evaluation. It presents five thumbnails (labelled A-E) from the five BLD algorithms on the left side of the screen. Each thumbnail is displayed in

TABLE 1. Overview of the scenes used for the ranking based subjective experiment.

| Thumbnail	Name	Dynamic Range (stops)	Thumbnail	Name	Dynamic Range (stops)
	507	44.32		BalancedRock	17.72
	BloomingGorse(2)	15.68		CadesCove	13.38
	CanadianFalls	44.22		Exploratorium(1)	21.17
	HDRMark	18.66		LasVegasStore	45.12
	LetchworthTeaTable(2)	15.97		MirrorLake	12.71
	PaulBunyan	15.18		Peppermill	43.91
	RedwoodSunset	48.13		TheGrotto	12.70
	WillySentinel	17.58			

full-screen on the HDR screen when participants *double − clicked* on it. Participants are tasked to view each image displayed in full screen mode then rank the images on the left side in the order of which image is most natural looking by dragging their preferred choice to its corresponding position (labelled 1-5) with 1 being the best one and 5 being the worst. The software permitted users to view all the BLD versions of a single scene and rank them using a drag and drop method for each image. Hardware resources included a SIM2 HDR display with a peak luminance 4,000 cd/m^2, an HP 22'' LED display with peak luminance 300 cd/m^2 and a computer with a solid state drive for quick loading of the images.

The subjective and objective evaluation in this work used a set of 15 HDR images selected from the Fairchild database [37]. Table 1 provides the details of the selected HDR images. The dynamic range of all the images spans between 12 - 48 stops, and the average dynamic range and the dynamic range distribution for the 15 HDR images were also considered to ensure that the selection is representative of the Fairchild database [37]. To adapt to the resolution and range of the SIM2 HDR display, these images were resized to 1920 × 1080 by padding zero pixels and their pixel values were scaled to [0, 4000] in this experiment.

These HDR images were processed by the methods described in Section III. 15 (HDR scenes) × 5 (BLD algorithms) images were generated in total.
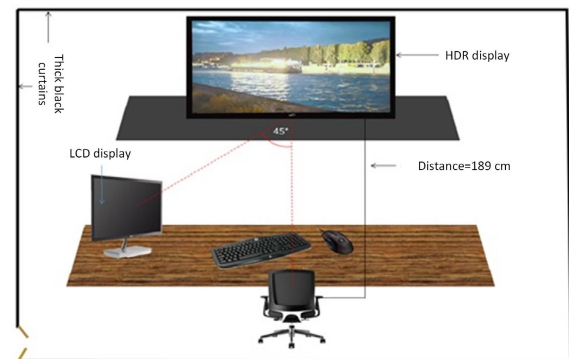
FIGURE 4. Subjective experiment setup.

C. ENVIRONMENT
The experiments were conducted in a typical dark room with minimal ambient lighting (below 25lux) which is within the recommended luminance levels according to ITU-R recommendations [38]. The distance between the HDR display and the participant was set to approximately 3.2 times the height of the HDR display; at a distance of 189cm with an LCD monitor placed at the angle of 45 degrees (see Figure 4).

D. PARTICIPANTS
A total of 24 participants (M = 14; F = 10) were involved in this experiment, with an age range of 18 to 50 years.

Participants were asked whether they have a normal version (or corrected to normal vision) during the recruiting process. Participants with self-reported normal vision (or corrected normal vision) were used to conduct the subjective experiment. These participants were recruited from university students and staff.

E. PROCEDURE

The participants were introduced to the goal of the experiment before the start followed by a brief training session using a scene which was subsequently discarded from the results. After completing the training session, the participants were asked to rank the images of the 15 scenes.

In order to view each image in full screen, each participant was asked to *double − click* the small images before they began to evaluate each set of images. Subsequently, the participants would view each of the five images for each scene and give their assessment as to which image is most natural looking. Based on their judgement, the participants positioned the corresponding thumbnails to one of the blank positions on the right labelled [1-5], as shown in Figure 3.

F. SUBJECTIVE RESULTS AND ANALYSIS

This section provides an overview and analysis of the results obtained from the subjective experiment.

Suppose that H_0 and H_1 indicate there are no significant differences and there are significant differences between the BLD algorithms, respectively. Assume that the statistical significance p is 0.05. The sample size is 24. If H_1 is true, it is important to determine the coefficient of concordance which measures the degree by which the participant mutually agree on choices.

In this work, a 3-dimensional data array $A(N, M, S)$ is assumed to represent the ranks given by each participant to each algorithm for each HDR scenes. In this data array, N (N = 24) denotes all participants, M denotes five BLD algorithms and S denotes all 15 HDR scenes, respectively. Then, the mean ranks that averaged across all participants and HDR scenes can be denoted by Eq.(7).

$$\frac{1}{K}\sum_{S=1}^{K}\overline{A}(\bullet, M, S) = \overline{A}(\bullet, M, \bullet) \qquad (7)$$

where $\overline{A}(\bullet, M, S)$ represents the ranks for each M and S, averaged across all participants and $\overline{A}(\bullet, M, \bullet)$ represents the mean ranks averaged across all the participants and scenes. K is the total number of HDR scenes.

