Objective and Subjective Evaluation of High Dynamic Range Video Compression

R. Mukherjee[1], K.Debattista[1], T. Bashford-Rogers[1], P. Vangorp[2], R. Mantiuk[3], M. Bessa[4], B. Waterfield[5], A. Chalmers[1]

Abstract

A number of High Dynamic Range (HDR) video compression algorithms proposed to date have either been developed in isolation or only-partially compared with each other. Previous evaluations were conducted using quality assessment error metrics, which for the most part were developed for qualitative assessment of Low Dynamic Range (LDR) videos. This paper presents a comprehensive objective and subjective evaluation conducted with six published HDR video compression algorithms. The objective evaluation was undertaken on a large set of 39 HDR video sequences using seven numerical error metrics namely: PSNR, logPSNR, puPSNR, puSSIM, Weber MSE, HDR-VDP and HDR-VQM. The subjective evaluation involved six short-listed sequences and two ranking-based subjective experiments with hidden reference at two different output bitrates with 32 participants each, who were tasked to rank distorted HDR video footage compared to an uncompressed version of the same footage. Results suggest a strong correlation between the objective and subjective evaluation. Also, non-backward compatible compression algorithms appear to perform better at lower output bit rates than backward compatible algorithms across the settings used in this evaluation.

Keywords: `HDR video`, `compression algorithm`, `quality assessment`, `ranking`, `rate-distortion`

1. Introduction

High Dynamic Range (HDR) imaging is able to capture, store, transmit and display the full range of real-world lighting with much higher precision than mainstream Low Dynamic Range (LDR) (also known as Standard Dynamic Range (SDR)) imaging. However, a significantly large amount of data needs to be stored and processed for HDR video. To allow HDR video to be handled practically, a number of pre/post processing (compression) algorithms have been proposed to convert HDR video data to an encoder friendly format. However, to date, these compression algorithms have only partially been compared with each other. This paper undertakes a comprehensive objective and subjective comparison of six previously published or patented HDR video compression algorithms and in doing so, follows a detailed methodology for evaluation and qualitative assessment of compressed HDR video content. In addition, a correlation is computed between the subjective and objective results for a better understanding of the shortcomings of current objective evaluation techniques for HDR video quality.

The primary contributions of this work are: a) An objective evaluation of six HDR video compression algorithms using seven full-reference quality assessment (QA) metrics, b) Two subjective evaluations of the compression algorithms at two different output bitrates, using a ranking method with hidden reference conducted with 32 participants each; and c) An assessment of the correlation between the objective and subjective evaluations.

In addition, the objective QA results are averaged over 39 sequences at 11 different quality settings to generalize the overall rate-distortion (RD) characteristics of the compression algorithms.

---

[1]WMG, University of Warwick
[2]Bangor University
[3]University of Cambridge
[4]UTAD, Portugal
[5]Jaguar Land Rover Automotive PLC, UK

2. Related work

The acquisition of HDR imagery results in a large amount of floating point data which needs to be compressed in order to be handled efficiently on existing image or video infrastructure. This problem has been tackled through a variety of image and video compression algorithms which convert input HDR data to video encoder input formats. A brief review of compression algorithms is available in Banterle et al. [7].

Substantial research has also been conducted on evaluation of QA metrics for LDR image/video. Avcıbaş et al. [1] and Sheikh et al. [41] evaluated a number of QA metrics on distorted still images and concluded that metrics based on spectral magnitude error, perception, absolute norm and edge stability are most suitable for detecting image artefacts. They also conclude that although multiple QA metrics perform well on multiple image datasets none of them performed at par with subjective quality assessment. Seshadrinathan et al. [39] conducted an objective and subjective video quality assessment (VQA) and concluded that dedicated VQA metrics such as Motion Based Video Integrity Evaluation (MOVIE) perform significantly better and have higher correlation with subjective results than still-image QA metrics.

In comparison, substantially less research has been conducted on development and evaluation of dedicated HDR QA metrics. Existing LDR QA metrics such as PSNR, SSIM [46] and VIF [40] have been extended to handle HDR values using Perceptually Uniform (PU) encoding [3]. Recently, a few full-reference HDR QA/VQA metrics have been proposed such as High Dynamic Range-Visible Difference Predictor (HDR-VDP) [28, 34], High Dynamic Range-Visual Quality Metric (HDR-VQM) [36] and Dynamic Range Independent-Visual Quality Metric (DRI-VQM) [2].

A few evaluations have been conducted to test the performance of the proposed metrics. Čadík et al. [10] conducted an evaluation of HDR-VQA metrics with a dataset consisting of six HDR sequences using an HDR display and concluded that although the predictions by DRI-VQM and HDR-VDP are most suited for HDR-HDR image pairs, executing DRI-VQM becomes prohibitively expensive for sequences with greater than VGA resolution. Azimi et al. [4] tested the correlation between seven QA metrics and subjective quality scores with a dataset of 40 HDR video sequences and five types of distortions. The work demonstrates that HDR-VDP-2 [28] outperforms all other QA metrics when measuring compression induced distortions and has the highest correlation with the subjective quality scores. However, VIF [40] using PU encoding produces the best overall (tested against all distortions) results. Similar benchmarking evaluations of QA metrics for HDR image/video content have been conducted by Valenzise et al. [45], Mantel et al. [26] and Hanhart et al. [15] and Minoo et al. [32].

2.1. Evaluation of HDR video compression methods

Despite the research conducted into development and evaluation of QA/VQA metrics for both LDR and HDR content, very little has been done to evaluate existing HDR video compression algorithms using both QA metrics and psychophysical experiments. Koz et al. [20] conducted a comparative survey on HDR video compression which compares the two different approaches to HDR video compression as explained later in Section 3.1. However, this work does not bring together objective and subjective evaluation techniques in order to provide a comprehensive evaluation of individual algorithms across a large set of sequences.

Recently, Hanhart et al. [17] conducted an evaluation of nine HDR video compression algorithms submitted in response to the Motion Pictures Experts Group (MPEG) committee's Call for Evidence (CfE) [25] to evaluate the feasibility of supporting HDR and Wide Color Gamut (WCG) content using the High Efficiency Video Codec (HEVC) [43] encoder. The paper concludes that the proposals submitted to MPEG can noticeably improve the standard HDR video coding technology and QA metrics such as PSNR-DE1000, HDR-VDP2 and PSNR-Lx can reliably detect visible difference. Azimi et al. [5] conducted a study to evaluate the compression efficiency of two possible HDR video encoding schemes (as defined in MPEG CfE [25]) based on the perceptual quantization of HDR video content [31] and tone mapping-inverse tone mapping with metadata. The paper concludes that for specific bitrates, subjective evaluation results suggest that HDR video generated by the perceptual quantization scheme were rated higher than the videos reconstructed using the inverse tone-mapping scheme. Similar evaluations on HDR video content have been conducted by Dehkrodi et al. [6], Dong et al. [11], Rerabek et al. [38], Hanhart et al. [16], Narwaria et al. [35].

