VISUAL ATTENTION IN LDR AND HDR IMAGES

Hiromi Nemoto, Pavel Korshunov, Philippe Hanhart, and Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Station 11, CH-1015 Lausanne, Switzerland

ABSTRACT

Recent advances in high dynamic range (HDR) capturing and display technologies attracted a lot of interest to HDR imaging. Many issues that are considered as being resolved for conventional low dynamic range (LDR) images pose new challenges in HDR context. One of such issues is human visual attention, which has important applications in image and video compression, camera and displays manufacturing, artistic content creation, and advertisement. However, the impact of HDR imaging on visual attention and on the performance of saliency models is not well understood. Therefore, in this paper, we address this problem by creating a publicly available dataset of 46 HDR and corresponding LDR images with varying regions of interests, scenes, and dynamic range. We conducted eye tracking experiments and obtained fixation density maps, which demonstrate a significant difference in the way HDR and LDR capture attention of the observers.

1. INTRODUCTION

High Dynamic Range (HDR) imaging is one of the most promising technologies to enhance our visual quality of experience. HDR images can reproduce more realistic and more visually appealing content because they can represent the amount of light in the scene that is close to reality. Although HDR imaging has been widely studied in terms of picture quality or fidelity [1–3], the influence of HDR image on human visual attention is not yet well understood. However, clear understanding of this property is very important, because visual attention information is often required in many image and video applications such as gaze-adaptive compression [4], objective quality metrics [5], and image retrieval [6]. Since luminance contrast significantly affects visual attention [7], HDR images may lead to different human visual attention patterns compared to LDR images.

To take advantage of visual attention information in practical applications, automatic salient region detection algorithms, i.e., computational models for predicting where humans look in images without any human interactions, have been extensively investigated. This research trend resulted in many computational models of visual attention [8], as well as different datasets with ground truth eye tracking data [9]. Despite a number of studies on visual attention and eye tracking tests, there are few reports about the effect of HDR image on human visual attention.

Therefore the main objective of this paper is to understand the influence on the human visual attention when a conventional LDR image is replaced with an HDR image. To this end, we first created a new HDR public dataset[1] that contains 46 HDR images together with their LDR versions and covers large variety of content. We conducted an eye tracking experiment involving 20 naïve subjects to collect eye tracking data for these images using a professional eye tracking system Smart Eye Pro 5.8 and a commercially available HDR SIM2 monitor. From the raw tracking data, we computed fixation density maps (FDMs) to analyze the difference between salient regions in HDR and LDR images and compared them using similarity score metric [8].

In summary, the main contributions of the paper are:

1. Dataset of LDR and HDR images with corresponding subjective eye tracking data;
2. Similarity analysis of FDMs for LDR and HDR images to understand if there is a difference in visual attention.

The remainder of this paper is organized as follows. Section 2 presents background of this work including related works, computation of FDM and metrics for comparison of FDMs. Section 3 describes contents creation and eye tracking experiments, whereas Section 4 presents the results of subjective experiments. Section 5 concludes the paper.

2. BACKGROUND

2.1. Related work

Although many researchers have conducted a number of eye tracking experiments for the purpose of investigating human visual attention mechanism, there are only a few studies related to HDR images.

[1] http://mmspg.epfl.ch/hdr-eye

Josselin *et al.* [10] carried out an eye tracking test using the projector-based HDR display and then proposed a new computational visual attention model for HDR image, which is derived from Itti & Koch (2000) model [11]. The proposed model is, to the authors knowledge, the only existing model that takes higher dynamic range of HDR images into account. The authors, however, used only one image content in the experiments. Therefore, a further studies on a larger HDR dataset are necessary to better understand the impact of HDR on visual attention.

