Evaluating HDR Pipeline Relative to Tone Mapped SDR Pipeline

Hamzeh Issa

Motivation

- In recent years, a lot of progress was made in the field of HDR (High Dynamic Range) as it gets closer to having a more widespread use by customers.

- However, little is known about the real feel of the HDR experience among most people since it’s still restricted to certain devices and technologies and hasn’t reached the mass market yet.

Motivation

- Having Tone Mapping solutions readily available for the vast customer base introduced questions about the worthiness of HDR compared to tone mapped SDR:
- Will HDR be better than SDR, and if so, by how much?
- Also, is it much better than the tone mapped versions and is it really worth to work on this technology?
- These questions need to be answered to provide the industry and researchers working on HDR with intelligible information on what the users actually want from that technology and how can it be improved or modified to make it truly worth the effort and the extra price in the eye of the customer.

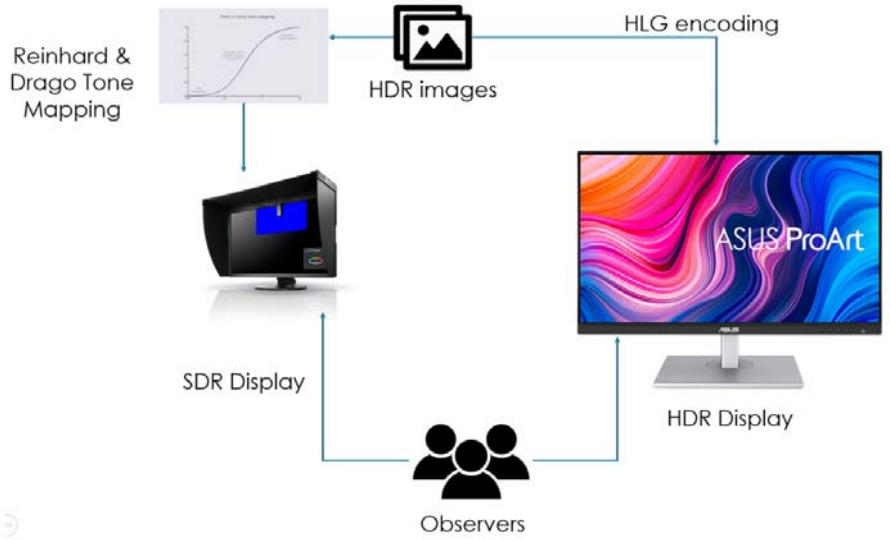
Motivation

- That produced a need for evaluating full HDR pipelines to Tone Mapped SDR pipelines.
- Previous works only delt with evaluating SDR tone mappings with each other [1][2].
- Or compared HDR to real scenes or other HDR pipelines [3].

- [1] Martin Cad´ık, Michael Wimmer, Laszlo Neumann, and Alessandro Ar- ˇ tusi. Evaluation of hdr tone mapping methods using essential perceptual attributes. Computers & Graphics, 32(3):330–349, 2008.
- [2] Rafa l Mantiuk, Scott Daly, and Louis Kerofsky. Display adaptive tone mapping. In ACM SIGGRAPH 2008 papers, pages 1–10. 2008.
- [3] Jiangtao Kuang, Hiroshi Yamaguchi, Changmeng Liu, Garrett M Johnson, and Mark D Fairchild. Evaluating hdr rendering algorithms. ACM Transactions on Applied Perception (TAP), 4(2):9–es, 2007.

Overview

- In this work, a psychophysical experiment is conducted between a full HDR pipeline that shows an HDR image on an HDR display and a tone mapping pipeline that shows tone mapped version of these images on an SDR display.



Methodology

- 20 naive observers with an age range of 22 to 34. each session lasted an average of 34 minutes.
- We needed the observer to not be affected by the high luminance of the HDR display while seeing and judging the SDR image on the SDR display, so both displays cannot be shown at the same time, so there was an adaptation time of 20 seconds between showing the user one screen or the other.
- The 20 seconds adaptation time was found using a heuristic approach with a small experiment done during the trials.

Methodology

- Since paired comparison can’t be done with an adaptation time, and since scaling or magnitude estimation is a proven and effective way for these types of experiments as discussed in the literature review section, low frequency magnitude estimation was chosen for this experiment.
- The observers were asked to judge the images on a scale of 1 to 4. Also, with the scales shown for the observers there were guidelines on what each scale represents (like category judgement) but these were mere guidelines to help the observer understand the scales correctly and not get confused.