3. Method and Materials

This section introduces the compression algorithms, sequences and overall research method followed for preparing the materials for the objective and subjective evaluation. The individual aspects of the objective and subjective evaluations are presented in Section 4 and Section 5 respectively.
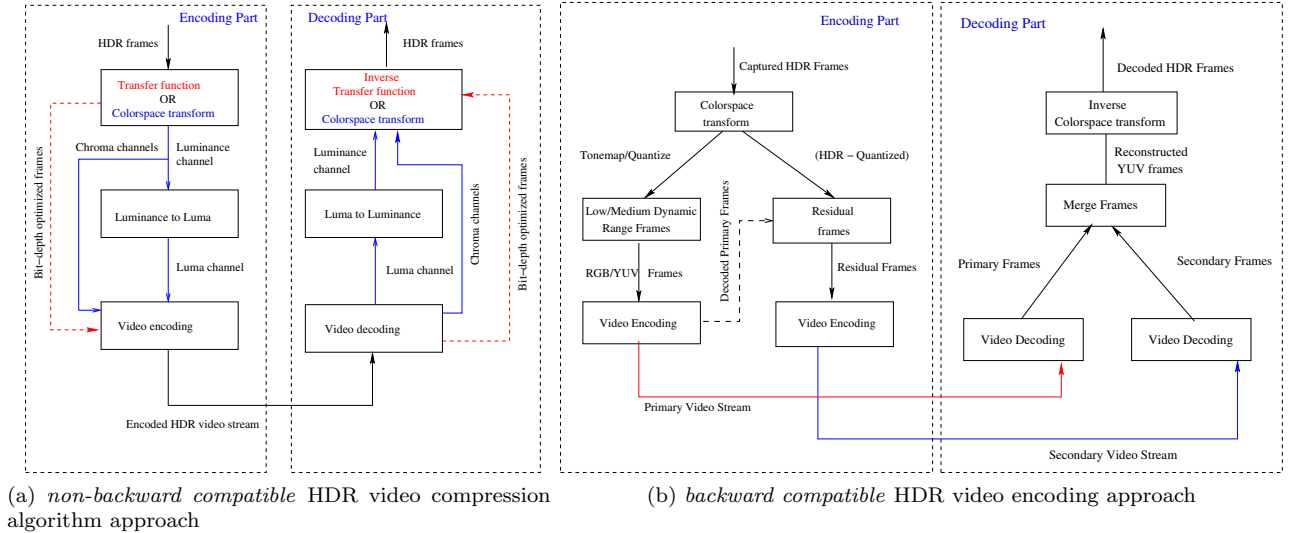
3.1. HDR video compression algorithms

Figure 1: Two generic approaches to HDR video encoding and decoding

HDR video compression algorithms can be classified into two approaches: *non-backward compatible* and *backward compatible*. The *non-backward compatible* approach, shown in Figure 1a, takes advantage of the higher bit-depth support (typically 10-14 bits) in state-of-the-art video encoders [44, 43]. Using a range of transfer functions which work on either linear RGB data or perform colorspace conversions followed by luminance to luma mapping [21, 29, 14] they manipulate captured data in order to pack as much information as possible within the specified bit-depth limit. The resultant optimized video stream is then passed to the encoder.

The *backward compatible* approach, as shown in Figure 1b, splits the input HDR video stream into a base and a residual stream, each with a fixed number of bits, typically 8 or 10 which can be encoded using any existing encoder such as H.263/MPEG-2, H.264/AVC [44], H.265 (HEVC) [43] etc. A few *backward compatible* algorithms [47, 27, 23] contain a tone-mapped (TM) stream which enables it to be played on legacy video players. The residual stream contains additional information to be used for reconstruction of HDR frames during decompression. These algorithms typically use *dual-loop* encoding where the base stream is first encoded and decoded back to create the residual stream. The resultant residual stream compensates for the distortions introduced by the encoder at a chosen quality level, thus minimizing the loss in quality during reconstruction of HDR frames.

Six, published, state-of-the-art algorithms, two following the non-backward compatible and four following the backward compatible approach have been considered in this work. A brief chronological description of each is provided below:

3.1.1. Mantiuk (hdrv)

Mantiuk et al. (2004) [29] proposed a non-backward compatible algorithm extending the existing MPEG-4 encoder to accommodate HDR video content. The algorithm maps linear RGB or XYZ channels, to an 11-bit perceptually uniform luma space L_p and 8-bit chroma channels, u' and v', similar to LogLuv encoding [21]. The method extends discrete cosine transforms (DCT) by encoding sharp contrast edges in spatial, rather than frequency domain. This avoids ringing artefacts due to the quantization of the high value frequency coefficients. The DCT extension in spatial domain was not implemented in this work, since it was deemed redundant due to the advances made in video encoders since 2004.

3.1.2. Ward and Simmons (hdrjpeg)

Ward and Simmons (2004) [47] proposed a *backward compatible* algorithm to encode HDR images. The proposed algorithm extends the standard JPEG image format by tone mapping input HDR frames using the photographic TMO [37] and storing multiplicative HDR information, defined as the ratio between luminance of the HDR and the tone mapped frame: $\mathrm{RI}(x, y) = \frac{Y_{\mathrm{hdr}(x,y)}}{Y_{\mathrm{ldr}(x,y)}}$. The ratio image is down-sampled to reduce storage requirements. The implementation in this work extends the image compression algorithm to HDR video by using a temporally coherent version of the photographic TMO [19] and a dual-loop encoding scheme wherein the TM stream is encoded and decoded back to create a ratio stream. Down-sampling of the ratio frame has not been implemented as that would result in frame quality reduction.

3.1.3. Mantiuk (hdrmpeg)

Mantiuk et al. (2006) [27] proposed a *backward compatible* algorithm. The base stream is a tone-mapped LDR stream using the photographic TMO [37]. Using the dual-loop encoding scheme, the TM stream is encoded and decoded followed by a colorspace conversion of both the HDR and decoded LDR frames in order to compute a monotonically increasing reconstruction function (RF). The RF is required to construct a predicted HDR frame (*pHDR*) which is then subtracted from the original to create a residual frame: $\mathrm{Res}_{\mathrm{frame}} = \mathrm{HDR} - \mathrm{pHDR}$. The residual frame undergoes noise reduction using a visual model which removes invisible noise. The residual stream is then passed through the encoder, thus, completing the second loop. The RF and other additional information are stored as auxiliary meta data.

3.1.4. Lee and Kim (rate)

Lee et al. (2008) [23] proposed another *backward compatible* algorithm where the base 8-bit TM stream is created using a temporally coherent TMO [22]. Using the dual-loop encoding scheme, the TM stream is encoded and decoded to create a ratio stream similar to [47]. Subsequently, a bilateral filter is applied to the ratio frames to reduce noise artefacts. The ratio stream is then passed through the encoder completing the second loop. A Lagrangian optimization technique encodes the secondary stream at a higher quantization parameter (QP) value resulting in a reduced file size compared to other *backward compatible* algorithms where the same QP values are allocated to both base and residual streams. A block-based extension of this algorithm was later proposed by Lee et al. (2012) [24] which uses a perceptual quantization similar to *hdrmpeg*. However, this extension is a not a part of this evaluation.

3.1.5. goHDR (gohdr)

goHDR Ltd. (2009) [8] proposed a *backward compatible* algorithm which uses two 8-bit streams to encode luminance and chroma information separately. The base stream is created by calculating the luminance Y_{HDR} from the HDR frame followed by bilateral filtering, log encoding and sigmoidal tone compression of the luminance stream. Using a dual-loop encoding scheme, the residual stream is created by decoding and inverting the TM luminance frame (Y_{inv}) and dividing the input HDR frames with the inverted luminance such that $\mathrm{Res}_{\mathrm{frame}} = \frac{\mathrm{HDR}}{Y_{\mathrm{inv}}}$. The residual stream is again log encoded and tone mapped using sigmoidal tone compression. The algorithm also produces a meta data which stores the minimum and maximum luminance values at every frame and is used during reconstruction.

3.1.6. Garbas and Thoma (fraunhofer)

Garbas and Thoma (2011) [14] proposed a *non-backward compatible* algorithm by modifying the Adaptive LogLuv algorithm [33] and adding temporal coherence to minimize flickering artefacts in HDR video. The algorithm converts linear RGB values to Lu'v' colorspace by converting real-world luminance to 12-bit luma and adjusting color information into 8-bit u' and v' channels similar to LogLUV encoding[21]. The $L_{12}u'v'$ frame is passed to the encoder and an auxiliary stream stores the minimum and maximum luminance value of every frame which is used during reconstruction.

Unless otherwise stated, the reference H.264/AVC [44] encoder was used in this work for all encoding purposes.

Disclaimer: Although, every effort has been made to faithfully reimplement the algorithms, it cannot be guaranteed that the re-implementations will produce the exact results as the original implementations.

3.2. HDR video footage

The objective evaluation in this work uses a large set of 39 HDR video sequences with an average dynamic range spanning between $14 - 23$ stops ($\approx 42 - 69$ dB). Subsequently, the average dynamic

range for the each of 39 sequences are then measured and after careful consideration six sequences were short-listed for the subjective evaluation. The selection ensured that the sequences represent a variety of HDR video production techniques and contain a large variation in dynamic range. All sequences used in this work have full HD (1920×1080) resolution. Table 1 provides a brief description of the six sequences along with a tone mapped frame, overall dynamic range and production techniques.

