The effect of viewing distance and display peak luminance — HDR AV1 video streaming quality dataset

Dounia Hammou
University of Cambridge
United Kingdom
dh706@cl.cam.ac.uk

Lukáš Krasula
Netflix Inc.
Los Gatos, CA, USA
lkrasula@netflix.com

Christos G. Bampis
Netflix Inc.
Los Gatos, CA, USA
christosb@netflix.com

Zhi Li
Netflix Inc.
Los Gatos, CA, USA
zli@netflix.com

Rafał K. Mantiuk
University of Cambridge
United Kingdom
rafal.mantiuk@cl.cam.ac.uk

***Abstract*—While it is well recognized that the visibility of distortions is affected by the viewing distance and display peak luminance, very few datasets control those conditions, and also few video quality metrics can account for them. To address this gap, we collected a new video quality dataset, HDR-VDC, which captures the quality degradation of HDR content due to AV1 coding artifacts and the resolution reduction. The quality drop was measured at two viewing distances, corresponding to 60 and 120 pixels per visual degree, and two display mean luminance levels, 51 and 5.6 nits. In contrast to the existing datasets that use direct rating protocol, we employ a highly sensitive pairwise comparison protocol with active sampling and comparisons across viewing distances to ensure possibly accurate quality measurements. We also provide the first publicly available dataset that measures the effect of display peak luminance and includes HDR videos encoded with AV1. Our results indicate that the effect of both viewing distance and display luminance is significant, and it reduces the visibility of coding and upsampling artifacts on dimmer displays or those seen from a further distance. The dataset is available at https://doi.org/10.17863/CAM.107964 and the code at https://github.com/gfxdisp/HDR-VDC.**

***Index Terms*—subjective video quality assessment, video streaming, AV1 encoding, high dynamic range, display peak luminance, viewing distance, effective resolution, pairwise comparison**

I. Introduction

It is recognized that viewing distance and display peak luminance affect the visibility of video streaming distortions [1]–[3]. High-frequency distortions become less visible as the viewing distance increases and the display luminance is reduced — both effects predicted by the contrast sensitivity function [4]. Yet, very few datasets control those conditions, and also few video quality metrics can account for them.

In this paper, we introduce the "Viewing and Display Conditions Dataset" (HDR-VDC), a new HDR video quality assessment dataset designed to capture the effect of viewing distance and display luminance on the visibility of video streaming distortions (we consider the visibility of distortions as the only indicator of quality in this study). Here, we consider distortions due to video compression and lower resolution of the streamed content — two dominant distortions in video streaming. The dataset is collected for modern 4K HDR content, both camera-captured and computer-generated (video games and animation), encoded with AV1 codec [5].

In contrast to existing datasets exploring the effect of viewing distance [1], [2], our goal is to measure the quality as accurately as possible so that even if both effects are subtle, they can be reliably captured. For that reason, we employ a pairwise comparison protocol with active sampling (ASAP [6]), which provides much better sensitivity than direct rating methods [7]. Moreover, we include cross-condition comparisons, in which content shown at two different distances is directly compared. Those measures should provide accurate measurements. The dataset is also the first to measure the effect of display luminance on the visibility of streaming distortions and is one of the few publicly available datasets that measure the quality of AV1-encoded 4K HDR10 videos.

This paper is meant to introduce the dataset. A follow-up publication will provide a complete analysis of the video metric performance on the dataset and considerations for modeling both effects.

II. Related Work

A. HDR Video quality assessment datasets

In recent years, the adoption of HDR video formats has led to numerous studies aimed at assessing the quality of HDR videos to benchmark the various HDR video quality metrics.

Azimi et al. [8], [9], Rerabek et al. [10], and Narwaria et al. [11] were among the first to do so. They studied the effect of multiple distortions, such as noise, blur, and compression, on the perception of distortions in 2048×1080 / 1920×1080 HDR video content. Furthermore, Pan et al. [12] and Athar et al. [13] were the first to study the effect of compression artifacts on 4K HDR (PQ or HLG encoded) high framerate videos. However, none of these datasets were made publicly available. The LIVE-HDR [14] presented the first large-scale quality dataset of 4K HDR10 videos, in which HEVC compression and upscaling artifacts were considered — the two main distortions in video streaming. Following that, Shang et al. [15] studied the effect of different HDR live streaming scenarios on sports content. Krasula et al. [16], on the other hand, presented the first at home-study for collecting a 4K HDR10 subjective quality dataset. Nonetheless, the two latter datasets were also not made publicly available.

In contrast to these works, our dataset is the first publicly available 4K HDR10 dataset that studies the AV1 compression artifacts, as well as four display and viewing conditions: two viewing distances and two display luminance levels.

B. The effect of viewing distance and display luminance on perceptual quality

Many studies investigated the effect of viewing distance on the perceptual quality of both images and videos, notably, the visibility of distortions due to compression, noise, and others.

