IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Learning Preferential Perceptual Exposure for HDR Displays

THOMAS BASHFORD-ROGERS[1], KURT DEBATTISTA[2], MIGUEL MELO[3], DEMETRIS MARNERIDES[4], MAX BESSA[5], AND ALAN CHALMERS.[6]
[1]University of the West of England, UK (e-mail: Tom.Bashford-Rogers@uwe.ac.uk)
[2]University of Warwick, UK (e-mail: K.Debattista@warwick.ac.uk)
[3]Universidade de Tráás-os-Montes e Alto Douro, Portugal (e-mail: emekapa@sapo.pt)
[4]University of Warwick, UK (e-mail: D.Marnerides@warwick.ac.uk)
[5]Universidade de Tráás-os-Montes e Alto Douro, Portugal (e-mail: maxbessa@utad.pt)
[6]University of Warwick, UK (e-mail: Alan.Chalmers@warwick.ac.uk)

Corresponding author: Thomas Bashford-Rogers (e-mail: Tom.Bashford-Rogers@uwe.ac.uk).

ABSTRACT High Dynamic Range (HDR) displays are capable of displaying a wider dynamic range of values than conventional displays. As HDR content becomes more ubiquitous, the use of these displays is likely to accelerate. As HDR displays can present a wider range of values, traditional strategies for mapping HDR content to Low Dynamic Range (LDR) displays can be replaced with either directly displaying values, or using a simple shift mapping (exposure adjustment). The latter approach is especially important when considering ambient lighting, as content viewed in a dark environment may appear substantially different to a bright one. This work seeks to identify an exposure value which is suitable for displaying specific HDR content on an HDR display under a range of ambient lighting levels. Based on data captured with human participants, this work establishes user preferred exposure values for a variety of maximum display brightnesses, content and ambient lighting levels. These are then used to develop two models to predict the preferred exposure. The first is based on linear regression using straightforward image statistics which require minimal computation and memory to be computed, making this method suitable to be directly used in display hardware. The second is a model based on Convolutional Neural Networks (CNN) to learn image features which best predict exposure values. The CNN model generates better results than the first model at the cost of memory and computation time.

INDEX TERMS High Dynamic Range, Machine Learning, Perception.

I. INTRODUCTION

High Dynamic Range (HDR) imaging grants the ability to capture, manipulate and display real-world lighting, as opposed to Standard or Low Dynamic Range (LDR) imaging, which is limited to 8-bit content and may result in over and/or under exposed pixels. HDR is quickly becoming the future of imaging with commercial availability becoming more popular; yet, commercial displays are limited to only a fraction of real-world lighting. In this work we present a study using a $10,000cd/m^2$ display which has a much wider dynamic range than commercial HDR displays and is significantly brighter; displays with these specs will likely form the basis of the next generation of HDR displays. This is of increasing importance as larger amounts of HDR content are being generated and applications ranging from multimedia video processing pipelines and entertainment software [1], inverse tone mapping [2], HDR super-resolution [3], quality assessment [4] and HDR coding [5] rely on HDR values, and display technology is able to reproduce these values [6]. However, the range of most HDR displays is still substantially lower than the values encoded in the HDR content. In order to be able to display this content on an HDR display, a remapping process, known as Tone Mapping, is required to adapt the HDR values to the displayable range.

While there are many Tone Mapping Operators (TMOs) which are capable of mapping HDR content to HDR or LDR displays, see Banterle et al. [7] for an overview, many of these operators produce artifacts such as color shifts and loss of contrast, see Akyuz et al. [8], Mantiuk et al. [9], Rana et al. [10] and Shirley et al [11]. While there are approaches to address these problems, for example Pouli et al. [12], Akyuz et al. [13] showed via a user study that there appears to be a preference for one of the simplest operators: exposure adjustment. This does not lead to color shifts or loss of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2898910, IEEE Access

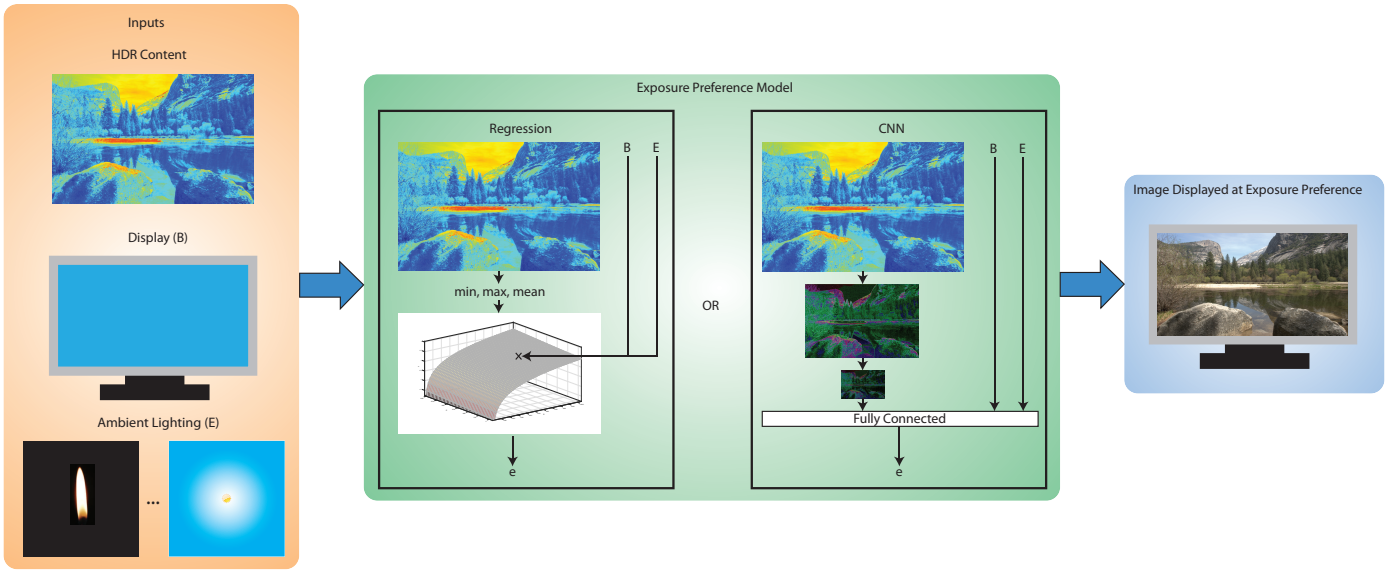

FIGURE 1: Overview of our work. To generate an exposure preference, the input HDR image, maximum display brightness and ambient lighting (orange box) are input into either the regression or the CNN model (green box), and the resulting exposure value for the image is generated for display (blue box).

contrast; however this clips under and over exposed values to the minimum and maximum displayable values, leading to a loss of information at the extremes. While this is significant for LDR displays, HDR display technology is capable of displaying a significantly wider dynamic range, with the result that fewer values are clipped. Exposure adjustment is a simple operation which scales the value of each pixel by an exposure value e: $D_p = 2^e I_p$, where D_p is the output pixel value, i.e. the value sent to the screen, and I_p is the input pixel value.

