Uniform Color Space based High Dynamic Range Video Compression

R. Mukherjee[1, 4], K. Debattista[1], T. Bashford-Rogers[2], M. Bessa[3, 4] and A. Chalmers[1]

[1]University of Warwick, UK
[2]University of the West of England, UK
[3]Universidade de Trás-os-Montes e Alto Douro, Portugal
[4]INESC TEC, 4200-465 Porto, Portugal

***Abstract*—Recently, there has been significant progress in the research and development of High Dynamic Range (HDR) video technology and state-of-the-art video pipelines are able to offer a higher bit depth support to capture, store, encode and display HDR video content. In this paper, we introduce a novel HDR video compression algorithm which uses a perceptually uniform color opponent space, a novel perceptual transfer function to encode the dynamic range of the scene and a novel error minimization scheme for accurate chroma reproduction. The proposed algorithm was objectively and subjectively evaluated against four state-of-the-art algorithms. The objective evaluation was conducted across a set of 39 HDR video sequences, using the latest x265 10-bit video codec along with several perceptual and structural quality assessment metrics at 11 different quality levels. Furthermore, a rating-based subjective evaluation (n=40) was conducted with six sequences at two different output bitrates. Results suggest that the proposed algorithm exhibits the lowest coding error amongst the five algorithms evaluated. Additionally, rate-distortion characteristics suggest that the proposed algorithm outperforms the existing state-of-the-art at bitrates ≥ 0.4 bits/pixel.**

I. Introduction

The capability of High Dynamic Range (HDR) video to capture, store and display a much larger magnitude of real-world lighting with floating point precision requires significantly higher storage and transmission costs when compared to Low/Standard Dynamic Range (LDR/SDR) video. Thus, several HDR video compression algorithms have been proposed which transform floating point RGB frames to file formats suitable for LDR video codecs. The resultant files are subsequently encoded into compressed HDR video streams.

Existing HDR video compression algorithms can be classified into two broad groups. The first group follows the *non-backward* compatible approach that produces a single video stream using the higher bit-depth support (typically 10-12 bits) provided by modern video codecs. The second group follows the more conventional *backward* compatible approach that produces at least two 8-bit legacy video streams, typically termed as base and residual streams, such that both streams can be encoded and decoded using legacy hardware and at least one stream can be played back using legacy hardware/software video players. Previous evaluations [1] have shown that the first approach facilitates high-fidelity HDR video reconstruction with lower storage and transmission requirements compared to the *backward* compatible algorithms.

In this paper, we introduce a novel *non-backward* compatible HDR video compression algorithm which uses the state-of-the-art perceptually uniform *Intensity, Protan and Tritan* (IPT) color opponent space [2], a novel perceptual transfer function with an analytical solution (henceforth labeled as PATF) and a novel quantization error minimization function (EMF) to non-linearly encode achromatic and chroma components, respectively. The proposed PATF and EMF in conjunction is able to achieve superior HDR video reconstruction performance compared to existing solutions. A visual description of the algorithm's overall work-flow is shown in Figure 1.

The proposed algorithm has been comprehensively evaluated against four state-of-the-art published and/or patented *non-backward* compatible HDR video compression algorithms using both objective and subjective quality assessment techniques. The objective evaluation was conducted using a set of 39 HDR video sequences and the x265 codec [3] (an HEVC [4] implementation) at 11 different quality levels where the compression performance (image reconstruction quality and transmission requirements) of the five algorithms were evaluated against several energy difference, structural and perceptual quality assessment (QA) metrics generating a set of generalized rate-distortion (RD) characteristics (see Section V for details). Additionally, a rating-based subjective evaluation was also conducted at two different output bitrates and the correlation between objective and subjective evaluation results were also computed (see Section VII for details).

The primary contributions of this work are; a) a novel *non-backward* compatible HDR video compression algorithm which uses a combination of IPT color opponent space, a novel non-linear analytical PATF to encode the *intensity* information and a novel EMF to non-linearly encode the chroma information b) a generic framework (modular structure) to plug-in existing contrast sensitivity based PTFs inside the algorithm to map real-world intensity values to luma and finally; c) a comprehensive objective and subjective evaluation of the proposed algorithm against existing state-of-the-art solutions.

II. Related work

A number of HDR video compression algorithms have been proposed to date. Additionally, several objective and subjective evaluations have been conducted to compare the compression

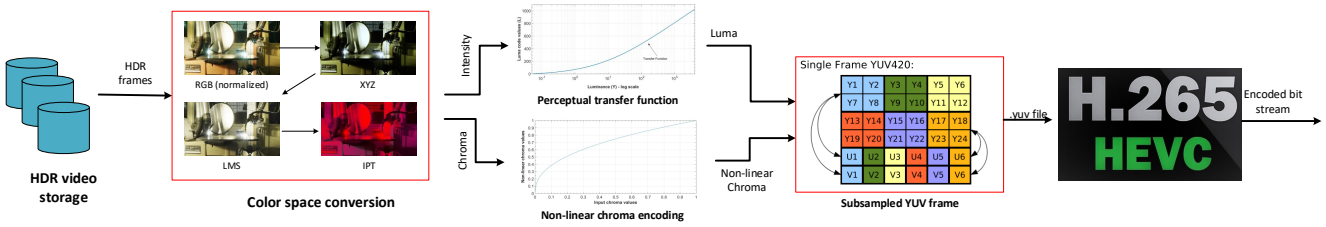


Fig. 1: An overall work-flow of the proposed HDR video compression algorithm.