Authors in [17]–[19] studied the effect of viewing distance on the visibility of image distortions. Similarly, Amirpour et al. [2] and Sugito et al. [1] studied the effect of viewing distance on video distortions. All studies reported similar findings: the distortions are less visible (affect the quality less) at larger viewing distances. Conversely, Kufa et al. [20] and Keller et al. [21] studied the effect of viewing distance on the perceptual preference of high-resolution over lower-resolution videos. The results from both studies report a decrease in the preference for high-resolution videos (decrease in the visibility of upscaling distortions) with the increase in viewing distance.

All the studies above employed direct rating protocols. In contrast, Mikhailiuk et al. [3] measured the visually lossless thresholds for JPEG and WebP distortions across two viewing distances and luminance levels using the adaptive QUEST procedure and the 4-alternative-force-choice protocol. Such a protocol gives accurate measurements of the compression settings at which the distortions become imperceptible. Nonetheless, their study was limited to images, and the measured visually lossless thresholds cannot be easily transferred to the full range of quality degradations.

Compared to other works, our study focuses on video streaming applications in the context of modern and diverse 4K HDR content, as seen at varying viewing conditions. We employ the pairwise comparison protocol and perform comparisons across viewing distances for accurate measurements of video quality.

III. Subjective Experiment

The main goal of this subjective study is to measure the effect of display and viewing conditions on the distortions' visibility for high-resolution and high-dynamic range videos. Thus, we collect a new 4K HDR video quality dataset under four display and viewing conditions — the combination of two viewing distances and two display luminance levels.

A. Source Sequences

It is crucial to select a balanced and diverse set of HDR source sequences to capture the content-dependent effects in addition to the effects of viewing conditions. Hence, we collected 50 HDR10, 4K (3840 × 2160) / 1080p (1920 × 1080) video sequences from various sources [22]–[25], which included natural camera-captured, animated and computer-generated (gaming) content. Although we focused on modern 4K content, we also selected a few full-HD sequences from the color-graded Stuttgart dataset [25] as those contained a much larger dynamic range than other available content.

All sequences are characterized by a bit depth of 10 bits, a YUV 4:2:0 chroma sampling, SMPTE ST2084 (PQ) EOTF, and BT.2020 primaries. Moreover, all sequences were clipped into 5 to 11-second videos while avoiding abrupt or awkward end cuts that may disturb the viewing experience. Furthermore, all videos were stored in a raw format to calculate the video information indices defined below.

Under the recommendation of ITU-T P.910 [26], for each source sequence, we calculated the spatial and temporal information indices of the videos. However, the standard indices alone do not capture HDR characteristics. Thus, we included two other indices: the dynamic range (computed as the median of the differences between the logarithm of the brightest luminance value and the darkest luminance value from the low-pass filtered luma channel of each frame) and median luminance. These four indices were used to select 16 reference videos from the 50 collected source sequences by employing the k-means algorithm [27], where the input of the algorithm is the video sequences' indices, and the number of clusters is 16, as the number of reference videos. Then, one video was selected from each cluster in order to provide a diverse set of content. The thumbnails for all selected reference videos are shown in Figure 1, and index plots for those are shown in Figure 2.

B. Encoding Space

A careful design of the encoding space is imperative for our specific study objectives. Because our focus is not solely on the effect of distortions on perceived quality but rather on the effect of display and viewing conditions on the visibility of distortions arising from video streaming, we design a concise distortion space that spans a broad range of perceptual quality.

The encoding space consists of two distinct types of distortions: compression artifacts and upscaling artifacts. The SVT-AV1 codec (v1.5.0) [5], [28], [29] was selected for introducing compression artifacts, while the Lanczos filter ($a = 3$) [30] was employed to upscale videos to the display full resolution (4K). SVT-AV1 was selected as a representative modern video codec offering good performance.

Each video sequence was first downscaled to one of the three resolutions, namely 4K (3840 × 2160 / unchanged), 1080p (1980×1080), and 720p (1280×720), and then encoded with AV1 (using a preset of four) to produce videos at three bit rates. The compression levels were determined by employing 3 constant rate factors (CRF) values separately for each source sequence, resulting in three quality levels: "High", "Medium", and "Low".

To ensure a perceptual separation between the three bit rates, we compressed each source video at its native resolution using CRF values ranging from 1 to 63. The ColorVideoVDP metric (v0.3) [31] was employed to guide the selection of the final CRF values to ensure a similar distance in quality between the low and medium quality and the medium and high quality.