Although picking an exposure which maximizes the displayable content is feasible, see Debattista et al. [14], this approach neglects the viewing conditions of the user. For example, if the best exposure in terms of image content leads to bright values which are then displayed in a dark environment, visual discomfort may be experienced [15], and likewise dark content viewed in a bright environment may not be visible. Therefore, this indicates that the exposure needs to be augmented with knowledge of the ambient illumination.

The aim and novelty of this work lies in the development of models which predict preferential exposure of HDR content based on the display capabilities of the device, the ambient illumination on the display and the image or frame content. These models are designed to produce exposure preference values which are generalized across differing HDR displays, ambient illumination scenarios, and HDR content. This means our models are applicable to any display, under any ambient lighting, viewing any content.

In order to generate these models, we conducted a participant-based experiment to gather preferred exposures from a series (N = 30) of HDR images with varying dynamic ranges (10.96 - 29.15) across a range of maximum display brightnesses ($500cd/m^2$ - $10,000cd/m^2$) and ambient lighting values ($0lux$ - $4,000lux$). This was used to develop two models that predict preferred exposure. The first uses regression on image statistics in combination with the display and ambient lighting properties, and the second uses a Convolutional Neural Network (CNN) to predict the exposure value. Two models are proposed as the first is aimed at low computational and memory requirements, as might be found in display hardware, and the CNN approach gives more accurate results with a larger computational and memory overhead.

Our approach is summarized in Figure 1, and the major contributions of this work are as follows:

- Data capture of preferred exposure values for multiple display brightness levels, ambient lighting values for a series of HDR images, using human participants.
- A statistical analysis of the results that shows that the display brightness, ambient illumination and content are all significant factors to be considered when displaying HDR images.
- Predicting preferential exposure of HDR media from the scene content, the display illumination capabilities, and ambient lighting in the environment through the use of two models a linear regression model and a CNN.
- Results showing that both these models are able to predict preferred exposure values.

II. RELATED WORK

Ambient Lighting on LDR Displays

The effects of ambient lighting on LDR displays has been the subject of much research. Increased ambient illumination levels lead to a decrease in perceived contrast [17]. Multiple methods have been developed to mitigate this. Ware [18] proposed gamma correction to mitigate contrast loss. Devlin

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2898910, IEEE Access

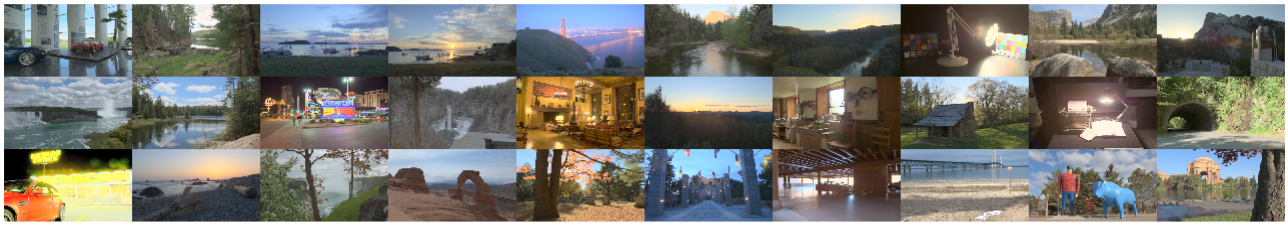

FIGURE 2: The set of HDR images used in the experiment from the Fairchild database [16].

et al. [19] measured JNDs for contrast in dark and bright environments, and from this developed a model for contrast enhancement for displays under varying ambient lighting conditions. Lebedev et al. [17] minimized the difference between the original image and the image viewed under a different ambient illumination level based on a user study. Other work has developed models for the relationship between display brightness and ambient lighting [20]. Guterman et al. [21] studied the effect of ambient illumination on digital LDR signage and found a weak effect of ambient lighting. Mantel et al. [22] performed a study which investigated users preferences for using backlight dimming for displaying videos on LCD screens. This used LDR content on an LDR device, and did not find a significant effect of ambient illumination by itself, but had an effect when considering image content.

Kane and Bertamío [23] investigated the preference of gamma when viewing HDR content on LDR displays with dark and light backgrounds. Similar to previous work [24], [25], they found that both the background and image content had an impact on chosen gamma values, with users preferring a linear gamma where image content permits it.

There has also been a significant amount of work on adaptive brightness control for small screen devices. Swinkels et al. [26] developed an approach to adjust screen brightness while avoiding flicker in changing lighting environments. Other work has investigated and developed models for backlighting preferences on small screen displays. Na et al. [27], [28] who performed user studies on adjusting backlighting of small screen devices based on ambient lighting levels to find optimal viewing preferences. Schuchhardt et al. [29] developed a personalized approach to adaptive backlighting levels on small screen devices, and showed that backlighting preferences on these devices vary amongst users.

Tone mapped HDR content presented on LDR display devices under varying levels of ambient illumination was investigated by Melo et al. [30], [31] investigated reflections from small screen devices when displaying tone mapped HDR video. They found that reflections as well as the HDR content significantly affect the viewing experience. Wang et al. [32] improved the contrast of tone mapped images using the Bartleson-Breneman [24] equations.

Krawczyk et al. [33] developed a linear model for brightness adjustment for LDR displays based on user studies. Similar to our work, they recognized the importance of image content in the brightness preference, specifically the importance of image key preferred brightness of an image. Our work differs from theirs in building a model for HDR displays, takes ambient lighting into account, and exploits more image features.

Ambient Lighting on HDR Displays

The effect of ambient illumination on HDR displays was first investigated by Rempel et al. [34]. They performed a study which measured contrast preferences and visual fatigue when viewing HDR content on HDR display devices. They found a sublinear relationship between ambient lighting and preferred display brightness, and no significant effect on visual fatigue.

