Evaluation of High Dynamic Range Content Viewing Experience Using Eye-Tracking Data

(Invited Paper)

Eleni Nasiopoulos[1], Yuanyuan Dong[2,3] and Alan Kingstone[1]
[1]Department of Psychology, University of British Columbia
[2]Department of Electrical & Computer Engineering, University of British Columbia, Canada
[3]Institute for Computing, Information and Cognitive Systems, University of British Columbia, Canada
enasiopoulos@psych.ubc.ca, yuand@ece.ubc.ca, and alan.kingstone@ubc.ca

***Abstract*—High Dynamic Range (HDR) technologies have demonstrated that they can play an influential role in the design of camera and consumer display products. Understanding the human visual experience of viewing HDR content is a crucial aspect of such systems. Although the visual experience of Low Dynamic Range (LDR) technologies have been well explored, there are limited comparable studies for HDR content. In this paper, we present a study that evaluates the viewing experience of HDR and LDR content as measured both subjectively, and objectively vis-a-vis eye-tracking data. The eye-tracking data was collected while individuals viewed HDR or LDR videos in a free-viewing task. Our study shows a clear subjective preference for HDR content when individuals are given a choice between HDR and LDR displays, but this preference does not translate into a reliable difference in the subjective or objective eye movement measures when the displays are viewed sequentially, suggesting that objective performance measures are not the foundation upon which subjective preferences are based. Our findings should help in developing new visual attention models for the role of HDR and LDR content on subjective and objective experience and performance.**

Keywords-High dynamic range; HDR; Eye-Tracking; Visual attention model.

I. INTRODUCTION

In the real world, luminance ranges many orders of magnitude (100 000 000:1). The two types of photo receptors in the eye, rods and cones, allows the human eye to accommodate a range of 10 000:1 in a single view. However, conventional digital images and display devices only represent a dynamic range of 100:1. To provide a life-like visual experience with significantly higher contrast and richer color, High Dynamic Range (HDR) imaging was proposed. With a truthful representation of the real world, more details and wider color gamut, HDR images and videos have many important applications in various areas such as surveillance, medical imaging, digital photography, and computer graphics. Over the last decade, there has been much research conducted on all stages of the HDR pipeline from capturing, compression and transmission, to HDR display and tone mapping techniques enabling HDR content to be shown on traditional devices.

This work was supported in part by NSERC under Grant STPGP 447339-13, the Institute for Computing, Information and Cognitive Systems (ICICS) and TELUS People & Planet Friendly Home Initiative at UBC.

Camera sensors have been developed with the capability to capture HDR images and videos, though they are still relatively expensive. In the solutions with only low dynamic range (LDR) sensors used, it is also possible to generate HDR content by computationally fusing multiple exposures of the scene. Existing compression standards are designed for LDR content, which is usually represented in the manner of 8-bits per pixel integers. However, HDR content has higher color bit-depth and is mostly formatted in floating-point to store more details in a wider intensity range. This also introduces more information, more details in both dark and bright areas of the image, and also larger amounts of data. In terms of displaying HDR contents, significant progress has been made as well. A few HDR display prototypes have been designed in academia with dynamic range beyond 50000:1 [1] and are becoming commercially available. Since the HDR display is not widely available yet, displaying HDR content on conventional devices during the transition from LDR to HDR still poses a problem. Thus, tome-mapping techniques have been designed to convert the dynamic range of HDR to dynamic range of LDR.

With all kinds of emerging techniques making HDR more readily available, some subjective studies have been published with regards to human preferences for HDR content. A psychophysical experiment conducted by Akyüz et al. [2] evaluated the effects of dynamic range and mean luminance, finding that participants prefer HDR displays to LDR displays. It also found that for LDR images, simply boosting the intensity range linearly to fit the HDR display could give a better viewing experience. In another subjective study conducted by Masia et al. [3] reverse tone-mapping operators were evaluated with input images at different exposure conditions. The results indicated that existing inverse tone-mapping operates well with correctly-exposed and dark images, while the performance decreases for input images containing large bright areas.