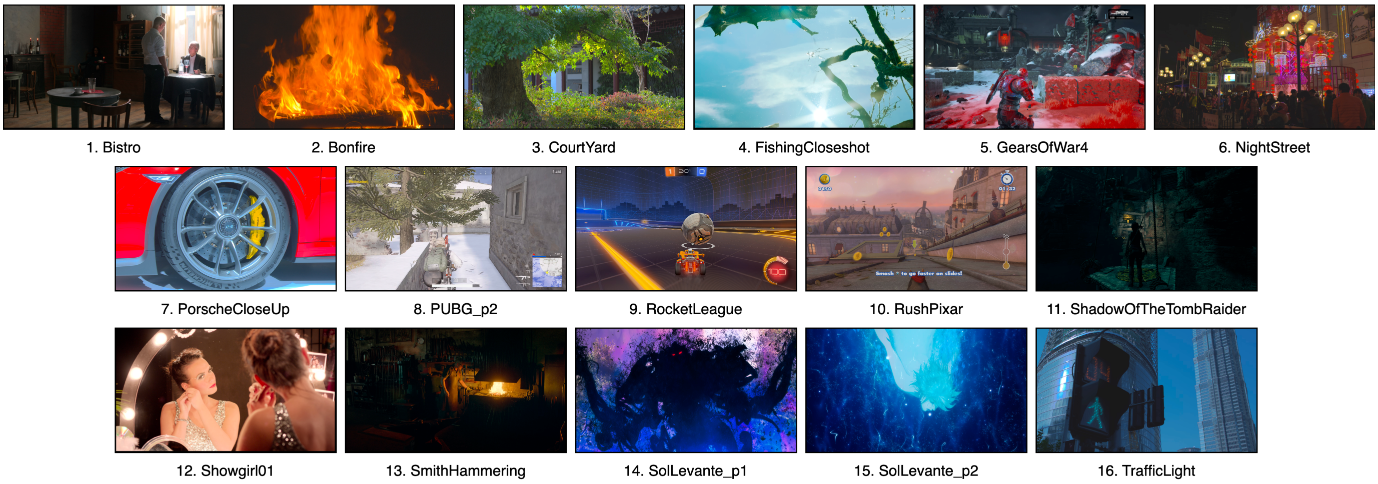

Fig. 1: Screenshots of the selected reference videos (The sequences are tone mapped to BT.709 for visualization).

TABLE I: CRF settings of the test videos. The same CRF value was used across all resolutions. The reference videos used high-quality CRF at the highest resolution.

Video sequence index	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
High quality CRF	7	9	9	5	9	9	9	9	7	7	11	7	9	7	7	7
Medium quality CRF	43	37	43	33	35	41	49	37	43	33	45	27	37	37	36	55
Low quality CRF	61	58	62	62	53	59	62	60	60	51	59	55	58	56	52	62

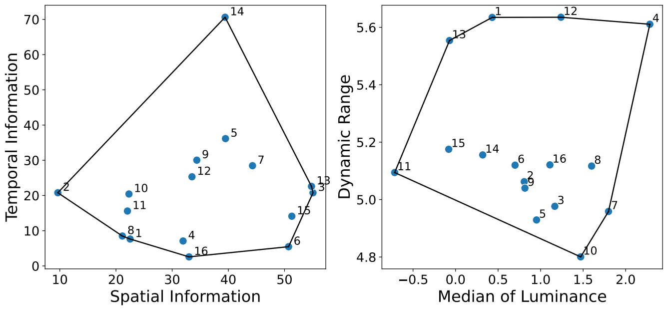

Fig. 2: Scatter plots and corresponding convex hulls of Spatial Information / Temporal Information and Dynamic Range / Luminance Median of the selected reference videos. The indexes reported next to each scatter point correspond to the indexes in Figure 1.

Moreover, we ensured that the high-quality sequence attained a quality exceeding 9.5 (on a 10-point scale) when compared to the uncompressed video while concurrently ensuring a real-time playback of the videos using our hardware. The CRF values were selected for the native resolution and used across all other resolutions. The selected CRF values for each source video are reported in Table I.

Our experiment design required real-time playback of three 4K HDR videos at the same time and without dropping frames. To ensure that, we had to use 4K video compressed at high-quality settings (reported in Table I) as the reference. We ensured that the compression distortions were imperceptible in the reference videos.

C. Experimental procedure

Following the BT.2100-2 recommendations [32] for HDR viewing, two viewing distances were considered in the experiment: 1.6 and 3.2 of display height (an effective resolution of approximately 60 ppd and 120 ppd), as well as two display luminance levels: a bright display (peak luminance of 700 cd/m^2 and a mean luminance of 51 cd/m^2 as measured from a color patch covering 5% of the display using the Konica Minolta Chroma Meter CS-200) and a dim display (peak luminance of 600 cd/m^2 and a mean luminance of 5.6 cd/m^2). The dimmer screen was simulated using an OpenGL fragment shader that reduced the luminance of all pixels by a factor of 8 (in a linear RGB color space).