Mantiuk et al. [35] presented a tone mapping operator which adapted to display conditions including ambient lighting. They minimized the difference between a modeled response of the Human Visual System for viewing an HDR image, and the image presented on the display. This model takes into account the various aspects of the display, such as maximum and minimum brightness, ambient lighting and display panel reflectance. Our work is different in that we seek to avoid the non-linear tone mapping step, and investigate a wider range of ambient lighting and display brightnesses.

Borer [36] performed a similar experiment to this work. The effect of ambient lighting and display brightness was investigated, and a model which adapted a gamma function given viewing conditions was presented. The experiment aimed to match the perception of images on HDR displays under different viewing conditions, and adjusted a custom gamma function to achieve this. This work used a maximum display brightness of $2,000cd/m^2$ and maximum ambient brightness of $500cd/m^2$. Our work has a different objective, and takes into account a wider range of both display brightness and ambient lighting conditions.

Other related work has considered aspects such as as brightness comfort for tone re-targeting applications [37], modelling brightness perception for content on HDR displays [38], calibrating HDR displays [39] and perception of HDR images [40].

Unlike previous work, the presented work is, to the best of our knowledge, the first of its type to analyze and evaluate the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2898910, IEEE Access

effect of ambient lighting, HDR scene content and display illumination on the user experience. The effect is shown to be significant and hence novel models are proposed that take into account these characteristics in order to predict preferred exposure for HDR media.

III. MOTIVATION

The goal of this paper is to develop a model of display exposure for HDR images. As the link between image content, display brightness and ambient lighting is unclear, this work studies the relationship via an experiment presented in Section IV, and found to be significant. This serves as motivation to use ambient lighting and display brightness as input into a model in Section V.

This work uses exposure as a method for adapting display viewing. In this work we choose to use exposure adjustment as the method of adapting the content for the display conditions. While there are many other tone mapping operators which could perform this task, we used the simpler method of exposure adjustment for two main reasons. The first is that tone mapping operators shift the image content, for instance by altering contrast, which we wanted to avoid in the experiment. Secondly, the method of exposure adjustment only requires one parameter to be controlled by participants. Other operators which depend on more parameters introduce methodological problems in the experiment in that users would have to adjust more variables, leading to infeasibly long experiments. Furthermore, this serves to not to introduce another competing tone mapping operator, but to develop a model for a simple operator to adapt to a wide range of display, image, and ambient lighting conditions.

Subsequently, we determine via the experiment which of these factors are perceptually significant. Based on the data gathered during the experiment, the impact of these factors can be analysed to determine whether, and to what extent, each factor influences the exposure value chosen. Analysis of experimental data motivates the development of the model based on the statistically significant factors. The data captured during the experiment is used again here in order to build and test the model. Two models are proposed, a straightforward one using linear regression and suitable for fast computation and a more complex method based on a deep learning architecture. The models demonstrate that they can suitably predict exposure values based on the three parameters of image content, display brightness and ambient lighting.

IV. EXPERIMENT

This section describes the methodology used to gather exposure preferences under varying display brightnesses and ambient lighting conditions. The results from this study are analyzed in this section, and these results are subsequently used to develop the models of preferential exposure values in Section V.

A. DESIGN

The goal of this experiment is to find exposure preferences for HDR content displayed under varying display brightnesses and ambient lighting conditions. We hypothesize that the exposure e is a function of maximum display brightness B, ambient lighting E and the HDR image I_x. For this experimental design e is the dependent variable, and B, E, I, and the response of each participant $part$ are independent variables.

The parameters $B \in [B_{min}..B_{max}]$, $E \in [E_{min}..E_{max}]$ and $I_x, x \in [0..N]$ form a three dimensional space, with an exposure preference associated with each point in that space $e(B, E, I_x)$. This experiment seeks to collect data for the values for the points in this space. In order to acquire these values, we sample this space to capture extrema as well as mid-points within this space, under the assumption that the mean exposure preference will be smoothly varying. The sampling points in each dimension are discussed below.

Display brightness B is a within participants independent variable. As the maximum display brightness is limited by display technology, we chose samples in the brightness dimension to be representative of current and future HDR display technology. The maximum display brightness B is divided into four levels representative of the types of HDR display currently available. These brightnesses were chosen to reflect a bright LDR display ($B_{min} = 500cd/m^2$), commercial and research HDR displays ($1,000cd/m^2$ to $4,000cd/m^2$), and future displays ($B_{max} = 10,000cd/m^2$ [41]).

E, ambient lighting, is a between-participants independent variable. Four representative lighting conditions of typical use cases of HDR displays are chosen to represent E. The first is a dark environment with a mean value of $E_{min} = 0.005lux$, followed by room lighting $394lux$, bright indoor lighting $1,470lux$, and finally outdoor lighting $E_{max} = 2,941lux$. For the final outdoor lighting condition, the experiments were performed over multiple days, between 11am and 4pm in the summer in a location without clouds. However, there was a small amount of variability in the ambient lighting in these results, which is considered when developing the model in Section V. A photograph of the indoor setup can be seen in Figure 3. The spotlights in this image were used to adjust the lighting levels around the screen to the required values and the rest of the room was also brightly lit. Based on recommendations from ITU-R BT.2022 [42], the participants were seated $1.8m$ from the screen.

Finally, for the HDR image I, also a within-participants variable, a selection of 30 HDR images were used. The order at which the images were presented was randomized for each participant, as well as the ordering of ambient lighting conditions and display brightnesses.

The dependent variable is the exposure chosen by each participant for all possible sets of (b, E, I).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2898910, IEEE Access

FIGURE 3: Photograph of the experiment setup.

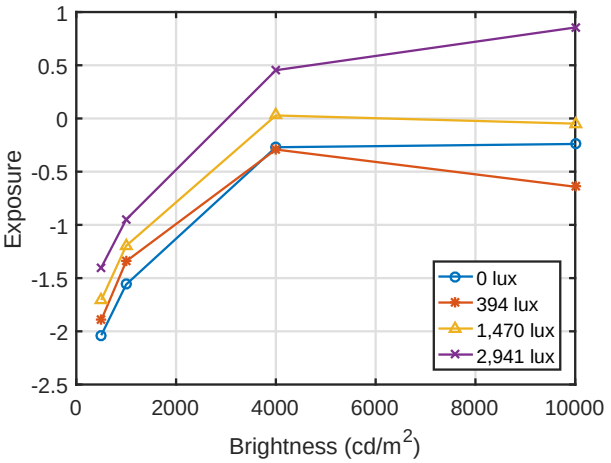

FIGURE 4: Descriptive statistics showing means for brightness B and ambient lighting E across all conditions, collapsed across images.