Methodology

- The observer would look at the HDR image, and on their mark that they are quite content with it, the HDR display is covered with the dark sheet and the observer is asked to wait for 20 seconds, then the tone mapped SDR image would be shown on the SDR display.

Methodology

- The observer is asked to rate ”how much they liked the previous image compared to this image”. Meaning, how much do they think the HDR image is more pleasing compared to the SDR image.
- Judgement of the images was based on general quality, pleasantness and subjective preference of the observer.
- 4 would be the highest ranking (meaning that the HDR image was much better than the SDR one), and 1 would be the lowest, meaning that the HDR image was not good compared to the SDR image (that the SDR image is better).

Methodology

- However, after experimenting with 13 observers with this setup, we decided to increase the scale to (1 - 6) instead of (1 - 4) based on expert feedback that suggested the 4 scale was too coarse, and increasing it will provide more accuracy to the readings. So, the other 7 observers were tested with a 1 to 6 scale instead.

Methodology

- For the rating and displaying the images on the SDR display, Quickeval [4] was used since it offers a simple interactive framework. The background was gray with 20% of the luminance of the adapting white point and the observers were set around 60 cm from the display. Same dimensions were used for the HDR display.
- the software used to render the HDR images on the HDR display, a special Matlab library called Psychtoolbox [5] was used to render the images in their full 10-bit range.

- [4] Khai Van Ngo, Christopher Andr´e Dokkeberg, Ivar Farup, Marius Pedersen, et al. Quickeval: a web application for psychometric scaling experiments. In Image Quality and System Performance XII, volume 9396, page 93960O. International Society for Optics and Photonics, 2015.
- [5] Mario Kleiner, David Brainard, and Denis Pelli. What's new in psychtoolbox-3? 2007.

Methodology

- Finally, a small last test was conducted by showing the observers a full white blank screen on the HDR display and recording their reaction for about a minute.

Stimuli

- 40 images, 20 Reinhard, 20 Drago.
- Same images on HDR pipeline.
- Some images had good quality while others had a lot of noise.
- Some images had a lot of lightness and saturation and some had dark regions.

Stimuli















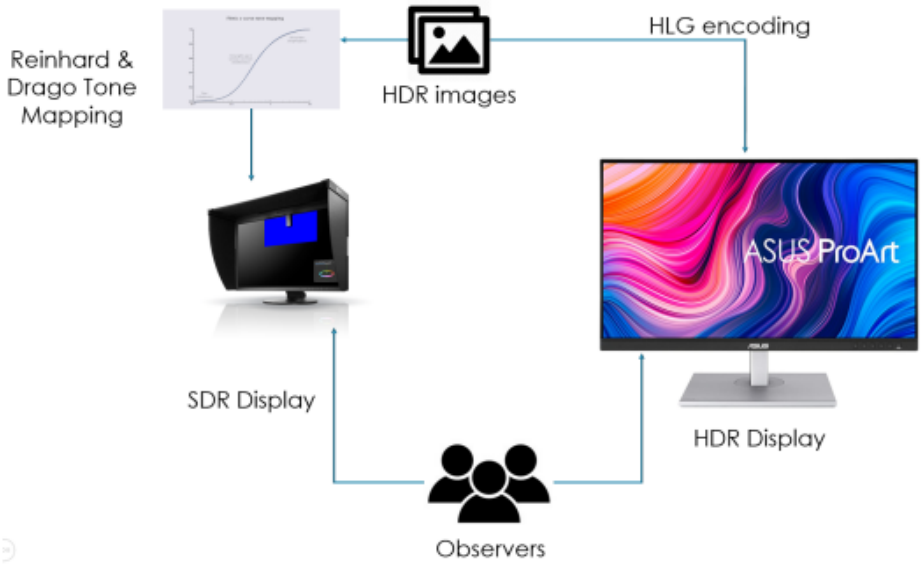




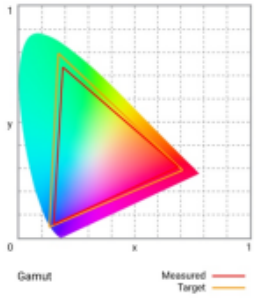
Experimental Setup

- SDR: The Eizo color edge cg248 professional display, maximum luminance of 173 cd/m2 . The area shown is 714x340 pixels with maximum brightness.
- HDR: Asus ProArt with maximum luminance of 1672 cd/m2 or nits set to Hybrid Log Gamma (HLG) with shown area of also 714x340 and maximum brightness. The display uses a BT2020 colour gamut and an HLG Electro-Optic Transfer Function (EOTF)
- 1 meter apart with dark cloth covering the other display.