Narwaria *et al.* [12] investigated the impact of tone-mapping operators (TMOs) on human visual attention in HDR images. The authors performed an eye tracking experiment using both original HDR images and their tone-mapped versions. The results have shown that TMOs have a significant impact on visual attention patterns. While this study focused on the influence of TMOs on salient regions, the authors did not compare the difference in visual attention between HDR and typical LDR images (such images can be obtained by using the automated exposure mode of the camera). However, without clear understanding of this property, it is hard to understand whether HDR image has an impact on various computational models of visual attention, which were originally developed for LDR images. If HDR images have little effect on human visual attention when compared to LDR images, we might be able to use existing vision models for LDR images to estimate human response for HDR images.

2.2. Computation of fixation density maps

Fixation density maps (FDMs), which represent the level of attention at certain locations, are computed by convolving the recorded gaze points with a Gaussian filter, and then normalizing the result by the peak amplitude of the map into the range of 0 to 1.

To compute FDMs accurately, it is important to exclude gaze points associated with saccades and blinks, since visual fixation is not reflected during these eye movements. The eye tracking system used in our experiments (see Section 3.3) automatically discriminates between saccades and fixations based on the gaze velocity information. More specifically, during a time frame, all gaze points associated with gaze velocity below a fixation threshold are classified as fixation points, whereas saccades are detected when the gaze velocity lies above the fixation threshold. Blinks are also detected automatically by the eye tracking system based on the distance between the two eyelids of each eye. The remaining gaze points, which are classified into fixation, are then filtered with a Gaussian kernel to compensate the eye tracker inaccuracies and to simulate the foveal point spread function of the human eye.

As suggested in the state of the art [8, 13, 14], the standard deviation of the Gaussian filter used for computing the FDMs is set to 1 degree of visual angle, which corresponds to $\sigma = 60$ pixels in our experiments. This standard deviation value is based on the assumption that the fovea of the human eye covers approximately 2 degrees of visual angle.

2.3. Similarity score

The similarity score is a distribution-based metric of how similar two saliency maps are. The similarity score S between two normalized maps P and Q is

$$S = \sum_{i,j} min(P_{i,j}, Q_{i,j}), where \sum_{i,j} P_{i,j} = \sum_{i,j} Q_{i,j} = 1 \quad (1)$$

If a similarity score is one, the two saliency maps are the same, if it is zero, the maps do not overlap at all.

3. EXPERIMENTAL PROTOCOL

To investigate the difference in visual attention for LDR and HDR contents, we conducted an eye tracking experiment to acquire eye movements for both LDR and HDR still images. In this section, we first describe the strategy for content creation and then provide the details of the evaluation.

3.1. Content generation

Although there are several publicly available HDR image dataset, most of them contain only the resulted HDR images without providing the original bracketed LDR images. A few datasets that include original LDR images contain also color artifacts caused by image fusion, visible camera noise, or blurring artifacts caused by moving objects such as cars, moving trees, or walking people. For focus of attention experiments, to obtain practically useful results, a large variety of content is also desirable.

Therefore, in addition to a few selected images from the existing datasets (several images from EMPA HDR images dataset[2] and a few frames from 'Tears of Steel' short film[3], we have built a new public HDR dataset by combining nine bracketed images acquired with several cameras, including Sony DSC-RX100 II, Sony NEX-5N, and Sony α6000, with different exposures settings (-2.7, -2, -1.3, -0.7, 0, 0.7, 1.3, 2, 2.7 [EV]). We also used several images (obtained with Nikon D70 camera) from PEViD-HDR dataset [15] that shows different people under different lighting conditions.

To avoid ghost artifacts in fused HDR images due to camera shake and moving items, the cameras were placed on a tripod and special care was taken to avoid moving objects appearing in the pictures during the shooting. Open sourced Pictureaunt 3.2 software[4] was used for linearizing

[2]http://www.empamedia.ethz.ch/hdrdatabase/index.php
[3]https://media.xiph.org/mango/
[4]http://www.hdrlabs.com/picturenaut/

Table 1: Dynamic range of the scenes in the dataset

Dynamic range [dB][5]	Number of scenes
<48	13
48-60	7
60-72	8
72-84	11
>84	7

the bracketed exposures with the inverse of the camera response, and combining them into a single radiance map. For the better picture quality of fused images, we also used ghost removal and image alignment provided by the software.