B. MATERIALS AND STIMULI

For the four conditions in B, the same display is used. To achieve this the maximum brightness is thresholded depending on the value of B. To capture the brightest display condition, we used a prototype HDR display from Sim2 with a maximum display luminance of $10,000 cd/m^2$. This has a significantly higher brightness and dynamic range than other current HDR display hardware, and allows us to achieve the goal of developing a model that is suitable for a wide range of displays.

For E, the first three conditions were performed in an indoor environment with controllable and stabilized lighting, and the fourth used outdoor lighting. For the outdoor lighting condition the display was moved to a shaded location in a sunny outdoor environment. For each participant, we measured the ambient lighting at the display using a Sekonic L-758D DigitalMaster. These measurements were averaged over several positions on the screen and were captured facing the participant.

The images I were chosen from the Fairchild database [16]. These were selected to have a range of image statistics which are additionally used to develop the models. A list of images used are shown in Figure 2. These images are purposely all high quality and uncompressed, chosen to avoid any artifacts stemming from image compression or poor capture. The original images are used without grading. All images were resized to display resolution 1920×1080.

Software was developed to display the HDR images on the HDR display in such a way that the brightest values are clamped to the selected maximum brightness. Backlight modulation was used regardless of the maximum simulated brightness of the display, i.e. this work is only considering HDR displays. The software permitted participants to use the "left" and "right" arrow keys to adjust the current exposure value of the HDR image down and up respectively in coarse steps of $0.01ev$, and the "up" and "down" arrow keys were used for finer adjustment with a step size of $0.0001ev$. The initial exposure for each image was set to a random value between $\pm 10ev$ of the 0 exposure. This range was determined from a pilot study where we found higher initial exposure values caused visual discomfort in the case of the brightest display value.

C. PARTICIPANTS AND PROCEDURE

80 participants, consisting of 26 females and 54 males with an average age of 24.7, from a local university volunteered for this experiment. They all exhibited normal or corrected to normal vision.

The participants were randomly allocated to one of the four conditions in E, and after a brief introduction and overview were asked to begin the experiment. Each participant was asked to adjust a slider which changed the exposure of the image on the HDR display to find an exposure which best displayed the image content. We also asked participants to not sacrifice visual discomfort for the sake of more details, which aimed to prevent eye strain during the experiment. This question was designed to explain to participants that the aim was to adjust the image to a consistent standard, i.e. a similar amount of visible detail. Other questions could have been asked which may have achieved different results, such as overall image quality, image aesthetic quality or fidelity to original image content. However, these questions would have led to further subjective aspects being included in the results, which is not the aim of the experiment.

Each participant performed the exposure preference selection for each of the four display brightnesses B, for 30 images I, under one ambient lighting E. such that 20 volunteers participated in each of the E conditions. The total time taken per participant was about 15 minutes.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2898910, IEEE Access

E / B	500 cd/m^2	1,000 cd/m^2	4,000 cd/m^2	10,000 cd/m^2	B_μ
0 lux	-2.0370 ($\pm$1.3601)	-1.5595 ($\pm$1.3917)	-0.2704 ($\pm$1.6644)	-0.2397 ($\pm$1.7132)	-1.0267 ($\pm$1.5323)
394 lux	-1.8947 ($\pm$1.1937)	-1.3416 ($\pm$1.1384)	-0.2919 ($\pm$1.5494)	-0.6416 ($\pm$1.6053)	-1.0425 ($\pm$1.3717)
1,470 lux	-1.7065 ($\pm$1.1205)	-1.1960 ($\pm$1.0581)	0.0282 ($\pm$1.5111)	-0.0487 ($\pm$1.3115)	-0.7308 ($\pm$1.2503)
2,941 lux	-1.4012 ($\pm$1.0405)	-0.9494 ($\pm$0.9816)	0.4534 ($\pm$1.1954)	0.8548 ($\pm$1.2639)	-0.2606 ($\pm$1.1204)
E_μ	-1.7599 ($\pm$1.1787)	-1.2617 ($\pm$1.1425)	-0.0202 ($\pm$1.4801)	-0.0188 ($\pm$1.4735)	-0.7651 ($\pm$1.3187)

TABLE 1: Descriptive statistics showing means for brightness B and ambient lighting E across all conditions, collapsed across images. The term in brackets denotes the variance of the participant responses, averaged across all images. B_μ and E_μ denote the mean values for brightness and ambient lighting respectively.

D. INITIAL RESULTS AND ANALYSIS

In order to identify whether it is worth building a model to express $e(B, E, I_x)$ an initial analysis is conducted to verify differences between the conditions. In order to do so the results are collapsed across the image I variable and descriptive and inductive statistics are analyzed.

Descriptive statistics for the means across all E and B conditions can be seen in Table 1. Figure 4 graphs the results. The results show an increasing overall trend in the chosen exposure values as brightness B increases and ambient lighting increases E.

Inductive statistics are analyzed using a 4 (E) $\times$ 4 (B) repeated measures factorial ANOVA. The main within-participants effect of brightness (B) is found to be significant, using Greenhouse-Geisser corrections (due to Maulchy's spherecity being violated $p < 0.05$) $F(2.2, 167.54) = 234.257$, $p < 0.01$. The between-participants main effect of ambient lighting (E) was also found to be significant $F(1, 76) = 96.63$, $p < 0.01$.

Pairwise comparisons using a t-test with Bonferroni corrections were also conducted in order to identify the significance of the difference among all the conditions. For B, significant differences were found between five of all the six pairs. The only condition were significant differences were not found was between $4,000cd/m^2$ and $10,000cd/m^2$ conditions. For E significant differences were found between the conditions of $0lux$ and $2,941lux$, and, $394lux$ and $2,941lux$.

The above results indicate sufficient differences between the data to build a model taking into account all the above parameters.

V. MODELS

We propose two models for preferred exposure prediction e. The models both take the maximum display brightness B and ambient lighting E as input, and combine this with the HDR frame I in order to perform the prediction. Our models are designed to work in two cases. The first is the case of streaming HDR content, as may be seen in display hardware, and so therefore this uses quantities which can be computed via a single, efficient, pass over the image. This significantly limits the image features used for prediction, but allows for a computationally efficient method. The second approach uses a multi-layer CNN with the HDR image as an input. This achieves lower error, but has higher computational and memory requirements.

The following sections describe both models, outline the rationale behind their construction, and present results from the models.