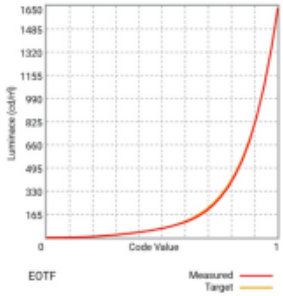
Experimental Setup



Target Name:HDR HLG Rec2020

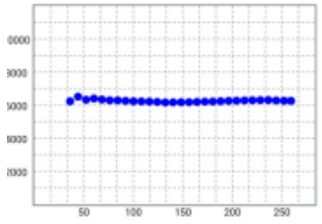


	Target	Result
Color Gamut	BT 2020/Rec.2020	-
Maximum Brightness	-	1627.92 nits
Minimum Brightness	-	0.00 nits
Contrast Ratio	-	> 1000000:1
Color Temperature	6500K	6515K
Gamut	100%	83.91%
EOTF	HLG	-
R	(0.71, 0.29)	(0.6884, 0.3058)
G	(0.17, 0.80)	(0.1892, 0.7368)
B	(0.13, 0.05)	(0.1499, 0.0584)



Summary

The average Δ E result is 0.71, your monitor quality meets the standard.



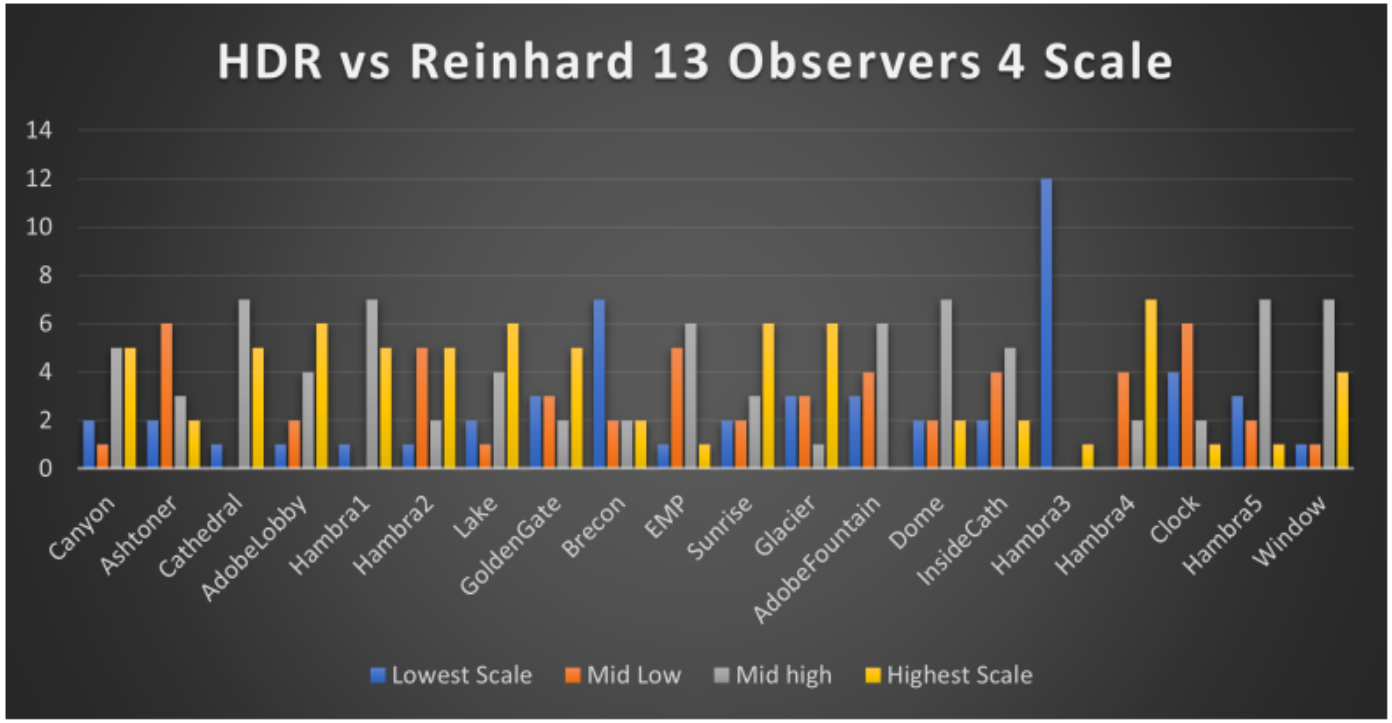
Overview

HDR;TargetGamut:BT 2020/Rec.2020;EOTF:HLG;WhitePoint:6500K

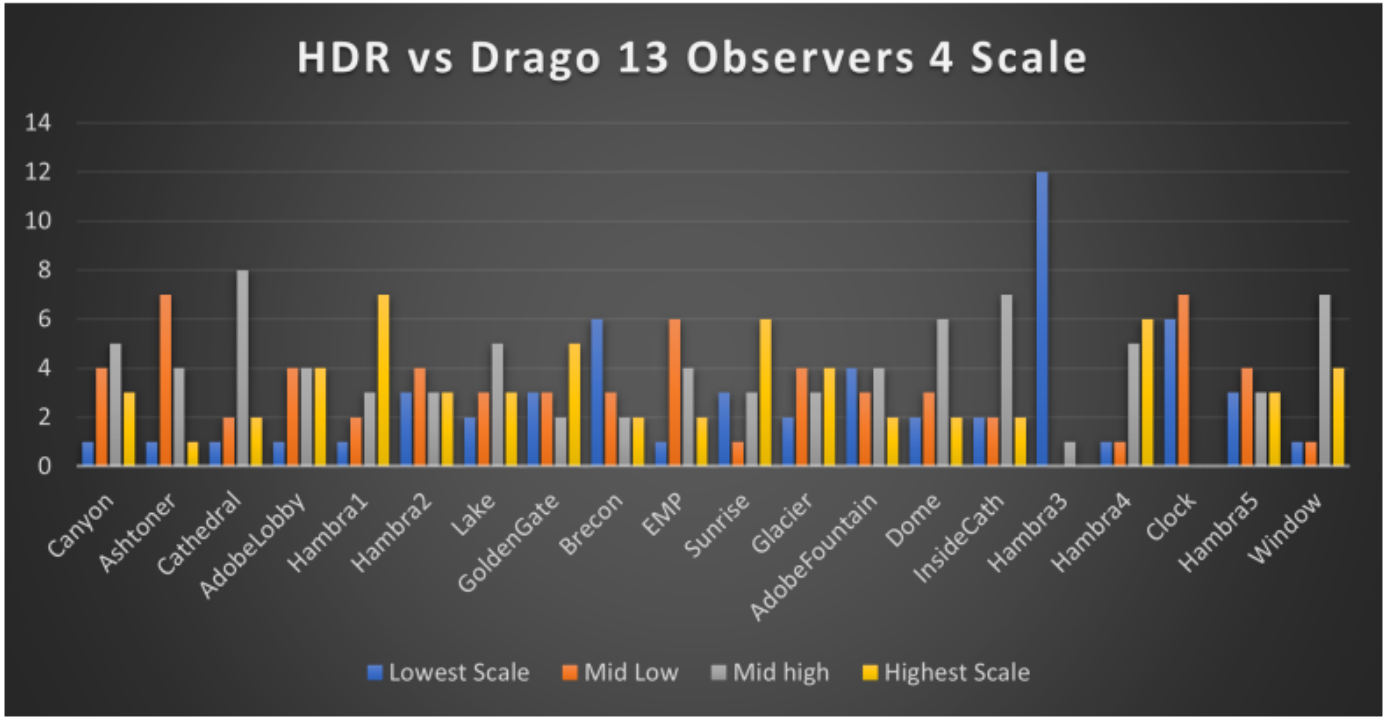
0-2 2-4 4-10

	Device Values			Measured Values			
	R	G	B	a*	b*	L	ΔE*
	8	8	8	0.00	0.00	0.00	0.02
	16	16	16	0.04	-0.01	0.16	0.06
	24	24	24	0.04	-0.03	0.48	0.06