To compare the significant differences between five BLD algorithms, the Kendall's W (Kendall's co-efficient of Concordance W) is employed. It can be computed by Eq.(8), Eq.(9), Eq.(10) and Eq.(11).

$$W = \frac{12R}{N^2(M^3 - M)} \qquad (8)$$

$$R = \sum_{i=1}^{M}(R_i - \overline{R})^2 \qquad (9)$$

FIGURE 5. Subjective ranks with Kendall's W, averaged across participants.

Images	BLD Algorithms					Kendall (W)	Sign. μ
507	Zerman (1.58)	IMF (2.79)	Max (3.04)	Cho (3.38)	Avg (4.21)	0.365	$p < 0.05$
BalancedRock	Zerman (2.42)	IMF (2.92)	Max (2.96)	Cho (3.12)	Avg (3.58)	0.070	0.15
BloomingGorse (2)	Zerman (1.54)	IMF (1.60)	Max (3.46)	Cho (3.67)	Avg (4.73)	0.774	$p < 0.05$
CadesCove	IMF (1.46)	Zerman (1.98)	Max (3.12)	Cho (3.81)	Avg (4.63)	0.675	$p < 0.05$
CanadianFalls	Zerman (1.67)	IMF (1.86)	Max (3.33)	Cho (3.79)	Avg (4.33)	0.556	$p < 0.05$
Exploratorium (1)	IMF (1.50)	Zerman (1.58)	Max (3.25)	Cho (3.88)	Avg (4.79)	0.830	$p < 0.05$
HDRMark	Max (2.13)	Zerman (2.21)	Cho (2.61)	Avg (3.81)	IMF (4.75)	0.488	$p < 0.05$
LasVegasStore	Zerman (2.54)	Cho (2.75)	Max (2.79)	IMF (3.12)	Avg (3.79)	0.096	0.06
LetchworthTea Table(2)	Zerman (1.04)	Max (2.46)	Cho (3.13)	Avg (3.50)	IMF (4.88)	0.791	$p < 0.05$
MirrorLake	Zerman (1.40)	IMF (1.75)	Max (3.25)	Cho (4.08)	Avg (4.52)	0.770	$p < 0.05$
PaulBunyan	Zerman (1.83)	IMF (2.29)	Max (2.83)	Cho (3.71)	Avg (4.33)	0.417	$p < 0.05$
Peppermill	Max (2.58)	IMF (2.63)	Zerman (2.67)	Cho (2.96)	Avg (4.17)	0.179	$p < 0.05$
RedwoodSunset	Zerman (2.71)	IMF (2.77)	Cho (3.02)	Max (3.12)	Avg (3.38)	0.029	0.59
TheGrotto	Zerman (1.54)	IMF (1.71)	Max (3.33)	Cho (3.71)	Avg (4.75)	0.733	$p < 0.05$
WillySentinel	Max (1.42)	Zerman (2.46)	Cho (2.63)	Avg (3.12)	IMF (4.38)	0.268	$p < 0.05$

FIGURE 6. Subjective mean ranks with Kendall's W across all images and participants.

Average results of ranking across 24 participants	BLD Algorithms					Kendall's coefficient of concordance (W)	Sign. μ
	Zerman (1.29)	IMF (2.27)	Max (2.88)	Cho (3.81)	Avg (4.75)	0.720	$p < 0.05$

$$R_i = \sum_{j=1}^{N} r_{i,j}, \, i \in M, j \in N \qquad (10)$$

$$\overline{R} = \frac{R_i}{N} \qquad (11)$$

where $r_{i,j}$ is the rank for each algorithm by each participant. $\overline{R}$ is the mean ranks and R is the standard squared deviation [35].

Kendall's W is a non-parametric model which used for assessing agreement amongst the participans' choices. It provides a value from 0 to 1, where 0 means complete disagreement and 1 means completely in agreement. Figure.5 provides $\overline{A}(\bullet, M, S)$ along with W scores sorted in order of the most preferred method on the leftmost side. The groupings in the results show that groups formed by two or more algorithms were not considered significantly different. However, algorithms outside of groups are statistically significantly different from other groups and non-grouped methods. The results show that there are statistically significant differences between the five BLD algorithms for most HDR images with high Kendall's W scores and at a significance $p < 0.05$. It can be seen that Zerman's method is frequently first and first overall, followed generally by IMF and Max, and finally Cho and Avg.

Figure 6 shows $\overline{A}(\bullet, M, \bullet)$ along with W score. For these five BLD algorithms, there is a significant difference in-between separate groups with a relatively high Kendall's
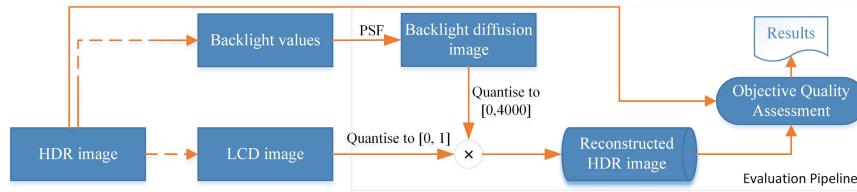
FIGURE 7. Schematic diagram of the evaluation methodology for displaying HDR images.