To ensure the high accuracy of our measurements across the display and viewing conditions, we employed a pairwise comparison protocol [33], combined with an active sampling method (ASAP [6]), which can maximize the information gain from each comparison and reduce the number of required comparisons. The pairwise comparison protocol was shown to provide higher sensitivity than direct rating (e.g., ACR) [7]. This is especially important when measuring the effect of viewing conditions, which could be relatively small compared to the measurement noise associated with direct rating.

In each trial of the experiment, the participants were presented with a pair of test videos of the same content, played simultaneously on two displays (LG OLED Evo G2, see the details below). The participants could view the corresponding reference videos by holding the space key. A short blank screen was shown when switching between test and reference videos so that the temporal change could not be used to

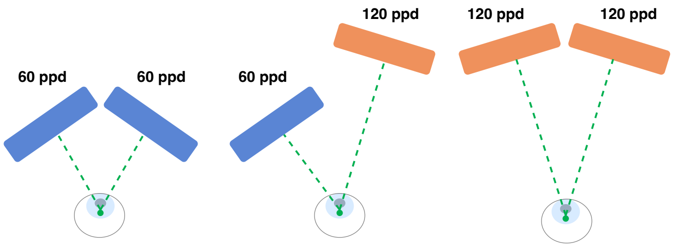

Fig. 3: The viewing configurations corresponding to the combination of the two viewing distances, shown on either left or right display, where each display was rotated to ensure a perpendicular viewing angle to the screen center. The cross-viewing distance configurations had the far display either on the right (middle column) or left (not shown). The order of the viewing configurations was randomly selected for each participant.

Fig. 4: The experimental setup for the near-near viewing setting session and bright display luminance block.

identify the distortions. The participants were instructed to select the test video that is closer to the reference rather than explicitly selecting the video that looks better. This criterion allowed us to measure the visibility of distortions rather than the overall perceived quality.

We split the experiment into four sessions, each lasting one hour. In each session, we present the test videos in a different viewing setting to guarantee that participants compare the test conditions across all four combinations of compared viewing distances and display positions (see Figure 3). A chinrest was used to control viewing distance. Furthermore, the screens were rotated so that the viewing angle was always perpendicular to the screen center (as shown in Figure 3).

Each session was composed of two blocks, each with a different display luminance: either bright or dim. To adapt the eye to the new luminance level, a random noise texture with the mean luminance corresponding to the next luminance level was displayed for 20 seconds during the transition between the two blocks of the session. Note that we did not allow for comparisons or pairs shown at different display luminance levels. We experimented with such cross-peak-luminance comparisons in a pilot experiment, but we found the observers' selection was strongly biased towards the brighter display. To minimize the learning effect, the order of the sessions was randomized between the participants, as well as the order of the display luminance blocks in each session. An overview of the setup of the experiment is shown in Figure 4.

D. Display and Viewing Conditions

Two LG OLED Evo G2 55-inch 4K HDR10 displays were used to conduct the experiment, each with firmware version 3.0.10. The HDR mode selected for both displays was "Cinema Home (User Settings)". To prevent any unintended alterations to the video content, all auto adjustments and enhancements were disabled. Furthermore, global sticky reduction (GSR) and temporal peak luminance control (TPC) were turned off in the service menu to maintain consistent luminance throughout the experiment and across participants.

Both displays were connected to the workstation equipped with an NVIDIA GeForce RTX 3080 GPU using a 4K HDMI 2.1 cable capable of transferring HDR content at their respective framerates for real-time video playback. The workstation operated on Windows 10, with HDR mode enabled for both display connections. The GStreamer 1.22.6 player was used for the hardware decoding and playback of the videos within the Psychtoolbox 3.0.19 framework [34]. Lanczos ($a = 3$) filter was implemented as an OpenGL fragment shader to upsample videos to 4K resolution.

The experiment was conducted in a controlled environment within a dim room, maintaining an ambient illumination of approximately 0 lux throughout the experiment.

E. Participants

A total of 30 volunteers (16 females and 14 males) participated in the experiment, aged between 13 and 52 years (median age of 24 years and mean age of 26 years), where 8 of them were familiar with HDR. 25 of the participants completed all four sessions, resulting in a total number of 25762 pairwise comparisons. All participants had normal color vision, as indicated by the Ishihara test, and all of them had normal or corrected to normal visual acuity of at least 20/25, as measured by the Snellen test. Preceding the experiment, participants were provided with a briefing form outlining the experiment objectives and instructions, alongside a consent form, which they signed on their first session. Furthermore, a brief training session was provided during their first session to familiarize participants with the experiment setup and to ensure their full understanding of the provided instructions. During each session, the participants were given a break lasting 5 to 10 minutes after running the experiment for 25 minutes. Additionally, participants could initiate a break at any point by pressing the 'F' key. This setup allowed the participants to run each session with no noticeable fatigue. The experiment was approved by the departmental ethics board, and all participants were rewarded for their participation.