	32	32	32	0.06	-0.03	1.08	0.10
	40	40	40	0.07	-0.17	1.94	0.21
	48	48	48	0.25	-0.22	3.23	0.46
	56	56	56	0.26	-0.35	5.05	0.64
	64	64	64	0.36	-0.39	7.20	0.89
	72	72	72	0.54	-0.44	9.33	1.07
	80	80	80	0.53	-0.43	11.41	0.98
	88	88	88	0.48	-0.34	13.82	0.94
	96	96	96	0.52	-0.30	15.67	0.86
	104	104	104	0.43	-0.22	17.73	0.69
	112	112	112	0.49	-0.22	19.62	0.75
	120	120	120	0.43	-0.12	21.70	0.65
	128	128	128	0.50	-0.05	23.70	0.77
	136	136	136	0.51	-0.07	26.00	0.79
	144	144	144	0.49	-0.06	28.48	0.76
	152	152	152	0.61	-0.14	30.98	1.03
	160	160	160	0.66	-0.22	33.67	1.26
	168	168	168	0.71	-0.31	36.91	1.42
	176	176	176	0.80	-0.39	40.48	1.61
	184	184	184	0.67	-0.39	44.56	1.47
	192	192	192	0.41	-0.34	49.04	1.17
	200	200	200	0.25	-0.30	53.82	0.98
	208	208	208	0.02	-0.26	59.17	0.66
	216	216	216	-0.20	-0.23	64.85	0.60
	224	224	224	-0.20	-0.31	71.04	0.54