A. REGRESSION MODEL

The first model aims to satisfy the constraints of computing the exposure with a minimal computational overhead. Therefore, this model uses statistics of the HDR content which can be gathered over a single pass of the data, for example functions such as min, max, mean, log mean and their combinations can be gathered efficiently in an online fashion.

1) Overall Model

To develop this model, we first fit a surface to the experimental exposure preference results for each image. This surface is parameterized by the brightness B and ambient lighting E, and is fit using the Levenberg-Marquardt algorithm to the means of the experimentally determined exposure preferences. We explored a series of models of B and E, and empirically chose a surface of the form in Equation 1.

$$e(B, E, I_x) = \alpha_x + \beta_x log(B) + \gamma_x E, \qquad (1)$$

where α_x, β_x and γ_x represent the coefficients of the model, and the subscript refers to each HDR image. This fits the data well, with an average R^2 value of 0.8440 over the dataset. An example fit is shown in Figure 5. At this point, an exposure preference can be predicted from the brightness and ambient lighting for each image. However, the coefficients α_x, β_x and γ_x are unique to each image, and so we therefore correlate these values to statistics from the image in order to form a complete model. The coefficients α_x and β_x predominantly cover the majority of the variance, so γ_x is replaced with the mean value over the dataset $\gamma = 0.000274$.

2) Fitting the coefficients:

In order to identify which single pass statistics are most suitable for predicting α_x and β_x a number of single pass statistics were run through a regression with stepwise backwards entry. The single pass statistics used are calculated on luminance L_p of a pixel p and the N pixels in the image; these are:

- min L_m
- max L_M
- mean $L_\mu = \frac{1}{N}\sum_i L_{p(i)}$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2898910, IEEE Access

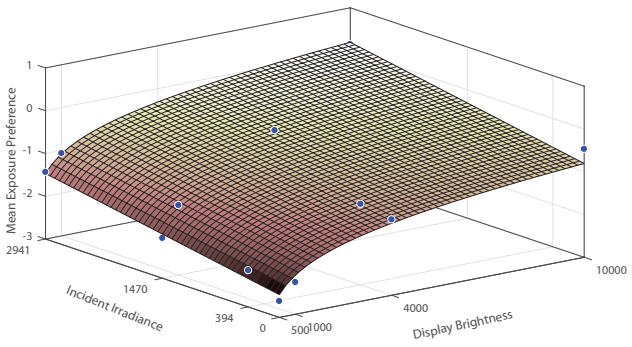

FIGURE 5: Plot of exposure preference for one of the tested images. The dots show the mean of the participants scores for the given brightness and ambient lighting, and the surface shows the resulting fit.

- log10mean $L_{log10\mu} = \frac{1}{N}\sum log_{10}(L_{p(i)})$
- geometric mean = $\sqrt[N]{\prod_i L_{p(i)}}$
- contrast = $\frac{L_M}{L_m+\epsilon}$ (ϵ is a small constant)
- dynamic range (DR) = $log_2\left(\frac{L_M}{L_m+\epsilon}\right)$,
- image key (IK) = $\frac{logL_\mu - logL_m}{logL_M - logL_m}$ (see [8])
- harmonic mean = $\frac{N}{\sum_i \frac{1}{L_{p(i)}}}$

Other statistics such as variance and kurtosis require multiple passes through the data and hence were not included in order to maintain the single pass nature of the method; the CNN model below is used as a more complex solution which learns the required statistics, whereas this model is designed for simplicity and efficiency. The stepwise backwards entry builds a regression model with all variables and removes them consecutively testing whether there are changes from the removal and thus helps identify which variables can be removed from the model without significant differences in the resulting model. For both α_x and β_x the variables log10mean, IK and DR are considered sufficient to represent the model and all other statistics are excluded without any significant differences in the model's predictive power.

Using these results, the expression for both α_x and β_x can be expressed as:

$$f_y(I) = c_y^0 + c_y^1 L_{log10\mu} + c_y^1 log_2\left(\frac{L_M}{L_m+\epsilon}\right) + c_y^1 IK, \quad (2)$$

where c refers to coefficients from the second regression, y refers to either α or β, and $IK = \frac{logL_\mu - logL_m}{logL_M - logL_m}$. The resultant final model is expressed as:

$$e(B,E,I) = f_\alpha(I) + f_\beta(I)log(B) + 0.000274E, \quad (3)$$

3) Results

We used 27 of the 30 input images to train the model, and validated the error on the remaining three. As the resulting model is dependent of the range of the simple statistics used to describe the training set, if images towards either extrema are excluded, as expected the model is less able to represent new data outside the trained ranges. Therefore, we always train with images including the smallest and largest min, max, log mean and dynamic range, and validate against the remaining images. We used a k-fold cross validation methodology, where except for the images with maximum and minimum values, every combination of images is validated against the model trained on the remainder. Using this methodology, the mean absolute error of this model is 1.15.

This model incurs a memory overhead of nine floating point numbers for the coefficients, totaling 36 bytes. This model is executed in $O(N)$ time where N is the number of pixels in the image. This is guaranteed due to the choice of restricting the model to use image statistics which can be computed in a single pass. Also, note for a specific display, $log(B)$ can be pre-computed if the maximum display brightness does not change. We implemented this method in C++, and ran it on a single thread on a MacBook Pro laptop i7-3740 CPU @ 2.7 GHz. This method takes $3.61ms$ to compute the predicted exposure value, averaged over $10,000$ runs. This timing excludes image loading times, and therefore shows the timings for the regression model by itself. It should also be noted that our implementation of the method could likely be further optimized to improve timings.

B. CNN MODEL

The second model we propose is based on a Convolutional Neural Network (CNN). The proposed CNN model can give a more accurate prediction of exposure, at the expense of computational cost. This model is trained to predict exposure preference based on a (b, E, I) combination. A subset of exposure preferences and the corresponding (b, E, I) combination from Section IV is used for training, and then the model is tested on the remaining set of preferences gathered from the experiment.

A (Deep) Neural Network consists of multiple layers of non-linear transformations. The input $\mathbf{x}_l$ of each layer l is linearly transformed by a matrix of learnable weights W_l and biases $\mathbf{b}_l$ followed by a non-linear function $f(a)$ applied to each dimension separately. The output of each layer $\mathbf{o}_l = f(W_l\mathbf{x}_l + \mathbf{b}_l)$, becomes the input of the next $\mathbf{o}_l = \mathbf{x}_{l+1}$ in a feedforward fashion. $f(a)$ is known as the activation function, and similar to many CNN architectures we use a Rectified Linear Unit (ReLU) $f(a) = max(0, a)$. The layers between the input and output layer are known as hidden layers.