W at 0.72 and is significant at $p < 0.05$. The grouping surrounded by the red line shows that IMF and Max methods are similar. Avg, Cho and Zerman form a group of their own demonstrating that they have a significant difference.

V. OBJECTIVE EVALUATION

Objective evaluation of BLD algorithms is not straightforward as the output of the BLD are separated LCD and backlight images which cannot be evaluated separately. In this section a new evaluation methodology that combines the LCD and backlight images is proposed for HDR images. This enables the BLD methods to be compared using traditional metrics which are discussed and results presented. In the following section the objective results from this section are compared and evaluated against the subjective results from the previous section.

A. OBJECTIVE EVALUATION METHODOLOGY

To evaluate these five BLD algorithms using QA metrics, a new evaluation methodology combining the backlighting and LCD images is introduced, see Figure 7. To keep the consistency with subjective experiment, the same scenes as used in that experiment were initially adopted and evaluated, and the full Fairchild set is subsequently evaluated.

The images were scaled to [0, 4000] to adapt to the maximum luminance of the HDR display and as the reference HDR images. The backlight values and the corresponding LCD images can be obtained by different BLD algorithms as introduced in Section III-B. Figure 8 provides an example for these five BLD algorithms. To simulate the backlight source, a measured PSF was used to convolve with backlight values. The simulated backlight diffusion images were quantised to [0, 4000], and the LCD images were quantised to 8 bits after clipping pixels exceeding 1. Finally, new HDR images were constructed by combining backlight diffusion images and LCD images by Eq.(12). Most of the displays can be modelled by the following formula:

$$L_{HDR} = T \times B \qquad (12)$$

where L_{HDR} is the luminance that people can perceive, and T and B represent the Liquid Crystal (LC) transmittance (pixel grey level of LCD images) and backlight luminance [2]. The reconstructed HDR images were evaluated against the reference, original, HDR images via a series of QA metrics (see below).

(a) Average method

(b) Cho method

(c) IMF method

(d) Max method

(e) Zerman method
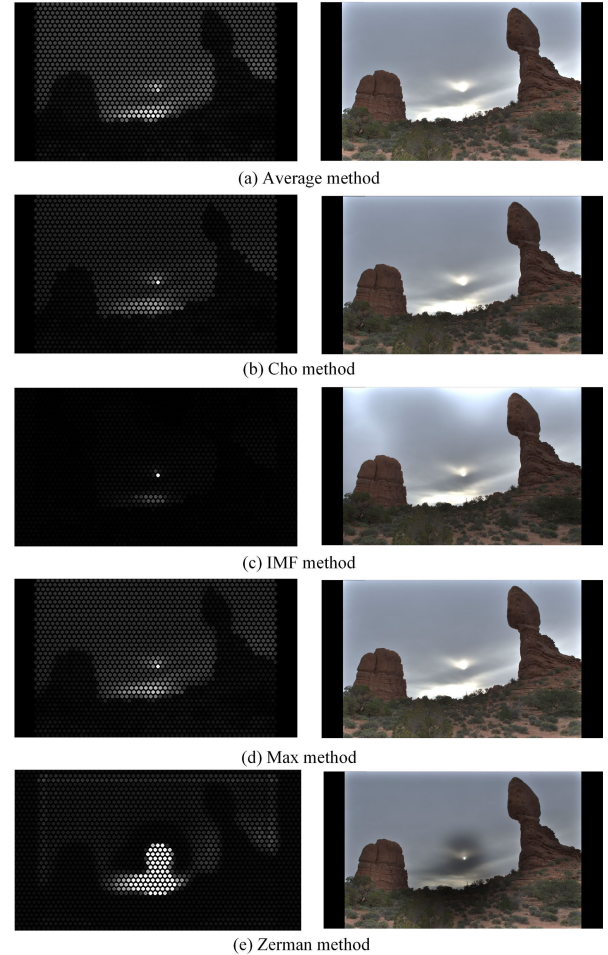
FIGURE 8. Backlight values and LCD images of five BLD algorithms.

B. QUALITY ASSESSMENT METRICS

The metrics used were:

- **HDR-VDP**: This metric was proposed by Mantiuk *et al.* [31] and based on a well calibrated visual model which can reliably predict visibility and quality differences between image pairs in a wide range of view conditions. It considers the visual system, including a board range of viewing conditions, intra-ocular light scatter, and contrast sensitivity across the full range of visible luminance and uses the Q score of version 2.2.1 to represent the image quality [39].
- **puPSNR**: Aydin *et al.* [29] proposed a QA extension to popular metrics, such as PSNR and SSIM and allowed

them to have the capability of handling a wide range of luminance levels via the introduction of a perceptual uniform (PU) curve to account for the human visual system's response to luminance. **puPSNR** is the version applied to the traditional PSNR metric.