The resulted dataset contains 46 images that cover a wide variety of content, e.g., natural scenes (both indoor and outdoor), humans, stained glass, sculptures, historical buildings, etc. Table 1 provides dynamic ranges of the scenes in the dataset.

3.2. Brightness adjustment

To reflect the real luminance of actual scenes, HDR images need to be reproduced with physically correct values using measured data, as suggested in [2]. However, most of the selected HDR pictures do not have this data, and the HDR monitor used in the test is not capable of yielding more than 4000 cd/m^2. This peak luminance of the monitor is not sufficient to display some of the bright scenes. Therefore, to make all HDR pictures look visually acceptable on the HDR monitor, we adjusted the brightness of the HDR images in accordance with the following equation proposed in [16]:

$$\log R_{new} = \log R - f(L) + c \quad (2)$$

$$f(L) = 0.28 \cdot L^{[9]} + 0.37 \cdot L^{[42]} + 0.35 \cdot L^{[100]} \quad (3)$$

where R and R_{new} are the original linear and adjusted luminance values, L represents the original logarithm luminance values, $L^{[p]}$ denotes the p-th percentile of the original logarithm luminance values, and c is a target logarithm luminance on a display. This approach can be interpreted more intuitively as the logarithm luminance of the original image being scaled in a way that matches the target logarithm luminance of the display. According to the literature, to estimate the best preferred brightness, the reference logarithm luminance of the original image $f(L)$ has to be computed based on the relative distribution of low, high, and mid-tones of the images, as shown in Equation 3, and 60% of the white luminance of the display is used as a target luminance c. We used 2000 cd/m^2 as white luminance, since this value was used for the color calibration of the monitor

To display LDR contents with the HDR monitor, we also converted the LDR images into radiance map representation. The images were linearized with a typical gamma

[5]HDR Toolbox for Matlab was used to compute dynamic range.

Table 2: Overview of the eye tracking experiments.

Category	Details	Specification
Participants	Number	20
	Age range (average age)	$18 - 56$ (25.3)
	Screening	Snellen and Ishihara charts
Viewing conditions	Environment	Laboratory
	Illumination	20 [lux]
	Color temperature	6500 [K]
	Viewing distance	1.89 [m]
	Task	Free-viewing
Display	Manufacturer	SIM2
	Model	SHDR47E S K4
	Type	LCD
	Size	47 [inch]
	Resolution	1920×1080 [pixels]
	Angular resolution	60 [pixel/degree]
Eye tracker	Manufacturer	Smart Eye
	Model	Smart Eye Pro 5.8
	Mounting position	0.7 [m] from the display
	Sampling frequency	60 [Hz]
	Accuracy	< 0.5 [degree]
	Calibration points	5 points on screen
Image presentation	Presentation order	Random
	Presentation time	12 [s]
	Grey-screen duration	2 [s]

Fig. 1: Experimental setup.

curve ($\gamma = 2.2$), then the pixel values were adjusted proportionally so that theoretical maximum pixel values of LDR image can match the peak luminance of common LDR monitor. ITU-R BT.2022 [17] specifies optimal peak luminance between 70 and 250 cd/m^2 in general viewing condition. We chose 120 cd/m^2 as the peak luminance because it is the default value in most monitor calibration software. Assuming that the LDR images taken with middle exposure setting of 0 [EV] are the most common LDR images, we used middle-exposed LDR images in radiance format in the eye tracking experiments.

3.3. Eye tracking experiments

The eye tracking experiments were conducted at the MM-SPG test laboratory, which fulfills the recommendations for subjective evaluation of visual data issued by ITU-R [18]. The viewing conditions were set according to recommendation ITU-R BT.2022 [17] and all subjects were naïve for the purpose of this study. Table 2 presents the detailed summary of the experiment and Figure 1 illustrates the physical experimental setup.