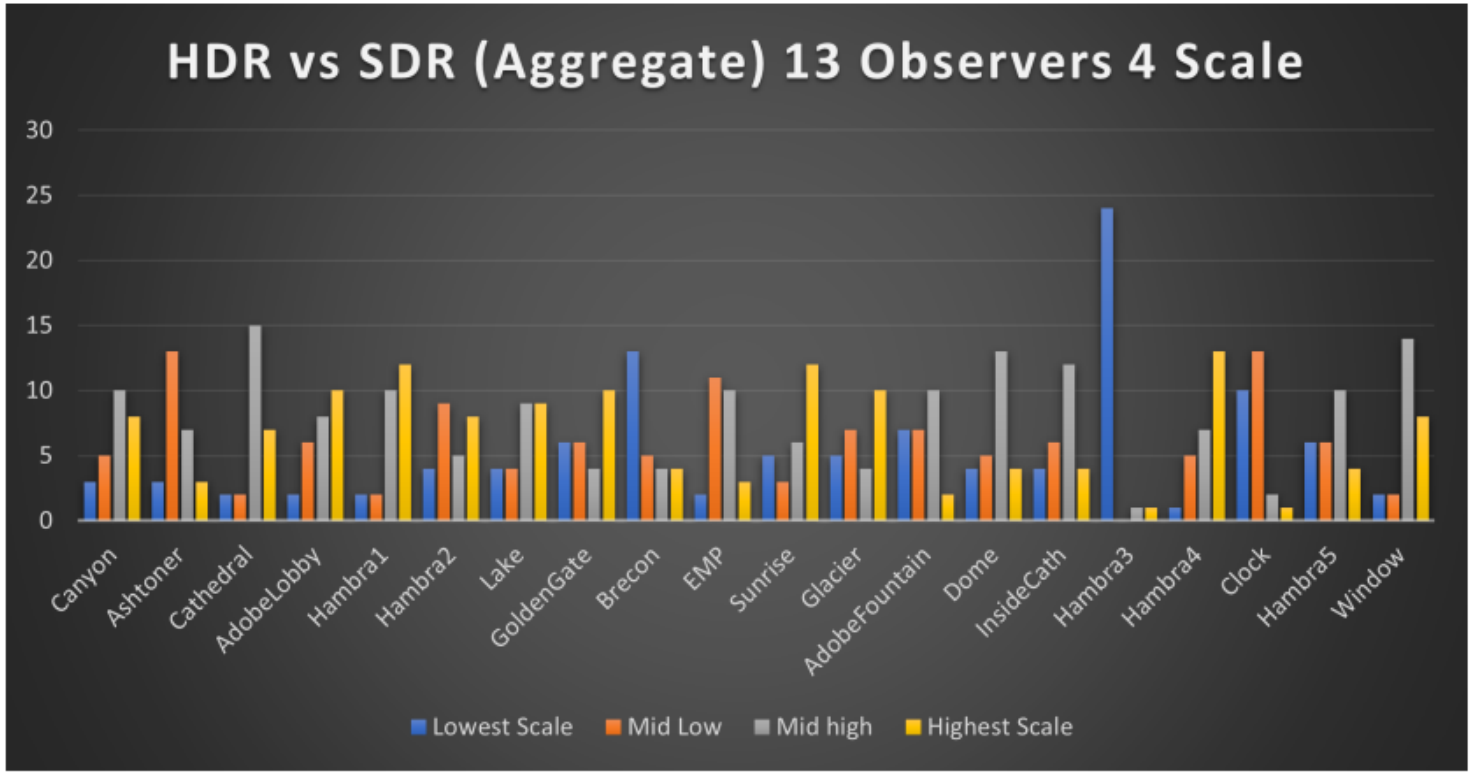
Results and Discussion



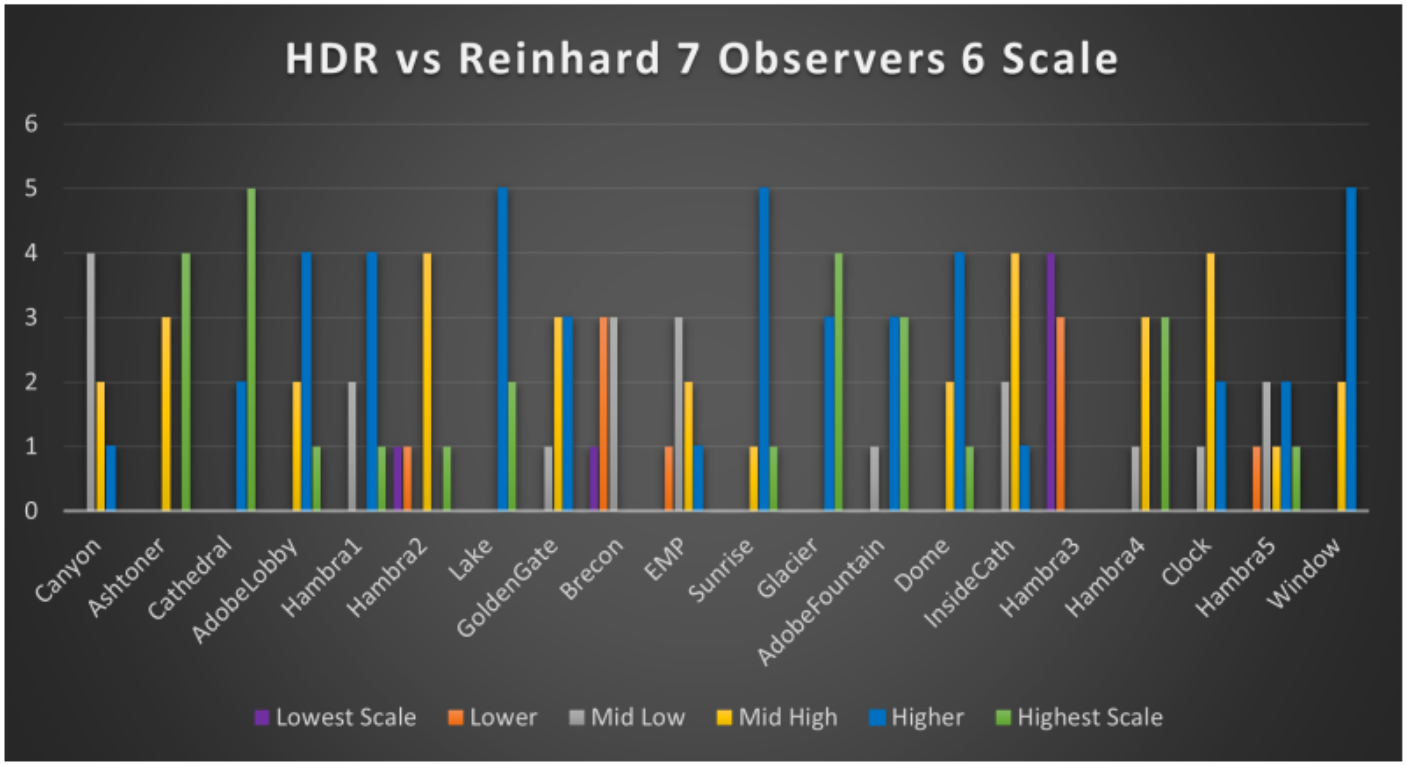
Results and Discussion



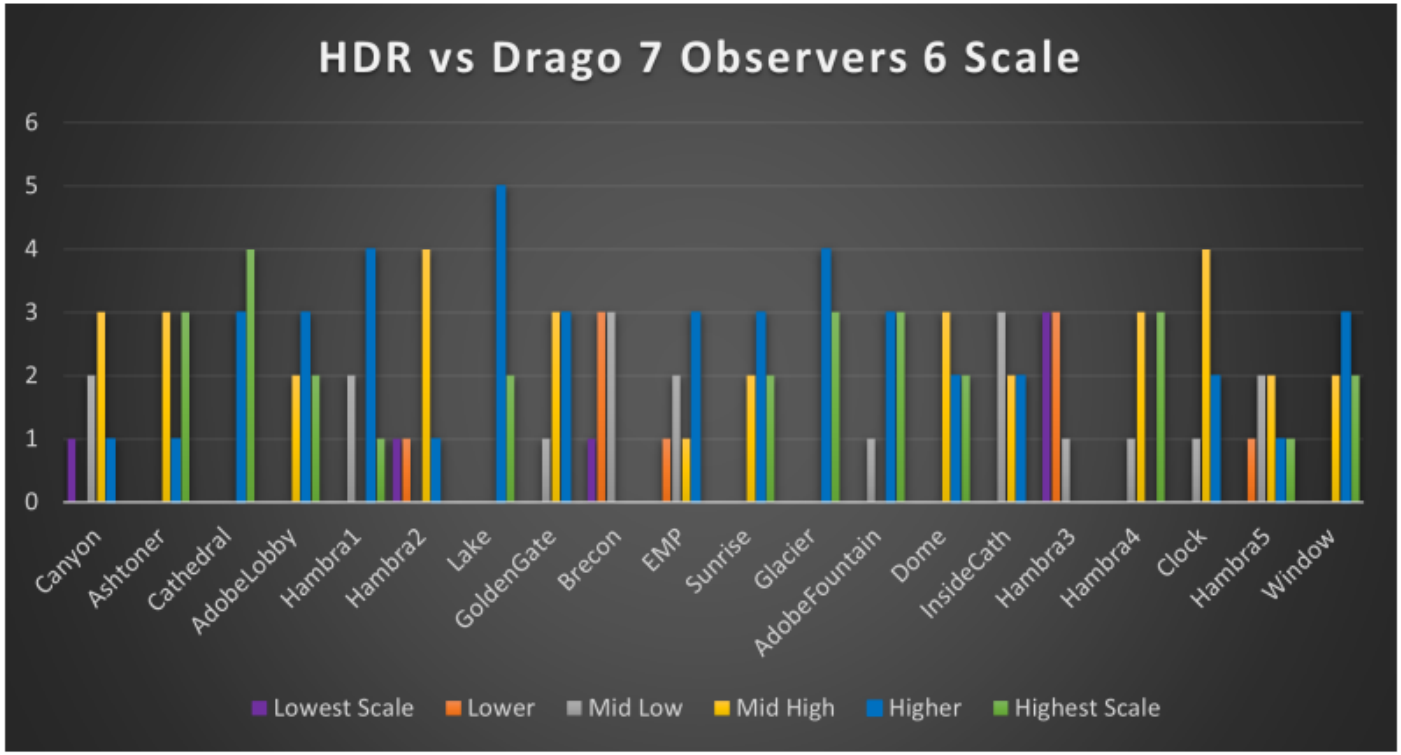
Results and Discussion



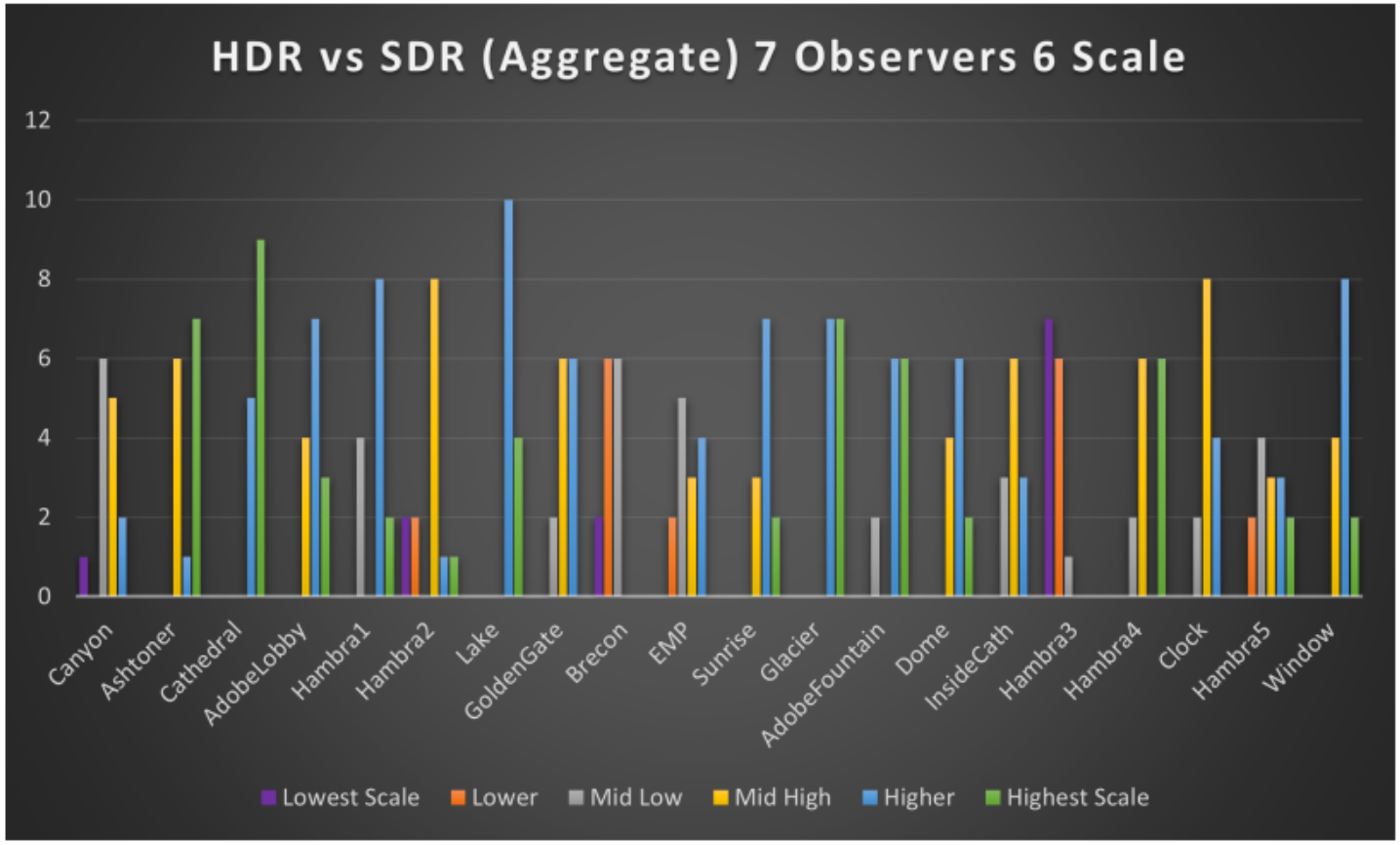
Results and Discussion



Results and Discussion



Results and Discussion



Results and Discussion

Scale	Lowest	Mid Low	Mid high	Highest
Reinhard Total	53	53	82	72
Reinhard %	20.38	20.38	31.54	27.69
Drago Total	56	64	79	61
Drago %	21.54	24.62	30.38	23.46
Aggregate Total	109	117	161	133
Aggregate %	20.96	22.5	30.96	25.58

Results and Discussion