- **puSSIM**: Similar to puPSNR, this metric is a dynamic range independent extension to the Structure Similarity Index Metric [30] using the PU curve. SSIM is used to measure the structural similarity between the reference and the reconstructed HDR images.

- **puVIFP**: Visual Information Fidelity (VIF) is a full reference image quality assessment metric which was proposed by Sheikh and Bovik [40]. It was developed based on natural scene statistics and the notion of image information extracted by the human visual system. To adapt to the high dynamic range of luminance of HDR images, puVIFP is used by adopting the PU curve to compress the luminance channel to account for the Human Visual System's response to luminance.

- **Power-Saving Rate (PR)**: Since one of the major goals of the BLD algorithms is to consume less power, the power-saving rate is estimated via the average of the LED values such that $p = \left(1 - \frac{\sum_1^N LED_N}{N}\right) \times 100\%$. This method is adopted as the LEDs are controlled with pulse width modulation (PWM) for which the duty cycle is proportional to the emitted luminance [26].

C. OBJECTIVE RESULTS FOR THE SELECTED 15 HDR SCENES

This section provides the objective results for the selected 15 HDR scenes. Figure 9(a), Figure 9(b), Figure 9(c) and Figure 9(d) show the HDR-VDP, puPSNR, puSSIM and puVIFP results respectively. Figure 9(e) shows the power-saving ratio of the five methods. In these figures, violin plots [41] are used to indicate the probability density distribution of all the scenes for the different metrics. The black lines and the dotted red lines within the violin plots indicate the mean and median values respectively. Figure 10 provides objective rank results for five BLD algorithms and their statistically differences across the selected 15 HDR scenes.

From the overall results we can see that Zerman performs best followed by IMF, Max, Cho and Avg. These are similar to the results from the subjective experiments.

For objective ranks of the five BLD algorithms, as shown in Figure 10, there are significant differences with relatively high Kendall's W scores and at a significance $p < 0.05$ in the perceptual QA metrics puPSNR, puSSIM and puVIFP. Zerman and IMF, IMF and Max form a group respectively in QA metric HDR-VDP, which means that they are similar in their own group. A fuller analysis of these correlations is presented in the next section. The power-saving ratio reveals an inverse trend with Avg being the most efficient and Zerman the least efficient, and Zerman and IMF perform similar.
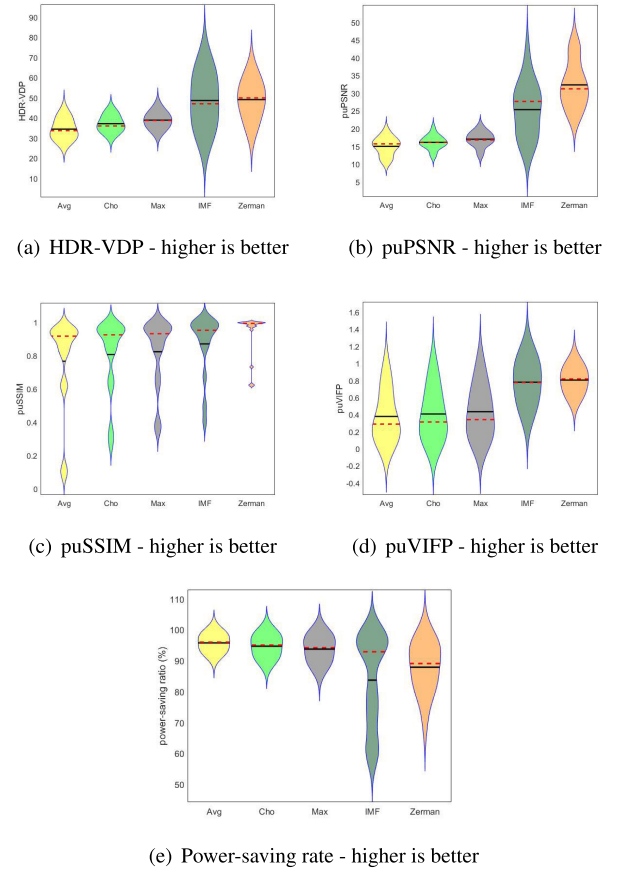

(a) HDR-VDP - higher is better   (b) puPSNR - higher is better

(c) puSSIM - higher is better   (d) puVIFP - higher is better

(e) Power-saving rate - higher is better

FIGURE 9. Objective evaluation results for the selected 15 scenes.

Objective methods	BLD Algorithms					Kendall (W)	Sign. μ
HDR-VDP	Zerman (4.27)	IMF (3.80)	Max (3.33)	Cho (2.33)	Avg (1.27)	0.580	$p < 0.05$
puPSNR	Zerman (4.87)	IMF (4.07)	Max (3.07)	Cho (1.93)	Avg (1.07)	0.950	$p < 0.05$
puSSIM	Zerman (4.93)	IMF (4.07)	Max (3.00)	Cho (2.00)	Avg (1.00)	0.988	$p < 0.05$
puVIFP	Zerman (4.33)	IMF (4.33)	Max (3.20)	Cho (2.07)	Avg (1.07)	0.820	$p < 0.05$
Power- Saving Rate	Avg (5.00)	Cho (3.87)	Max (2.87)	IMF (1.73)	Zerman (1.53)	0.852	$p < 0.05$

FIGURE 10. Objective ranks with Kendall's W across the selected 15 HDR scenes.