Each subject participated in two sessions of 13 minutes each with a 15 minutes break in between. All 46 contents were viewed by each subject in one session, and both HDR and LDR contents were displayed in the same session in a random order. Also, for half of the tested images, their HDR versions were displayed in the first session followed by the corresponding LDR versions in the second session. And for the other half of the images, the order was reversed: LDR versions were shown during the first session and HDR during the second. This approach was used to reduce the influence of potential memory effects on visual attention from viewing the same content twice. To reduce contextual effects, the stimuli orders of display were randomized applying different permutation for each subject. A training session (different images were used from those in the test) was organized to allow subjects to familiarize with the procedure before the test.

4. RESULTS OF EYE TRACKING EXPERIMENT

The resulting FDMs computed from the eye tracking data for LDR and HDR images were first inspected and compared visually. Figure 2 shows the LDR image, LDR FMD, tone mapped HDR image, and HDR FDM for contents exhibiting significant differences between LDR and HDR. In these examples, different FDM patterns can be observed, depending on scene characteristics. For example, it can be noted that viewers looked at more objects in some HDR images, e.g., the color chart in the dark part of content *C09* or the inscription below the statue on content *C40*.

While results show that viewers tend to look more at the bright objects in LDR images, details in the dark regions become more visible in HDR, resulting in the increased visual attention in these areas. This effect can be observed for content *C10* where viewers looked more attentively at the entrance door of the cathedral. Also, in some contents, focus of attention can shift from the bright areas of the LDR image to details in the darker areas of the HDR image. For example, in content *C16*, the attention was mostly focused on the building visible through the window in the LDR image, whereas the viewers mostly looked at the details of the statues located in the darker parts on the right and left side of the HDR image.

Table 3: Average similarity score between the FDMs of LDR and HDR.

Change in	Nb scenes	Similarity score	
		mean	std
Visual attention pattern	9	0.6742	0.1101
Fixation intensity	14	0.7447	0.0359
No change	23	0.7720	0.0379

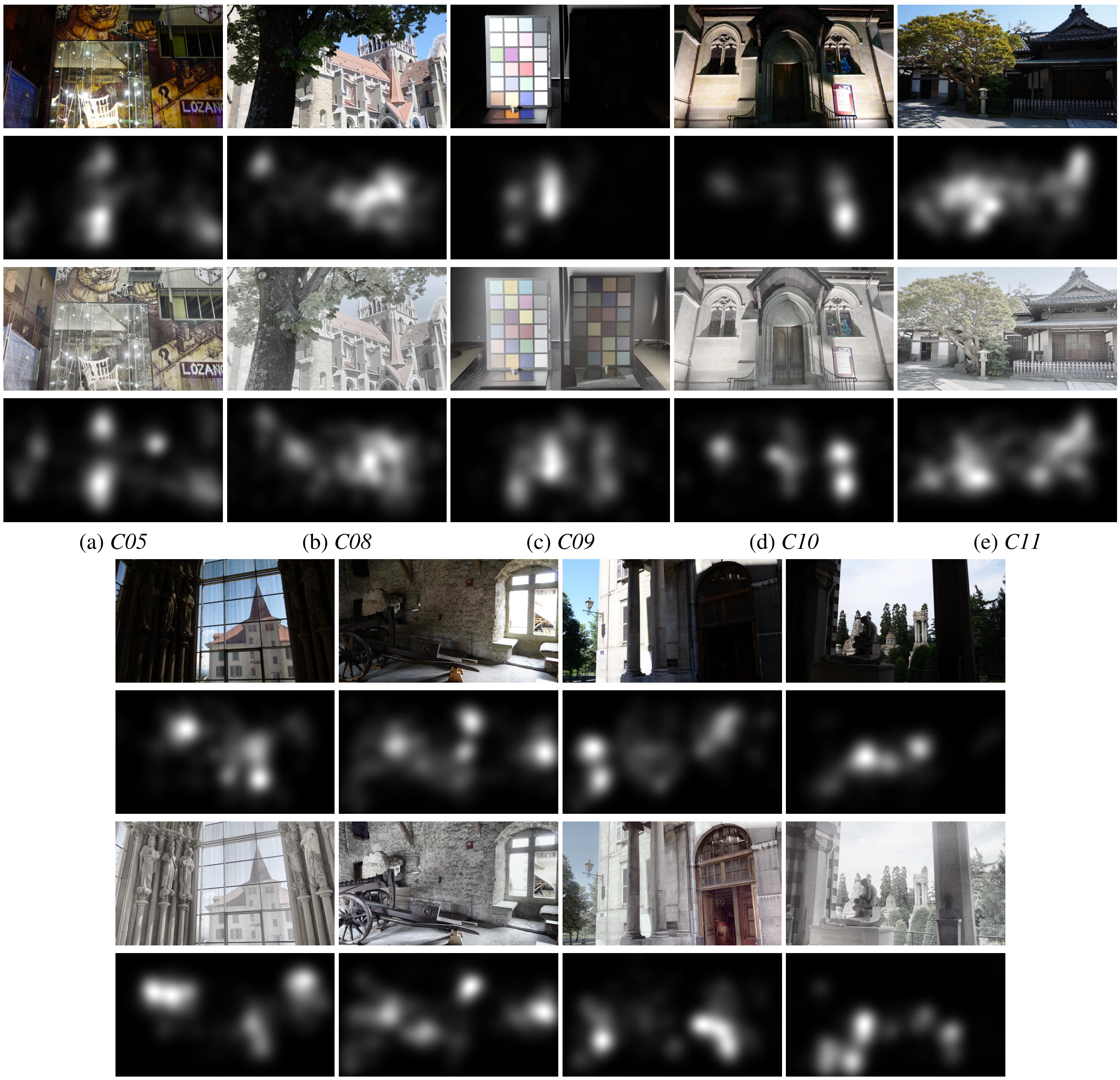
Table 4: p-value between each cluster of similarity score.