Thumbnail	Name	Resolution	Dynamic Range (stops)	Production Technique	Description
	Welding	1920×1080	21.36	Spheron VR	An indoor scene of a gas welding machine producing intermittent sparks of very high luminance.
	CGRoom	1920×1080	23.26	Rendered	An artificially rendered scene of the dark basement with an overhead lamp swinging as barrels fall from an overhead shelf.
	Jaguar	1920×1080	28.45	Canon EOS 1Ds Mark-II	An side profile indoor shot of a Jaguar E-Type. Bright lights are placed in the room to artificially expanding the scene dynamic range.
	Seine	1920×1080	24.55	Arri Alexa	Night outdoor scene of the river Seine in Paris with a brightly lit ferry producing the high luminance region of the scene. Scene post-processed.
	Tears of Steel	1920×1080	17.24	N.A.	A clip extracted from the short film produced as a part of the Open Movie project by Blender Foundation.
	Mercedes	1920×1080	19.36	Arri Alexa	An outdoor daylight scene of a Mercedes showroom with a parking lot. Scene post processed.

Table 1: Overview of the scenes used for the rating based psychophysical experiment.

For the purpose of this evaluation the sequences were graded (in absolute luminance terms) such that the pixel values are in the range $\in (10^{-3}, 4000)$ cd/m^2; within the range of the SIM2 HDR display [42]. A tone-mapped frame of each sequence along with the overall dynamic range is given in the supplementary materials and sequences common to both objective and subjective evaluations are marked suitably.
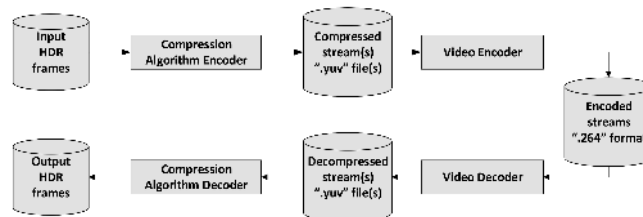
3.3. Preparation of HDR videos (HDRVs)

Figure 2: Compression Protocol used for the evaluation.

Input HDR frames were converted to a format, suitable for the encoder, using the described algorithms, creating intermediate '.yuv' files which are then passed to the encoder, in this case H.264/AVC creating a raw video format '.264'. This is subsequently decoded and decompressed - thus reconstructing the HDR frames. This pipeline allows to plot the RD characteristics of each algorithm when the '.yuv' files are encoded at different quality (QP) settings. Also, the duration of all HDRVs used in this work was fixed to five seconds i.e. 150 frames encoded at 30 frames per second. To preserve the best frame fidelity, the encoder sub-sampling format was set to *High 4:4:4* for 3-channel '.yuv' streams and *High 4:0:0* for luma (only) streams respectively. The Group of Pictures (GOP) structure was set to *I-P-P-P* with an Intra-frame period of 30 frames. Figure 2 provides a visual description of the pipeline.

4. Objective evaluation

This section introduces the QA metrics used for objective evaluation, describes the methodology followed and finally provides overall results for all HDR video compression algorithms across all sequences for each of the seven QA metrics.

4.1. Quality assessment (QA) metrics

- **PSNR**: The Peak Signal to Noise Ratio is the most widely used QA metric for judging image/frame quality in compression applications. Although, not always analogous to human perception, PSNR provides an indication of video quality, even in case of HDR video. For better suitability, this work modifies the PSNR by fixing a peak value $L_{peak} = 10^4$ and averaging the values across all three channels.

- **logPSNR**: This metric modifies the PSNR by taking into account the logarithmic perception of real-world luminance. Therefore, logPSNR was derived by log encoding each channel of the decoded and uncompressed frames as follows such that $r_1' = \log_{10}(r_1)$. Similarly, g_1' and b_1' are calculated for both decoded and uncompressed frames. Next, the Mean Square Error (MSE), for each channel is calculated and logPSNR is computed such that $logPSNR = 20\log_{10}(\frac{\log_{10}(L_{peak})}{\sqrt{MSE}})$. Again for better suitability, the peak value was fixed such that $L_{peak} = 10^4$ and the mean across three channels is reported.

- **puPSNR**: Aydin et al.(2008) [3] proposed a dynamic range independent extension to QA metrics such as PSNR and SSIM which is capable of handling luminance levels from 10^{-5} to 10^8 cd/m^2. The proposed metric maps real-world luminance to perceptually uniform code values derived from the contrast sensitivity function. The resultant code values are used in the QA metric instead of gamma corrected RGB or luma values.

- **puSSIM**: Similar to puPSNR, this metric is a dynamic range independent extension to the Structural Similarity Index Metric [46] which is used to measure the structural similarity between the reference and decoded image pairs.

- **Weber MSE**: This QA metric calculates the mean square error between the reference and decoded frames using Weber ratios and is defined as:

$$Q = \frac{1}{C} \sum_{j=1}^{C} \left(\frac{(i_1 - i_2)^2}{i_1^2 + i_2^2} \right) \tag{1}$$

where C represents the number of channels, i_1 and i_2 are the reference and decoded HDR frames respectively.

- **HDR-VDP**: Mantiuk et al (2011) [28] proposed a perceptual QA metric which estimates the probability at which an average human observer will detect differences between a pair of images in a psychophysical evaluation. This takes several aspects of the human visual system into account such as a wide range of viewing conditions, intra-ocular light scatter, photo-receptor spectral sensitivities and contrast sensitivity across the full range of visible luminance. This evaluation uses the Q correlate of version 2.2.1 [34].

6

- **HDR-VQM**: Narwaria et al. [36] proposed the first dedicated HDR video quality metric which uses perceptually uniform encoding, sub-band decomposition, short-term and long-term spatio-temporal pooling to predict the reconstruction quality of HDR video sequences. Unlike HDR-VDP which considers only physical luminance thus disregarding color information, this metric takes into account the color information of the video frames. Also, unlike other modified/dedicated HDR QA metrics this QA metric was primarily designed for HDR video and thus takes temporal coherence into account. This evaluation uses HDR-VQM - version 2.

4.2. Quality and bitrate selection

For the objective evaluation, the HDRVs were encoded at 11 different quality settings which were controlled by the quantization parameter (QP) of the encoder such that $QP = 1, 5, 10, 15, 20..., 50$. While $QP = 1$ represents near lossless compression and maximum output bpp, $QP = 50$ results in a highly lossy compression and minimum output bpp.

For the *backward compatible* algorithms except *rate*, the same QP was set for both the base and residual streams. For the *rate* algorithm, the Lagrangian optimization formula $QP_{residual} = 0.77 \times QP_{base} + 13.42$ mentioned in [23] was used with automatic RD correction switched off.

4.3. Objective results

This section provides two sets of objective results. First, the six algorithms, were evaluated against a larger set of 39 sequences. Figure 3 demonstrates the RD characteristics i.e. quality vs. output bitrate for each of seven QA metrics.

Second, the algorithms were evaluated against the six short-listed sequences which were to be used for the subjective evaluation. This facilitates the correlation between the objective and subjective evaluation results. Figure 4 demonstrates the *puPSNR* and *HDR-VDP* and *HDR-VQM* results. To average the results across all scenes while providing a fair comparison, the same set of scenes must be averaged for each method and the data must be provided for a fixed set of bitrates or quality levels. Although, quality variation is more commonly reported at fixed bitrates, this was difficult for the given set of scenes as each scene required significantly different bitrate to encode. Therefore, the RD curves did not overlap and the data could not be averaged properly. However, it was possible to average across the same set of scenes if the quality levels were fixed and bitrate values were interpolated.

Finally, the supplementary material to this paper contains; a) The RD graphs (for 11 QP values), plotted from *raw-data* averaged over 39 sequences, b) The *HDR-VDP* and *puPSNR* RD graphs for six short-listed sequences plotted from *raw-data* points, c) *puPSNR* and *HDR-VDP* RD graphs with quality variation on interpolated bitrates and d) QP vs. bitrate with 95% confidence interval bounds.