A CNN contains specialized convolution layers [43], which exploit the spatial correlations in image data through the application of a convolution kernel. The kernel can again be expressed as a linear transformation matrix W_l, but with high sparsity and repeated weights. CNNs, usually contain subsampling layers such as max pooling, which returns the maximum of an $N \times N$ region of its input. Subsampling layers are used to reduce the dimensionality of the data at low computational cost and with no added parameters.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2898910, IEEE Access

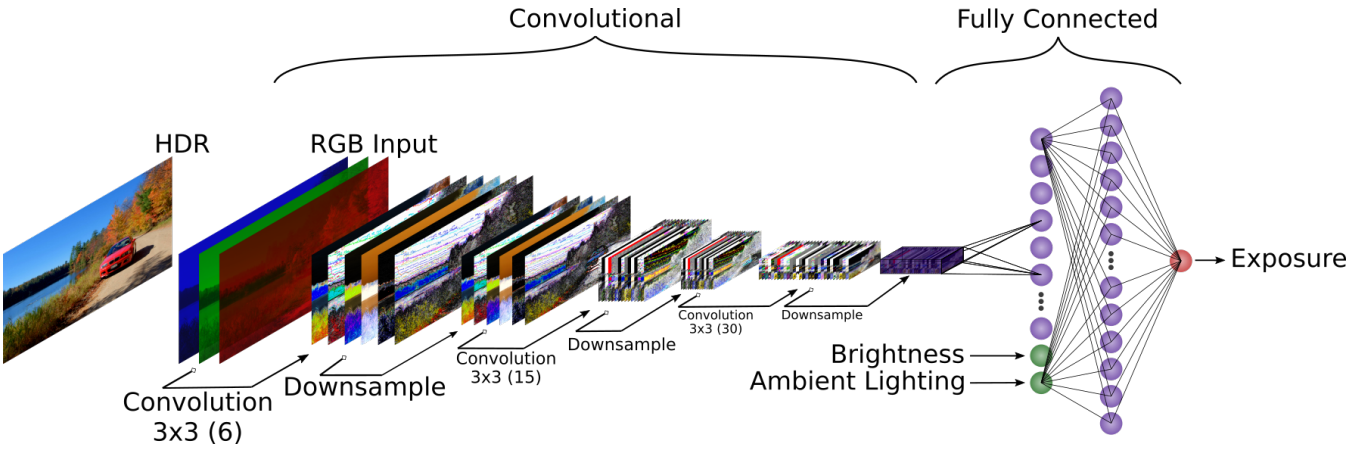

FIGURE 6: Diagram showing the structure of our CNN. The edges connecting the hidden layers represent parameters from the dense matrices W_l.

The weights and biases are learnt from data in a supervised fashion such that a chosen loss function is minimized. In our case we use the L1-Norm between the network output and the data. The loss is minimized by stochastic gradient descent (SGD) using the backpropogation algorithm [44]. The algorithm computes the gradient of the loss with respect to the network parameters using the chain rule. The gradient is then used to update the parameters of the network in the direction that minimizes the loss, one mini-batch at a time.

The input to our CNN is a full resolution unnormalized RGB HDR image I. It is first passed through three convolution and max pooling layers. Each layer uses a 3×3 kernel with stride 2, which halves the width and height of the layer. The kernel size of 3×3 is chosen due to it's efficient implementation in modern deep learning libraries for NVIDIA GPUs. It can also be shown that stacks of 3×3 kernels are equivalent to larger sized kernels. The first layer consists of 6 filters followed by 15 and 30 filters in the second and third layers. The filter bank sizes of the layers are increased as the network gets deeper in order to increase the dimensionality of the each pixel vector. This makes it easier for the network to disentangle different features of the input since points are more easily linearly separable if projected into higher dimensions.

We use max pooling with a kernel size of 2 and stride 2, as well as Batch Normalization [45] after each convolution. Max pooling lowers the dimensionality of the data and at the same time increases the receptive fields of the layers as the network gets deeper. Batch Normalization speeds up and stabilizes stochastic gradient descent by reducing internal covariate shift. The result of the first three layers is a set of high level features which describe the type of content in the HDR image. These features are then fully connected (dense, non convolutional W_l) to a hidden layer with 64 units which is then combined with the maximum display brightness B and ambient illumination values E. The resulting layer of size 66, is subsequently fully connected to a hidden layer of size 512. The last layer produces a single output: the exposure value. The full architecture is illustrated in Figure 6.

Our dataset of $9,600$ exposure preferences was split into a training set of $8,640$ exposure preferences, and a test set consisting of the remaining 960 values. We trained the network with PyTorch [46] for 100 epochs, using the ADAM optimizer [47] with a learning rate of 0.0001 ($\beta_1 = 0.9$, $\beta_2 = 0.999$). The learning rate was reduced by a factor of 0.25 every 20 epochs. We trained the network using a Nvidia Geforce GTX1080 and the total training time was 3 hours. Figure 7 shows the convergence of the CNN model as a function of training epochs. The dips show the change in learning rate.

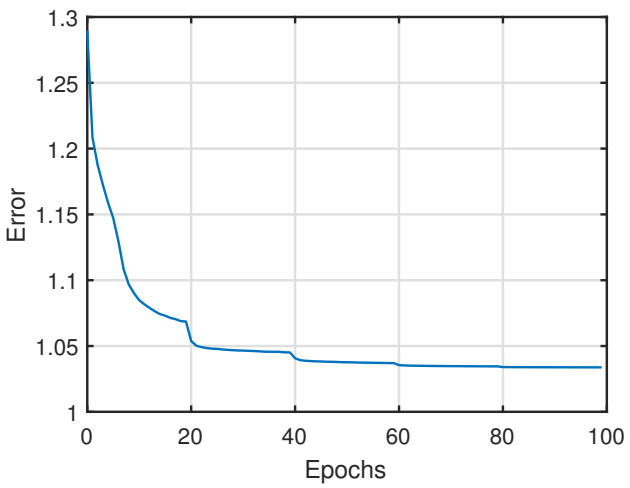

FIGURE 7: Convergence of the CNN model as a function of training epochs.

1) Results

For the CNN model, the mean absolute error of the test set, consisting of 10% of the dataset is lower at 1.03. The memory required for storing the $1,019,256$ weights and bias terms is $3.89MB$. The forward pass required to evaluate the

exposure value takes an average of $15.58ms$ to compute on at Nvidia GeForce GTX1080. This value was averaged over 10 runs, which like the regression model excludes image loading times.