Eye tracking is a well-established method of capturing an individual's looking behavior in order to measure visual attention. Eye trackers record eye movements to estimate gaze direction and have long been used as a technique for analyzing visual attention, studying the human visual system, and even assessing user interfaces [4]. An eye tracking system can provide data such as number of fixations, fixation duration, velocity, scan paths, saliency maps, etc., which make it

QSHINE 2014, August 18-20, RHODES, Greece

DOI 10.4108/icst.qshine.2014.256428

possible to measure the difference between two distinct gaze behaviors.

In this study, we use a SensoMotoric Instruments (SMI) eye tracking system to evaluate the gaze behaviors while participants view HDR content on an HDR display and LDR content (generated by tone mapping) on an LDR display.

II. Experiment

In this section we explain the HDR display system used to display the visual stimuli, the properties of our HDR content, and the procedure for conducting the experiments.

A. *Stimuli*

Nine HDR videos in total (see *Table 1*) were used in this study. The first two videos are from the JCT-VC - Joint Collaborative Team on Video Coding (JCT-VC) [5], Technicolor created them using a simultaneous low exposure/high exposure capture with a rig of two Sony F3 or F56 cameras; the following five sequences were captured at the Stuttgart Media University (HdM) by Froehlich et al. [6], and the final two were captured at the University of British Columbia with RED SCARLET cameras which capture dynamic range up to 18 stops.

Video sequences were chosen in order to span a wide variety of scenes that included day or night lighting conditions, different ranges of motion (i.e., minimal motion or fast moving objects) and a wide range of color spectrums. We limited the social context of the scenes, and those scenes that did include social components (i.e., people or human interaction) were kept neutral in nature, in order to effectively isolate different picture elements.

For each HDR video sequence, a corresponding LDR video was generated using photographic tone reproduction method proposed by Reinhard et al. [7]. Since the image based tone-mapping method processes each frame independently, flickering appears in some sequences as a result of abrupt changes in the mapping between consecutive frames. As such, an additional temporal coherency algorithm for video tone mapping proposed by Boitard et al. [8] was used, in combination with photographic tone reproduction to generate LDR videos with temporal coherency.

B. *Displays*

Tests were performed on two displays; one LDR display and one HDR display system (see *Fig.1*). The system modulates the image from a projector with a transmissive LCD. The display comprises a monochrome projector with a dynamic range of 800:1 in the back as the base, and a HD LCD screen with a dynamic range of 300:1 in the front. With the imperfections in the optical path, the measured dynamic range is 54000:1 and measured peak luminance is 2700 cd/m^2. Fig.1 illustrates the HDR display system used in our study.

To display HDR content, HDR videos are processed to generate two calibrated streams sent to the projector and LDR respectively, based on the procedure described in [1]. The stream sent to the projector consists of a monochrome intensity signal, while the other stream to the LCD screen contains the color content.

Besides the HDR display system, an LDR display is used to present tone-mapped sequences. The LDR display used in this study is a Hyundai 42" LCD display. The display was calibrated to white point 6500K and peak rightness 120 cd/m^2 before the study, to represent a typical LDR screen available on the market.

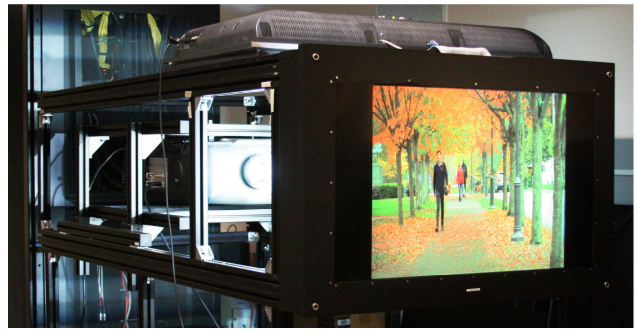


Fig. 1. HDR display system

C. *Eye Tracking System*

Eye movements of participants were tracked using the SensoMotoric Instruments (SMI) iView X RED system. The eye tracker was mounted on a tripod and participants were seated in a chair that allowed their eye height to be adjusted to meet the set up requirements of the SMI system. This setup is shown in *Fig. 2*. The sampling frequency of the SMI is 250 Hz and the resolution accuracy is 0.4 ± 0.03°.