IV. Experimental Results

A. Subjective score scaling

The results obtained from the pairwise comparison were scaled into Just-Objectionable-Difference (JOD) units using

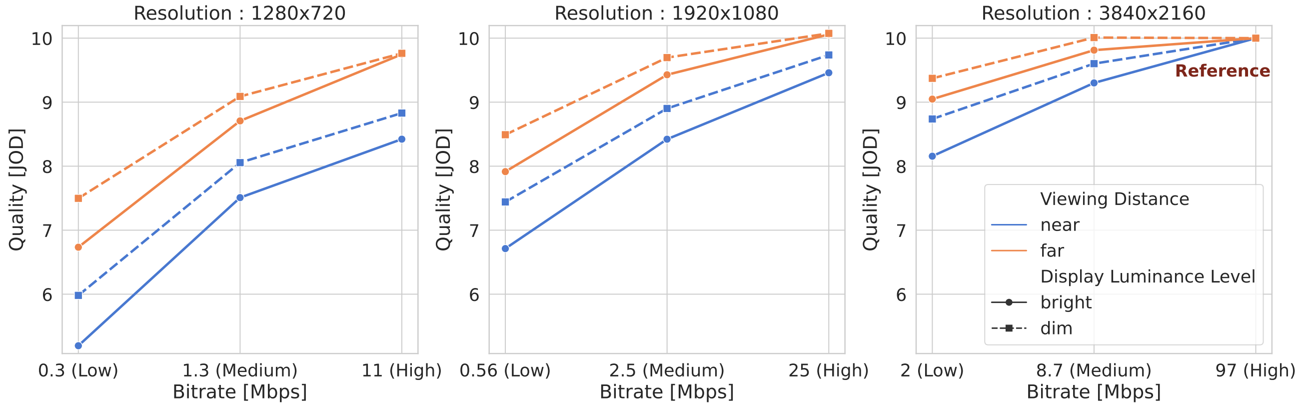

Fig. 5: The effect of viewing distance and display peak luminance on the perception of quality for different resolutions and bitrates (compression levels). The JOD scores are averaged over all contents. Refer to Table I for the CRF values selected for each compression level.

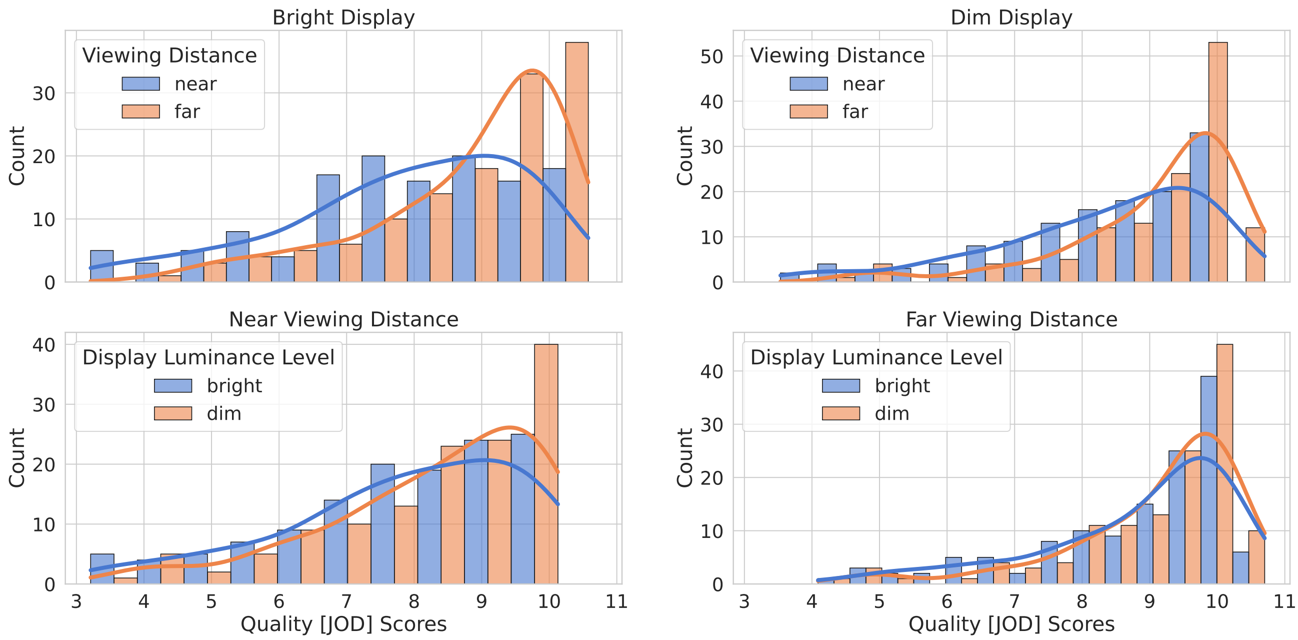

Fig. 6: Histograms of the JOD scores and their distribution (estimated using the kernel density estimate (KDE) for visualization purposes) across viewing distances and display luminance levels. The histograms on the top represent the effect of viewing distance for each display peak luminance. The histograms on the bottom represent the effect of display peak luminance for each viewing distance.