VI. DISCUSSION AND FUTURE WORK

Of the two proposed models, the CNN approach results in a lower error. This is unsurprising as it is able to extract features which are likely to be better than the manually selected parameters for the regression model. However, the difference between the two models is not large, which suggests the regression approach is competitive with the more sophisticated CNN approach, and the single threaded, unoptimized CPU implementation is 4.3 times faster than the CNN evaluated on a high-performance GPU.

Inspired by the similarity of the errors of the two approaches, we intend to investigate what features the CNN is learning, and see if they could be applied to the more intuitive and human understandable regression model which achieves lower error, yet can be evaluated at minimal cost. Additionally, we intend to perform an architecture search to determine how to improve the CNN model as future work.

Both models were trained on a subset of exposure preference values gathered from the experiment. However, the experimental data was used differently to construct both models. The regression model was built based on fitting parameters given an image, therefore all participant data for that image was used. The CNN model was trained on individual measurements of exposure preferences and associated (b, E, I) values. Therefore, the testing of both models had to be performed slightly differently; the regression model was based on images not used in training, and the CNN model was based on a set of exposure preferences not used in training. This is a methodological necessity given the different nature of the two models, and future work will alleviate this by capturing more exposure preference values with a wider set of images to more clearly compare both models. While our models have good predictive accuracy, gathering further data would likely improve the models substantially.

The results of both the regression model and the CNN model have a mean absolute error of above 1, which is of a similar magnitude to the variance in the dataset. This indicates that users have a broad tolerance to detail in images displayed on an HDR display, and that our models produce valid results. However, we intend to conduct further studies to intend to propose a personalised user exposure preference model as future work.

We have developed models for static imagery, and in the future we intend to validate our models against video sequences to test for temporal stability. We believe our models are likely to be stable from one frame to the next, as both the regression model and the CNN approach predict smoothly varying exposure preferences for all the imagery used in this work. However, as highlighted by Boitard et al. [48], this will require a separate experiment for video, and may require development of a new model. Additionally, Bist et al. [49] found that images used in experiments should cover a wider range of image aesthetics and colour graded images. Again, we plan to extend this work to a wider image database as future work. We also intend to extend our work to use full-reference image quality assessment to HDR displays, similar to Hadizadeh et al. [50] with the aim of developing new models for displaying HDR images on a range of HDR displays without requiring extensive user studies.

VII. CONCLUSION

This paper has presented an approach for establishing and modelling viewer preferences for exposure of HDR content on a range of display and ambient illumination conditions. These viewer preferences were then used to develop two models to predict exposure preferences; the first model is based on regression and is designed to be used in situations with limited computational resources, and the second model based on CNNs is designed for more available computational and memory resources, and provides results with lower error. These models are designed to be used across a range of maximum display brightnesses ranging from current consumer level HDR displays to prototype HDR displays with significantly increased peak brightness levels, a range of ambient lighting levels from dark environments to outdoor bright situations, and across a range of HDR image content.

REFERENCES

[1] G. McTaggart, C. Green, and J. Mitchell, "High dynamic range rendering in valve's source engine," in ACM SIGGRAPH 2006 Courses. ACM, 2006, p. 7.

[2] T.-H. Wang, C.-W. Chiu, W.-C. Wu, J.-W. Wang, C.-Y. Lin, C.-T. Chiu, and J.-J. Liou, "Pseudo-multiple-exposure-based tone fusion with local region adjustment," IEEE Transactions on Multimedia, vol. 17, no. 4, pp. 470–484, 2015.

[3] J. S. Park, J. W. Soh, and N. I. Cho, "High dynamic range and super-resolution imaging from a single image," IEEE Access, vol. 6, pp. 10 966–10 978, 2018.

[4] A.-F. Perrin, M. Řeřábek, W. Husak, and T. Ebrahimi, "Quality assessment of an hdr dual-layer backward-compatible codec compared to uncompromised sdr and hdr solutions," IEEE Transactions on Broadcasting, vol. 64, no. 2, pp. 422–431, 2018.

[5] O.-J. Kwon, S. Choi, and D. Shin, "Improvement of jpeg xt floating-point hdr image coding using region adaptive prediction," IEEE Access, vol. 6, pp. 3321–3335, 2018.

[6] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," ACM Transactions on Graphics (TOG), vol. 23, no. 3, pp. 760–768, 2004.

[7] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, Advanced high dynamic range imaging. CRC press, 2017.

[8] A. O. Akyüz and E. Reinhard, "Color appearance in high-dynamic-range imaging," Journal of Electronic Imaging, vol. 15, no. 3, pp. 033 001–033 001, 2006.

[9] R. Mantiuk, A. Tomaszewska, and W. Heidrich, "Color correction for tone mapping," in Computer Graphics Forum, vol. 28, no. 2. Wiley Online Library, 2009, pp. 193–202.

[10] A. Rana, G. Valenzise, and F. Dufaux, "Learning-based tone mapping operator for efficient image matching," IEEE Transactions on Multimedia, 2018.

[11] P. Shirley, A. Robison, and R. K. Morley, "A simple algorithm for managing color in global tone reproduction," Journal of Graphics, GPU, and Game Tools, vol. 15, no. 3, pp. 199–205, 2011.

[12] T. Pouli, A. Artusi, F. Banterle, A. O. Akyüz, H.-P. Seidel, and E. Reinhard, "Color correction for tone reproduction," in Color and Imaging Confer-

ence, vol. 2013, no. 1. Society for Imaging Science and Technology, 2013, pp. 215–220.

[13] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do hdr displays support ldr content?: a psychophysical evaluation," ACM Transactions on Graphics (TOG), vol. 26, no. 3, p. 38, 2007.

[14] K. Debattista, T. Bashford-Rogers, E. Selmanović, R. Mukherjee, and A. Chalmers, "Optimal exposure compression for high dynamic range content," The Visual Computer, vol. 31, no. 6-8, pp. 1089–1099, 2015.

[15] M. Luckiesh and S. K. Guth, "Brightness in visual field at borderline between comfort and discomfort (bcd)," Illuminating engineering, vol. 44, no. 11, pp. 650–670, 1949.

[16] "Mark fairchild's hdr photographic survey," http://rit-mcsl.org/fairchild/HDR.html, accessed: 2017-08-09.

[17] D. Lebedev, G. Rozhkova, V. Bastakov, C. Kim, and S. Lee, "Local contrast enhancement for improving screen images exposed to intensive external light," Proc. Graphicon, Se7/62, 2009.