D. OBJECTIVE RESULTS FOR 105 HDR IMAGES

To provide a more comprehensive result for the objective experiment, this section extends results of the perceptual QA metrics to all 105 HDR scenes in the Fairchild database [37]. Figure 11(a), Figure 11(b), Figure 11(c), Figure 11(d) and Figure 11(e) show the HDR-VDP, puPSNR, puSSIM, puVIFP and power-saving rate respectively. Compared with these results for the 15 HDR scenes, the results for the 105 HDR scenes broadly exhibit similar performance but, as expected, with smaller standard deviation.

Figure 12 shows objective rank results of the five BLD algorithms and their statistical differences across the 105 HDR scenes. The results show that the Zerman method performs best followed by IMF, Max, Cho and Avg in terms
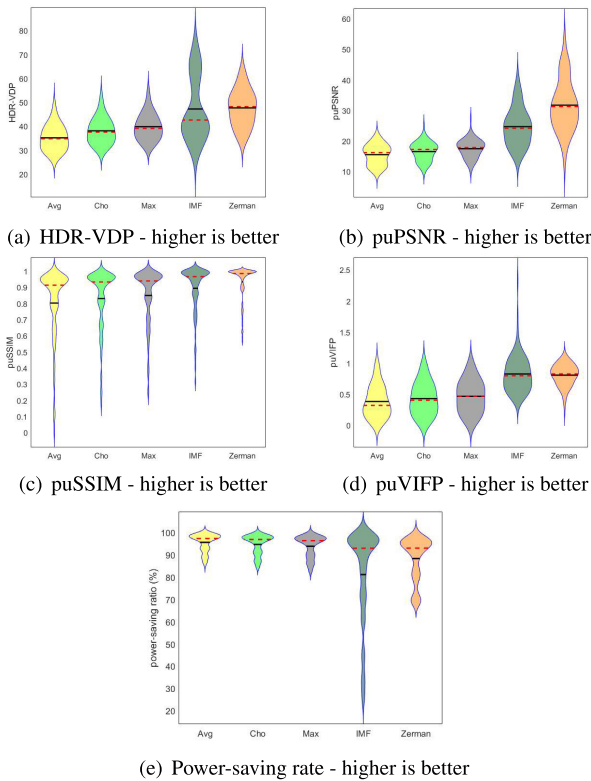
FIGURE 11. Objective evaluation results for the 105 HDR scenes.

FIGURE 12. Objective ranks for five evaluation methods across the 105 HDR scenes.

TABLE 2. Pearson's rank correlation between objective and subjective evaluation across the selected 15 HDR scenes.

	HDR-VDP	puPSNR	puSSIM	puVIFP	SR
HDR-VDP	-	0.943*	0.950*	0.989**	0.894*
puPSNR	0.943*	-	0.967**	0.957*	0.902*
puSSIM	0.950*	0.967**	-	0.926*	0.980**
puVIFP	0.989**	0.957*	0.926*	-	0.845
SR	0.894*	0.902*	0.980**	0.845	-

* denotes significance at $p < 0.05$ level,
** denotes significance at $p < 0.01$ level.

TABLE 3. Pearson's rank correlation between objective and subjective evaluation across the 105 HDR scenes.

	HDR-VDP	puPSNR	puSSIM	puVIFP	SR
HDR-VDP	-	0.925*	0.970**	0.985**	0.914*
puPSNR	0.925*	-	0.971*	0.921**	0.905*
puSSIM	0.970**	0.971**	-	0.935*	0.972**
puVIFP	0.985**	0.921*	0.935*	-	0.839
SR	0.914*	0.905*	0.972*	0.839	-

* denotes significance at $p < 0.05$ level,
** denotes significance at $p < 0.01$ level.

TABLE 4. Spearman's Rho rank correlation between objective and subjective evaluation across the selected 15 HDR scenes.

	HDR-VDP	puPSNR	puSSIM	puVIFP	SR
HDR-VDP	-	1.000**	1.000**	1.000**	1.000**
puPSNR	1.000**	-	1.000**	1.000**	1.000**
puSSIM	1.000**	1.000**	-	1.000**	1.000**
puVIFP	1.000**	1.000**	1.000**	-	1.000**
SR	1.000**	1.000**	1.000**	1.000**	-

* denotes significance at $p < 0.05$ level,
** denotes significance at $p < 0.01$ level.

TABLE 5. Spearman's Rho rank correlation between objective and subjective evaluation across the 105 HDR scenes.