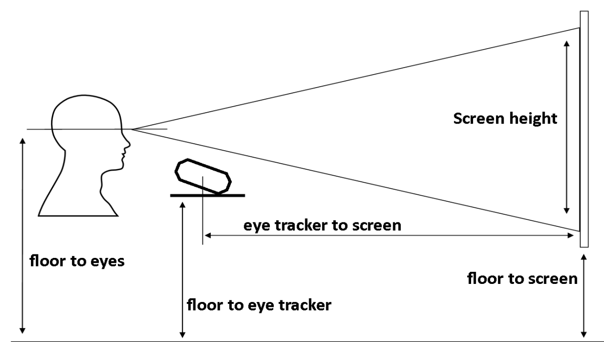


Figure 2. Setup of eye tracker[9]

D. *Participants*

18 individuals (distributed equally among males and females) participated in the study. All participants had normal or corrected to normal vision, and were screened for normal color vision. All subjects were naïve to the purpose of the experiment. Before each task, the participant's eye height and position was adjusted so that their eyes could be tracked accurately.

E. *Procedure*

Each participant completed two separate tasks during the study. In one task, participants were asked to 'free-view' the stimuli (meaning to watch the video naturally). Participants viewed the content on an HDR display and LDR display freely. In a second task, participants were given different viewing instructions, and were asked to pay attention to the detail in the video content in order to answer a short questionnaire about the video. The video content was identical in both tasks with only the instructions changing. *Table 1* outlines the video sequence (total viewing length of 115 seconds) that the participants viewed. To avoid a bias of viewing order, we performed a completely counterbalanced set up. The displays themselves were simply labeled as display A and display B.

Before each participant viewed the video sequence, a calibration was run which allows the eye tracker to determine where individuals are fixating on the screen, and ensures accuracy of the eye tracking data. The calibration stage was repeated if the quality of the calibration was not satisfactory. Before each video sequence, participants were asked to fixate on a dot presented at one of the four corners of a neutral gray background. This initial fixation triggered the start of a video trial. Note that by requiring participants to start each trial at one of the corners of the screen, we ensured that participants were free to choose where to first begin looking at the material presented on the displays, thereby avoiding any artificial center bias for viewing images [9]. The corner fixation dot was presented for 2s after each video and the corner location of the dot was randomized from one trial to the next.

At the end of the viewing tasks, participants filled out a short questionnaire where they answered how they thought the two different monitors were different with regards to the content they viewed, as well as which monitor they preferred viewing and why.

In order to allow for a more robust comparison, a separate questionnaire was administered to a new set of participants (n=5), where the monitors displayed identical content side by side. None of these participants had viewed HDR content before. The aim of this second set of participants was to allow for a direct comparison to be made of the two monitors.

F. Eye movement results

In order to extract the looking behavior data, we created Areas of Interest (AOIs) around key elements in the video stimuli that were considered especially eye catching (i.e., people, moving objects, etc). From these Areas of Interest we measured how many fixations participants made within the AOIs, how long their fixations were, and the number of times the AOIs were revisited.

For the free viewing condition (see *Fig. 3*), a comparison of the average fixations within the AOIs across the different video sequences yielded no significant difference between the LDR display (M=138.5, SD=38.4) and the HDR display (M=144.3, SD=29.5), t(17)=-.462, p=.65. Similarly, in the detailed viewing instructions task there was no significance between the LDR display (M=115.3, SD=41.1) and HDR display (M=122.3, SD=31.7) conditions, t(17)=-.591, p=.562.

We also extracted dwell time, the amount of time the fixations lasted within the AOIs, as shown in *Fig. 4*. In the free viewing condition, this yielded no significant difference between the LDR display (M=56917.8, SD=14646.3) and HDR display (M=57598.3, SD=15308.4), t(17)=-.222, p=.827. The detailed viewing instructions task also produced no significant difference between the LDR display (M=45081.8, SD=15365.4) and the HDR display (M=45751.9, SD=13579.3), t(17)=-.172, p=.865.

An examination of the number of times participants revisited AOIs (how many times they came back to look at a previously attended to AOI) yielded a similar set of findings (see *Fig. 5*). There was no significant difference for the LDR display (M=45.9, SD=16.6) and the HDR display (M=45.1, SD=11.0) for the free viewing condition, t(17)=.195, p=.848; and no significance between the LDR display (M=37.8, SD=16.7) and the HDR display (M=40.3, SD=12.9), t(17)=-.755, p=.461, for the detailed viewing instructions.