	Fixation intensity	No change
Visual pattern	0.0358	0.0007
Fixation intensity		0.0372

For some contents, the HDR FDM is mostly a modulated version of the LDR FDM, i.e., viewers looked at the same objects in both cases but with a different intensity. On the other hand, some contents did not show any significant difference between LDR and HDR FDMs. In particular, scenes containing human faces do not show any difference, as humans are very sensitive to human faces and are able to detect silhouettes easily, even in the dark regions.

Based on these observations, three clusters were manually created: (i) scenes that induce a change in visual attention pattern, (ii) scenes that induce a change in fixation intensity, and (iii) scenes that induce similar visual attention between LDR and HDR. Table 3 reports mean similarity score and its deviation computed on the images from these three clusters. From the table, it can be noted that the similarity score is lower when a change in the visual attention pattern or fixation intensity is observed in the FDMs. However, the difference between similarity scores for different clusters is not very large, which also indicates that similarity metric may not be the most suitable metric (note that FDMs in Figure 2) are visually different for LDR and HDR versions) to measure the changes in FDM that are caused by HDR.

To determine whether the difference between the three clusters is statistically significantly, we performed an analysis of variance (ANOVA) on the similarity scores. As shown in Table 4, the computed p-values indicate that the similarity scores are significantly different between the three clusters, in particular, between the scenes corresponding to visual attention pattern cluster and the scenes from the ‘no change’ cluster with LDR and HDR having similar FDMs. These findings show that, for some contents, HDR imaging impacts visual attention significantly, but it is not clear whether existing measurement tools can adequately measure this impact.



(a) *C05* (b) *C08* (c) *C09* (d) *C10* (e) *C11*

(f) *C16* (g) *C22* (h) *C32* (i) *C40*

Fig. 2: Examples showing significant visual differences between FDMs for HDR images and FDMs for LDR versions. First row: LDR version, second row: FDM of LDR, third row: tone-mapped HDR image, fourth row: FDM of HDR.