	HDR-VDP	puPSNR	puSSIM	puVIFP	SR
HDR-VDP	-	1.000**	1.000**	0.900*	1.000**
puPSNR	1.000**	-	1.000**	0.900*	1.000**
puSSIM	1.000**	1.000**	-	0.900*	1.000**
puVIFP	1.000**	1.000**	1.000**	-	1.000**
SR	1.000**	1.000**	1.000**	0.900*	-

* denotes significance at $p < 0.05$ level,
** denotes significance at $p < 0.01$ level.

of puPSNR with a high Kendall's W at 0.950 in Figure 10. A similar trend appears in Figure 12 in terms of puPSNR with a Kendall's W at 0.914. The puSSIM, puVIFP and power-saving rate show similar results. HDR-VDP presents a smaller difference between the selected 15 HDR scenes and 105 HDR images. The results being similar across both the scenarios corresponding to 15 and 105 indicate that the 15 scenes selected from the Fairchild database [37] used for the subjective experiment where representative of the whole.

VI. OVERALL ANALYSIS AND DISCUSSION

This section discusses the overall results combining the subjective and objective experiments in order to establish a correlation between them and analyse the overall performance of the five BLD algorithms.

The reconstructed HDR images from each algorithm were evaluated against the reference HDR images using the previously mentioned QA metrics. Correlations are computed by combining the objective and subjective results using Pearson's correlation test and Spearman's rank correlation test. Table 2 and Table 3 show the Pearson's rank correlation results across the selected 15 HDR scenes and the 105 HDR scenes respectively. In addition, Table 4 and Table 5 show the Spearman's rank correlation results across the selected 15 HDR scenes and the full 105 HDR set respectively. SR represents subjective evaluation method.

The results given in Table 2, Table 3, Table 4 and Table 5 indicate a strong correlation between the perceptual QA metrics and the subjective results. Moreover, the correlation results within the objective metrics are also very strong. For Pearson's rank correlation across both the selected 15 HDR scenes and the 105 HDR scenes, the values (0.980 vs 0.972) between subjective rankings and puSSIM are high with statistical significance at $p < 0.01$, which performs better than that with HDR-VDP (0.849 vs 0.914), puPSNR (0.902 vs 0.905) and puVIFP (0.845 vs 0.839) with statistical significance at

the $p < 0.05$ level. For Spearman's rank correlation across the selected 15 HDR scenes, the values (1.000) show a strong correlation amongst the objective metrics and the subjective evaluation method. Spearman's rank correlation coefficient also produced high values (1.000) when testing across all 105 HDR images, except for the correlation value (0.900) between puVIFP and the other evaluation methods owing to IMF performing better than Zerman's method when using puVIFP to evaluate them.

In both the subjective and objective experiments, across all QA metrics, a similar result can be seen. Zerman and IMF methods perform better than the other methods. The power-saving rate demonstrates an opposite trend to the results compared to the perceptual QA metrics. Comparing with Avg, Cho and Max methods, Zerman and IMF methods show lower values and a higher fluctuation range in power consumption.

As discussed in III-B, the LCD image was obtained by the division of the original HDR image and the backlight image. Considering the display model provided by Eq.(12), one "trade-off" relationship is formed between backlight values and pixel values of the LCD image. In one specific HDR image area, small backlight values are more likely to result in higher pixel values which will be clipped when they are larger than 1. Clipping artefacts inevitably affect image quality. Conversely, the higher the backlight values the less likely pixel values will exceed 1. As a result, the image quality of Avg, Cho and Max methods is worse compared with IMF and Zerman methods although these methods show a better power-saving rate. In addition, due to the better content adaption of IMF and Zerman methods compared with other methods, they produce a wider range or quality results in terms of perceptual QA metrics and power-saving rate.

The objective and subjective evaluation results suggest that the Zerman method performs best although its power-saving rate is worse when compared with other methods. While it performs best in terms of quality it suffers relative to the other from higher power consumption. However, this still limits maximum power consumption as well as taking clipping artefacts into consideration. IMF also exhibits a good performance amongst these five algorithms. Due to the high dynamic range of the tested images (from around 12 stops to 48 stops) and the better content adaptation, IMF experiences wide fluctuation on the perceptual QA metrics, such as HDR-VDP (Q) and puPSNR, and power-saving rate. The other BLD methods, such as Avg, Cho and Max, show worse results for the perceptual QA metrics although they perform a better compared with IMF and Zerman in power-saving.

VII. CONCLUSION AND FUTURE WORK

This work provides a detailed subjective and objective comparison of five BLD algorithms and explores the relationship between LCD images and backlight values further when using an HDR display. The results show that the Zerman method, which was developed particularly for HDR images, performs better than the other four methods.

This work has provided insights for developing new BLD algorithms. In practice, there are few BLD algorithms developed for HDR images. For BLD algorithms, clipping artefacts appear mostly in the bright areas, which is one of the main reasons for reduced image quality. In addition, a "trade-off" relationship, a clear inverse correlation, appears between backlight values and the LCD image. To get a higher image quality, power appears to be sacrificed. Therefore, a method that adapts the values of the backlight to the content of HDR images better, especially for brighter areas, while keep the balance between quality and power-saving rate could potentially prove very successful. How to identify and optimise for the above is the subject of future work. In addition, other methods such as [3] will be also considered in future work.