4.4. Objective Analysis and Discussion

A few salient points can be inferred from the objective evaluation results:

1. Perceptual QA metrics such as *puPSNR, HDR-VDP* and *HDR-VQM*, as seen in Figures 3 and 4 demonstrate that *non-backward compatible* algorithms using higher bit-depth outperform their *backward compatible* counterparts at lower output bitrates.
2. The RD characteristics demonstrated by the perceptual QA metrics such as *puPSNR, HDR-VDP* and *HDR-VQM* are similar to each other.
3. The inherent design of the *backward compatible* algorithms require a much higher bitrate to reproduce an acceptable image quality (without H.264 blocking artefacts). The mean output bitrate for *hdrjpeg* and *gohdr* (*backward compatible* algorithms - with residual streams containing the luma channel only) are similar to each other. The exceptions are *hdrmpeg* and *rate*. In *hdrmpeg*, both the base and residual streams contain 3-channels and are encoded with *High 4:4:4* sub-sampling. Again in *rate*, the Lagrangian optimization applied to the residual stream reduces the overall output bitrate, albeit at the cost of image quality.

5. Subjective evaluation

Most full reference QA metrics, were designed to evaluate image pairs without taking psychophysical aspects of the human visual system into consideration. Although perceptual QA metrics are good indicators of perceived image quality, the variation in objective results emphasizes the requirement for a comprehensive subjective evaluation.
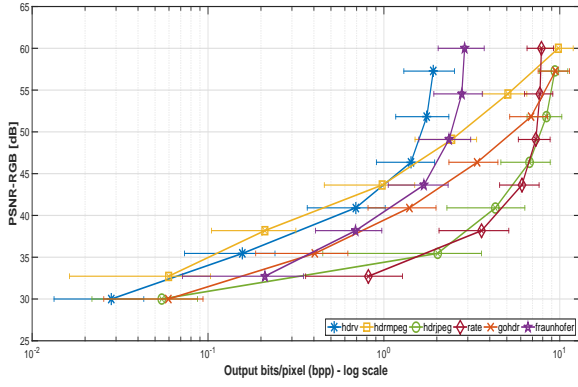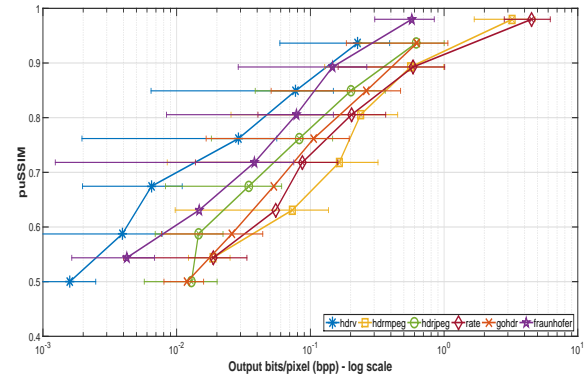
(a) PSNR results (higher PSNR - better quality)
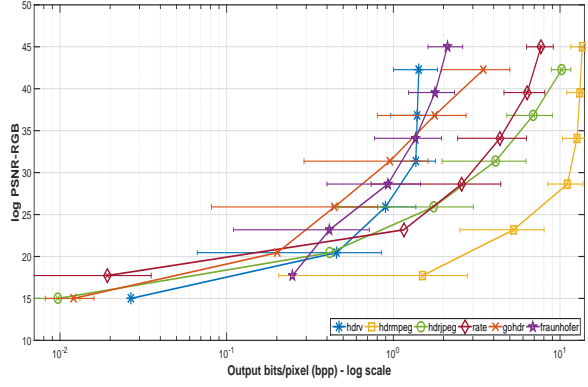
(b) logPSNR results (higher logPSNR - better quality)
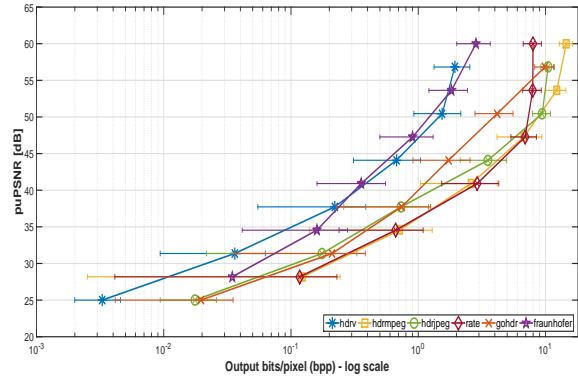
(c) puPSNR results (higher puPSNR - better quality)

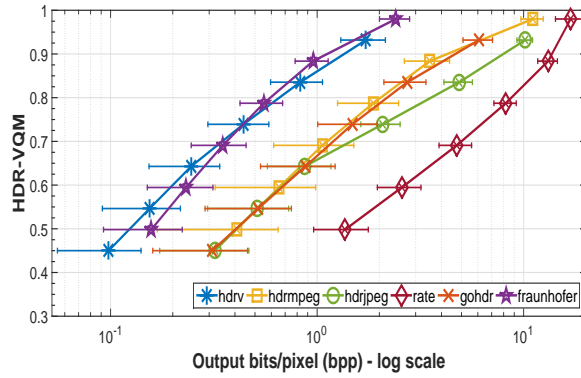(d) puSSIM results (higher puSSIM - better quality)
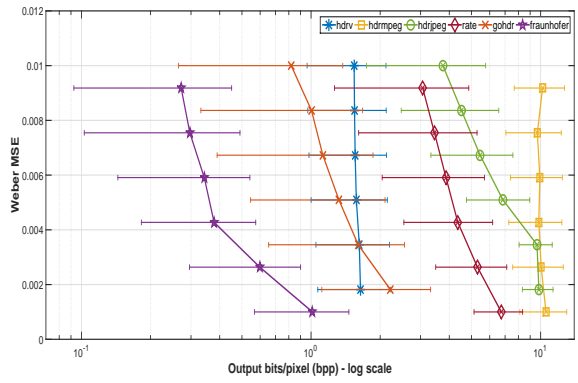
(e) Weber MSE results (lower Weber MSE - better quality)
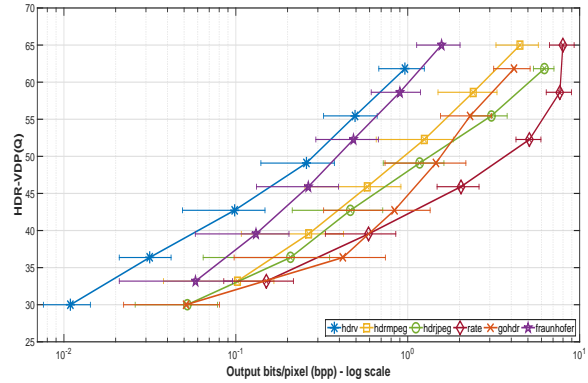
(f) HDR-VDP(Q) results (higher Q-values - better quality)

(g) HDR-VQM results (higher VQM - better quality)