G. Questionnaire Results

Results of the questionnaire revealed that most participants, 14 out of the 18 tested, perceived the two monitors (LDR vs HDR) as noticeably different. Because 5 participants responded that they had seen HDR content before, we removed them from the analysis of the monitor preference question, leaving us with 13 responses for our analysis. Out of the 13 participants, 7 stated that they preferred the HDR monitor compared to 6 participants preferring the LDR one. This difference was no greater than what one would expect by chance X^2=.077, p=.782. Note that a similar pattern emerged even when the 5 excluded participants were included X^2=.222, p=.637.

The separate questionnaire administered to the new set of participants, where the monitors were displaying identical content side by side, resulted in all participants reporting a preference for the HDR monitor. Participants reported preferring the HDR monitor over the LDR monitor for its superior 'clarity and life-like detail' or because objects were 'looking more real'.

III. CONCLUSIONS

The present study examined the impact of LDR and HDR displays on subjective and objective measures. Objective eye movement data indicated that naive observers were unaffected by the HDR versus LDR displays. Both displays had a similar effect on frequency and temporal eye movement data, specifically, the number of fixations to areas of interest (AOIs), the time spent fixating those AOIs, and the frequency of revisits to those AOIs. Furthermore, these objective measures were echoed by a non-significant difference in the preference for the HDR and LDR displays. Critically, however, when the two displays were pitted directly against each other, there was a consistent and unanimous preference for the HDR displays. Collectively these data indicate the a reliable subjective preference HDR displays emerges when an direct contrast to an LDR display is available, and does not arise from or lead to an objective performance difference in visual attention as measured by eye movement behavior.

REFERENCES

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

[2] A.O. Akyüz, R. Fleming, B. Riecke, E. Reinhard, and H. Bülthoff. "Do HDR displays support LDR content? a psychophysical evaluation," In ACM Transactions on Graphics (TOG), vol. 26, no. 3, pp. 38. ACM, 2007.

[3] B. Masia, S. Agustin, R. W. Fleming, O. Sorkine, and D. Gutierrez. "Evaluation of reverse tone mapping through varying exposure conditions," In ACM Transactions on Graphics (TOG), vol. 28, no. 5, pp. 160. ACM, 2009.

[4] I. Laurent and C. Koch. "Computational modelling of visual attention." Nature reviews neuroscience, vol. 2.3. pp. 194-203. 2001.

[5] S. Lasserre, F. Le Léannec, E. François, "Description of HDR sequences proposed by Technicolor," ISO/IEC JTC1/SC29/WG11 MPEG2014/ m31957, Oct. 2014, San Jose, USA

[6] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel. "Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays," In IS&T/SPIE Electronic Imaging, pp. 90230X-90230X. International Society for Optics and Photonics, 2014.

[7] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. "Photographic tone reproduction for digital images," In ACM Transactions on Graphics (TOG), vol. 21, no. 3, pp. 267-276. ACM, 2002.

[8] R. Boitard, K. Bouatouch, R. Cozot, D. Thoreau, and A. Gruson. "Temporal coherency for video tone mapping," In SPIE Optical Engineering+ Applications, pp. 84990D-84990D. International Society for Optics and Photonics, 2012.

[9] SensoMotoric Instrucments (SMI), "Experiment center 2 manual," Version 2.4, 2010.

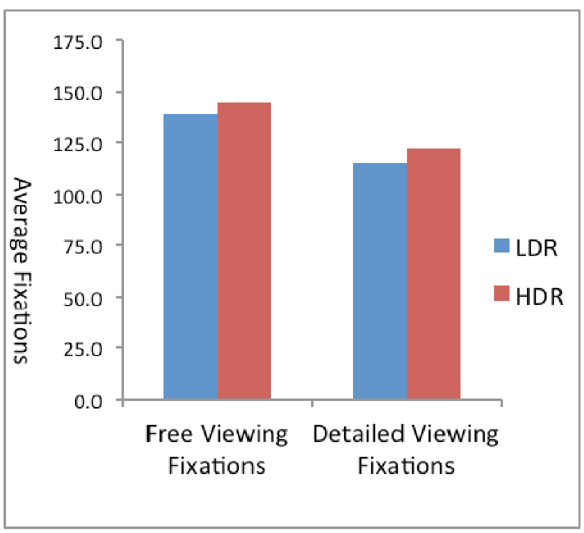
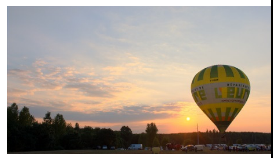