REFERENCES

[1] S. Kobayashi, S. Mikoshiba, S. Lim, *LCD Backlights*, 1th ed. New York, NY, USA: Wiley, 2009, ch. 4, sec. 4, pp. 49–58. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470744826

[2] S.-E. Kim, J.-Y. An, J.-J. Hong, T. W. Lee, C. G. Kim, and W.-J. Song, "How to reduce light leakage and clipping in local-dimming liquid-crystal displays," *J. Soc. Inf. Display*, vol. 17, no. 12, pp. 1051–1057, Dec. 2009.

[3] M. Narwaria, M. P. Da Silva, and P. Le Callet, "Dual modulation for LED-backlit HDR displays," in *High Dynamic Range Video*. New York, NY, USA: Academic, 2016, pp. 371–388.

[4] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, "Evaluation of tone mapping operators using a high dynamic range display," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 640–648, Jul. 2005.

[5] M. Melo, M. Bessa, K. Debattista, and A. Chalmers, "Evaluation of HDR video tone mapping for mobile devices," *Signal Process., Image Commun.*, vol. 29, no. 2, pp. 247–256, Feb. 2014.

[6] T. Funamoto, T. Kobayashi, and T. Murao, "High picture quality technique for LCD television: LCD-AI," in *Proc. Int. Display Workshops (IDW)*, Kobe, Japan, 2000, pp. 1157–1158.

[7] H. Cho and O.-K. Kwon, "A backlight dimming algorithm for low power and high image quality LCD applications," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, pp. 839–844, May 2009.

[8] X.-B. Zhang, R. Wang, D. Dong, J.-H. Han, and H.-X. Wu, "Dynamic backlight adaptation based on the details of image for liquid crystal displays," *J. Display Technol.*, vol. 8, no. 2, pp. 108–111, Feb. 2012.

[9] H. Nam, "Low power active dimming liquid crystal display with high resolution backlight," *Electron. Lett.*, vol. 47, no. 9, pp. 538–540, May 2011.

[10] S. I. Cho, H.-S. Kim, and Y. H. Kim, "Two-step local dimming for image quality preservation in LCD displays," in *Proc. Int. SoC Design Conf.*, Jeju, South Korea, Nov. 2011, pp. 274–277.

[11] S.-J. Kang and Y. H. Kim, "Multi-histogram-based backlight dimming for low power liquid crystal displays," *J. Display Technol.*, vol. 7, no. 10, pp. 544–549, Oct. 2011.

[12] E. Nadernejad, N. Burini, J. Korhonen, S. Forchhammer, and C. Mantel, "Adaptive local backlight dimming algorithm based on local histogram and image characteristics," in *Proc. 18th Color Imag., Displaying, Process., Hardcopy, Appl.*, Burlingame, CA, USA, Feb. 2013, pp. 237–249.

[13] H. Chen, J. Sung, T. Ha, and Y. Park, "Locally pixel-compensated backlight dimming on LED-backlit LCD TV," *J. Soc. Inf. Display*, vol. 15, no. 12, pp. 981–988, Jun. 2007.

[14] F.-C. Lin, Y.-P. Huang, L.-Y. Liao, C.-Y. Liao, H.-P.-D. Shieh, T.-M. Wang, and S.-C. Yeh, "Dynamic backlight gamma on high dynamic range LCD TVs," *J. Display Technol.*, vol. 4, no. 2, pp. 139–146, Jun. 2008.

[15] X. Shu, X. Wu, and S. Forchhammer, "Optimal local dimming for LC image formation with controllable backlighting," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 166–173, Jan. 2013.

[16] T. Zhang, X. Zhao, X. Pan, X. Li, and Z. Lei, "Optimal local dimming based on an improved shuffled frog leaping algorithm," *IEEE Access*, vol. 6, pp. 40472–40484, 2018.

[17] S. Cha, T. Choi, H. Lee, and S. Sull, "An optimized backlight local dimming algorithm for edge-lit LED backlight LCDs," *J. Display Technol.*, vol. 11, no. 4, pp. 378–385, Apr. 2015.

[18] M. Albrecht, A. Karrenbauer, T. Jung, and C. Xu, "Sorted sector covering combined with image condensation—An efficient method for local dimming of direct-lit and edge-lit LCDs," *IEICE Trans. Electron.*, vol. 93, no. 11, pp. 1556–1563, Nov. 2010.

[19] J.-J. Hong, S.-E. Kim, and W.-J. Song, "A clipping reduction algorithm using backlight luminance compensation for local dimming liquid crystal displays," *IEEE Trans. Consum. Electron.*, vol. 56, no. 1, pp. 240–246, Feb. 2010.

[20] N. Burini, E. Nadernejad, J. Korhonen, S. Forchhammer, and X. Wu, "Image dependent energy-constrained local backlight dimming," in *Proc. 19th IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 2012, pp. 2797–2800.