[18] C. Ware, Information visualization: perception for design. Elsevier, 2012.

[19] K. Devlin, A. Chalmers, and E. Reinhard, "Visual calibration and correction for ambient illumination," ACM Transactions on Applied Perception (TAP), vol. 3, no. 4, pp. 429–452, 2006.

[20] R. Merrifield and L. Silverstein, "The abc 's of automatic brightness control," in Society for Information Display Internatiational Symposium Digest of Technical Papers, 1988, pp. 178–181.

[21] P. S. Guterman, K. Fukuda, L. M. Wilcox, and R. S. Allison, "Is brighter always better? the effects of display and ambient luminance on preferences for digital signage," in SID Symposium Digest of Technical Papers, vol. 41, no. 1. Wiley Online Library, 2010, pp. 1116–1119.

[22] C. Mantel, J. Korhonen, J. M. Pedersen, S. Bech, J. D. Andersen, and S. Forchhammer, "Subjective quality of video sequences rendered on lcd with local backlight dimming at different lighting conditions," in SPIE/IS&T Electronic Imaging. International Society for Optics and Photonics, 2015, pp. 93 960S–93 960S.

[23] D. Kane and M. Bertamío, "Is there a preference for linearity when viewing natural images?" in Image Quality and System Performance XII, vol. 9396. International Society for Optics and Photonics, 2015, p. 939614.

[24] C. Bartleson and E. Breneman, "Brightness perception in complex fields," Josa, vol. 57, no. 7, pp. 953–957, 1967.

[25] C. Liu and M. D. Fairchild, "Re-measuring and modeling perceived image contrast under different levels of surround illumination," in Color and Imaging Conference, vol. 2007, no. 1. Society for Imaging Science and Technology, 2007, pp. 66–70.

[26] S. Swinkels, I. Heynderickx, D. Yeates, and M. Essers, "Ambient light control for mobile displays," in SID Symposium Digest of Technical Papers, vol. 39, no. 1. Wiley Online Library, 2008, pp. 1006–1009.

[27] N. Na and H.-J. Suk, "Adaptive display luminance for viewing smartphones under low illuminance," Optics express, vol. 23, no. 13, pp. 16 912–16 920, 2015.

[28] N. Na, K. Choi, and H.-J. Suk, "Adaptive luminance difference between text and background for comfortable reading on a smartphone," International Journal of Industrial Ergonomics, vol. 51, pp. 68–72, 2016.

[29] M. Schuchhardt, S. Jha, R. Ayoub, M. Kishinevsky, and G. Memik, "Caped: Context-aware personalized display brightness for mobile devices," in Proceedings of the 2014 International Conference on Compilers, Architecture and Synthesis for Embedded Systems. ACM, 2014, p. 19.

[30] M. Melo, M. Bessa, L. Barbosa, K. Debattista, and A. Chalmers, "Screen reflections impact on hdr video tone mapping for mobile devices: an evaluation study," EURASIP Journal on Image and Video Processing, vol. 2015, no. 1, p. 44, 2015.

[31] M. Melo, M. Bessa, K. Debattista, and A. Chalmers, "Evaluation of tone-mapping operators for hdr video under different ambient luminance levels," in Computer Graphics Forum, vol. 34, no. 8. Wiley Online Library, 2015, pp. 38–49.

[32] L. Wang and C. Jung, "Surrounding adaptive tone mapping in displayed images under ambient light," in Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on. IEEE, 2017, pp. 1992–1996.

[33] G. Krawczyk, R. Mantiuk, D. Zdrojewska, and H.-P. Seidel, "Brightness adjustment for hdr and tone mapped images," in Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on. IEEE, 2007, pp. 373–381.

[34] A. G. Rempel, W. Heidrich, H. Li, and R. Mantiuk, "Video viewing preferences for hdr displays under varying ambient illumination," in Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization. ACM, 2009, pp. 45–52.

[35] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," in ACM Transactions on Graphics (TOG), vol. 27, no. 3. ACM, 2008, p. 68.

[36] T. Borer, "Display of high dynamic range images under varying viewing conditions," in Applications of Digital Image Processing XL, vol. 10396. International Society for Optics and Photonics, 2017, p. 103960H.

[37] C. Bist, R. Cozot, G. Madec, and X. Ducloux, "Qoe-based brightness control for hdr displays," in Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on. IEEE, 2017, pp. 1–6.

[38] K. C. Noland, M. Pindoria, and A. Cotton, "Modelling brightness perception for high dynamic range television," in Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on. IEEE, 2017, pp. 1–6.

[39] E. Pieri and J. Pytlarz, "Hitting the mark—a new color difference metric for hdr and wcg imagery," SMPTE Motion Imaging Journal, vol. 127, no. 3, pp. 18–25, 2018.

[40] Y. Dong, M. T. Pourazad, and P. Nasiopoulos, "Human visual system-based saliency detection for high dynamic range content," IEEE Transactions on Multimedia, vol. 18, no. 4, pp. 549–562, 2016.

[41] A. Chalmers and K. Debattista, "Hdr video past, present and future: A perspective," Signal Processing: Image Communication, vol. 54, pp. 49–55, 2017.

[42] "Recommendation itu-r bt.2022," https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2022-0-201208-I!!PDF-E.pdf, accessed: 2017-10-05.

[43] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[44] D. E. Rumelhart, G. E. Hinton, R. J. Williams et al., "Learning representations by back-propagating errors," Nature, vol. 323, no. 6088, pp. 533–538, 1986.

[45] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proceedings of the 32nd International Conference on Machine Learning (ICML-15), D. Blei and F. Bach, Eds. JMLR Workshop and Conference Proceedings, 2015, pp. 448–456. [Online]. Available: http://jmlr.org/proceedings/papers/v37/ioffe15.pdf

[46] "pytorch," http://pytorch.org/.

[47] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[48] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch, "Zonal brightness coherency for video tone mapping," Signal Processing: Image Communication, vol. 29, no. 2, pp. 229–246, 2014.

[49] C. Bist, R. Cozot, G. Madec, and X. Ducloux, "Tone compatibility between hdr displays," in Applications of Digital Image Processing XXXIX, vol. 9971. International Society for Optics and Photonics, 2016, p. 99710D.

[50] H. Hadizadeh and I. V. Bajić, "Full-reference objective quality assessment of tone-mapped images," IEEE Transactions on Multimedia, vol. 20, no. 2, pp. 392–404, 2018.