Scale	Lowest	Lower	Mid Low	Mid high	Higher	Highest
Reinhard Total	6	9	20	33	45	27
Reinhard %	4.29	6.43	14.29	23.57	32.14	19.29
Drago Total	6	9	19	34	44	28
Drago %	4.29	6.43	13.57	24.29	31.43	20.00
Aggregate Total	12	18	37	66	88	53
Aggregate %	4.38	6.57	13.50	24.09	32.12	19.34

Conclusion

- A psychophysical experiment to evaluate the proficiency of HDR on an HDR display over tone mapped SDR and to evaluate the performance and differences between the 2 most successful tone mapping techniques (Reinhard and Drago operators) compared to a full HDR pipeline reference was conducted on a group of 20 observers.

Conclusion

- The images were varied to study different image properties' effect on the selection process and indeed a correlation between image colour saturation and luminance and HDR preference was found. For images with less luminance and colour saturation, the difference between HDR and tone mapped SDR decreases significantly.

Conclusion

- The results show that HDR outperforms SDR but only when specific circumstances are taken care of. These circumstances include noise levels that have to be much less in HDR since they are more prominent in it.
- Also, the variation of the area of maximum (or very high) luminance pixels in HDR with time is very important and has to be mitigated to prevent sudden changes that will affect the viewing experience

Conclusion

- For the tone mapping techniques, the Reinhard operator performed better than Drago operator with most images as expected by previous literature that did a similar comparison.
- Finally, results show that there is a small minority of people who don’t prefer HDR over SDR even at the HDR’s best performance, but more work has to be done to get a better estimate on the ratio of these opinions to the more general opinion of preferring HDR.