Figure 3: Averaged RD characteristics (quality vs output bitrate) of six HDR video compression algorithms against seven QA metrics over 39 sequences. Output bitrates shown in logarithmic scale.
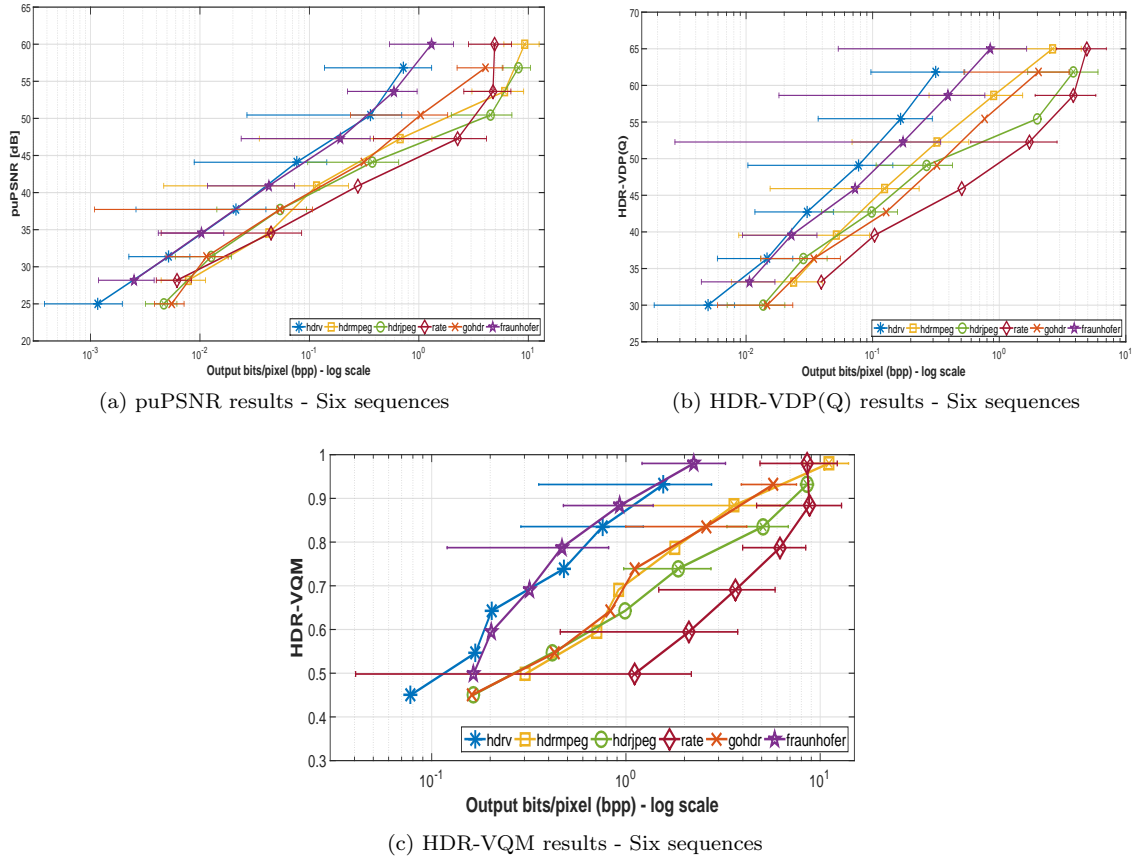
(a) puPSNR results - Six sequences

(b) HDR-VDP(Q) results - Six sequences

(c) HDR-VQM results - Six sequences

Figure 4: Average RD characteristics for the six short listed HDR video sequences.

5.1. Design

Multiple subjective evaluations at different image quality levels are ideally required to verify and correlate the results with objective evaluation. However, such an undertaking is very time consuming. Therefore, this work presents the results of two ranking-based psychophysical evaluations at two different quality levels. A ranking-based evaluation was chosen since it requires only one HDR display and guarantees that each ranked compression technique has a unique value, thereby ensuring quick and decisive results as opposed to a full-pairwise comparison experiment. Also, the relative rapidity of the process, approximately 20 minutes per participant, reduces fatigue.

The primary goal of the experiments was to rank and identify the order of each algorithm, across the six short-listed sequences, at two different quality levels. Participants were tasked to rank six algorithms for each of the six sequences, one at a time. For each sequence they had to view HDRVs from each algorithm at least once. They were tasked to identify and rank the given HDRVs in order of their resemblance to the clearly labeled reference HDRV. Also a *hidden reference*, identical to the labeled reference was mixed with the algorithms.

The sequences and algorithms were randomly presented in order to avoid bias. While ranking the sequences, participants were allowed to view the HDRVs as many times as required. The motivation behind this was to be able to distinguish between HDRVs that are relatively close in quality without the exhaustive full-pairwise comparisons.

5.2. Materials

Software resources included HDRVs from six compression algorithms, uncompressed reference HDRVs and a graphical user interface (GUI) for the ranking-based experiment. Hardware resources included a SIM2 HDR display [42] with a peak luminance rating of 4000 cd/m^2, an LG 22″ LED display with peak luminance rating of 300 cd/m^2 and a computer with a solid state drive for quick loading of HDRVs.

5.2.1. HDRVs for psychophysical experiment

Two fixed bpp(s) representing two quality levels were chosen based on the objective results as shown in Figure 4. The lower quality (LQ) level was chosen at 0.15 bpp (approximately 8.8 Mbps - similar to online streaming quality), such that image-quality distortions are clearly visible but not obscured by

H.264 blocking artefacts. The higher quality (HQ) level was chosen at 0.75 bpp (approximately 44.49 Mbps - similar to blu-ray quality).

Following the chosen quality levels, the six sequences were encoded at different QP settings for each algorithm to achieve the closest possible match to the target bitrate (within 5% error margin). Subsequently, the reconstructed HDR frames were converted to a custom file format suitable for displaying the HDR frames at 30 fps on a SIM2 HDR display.

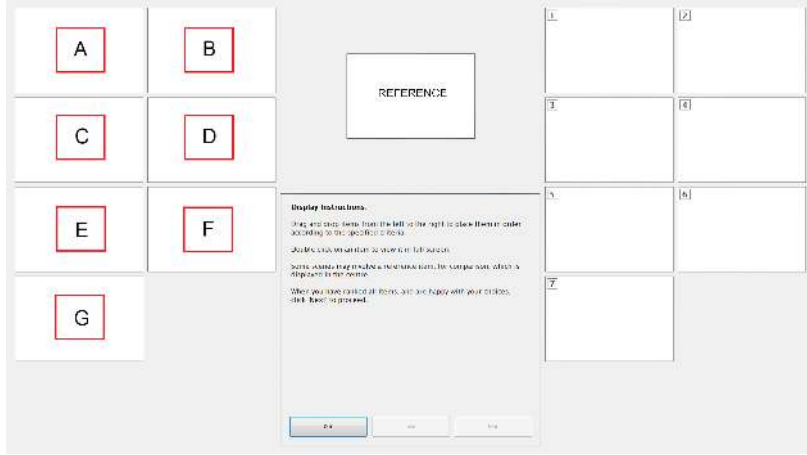### 5.2.2. Software for psychophysical experiment

Figure 5: Screenshot of the evaluation software

A custom GUI application, shown in Figure 5, was specifically built for the ranking-based subjective evaluation. It presents seven thumbnails each linked to an HDRV (labeled A-G), six from different algorithms and a hidden reference for each sequence, on the left side of the screen. The clearly marked reference HDRV (or ground truth) thumbnail is presented in the center. Each thumbnail, when *double-clicked* plays the linked HDRV on the HDR screen. Participants are tasked to view the reference HDRV first and subsequently rank the HDRVs on the left side in order of resemblance with the reference by dragging their preferred choice to its corresponding position (labeled 1-7) on the right side. The instructions for carrying out the experiment is clearly described in a text box below the reference thumbnail.

5.3. Participants

A total of 64 participants were divided into two groups, 32 for each experiment (LQ and HQ), with an age range of 20 to 50 years and from various academic and corporate backgrounds took part in the experiments. The participants reported normal or corrected to normal vision.

5.4. Environment

Following ITU-R recommendations [18], the experiments were conducted in a room with minimal ambient lighting (below 25 lux) which is within the recommended luminance levels for a typical dark environment [12]. The distance between the HDR display and the participant was set to approximately 3.2 times the height of the HDR display; at a distance of ≈ 189 cm with an LCD monitor placed at an angle of 45° (see Figure 6). In order to minimize glaring, the brightness and contrast of the LCD monitor was reduced to 25%.

5.5. Procedure

The participants were introduced to the objectives of the experiment prior to the start followed by a brief training session using a particular sequence which was subsequently discarded from the results. Upon completion of the training, the participants were asked to proceed further and rank the decoded HDRVs for the six short-listed sequences.

Each participant had to first view the reference HDRV on the HDR screen. Subsequently, the participant had to view each of the seven decoded HDRVs including the hidden reference and perform a qualitative assessment as to how much the decoded HDRVs resembled the ground truth HDRV in the center. Based on their judgement, the participants positioned the corresponding thumbnails to one of the blank positions on the right labeled [1-7] where 1 being an HDRV with least distortion compared to the reference and 7 being the HDRV with most visible distortions.

Figure 6: Psychophysical experiment setup.

5.6. Subjective results

This section provides an overview of the results obtained from the psychophysical experiments and analyzes the same.

Let the null Hypothesis H_0 be that there are no significant differences between the compression algorithms for both LQ and HQ. The alternative H_1 states that there are significant differences between the algorithms. The statistical significance p is assumed to be 0.05. The sample size for both LQ and HQ is 32. Also, if H_1 is true, it is important to determine the coefficient of concordance which measures the degree by which the participants mutually agree on choices.