[21] C. Mantel, N. Burini, E. Nadernejad, J. Korhonen, S. Forchhammer, and J. M. Pedersen, "Controlling power consumption for displays with backlight dimming," *J. Display Technol.*, vol. 9, no. 12, pp. 933–941, Dec. 2013.

[22] S. Forchhammer and C. Mantel, "Viewpoint adaptive display of HDR images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 1177–1181.

[23] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 760–768, Aug. 2004.

[24] E. Zerman, G. Valenzise, F. De Simone, F. Banterle, and F. Dufaux, "Effects of display rendering on HDR image quality assessment," in *Proc. SPIE*, vol. 9599, Sep. 2015, Art. no. 95990R.

[25] SIM2. *HDR47ES4MB*. Accessed: 2015. [Online]. Available: http://hdr.sim2.it/hdrproducts/hdr47es4mb

[26] N. Burini, E. Nadernejad, J. Korhonen, S. Forchhammer, and X. Wu, "Modeling power-constrained optimal backlight dimming for color displays," *J. Display Technol.*, vol. 9, no. 8, pp. 656–665, Aug. 2013.

[27] C. Mantel, J. Sogaard, S. Bech, J. Korhonen, J. M. Pedersen, and S. Forchhammer, "Modeling the quality of videos displayed with local dimming backlight at different peak white and ambient light levels," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3751–3761, Aug. 2016.

[28] C. Mantel, S. Bech, J. Korhonen, S. Forchhammer, and J. M. Pedersen, "Modeling the subjective quality of highly contrasted videos displayed on LCD with local backlight dimming," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 573–582, Feb. 2015.

[29] T. O. Aydın, R. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full luminance range images," *Proc. SPIE*, vol. 6806, Feb. 2008, p. 68060B.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[31] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, Jul. 2011, Art. no. 40.

[32] K. Karađuzović-Hadžiabdić, J. H. Telalović, and R. K. Mantiuk, "Assessment of multi-exposure HDR image deghosting methods," *Comput. Graph.*, vol. 63, pp. 1–17, Apr. 2017.

[33] K. Karađuzović-Hadžiabdić, J. H. Telalovic, and R. Mantiuk, "Subjective and objective evaluation of multi-exposure high dynamic range image deghosting methods," in *Proc. Eurographics*, 2016, pp. 29–32.

[34] J. R. C. Chien, M. H. Sheu, S. K. Wang, and S. C. Hsia, "High-performance local dimming algorithm and its hardware implementation for LCD backlight," *J. Display Technol.*, vol. 9, no. 7, pp. 527–535, Mar. 2013.

[35] R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, M. Bessa, B. Waterfield, and A. Chalmers, "Objective and subjective evaluation of high dynamic range video compression," *Signal Process., Image Commun.*, vol. 47, pp. 426–437, Sep. 2016.

[36] R. Boitard, "Background in high dynamic range imaging," in *Temporal Coherency Video Tone Mapping*, vol. 1. Paris, France: Univ. Rennes, 2014, ch. 2, sec. 2, pp. 18–31. [Online]. Available: https://tel.archives-ouvertes.fr/tel-01127358/document

[37] M. D. Fairchild, "The HDR photographic survey," *Color Imag. Conf.*, vol. 2007, no. 1, pp. 233–238, Jan. 2007.

[38] B. T. Series, "Methodology for the subjective assessment of the quality of television pictures," Int. Telecommun. Union, Geneva, Switzerland, Tech. Rep. ITU-R BT.500-13, 2012.

[39] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images," *J. Electron. Imag.*, vol. 24, no. 1, Jan. 2015, Art. no. 010501.

[40] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[41] Holger Hoffmann. (2020). *Violin Plot*. MATLAB Central File Exchange. Accessed: Jan. 27, 2020. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/45134-violin-plot

LVYIN DUAN is currently pursuing the Ph.D. degree with Tianjin University. She is also visiting the Visualization Group, University of Warwick. Her research interests include display technologies and HDR image/video processing.

KURT DEBATTISTA received the B.Sc. degree in mathematics and computer science, the M.Sc. degree in psychology, the M.Sc. degree in computer science, and the Ph.D. degree from the University of Bristol. He is currently a Reader with WMG, University of Warwick. He has published over 100 peer-reviewed articles in journals and international conferences on high-fidelity rendering, HDR imagery, HPC, machine learning, and applied perception.

ZHICHUN LEI received the Ph.D. degree in communication engineering from University Dortmund, Dortmund, Germany, in 1998. He is currently a Professor with the Tianjin University, China, and the Ruhr West University of Applied Sciences, Mülheim, Germany. His current research interests include image processing, multimedia technology, and electromagnetic pulse data acquisition and imaging technology.

ALAN CHALMERS received the M.Sc. degree (Hons.) from Rhodes University, South Africa, in 1984, and the Ph.D. degree in computer science from the University of Bristol, in 1991. He is currently a Professor of visualization with the University of Warwick. He has successfully supervised 49 Ph.D. students and published over 250 articles in journals and international conferences on high-fidelity virtual environments, multisensory perception, and HDR imaging.

• • •