Thurstone's Case V observer model [33]. The difference of 1 JOD unit signifies that one condition is chosen as closer to the reference 75% of the time. All reference videos (highest resolution and bitrate) were assumed to be the same node during scaling to ensure a common quality scale across contents and viewing conditions. Given that the JOD scores are relative, reference conditions were assumed to have a nominal quality of 10. The scaling was performed within the sets of conditions that were compared in the experiment — within the same content and within the same display luminance level. Furthermore, to quantify the precision of our quality scores, confidence intervals were computed using bootstrapping with 500 samples, as outlined in [33].

B. The effect of viewing distance and display luminance

The results across all participants and all contents are reported in Figure 5. The figure shows the JOD scores averaged over all video contents but plotted separately for each viewing condition. The figure reveals a consistent pattern, wherein observers are less likely to notice the drop of quality at lower bitrates or lower resolution when the video is shown from a further distance or at a lower luminance level. Notably, the distortions from AV1 compression at a medium level on 4K videos and the downscaling of videos to 1080p are

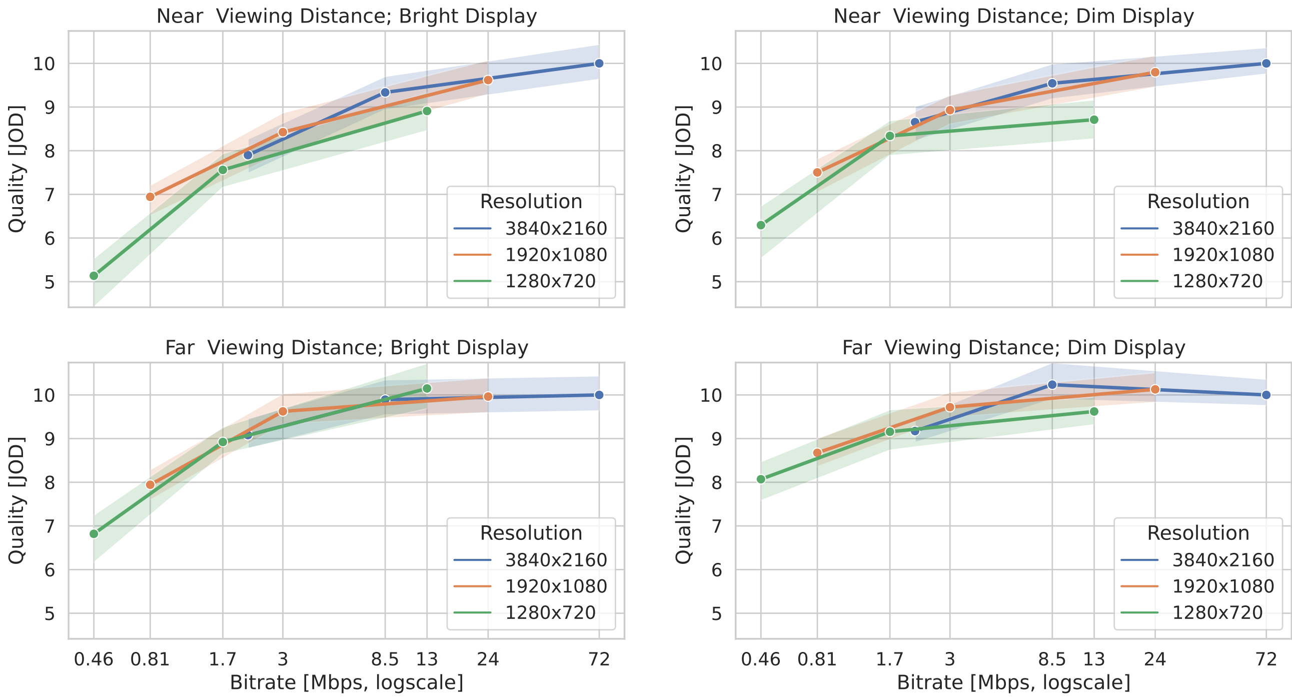

Fig. 7: Quality-bitrate curves for the RushPixar content (index 10 in Figure 1) for each viewing distance and display luminance level. The shaded areas represent 95% confidence intervals. A few conditions exceed the score of 10 JODs due to the measurement noise.

imperceptible (≈10 JOD) when observed on a dim display seen from a further distance. The difference between the viewing conditions is substantial, especially at low resolution and low bit-rate — at 720p resolution and low bit rate, there is over 2 JOD difference in quality between a bright display seen from near and a dim display seen from far. This clearly shows that video quality cannot be accurately assessed when the display and viewing conditions are not controlled.