Now, let A(N, M, S) be a 3-dimensional data array where N denotes all participants, M denotes all compression methods (algorithms) and S denotes all six sequences. Therefore, A(N,M,S) represents the ranks given by each participant to each method for each of the six sequences.

This implies that $\bar{A}(\bullet, M, S)$ represents the mean ranks for each M and S, averaged across all participants. Also, $\bar{A}(\bullet, M, \bullet)$ represents the mean ranks averaged across all participants and sequences keeping M fixed. The grand average should be equal such that:

$$\frac{1}{K} \sum_{S=1}^{K} \bar{A}(\bullet, M, S) = \bar{A}(\bullet, M, \bullet) \text{ ,where K = total number of sequences.} \tag{2}$$

Furthermore, for each sequence S, assuming N = 32 being the number of participants and M = 7 being the number of methods (algorithms), let $r_{i,j}$ be the rank assigned to each algorithm $i \in M$ by each participant $j \in N$. Thus, for each algorithm i, $R_i = \sum_{j=1}^{N} r_{i,j}$ is the sum of the ranks assigned by N participants.

Therefore, $\bar{R} = \frac{1}{N} \sum_{j=1}^{N} r_{i,j}$ is the mean of the ranks assigned to algorithm i and $R = \sum_{i=1}^{M} (R_i - \bar{R})^2$ is the standard squared deviation. From this data, the Kendall's coefficient of concordance W can be computed as:

$$W = \frac{12R}{N^2(M^3 - M)} \tag{3}$$

The significance of W can be analysed using chi-squared statistics such that $\chi^2 = \frac{M(M-1)(1+W(N-1))}{2}$ where χ^2 is asymptotically distributed with $\frac{M(M-1)}{2}$ degrees of freedom. A significance between scores suggests that the perceived image quality of two algorithms when compared with each other are different although no conclusions can be drawn for cases of similarity.

The analyzed results show that there are statistically significant differences between the algorithms for the six sequences. All tests show a significance $p < 0.05$. Therefore, H_0 is rejected and H_1 is accepted. Table 2a and 2b represents each $\bar{A}(\bullet, M, S)$ and $\bar{A}(\bullet, M, \bullet)$ along with W score and χ^2 value for the LQ and HQ experiments, respectively. Furthermore, by combining the results of Tables 2a and

, a generalized $\bar{A}(\bullet, M, \bullet)$ can be computed for the entire subjective experiment across a total sample size of $N_{\text{total}} = 64$ as given in Table .

(a) @ 0.15 bpp

Sequence	Compression Algorithms with Mean Rankings (@ 0.15 bpp)							Kendall (W)	X²	Sign. μ
Welding	reference (1.53)	hdrjpeg (2.63)	fraunhofer (3.75)	hdrv (3.78)	gohdr (4.25)	rate (5.97)	hdrmpeg (6.09)	0.586	112.58	p < 0.01
CGRoom	reference (1.56)	hdrv (2.13)	fraunhofer (3.75)	gohdr (4.47)	hdrjpeg (5.28)	hdrmpeg (5.41)	rate (5.41)	0.548	105.16	p < 0.01
Jaguar	reference (1.37)	hdrv (2.75)	fraunhofer (3.68)	hdrjpeg (3.71)	hdrmpeg (4.59)	gohdr (5.43)	rate (6.43)	0.607	116.50	p < 0.01
Seine	reference (1.68)	fraunhofer (2.81)	hdrv (3.71)	gohdr (3.78)	hdrjpeg (4.34)	hdrmpeg (5.18)	rate (6.46)	0.518	99.48	p < 0.01
TOS	reference (1.75)	hdrv (2.56)	hdrmpeg (3.34)	fraunhofer (3.46)	hdrjpeg (5.00)	rate (5.71)	gohdr (6.15)	0.587	112.76	p < 0.01
Mercedes	reference (1.21)	hdrv (2.93)	fraunhofer (3.46)	gohdr (3.84)	hdrmpeg (4.56)	hdrjpeg (5.15)	rate (6.81)	0.669	128.46	p < 0.01
Ā(. , M , .)	reference (1.52)	hdrv (2.97)	fraunhofer (3.48)	hdrjpeg (4.35)	gohdr (4.65)	hdrmpeg (4.86)	rate (6.135)	0.783	150.26	P < 0.01

(b) @ 0.75 bpp

Sequence	Compression Algorithms with Mean Rankings (@ 0.75 bpp)							Kendall (W)	X²	Sign. μ
Welding	reference (2.75)	fraunhofer (3.06)	gohdr (3.12)	hdrv (3.28)	hdrmpeg (5.00)	hdrjpeg (5.34)	rate (5.43)	0.307	58.94	p < 0.01
CGRoom	reference (2.75)	fraunhofer (3.00)	hdrv (3.06)	gohdr (3.71)	hdrmpeg (4.93)	rate (5.00)	hdrjpeg (5.53)	0.277	53.10	p < 0.01
Jaguar	reference (2.28)	fraunhofer (2.72)	hdrv (3.00)	gohdr (3.53)	hdrmpeg (5.28)	rate (5.46)	hdrjpeg (5.71)	0.449	86.18	p < 0.01
Seine	hdrv (2.63)	reference (2.81)	fraunhofer (3.06)	gohdr (3.40)	hdrmpeg (4.93)	hdrjpeg (4.93)	rate (6.21)	0.400	76.88	p < 0.01
TOS	reference (2.91)	hdrv (3.41)	fraunhofer (3.46)	hdrmpeg (3.62)	gohdr (4.31)	hdrjpeg (5.00)	rate (5.28)	0.168	32.30	p < 0.01
Mercedes	reference (2.81)	hdrv (3.46)	gohdr (3.62)	hdrmpeg (3.78)	fraunhofer (3.93)	rate (5.15)	hdrjpeg (5.21)	0.168	32.27	p < 0.01
Ā(. , M , .)	reference (2.71)	hdrv (3.14)	fraunhofer (3.21)	gohdr (3.61)	hdrmpeg (4.59)	hdrjpeg (5.29)	rate (5.42)	0.511	98.20	P < 0.01

(a) Subjective ranks with Kendall W, averaged across participants at 0.15 bpp

(b) averaged across participants at 0.75 bpp

Sequence	Compression Algorithms with Mean Rankings (combined for 0.15 & 0.75 bpp)							Kendall (W)	X²	Sign. μ
Ā(. , M , .)	reference (1.66)	hdrv (2.62)	fraunhofer (3.02)	gohdr (4.16)	hdrjpeg (4.90)	hdrmpeg (5.22)	rate (6.42)	0.597	229.15	P < 0.01

(c) Subjective mean ranks with Kendall W, combined and averaged over HQ and LQ experiments

Table 2: Subjective results and groups for the LQ and HQ experiments

5.7. Analysis

Tables 2a and 2b illustrate the ranking of all methods. For those compression algorithms that are grouped together, no significant difference was found at $p < 0.05$. However, there are significant differences in-between separate groups. Larger groups, as seen mostly in Table 2b indicate ambivalence of participants in choosing one algorithm over another. The corresponding low Kendall W reaffirms the difficulty in comparing algorithms at higher output bitrates. However, large groups are less likely for the LQ experiment and higher Kendall W confirms the consistency in participants' choices as it is expected that more differences are noted at lower qualities. The combined results obtained from Table 2c further reduces the sizes of the groups providing a generalized subjective result with a moderately high degree of consistency $W = 0.597$ between the overall participants' choices.

Based on the mean ranks of each M from the three $\bar{A}(\bullet, M, \bullet)$ in Tables 2a, 2b and 2c, the algorithms can be assigned an ordinal rank. Such an ordinal ranking system as given in Table 3 presents a summarized information about the choices made by participants in the LQ and HQ experiments. Table 3 also presents the ordinal ranks when the LQ and HQ results are combined.

Algorithm	LQ ranking	HQ ranking	LQ + HQ
hdrv	1	1	1
fraunhofer	2	2	2
gohdr	4	3	3
hdrjpeg	3	5	4
hdrmpeg	5	4	5
rate	6	6	6

Table 3: Ordinal ranks for both LQ and HQ subjective experiments

6. Discussion

This discussion combines the objective and subjective results in order to establish a correlation between them and analyze the overall performance of the algorithms.