5. CONCLUSION

This paper investigated the impact of HDR imaging on human visual attention. For this purpose, a public HDR image dataset with images of wide variety of natural scenes was created. The dataset also contains original bracketed LDR images and fixation density maps (FDMs) from the eye tracking experiment. The eye tracking test demonstrated that FDMs of HDR images for some scenes are significantly different from the FDMs of the corresponding LDR versions.

Three clusters of HDR images were then identified: (i) with FDMs having different visual attention pattern compared to FDMs of LDR versions, (ii) with FDMs showing different distribution of fixation intensities compared to FDMs of LDR versions, and (iii) with FDMs that are similar to FDMs of LDR images. The applied similarity metric demonstrated that these clusters are dissimilar in statistically significant way. However, the similarity scores for clusters (i) and (ii) are not as small compared to cluster (iii) as it was expected, which means the metric did not capture the difference between FDMs adequately. Therefore, the impact of HDR on human visual attention is scene-dependent and it is hard to measure it using existing metrics.

Future work will focus on finding an automated way to classify scenes for better understanding of the influence of HDR on visual attention. Different metrics of visual attention need to be investigated to identify the metric that captures the differences in visual attention patterns caused by HDR. The impact of HDR imaging on computational models of visual saliency will also be considered.

Acknowledgments

This work has been conducted in the framework of Swiss SERI project "Compression and Evaluation of High Dynamic Range Image and Video", COST IC1005 The digital capture, storage, transmission and display of real-world lighting HDRi, EU Network of Excellence VideoSense, and FP7 EC EUROSTAR TOFuTV Project.

6. REFERENCES

[1] P. Hanhart, P. Korshunov, and T. Ebrahimi, "Subjective evaluation of higher dynamic range video," in *Proc. SPIE 9217*, 2014, Applications of Digital Image Processing XXXVII.

[2] 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*, vol. 26, no. 3, July 2007.

[3] F. Banterle, P. Ledda, K. Debattista, M. Bloj, A. Artusi, and A. Chalmers, "A psychophysical evaluation of inverse tone mapping techniques," in *Computer Graphics Forum*, 2009, vol. 28, pp. 13–25.

[4] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004.

[5] J. Redi, H. Liu, P. Gastaldo, R. Zunino, and I. Heynderickx, "How to apply spatial saliency into objective metrics for JPEG compressed images?," in *16th IEEE International Conference on Image Processing (ICIP)*, Nov. 2009, pp. 961–964.

[6] K. Vu, K. A. Hua, and W. Tavanapong, "Image Retrieval Based on Regions of Interest," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, pp. 1045–1049, 2003.

[7] W. Einhäuser and P. König, "Does luminance-contrast contribute to a saliency map for overt visual attention?," *European Journal of Neuroscience*, vol. 17, no. 5, pp. 1089–1097, 2003.

[8] T. Judd, F. Durand, and A. Torralba, "A Benchmark of Computational Models of Saliency to Predict Human Fixations," *MIT tech report*, Jan. 2012.

[9] S. Winkler and R. Subramanian, "Overview of Eye tracking Datasets," in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2013, pp. 212–217.

[10] J. Petit, R. Brémond, and J.-P. Tarel, "Saliency maps of high dynamic range images," in *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, 2009.

[11] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, vol. 40, no. 10, pp. 1489–1506, 2000.

[12] M. Narwaria, M. Perreira Da Silva, P. Le Callet, and R. Pepion, "Tone mapping based HDR compression: Does it affect visual experience?," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 257–273, 2014.

[13] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, May 2006.

[14] U. Engelke, A. Maeder, and H. Zepernick, "Visual attention modelling for subjective image quality databases," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2009, pp. 1–6.

[15] P. Korshunov, H. Nemoto, A. Skodras, and T. Ebrahimi, "Crowdsourcing-based Evaluation of Privacy in HDR Images," in *SPIE Photonics Europe 2014, Optics, Photonics and Digital Technologies for Multimedia Applications*, Brussels, Belgium, Apr. 2014.

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

[17] ITU-R BT.2022, "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays," International Telecommunication Union, Aug. 2012.

[18] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Jan. 2012.