This trend is also evident when analyzing the distribution of the JOD scores of all contents and test conditions for each viewing distance and display luminance level, as shown in Figure 6. The top row of the figure illustrates the effect of the viewing distance on the distribution of the JOD scores for each fixed display peak luminance. We can observe again that the quality scores are higher for a larger viewing distance (top row) or lower luminance (bottom row), confirming that the distortions are less noticeable when seen from a greater distance and at lower luminance.

To investigate this effect on a per-content basis, we plot the quality-bitrate curve, often referred to as the bitrate ladder, for each viewing distance and display luminance level in Figure 7. The results for this content confirm the overall trend, showing the significant effect of the display and viewing conditions. For example, if we want to deliver a quality of 9 JODs, we can stream at 1.7 Mbps for the content seen on a dim display from a far distance, but we need to increase the bitrate to 8.5 Mbps for a bright display seen from a near distance.

To analyze the statistical significance of the viewing distance and the display peak luminance independently of the content and distortions, we employ a 5-way analysis of variance (ANOVA) model on the JOD distribution of each condition, in which the five factors represent the content, bitrate, video resolution, viewing distance, and display luminance level. Because the quality scores distributions are derived by bootstrapping (with 500 samples), we selected 30 random samples from each distribution, corresponding to the number of observers. The results of the test revealed that both the viewing distance and display luminance level had a significant effect (p-value < 0.05) on the JOD scores. The effect size, eta squared (η^2), was 0.101 for the viewing distance factor and 0.015 for the display luminance level factor, indicating that the viewing distance has a medium effect on the perception of distortions, while the display luminance has a small effect.

V. Conclusion

In this paper, we introduced a new HDR-VDC dataset capturing how the quality of streamed 4K HDR10 AV1 video varies across two viewing distances and two display luminance levels. The most important observation is that both viewing distance and display peak luminance significantly affect the visibility of streaming distortions. This amounts to very substantial differences in bitrate required to deliver the desirable quality, which depends on the viewing conditions. We hope that this dataset will help to test and develop quality metrics that account for the viewing conditions and bring attention to the important problem of controlling display peak-luminance and viewing distance in video quality studies.

References

[1] Y. Sugito, Y. Kondo, D. Arai, and Y. Kusakabe, "Modeling Perceived Quality on 8K VVC Video Under Various Screen Sizes and Viewing Distances," *IEEE Access*, vol. 10, pp. 97 237–97 247, 2022.

[2] H. Amirpour, R. Schatz, C. Timmerer, and M. Ghanbari, "On the impact of viewing distance on perceived video quality," in *2021 International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 2021, pp. 1–5.

[3] A. Mikhailiuk, N. Ye, and R. K. Mantiuk, "The effect of display brightness and viewing distance: A dataset for visually lossless image compression," *Electronic Imaging*, vol. 33, pp. 152–1–152–8, 2021.

[4] P. G. J. Barten, *Contrast sensitivity of the human eye and its effects on image quality*. SPIE Press, 1999.

[5] F. Kossentini, H. Guermazi, N. Mahdi, C. Nouira, A. Naghdinezhad, H. Tmar, O. Khlif, P. Worth, and F. B. Amara, "The SVT-AV1 encoder: overview, features and speed-quality tradeoffs," *Applications of Digital Image Processing XLIII*, vol. 11510, pp. 469–490, 2020.

[6] A. Mikhailiuk, C. Wilmot, M. Perez-Ortiz, D. Yue, and R. K. Mantiuk, "Active sampling for pairwise comparisons via approximate message passing and information gain maximization," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 2559–2566.

[7] M. Perez-Ortiz, A. Mikhailiuk, E. Zerman, V. Hulusic, G. Valenzise, and R. K. Mantiuk, "From Pairwise Comparisons and Rating to a Unified Quality Scale," *IEEE Transactions on Image Processing*, vol. 29, pp. 1139–1151, 2020.

[8] M. Azimi, A. Banitalebi-Dehkordi, Y. Dong, M. T. Pourazad, and P. Nasiopoulos, "Evaluating the performance of existing full-reference quality metrics on high dynamic range (HDR) video content," *arXiv preprint arXiv:1803.04815*, 2018.

[9] A. Banitalebi-Dehkordi, M. Azimi, Y. Dong, M. Pourazad, and P. Nasiopoulos, "Quality assessment of High Dynamic Range (HDR) video content using existing full-reference metrics," *ISO/IEC JTC1/SC29/WG11, France*, 2014.

[10] M. Rerabek, P. Hanhart, P. Korshunov, and T. Ebrahimi, "Subjective and objective evaluation of HDR video compression," in *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2015.