The reconstructed HDR frames at 0.15 bpp and 0.75 bpp from each algorithm are evaluated against the reference sequences using the previously mentioned QA metrics. The overall results from the algorithms for the six QA metrics can be sorted using the same ordinal ranking system as discussed in the previous section. Finally, a correlation is computed by combining the objective and subjective ordinal rankings at 0.15 bpp and 0.75 bpp using Spearman's rank correlation test. Table 4 shows Spearman's correlation results for the combined LQ and HQ experiments.

	PSNR	logPSNR	puPSNR	puSSIM	Weber MSE	HDR-VDP	HDR-VQM	LQ	HQ
PSNR	-	-0.257	0.371	0.257	-0.232	0.493	0.257	0.371	0.371
logPSNR	-.257	-	.657	.771	.812*	.522	.771	.600	.657
puPSNR	.371	.657	-	.943**	.725	.986**	.943**	.829*	1.000**
puSSIM	.257	.771	.943**	-	.841*	.899*	1.00**	.943**	.943**
Weber MSE	-.232	.812*	.725	.841*	-	.632	.841*	.754	.725
HDR-VDP	.493	.522	.986**	.899*	.632	-	.812*	.899*	.986**
HDR-VQM	.257	.771	.943**	1.00**	.841*	.899*	-	.943**	.943**
LQ	.371	.600	.829*	.943**	.754	.812*	.943**	-	.829*
HQ	.371	.657	1.000**	.943**	.725	.986**	.943**	.829*	-

Table 4: Spearman's Rho rank correlation between objective and subjective evaluation for the LQ and HQ experiment respectively. '*' denotes significance at $p < 0.05$ level and '**' denotes significance at $p < 0.001$ level

First of all Table 4 shows that there are significant correlations between objective and subjective evaluation. For both the LQ and HQ experiments, the correlation between perceptual QA metrics such as puPSNR/puSSIM/HDR-VDP/HDR-VQM and subjective rankings is very high with statistical significance at $p < 0.001$ level. However, the correlation in-between the QA metrics varies with the image quality. While Table 4 shows a very high correlation in-between perceptual QA metrics such as puPSNR, puSSIM, HDR-VDP and HDR-VQM, the correlation in-between perceptual and mathematical metrics such as HDR-VDP and PSNR is significantly low. Finally, analogous to previous studies mentioned in Section 2, PSNR demonstrates a significantly low correlation with subjective rankings.

The objective results suggest that dedicated *non-backward compatible* algorithms tend to outperform their *backward compatible* counterparts at low to moderately high output bitrates. However, the differences are less clear to human participants. The $\bar{A}(\bullet, M, \bullet)$ groups in Table 2a include *fraunhofer* and *hdrjpeg* in a single group which suggests no significant difference between the algorithms at 0.15 bpp. Similarly, the $\bar{A}(\bullet, M, \bullet)$ groups in Table 2b include *gohdr* along with the *non-backward compatible* algorithms. Finally, the combined data in Table 2c shows that although *fraunhofer* and *gohdr* are part of the same sub-group, *hdrv* and *fraunhofer* are preferred over other *backward compatible* algorithms.

It is important to note that even though *non-backward compatible* algorithms perform well at lower bpp, the output streams cannot be played back using existing video players. Furthermore, in practice, hardware support for 10/12 bit encoders and decoders are currently quite rare. However, the flexibility and simplicity of the *fraunhofer* design facilitates an easier adaptation to upcoming 10-bit video pipelines. On the other hand, the *backward compatible* algorithms can use existing 8-bit video pipelines providing a distinct advantage in early adoption of HDR.

Out of the four *backward compatible* algorithms, *hdjpeg, hdrmpeg* and *rate* contain an 8-bit tone-mapped base stream making them truly backward compatible. *gohdr*, the only exception, is able to match the performance of *non-backward compatible* algorithms, albeit at the cost of true backward compatibility. Also, the Lagrangian optimization in *rate* saves output bitrate at the cost of reconstructed image quality. Although, Lee et al. [23] claimed, that *rate* performed better than *hdrmpeg* at lower bitrates, the algorithm was tested on sequences with VGA resolution and an older version of HDR-VDP [30]. Therefore, the claim might not always hold true for a large set of full HD resolution sequences.

The overall comparison results suggest that the choice of compression algorithm is largely application specific. The best possible HDR video quality at minimal output bitrates can be delivered by *non-backward* compatible algorithms. However, video pipelines with higher bit-depth support are required in that case. On the other hand, *backward compatible* algorithms can deliver HDR video content using legacy pipelines albeit at the cost of higher output bitrates. The choices are reaffirmed by the algorithms' RD characteristics against perceptual QA metrics which evidently has a high correlation with the subjective evaluation.

7. Conclusion and future work

This work endeavors to provide a detailed comparison of a number of published and patented HDR video compression algorithms and forms the foundation against which other HDR video compression algorithms can be evaluated in future. It establishes that *non-backward* compatible compression algorithms enjoy a distinct advantage over their *backward compatible* counterparts at lower bitrates. Also, *backward compatible* algorithms require *dual-loop* encoding to create the residual stream. This adds more complexity to any hardware design thereby making it less suitable for real-time deployment.

The work presented in this paper opens up avenues of future research. In practice, there are no 12 bit hardware encoders and decoders and this paper presents a comprehensive evaluation of compression algorithms which were proposed before the MPEG CfE for HDR/WCG compatibility with HEVC-Main-10 profile. Therefore, an interesting research area which has recently gained traction following the HDR/WCG call for proposals to would be to see how *non-backward compatible* compression algorithms (including the ones presented in this work) can be adapted to perform with HEVC Main-10 profile [43] for even lower bitrates. Subsequently, the modified non-backward compatible algorithms can be evaluated against the recently adopted Perceptual Quantizer algorithm (SMPTE ST 2084) [31] and Hybrid Log-Gamma algorithm [9] to test and compare their HDR reconstruction performance such as has been considered by François et al. [13]. Efficient use of available bandwidth might finally lead to the widespread commercial adoption of HDR.

8. Acknowledgements

References

[1] I. Avcıbaş, B. Sankur, K. Sayood, Statistical evaluation of image quality measures, Journal of Electronic Imaging 11 (2) (2002) 206–223.
URL http://dx.doi.org/10.1117/1.1455011

[2] T. O. Aydin, M. Čadík, K. Myszkowski, H.-P. Seidel, Video quality assessment for computer graphics applications, in: ACM SIGGRAPH Asia 2010 Papers, SIGGRAPH ASIA '10, ACM, New York, NY, USA, 2010, pp. 161:1–161:12.
URL http://doi.acm.org/10.1145/1866158.1866187

[3] T. O. Aydın, R. Mantiuk, H.-P. Seidel, Extending quality metrics to full dynamic range images, in: Human Vision and Electronic Imaging XIII, Proceedings of SPIE, San Jose, USA, 2008, pp. 6806–10.

[4] M. Azimi, A. Banitalebi-Dehkordi, Y. Dong, M. T. Pourazad, P. Nasiopoulos, Evaluating the performance of existing full-reference quality metrics on high dynamic range (HDR) video content, in: ICMSP 2014: International Conference on Multimedia Signal Processing, 2014, p. 811.
URL http://waset.org/abstracts/Computer-and-Information-Engineering

[5] M. Azimi, R. Boitard, B. Oztas, S. Ploumis, H. Tohidypour, M. Pourazad, P. Nasiopoulos, Compression efficiency of HDR/LDR content, in: Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on, 2015, pp. 1–6.

[6] A. Banitalebi-Dehkordi, M. Azimi, M. T. Pourazad, P. Nasiopoulos, Compression of high dynamic range video using the HEVC and H.264/AVC standards, in: Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine), 2014 10th International Conference on, 2014, pp. 8–12.

[7] F. Banterle, A. Artusi, K. Debattista, A. Chalmers, Advanced high dynamic range imaging: theory and practice, CRC Press, 2011.

[8] F. Banterle, A. Artusi, K. Debattista, P. Ledda, A. Chalmers, G. Edwards, G. Bonnet, Hdr video data compression devices and methods, eP Patent App. EP20,080,012,452 (Jan. 13 2010).
URL http://www.google.com/patents/EP2144444A1?cl=en

[9] T. Borer, A. Cotton, A "display independent" high dynamic range television system.