Figure 3. Comparison of the average fixations within the AOIs

Table 1. Sequences used in the study

Clip		Source	
Balloon 	200 frames 30 fps 1920 * 1080	JCT-VC	Exterior Medium color spectrum Slow global and local motion
Market 	400 frames 50 fps 1920 * 1080	JCT-VC	Exterior High illumination High color spectrum Static scene with slow motion
Bistro 01 	151 frames 30 fps 1920 * 1080	Froehlich et al.	Interior High contract scenery with local bright sunlight at the window Single moving object and slow motion
Bistro 02 	300 frames 25 fps 1920 * 1080	Froehlich et al.	Interior High contract scenery with local bright sunlight at the window
Bistro 03 	170 frames 30 fps 1920 * 1080	Froehlich et al.	Interior Medium illumination
Amusement Park 	439 frames 30 fps 1920 * 1080	Froehlich et al.	Exterior scene at night Wide color spectrum Fast motion
Fishing 	371 frames 30 fps 1920 * 1080	Froehlich et al.	Sunlight scene Sunlight reflection on water surface
Playground 	222 frames 30 fps 2048*10 80	captured	Sunlight exterior scene High illumination High color spectrum Fast motion
Mainmall 	241 frames 30 fps 2048*10 80	captured	Medium illumination Slow local motion

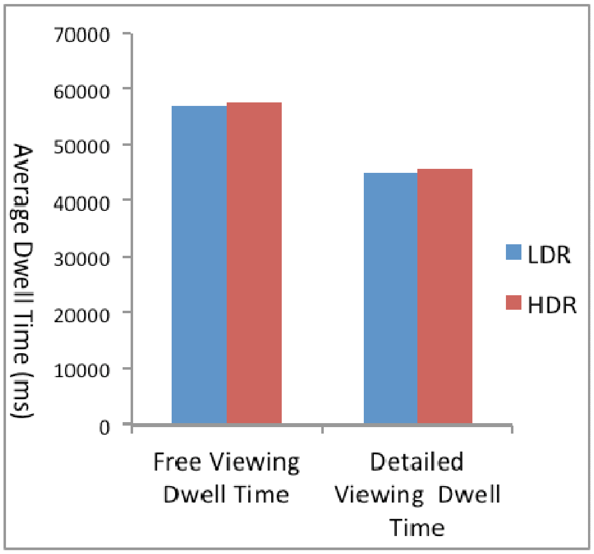


Figure 4. Comparison amount of time the fixations lasted within the AOIs

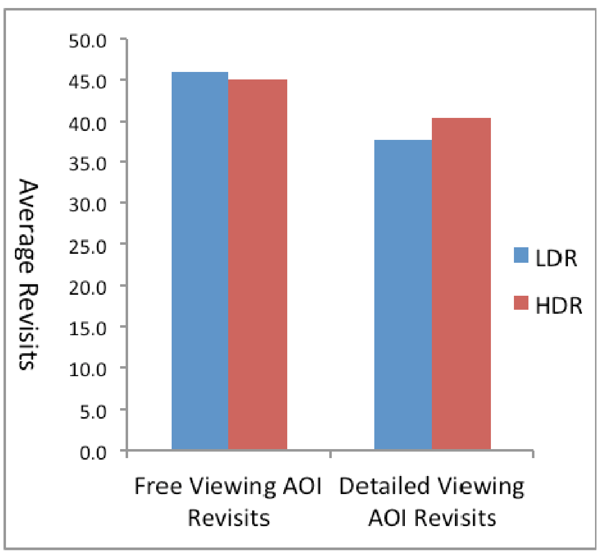


Figure 5. Comparison number of times participants revisited AOIs