[11] M. Narwaria, M. P. Da Silva, and P. Le Callet, "HDR-VQM: An objective quality measure for high dynamic range video," *Signal Processing: Image Communication*, vol. 35, pp. 46–60, 2015.

[12] X. Pan, J. Zhang, S. Wang, S. Wang, Y. Zhou, W. Ding, and Y. Yang, "HDR video quality assessment: Perceptual evaluation of compressed HDR video," *Journal of Visual Communication and Image Representation*, vol. 57, pp. 76–83, 2018.

[13] S. Athar, T. Costa, K. Zeng, and Z. Wang, "Perceptual quality assessment of UHD-HDR-WCG videos," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1740–1744.

[14] Z. Shang, J. P. Ebenezer, A. C. Bovik, Y. Wu, H. Wei, and S. Sethuraman, "Subjective assessment of high dynamic range videos under different ambient conditions," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 786–790.

[15] Z. Shang, Y. Chen, Y. Wu, H. Wei, and S. Sethuraman, "Subjective and objective video quality assessment of high dynamic range sports content," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 556–564.

[16] L. Krasula, A. Choudhury, S. Daly, Z. Li, R. Atkins, L. Malfait, and A. Mavlankar, "Subjective video quality for 4K HDR-WCG content using a browser-based approach for" at-home" testing," *Electronic Imaging*, vol. 35, pp. 263–1, 2023.

[17] X. Liu, M. Pedersen, and J. Y. Hardeberg, "CID: IQ–a new image quality database," in *Image and Signal Processing: 6th International Conference, ICISP 2014, Cherbourg, France, June 30–July 2, 2014. Proceedings 6*. Springer, 2014, pp. 193–202.

[18] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Transactions on Broadcasting*, vol. 61, no. 3, pp. 520–531, 2015.

[19] R. Fang, D. Wu, and L. Shen, "Evaluation of image quality of experience in consideration of viewing distance," in *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*. IEEE, 2015, pp. 653–657.

[20] J. Kufa and T. Kratochvil, "Visual quality assessment considering ultra HD, Full HD resolution and viewing distance," in *2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA)*. IEEE, 2019, pp. 1–4.

[21] D. Keller, F. von Hagen, J. Prenzel, K. Strama, R. Rao, R. Rao, and A. Raake, "Influence of Viewing Distances on 8K HDR Video Quality Perception," in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2023, pp. 209–212.

[22] "Netflix open content." [Online]. Available: https://opencontent.netflix.com

[23] N. Barman and M. G. Martini, "User Generated HDR Gaming Video Streaming: Dataset, Codec Comparison, and Challenges," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1236–1249, 2021.

[24] L. Song, Y. Liu, X. Yang, G. Zhai, R. Xie, and W. Zhang, "The SJTU HDR video sequence dataset," in *Proceedings of International Conference on Quality of Multimedia Experience (QoMEX 2016)*, 2016, p. 100.

[25] 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 *Digital photography X*, vol. 9023. SPIE, 2014, pp. 279–288.

[26] "Subjective video quality assessment methods for multimedia applications," in *Recommendation ITU-T P.910*, 2022.

[27] J. A. Hartigan, M. A. Wong *et al.*, "A k-means clustering algorithm," *Applied statistics*, vol. 28, no. 1, pp. 100–108, 1979.

[28] A. Norkin, J. Sole, M. Afonso, K. Swanson, A. Opalach, A. Moorthy, and A. Aaron, "SVT-AV1 open-source AV1 encoder and decoder," *Netflix Technology Blog*, 2020.

[29] P.-H. Wu, I. Katsavounidis, Z. Lei, D. Ronca, H. Tmar, O. Abdelkafi, C. Cheung, F. B. Amara, and F. Kossentini, "Towards much better SVT-AV1 quality-cycles tradeoffs for VOD applications," in *Applications of Digital Image Processing XLIV*, vol. 11842. SPIE, 2021, pp. 236–256.

[30] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of Applied Meteorology and Climatology*, vol. 18, no. 8, pp. 1016–1022, 1979.

[31] R. K. Mantiuk, P. Hanji, M. Ashraf, Y. Asano, and A. Chapiro, "ColorVideoVDP: A visual difference predictor for image, video and display distortions," *arXiv preprint arXiv:2401.11485*, 2024.

[32] "Image parameter values for high dynamic range television for use in production and international programme exchange," in *Recommendation ITU-R BT.2100-2*, 2018.

[33] M. Perez-Ortiz and R. K. Mantiuk, "A practical guide and software for analysing pairwise comparison experiments," *arXiv preprint arXiv:1712.03686*, 2017.

[34] M. Kleiner, D. Brainard, and D. Pelli, "What's new in Psychtoolbox-3?" 2007.