[10] M. Čadík, T. O. Aydin, K. Myszkowski, H.-P. Seidel, On evaluation of video quality metrics: an HDR dataset for computer graphics applications, in: B. E. Rogowitz, T. N. Pappas (eds.), Human Vision and Electronic Imaging XVI, vol. 7865, SPIE, 2011.

[11] Y. Dong, P. Nasiopoulos, M. T. Pourazad, HDR video compression using high efficiency video coding (HEVC), reproduction 6 (2012) 7.

[12] Engineering Toolbox, Illuminance - recommended light levels, http://www.engineeringtoolbox.com/light-level-rooms-d_708.html.

[13] E. François, C. Fogg, Y. He, X. Li, A. Luthra, A. Segall, High dynamic range and wide color gamut video coding in hevc: Status and potential future enhancements, IEEE Transactions on Circuits and Systems for Video Technology 26 (1) (2016) 63–75.

[14] J.-U. Garbas, H. Thoma, Temporally coherent luminance-to-luma mapping for high dynamic range video coding with H.264/AVC, in: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, 2011, pp. 829–832.

[15] P. Hanhart, M. Bernardo, P. Korshunov, M. Pereira, A. Pinheiro, T. Ebrahimi, HDR image compression: A new challenge for objective quality metrics, in: Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on, 2014, pp. 159–164.

[16] P. Hanhart, P. Korshunov, T. Ebrahimi, Y. Thomas, H. Hoffmann, Subjective quality evaluation of high dynamic range video and display for future tv, SMPTE Motion Imaging Journal 124 (4) (2015) 1–6.

[17] P. Hanhart, M. Rerabek, T. Ebrahimi, Towards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies, in: SPIE Optical Engineering+ Applications, 2015.

[18] ITU, Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures, Tech. rep., International Telecommunication Union (2012).
URL http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-I!!PDF-E.pdf

[19] C. Kiser, E. Reinhard, M. Tocci, N. Tocci, Real time automated tone mapping system for HDR video, in: In Proc. of the IEEE Int. Conference on Image Processing, IEEE, 2012.

[20] A. Koz, F. Dufaux, A comparative survey on high dynamic range video compression, in: SPIE Optical Engineering+ Applications, International Society for Optics and Photonics, 2012, pp. 84990E–84990E.

[21] G. W. Larson, LogLuv encoding for full-gamut, high-dynamic range images, Journal of Graphics Tools 3 (1) (1998) 15–31.

[22] C. Lee, C.-S. Kim, Gradient domain tone mapping of high dynamic range videos, in: Image Processing, 2007. ICIP 2007. IEEE International Conference on, vol. 3, 2007, pp. III – 461–III – 464.

[23] C. Lee, C.-S. Kim, Rate-distortion optimized compression of high dynamic range videos, in: Proceedings of the 16th European Signal Processing Conference, 2008.

[24] C. Lee, C.-S. Kim, Rate-distortion optimized layered coding of high dynamic range videos, Journal of Visual Communication and Image Representation 23 (6) (2012) 908 – 923.
URL http://www.sciencedirect.com/science/article/pii/S104732031200096X

[25] A. Luthra, E. Francois, W. Husak, Call for evidence (CfE) for HDR and WCG video coding, Geneva, Switzerland, February, 2015.

[26] C. Mantel, S. Ferchiu, S. Forchhammer, Comparing subjective and objective quality assessment of HDR images compressed with JPEG-XT, in: Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on, 2014, pp. 1–6.

[27] R. Mantiuk, A. Efremov, K. Myszkowski, H.-P. Seidel, Backward compatible high dynamic range MPEG video compression, in: ACM SIGGRAPH 2006 Papers, SIGGRAPH '06, ACM, New York, NY, USA, 2006, pp. 713–723.
URL http://doi.acm.org/10.1145/1179352.1141946

[28] R. Mantiuk, K. J. Kim, A. G. Rempel, W. Heidrich, HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions, ACM Trans. Graph. 30 (4) (2011) 40:1–40:14.
URL http://doi.acm.org/10.1145/2010324.1964935

[29] R. Mantiuk, G. Krawczyk, K. Myszkowski, H.-P. Seidel, Perception-motivated high dynamic range video encoding, in: ACM SIGGRAPH 2004 Papers, SIGGRAPH '04, ACM, New York, NY, USA, 2004, pp. 733–741.
URL http://doi.acm.org/10.1145/1186562.1015794

[30] R. Mantiuk, K. Myszkowski, H.-P. Seidel, Visible difference predicator for high dynamic range images, in: Systems, Man and Cybernetics, 2004 IEEE International Conference on, vol. 3, IEEE, 2004, pp. 2763–2769.

[31] S. Miller, M. Nezamabadi, S. Daly, Perceptual signal coding for more efficient usage of bit codes, SMPTE Motion Imaging Journal 122 (4) (2013) 52–59.

[32] K. Minoo, Z. Gu, D. Baylon, A. Luthra, On metrics for objective and subjective evaluation of high dynamic range video (2015).
URL http://dx.doi.org/10.1117/12.2187242

[33] A. Motra, H. Thoma, An adaptive LogLuv transform for high dynamic range video compression, in: Image Processing (ICIP), 2010 17th IEEE International Conference on, 2010, pp. 2061–2064.

[34] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, P. Le Callet, HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images, Journal of Electronic Imaging 24 (1) (2015) 010501.
URL http://dx.doi.org/10.1117/1.JEI.24.1.010501

[35] M. Narwaria, M. Perreira Da Silva, P. Le Callet, Study of high dynamic range video quality assessment, vol. 9599, 2015, pp. 95990V–95990V–13.
URL http://dx.doi.org/10.1117/12.2189178

[36] M. Narwaria, M. P. D. Silva, P. L. Callet, HDR-VQM: An objective quality measure for high dynamic range video, Signal Processing: Image Communication 35 (2015) 46 – 60.
URL http://www.sciencedirect.com/science/article/pii/S0923596515000703

[37] E. Reinhard, M. Stark, P. Shirley, J. Ferwerda, Photographic tone reproduction for digital images, ACM Trans. Graph. 21 (3) (2002) 267–276.
URL http://doi.acm.org/10.1145/566654.566575

[38] M. Rerabek, P. Hanhart, P. Korshunov, T. Ebrahimi, Subjective and objective evaluation of hdr video compression, in: 9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), No. EPFL-CONF-203874, 2015.

[39] K. Seshadrinathan, R. Soundararajan, A. Bovik, L. Cormack, Study of subjective and objective quality assessment of video, Image Processing, IEEE Transactions on 19 (6) (2010) 1427–1441.

[40] H. Sheikh, A. Bovik, Image information and visual quality, Image Processing, IEEE Transactions on 15 (2) (2006) 430–444.

[41] H. Sheikh, M. Sabir, A. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, Image Processing, IEEE Transactions on 15 (11) (2006) 3440–3451.

[42] SIM2 Multimedia, SIM2 HDR47, http://www.sim2.com/HDR/hdrdisplay/hdr47e_s_4k.

[43] G. Sullivan, J. Ohm, W.-J. Han, T. Wiegand, Overview of the high efficiency video coding (HEVC) standard, Circuits and Systems for Video Technology, IEEE Transactions on 22 (12) (2012) 1649–1668.

[44] A. M. Tourapis, A. Leontaris, K. Sühring, G. Sullivan, H.264/AVC reference encoder, http://iphome.hhi.de/suehring/tml/.

[45] G. Valenzise, F. De Simone, P. Lauga, F. Dufaux, Performance evaluation of objective quality metrics for HDR image compression, in: Proc. SPIE, vol. 9217, 2014, pp. 92170C–92170C–10.
URL http://dx.doi.org/10.1117/12.2063032

[46] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, Image Processing, IEEE Transactions on 13 (4) (2004) 600–612.

[47] G. Ward, M. Simmons, Subband encoding of high dynamic range imagery, in: Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization, APGV '04, ACM, New York, NY, USA, 2004, pp. 83–90.
URL http://doi.acm.org/10.1145/1012551.1012566