performance of the proposed algorithms using dedicated or modified HDR image/video quality evaluation metrics as well as rating-/ranking-/pairwise-comparison-based psychophysical experiments, respectively. In this section, we provide a brief overview of some of the significant works conducted on the proposal and evaluation of existing HDR video compression algorithms.

A. Existing HDR video compression algorithms

As mentioned in Section I, a number of HDR video compression algorithms following either the *non-backward* compatible approach or the *backward* compatible approach have been proposed to date. In this section we briefly discuss four state-of-the-art *non-backward* compatible algorithms which are relevant to the context and have been used later to compare and contrast the performance of the proposed algorithm.

1) ***Perception based HDR video encoding (hdrv)****:* Mantiuk et al. [5] proposed the first non-backward compatible algorithm (*hdrv*) which extends the MPEG-4 encoder to accommodate HDR video content. The algorithm maps linear RGB or XYZ channels, to an 11-12 bit perceptually uniform luma space L_p using a transfer function and 8-bit chroma channels, u' and v', similar to LogLuv encoding [6]. Additionally, it extends the Discrete Cosine Transform (DCT) by encoding sharp contrast edges in the spatial, rather than frequency domain. As opposed to the original work which assumes a luminance range of $\mathrm{Y} \in (10^{-5}, 10^{8}]\ \mathrm{cd/m^2}$, the sequences considered in this work were within the range of $\mathrm{Y} \in (10^{-5}, 10^{4}]\ \mathrm{cd/m^2}$. Thus, the modified implementation in this work uses a 10-bit look-up table (LUT) to map the physical luminance to integer luma L_p space. Also, the DCT extension was deemed redundant due to the progress made in video codecs since 2004.

2) ***Temporally coherent luminance to luma (fraunhofer)****:* Garbas and Thoma [7] proposed an HDR video compression algorithm by modifying the Adaptive LogLuv algorithm [8] and adding temporal coherence to minimize flickering artefacts. The algorithm maps linear RGB values to $Lu'v'$ color space by converting physical luminance to 12-bit luma and adjusting the chroma information into 8-bit u' and v' similar to LogLuv encoding [6]. Temporal coherence was added by deriving the scale and offset parameters. Additionally, an auxiliary stream stores the meta-data information which is used during reconstruction. Due to the flexibility of the adaptive logarithmic transform [8], it was fairly straightforward to adapt this algorithm for 10-bit encoding.

3) ***Perceptual signal encoding (PQ)****:* Miller et al. [9] proposed an Opto-Electronic Transfer Function (OETF) [10] based compression algorithm where the OETF is based on Barten's contrast sensitivity function (CSF) [11]. The algorithm maps input pixel values $\mathrm{V} \in [0.005, 10^{4}]\ \mathrm{cd/m^2}$ to a 10-bit perceptually quantized uniform space. The OETF works on normalized signal values thus providing the flexibility to convert, scale and discretize at any desired bit depth. In this work, the implementation was modified for 10-bits and the maximum value (normalization factor) of each frame is stored as auxiliary meta-data required during HDR reconstruction.

4) ***Hybrid log-gamma encoding (bbc-hlg)****:* Borer et al. [12] proposed another OETF based compression algorithm (*bbc-hlg*) using a Hybrid-Log-Gamma transform. This OETF is a combination of the De-Vries-Rose relationship [13] and a logarithmic transfer function applied to a normalized and scaled linear floating point values. The OETF maps relative pixel values $\mathrm{V} \in (0, 12]$ to a signal $\mathrm{S} \in [0, 1]$. This algorithm also provides the flexibility of scaling and discretization and was adapted for 10-bit encoding in this work.

B. Evaluation of HDR video compression algorithms

Typically HDR video compression algorithms were developed in isolation or only partially compared with each other. A preliminary work was conducted by Koz et al. [14] where the two different approaches to HDR video compression i.e. the *non-backward* and *backward* compatible approaches were compared against each other. However, this work does not provide a comprehensive objective and subjective evaluation of individual algorithms. To that end, Mukherjee et al. [1] conducted a comprehensive evaluation of six compression algorithms (following either of the two approaches) and concluded that *non-backward* compatible algorithms deliver superior reconstruction quality compared to *backward* compatible algorithms at feasible output bitrates. However, this work does not consider the recently proposed standards PQ [9] and bbc-hlg [12]. Hanhart et al. [15] conducted an evaluation of nine compression algorithms submitted in response to the CfE and concluded that the submitted proposals can noticeably improve the standard HDR video coding technology and QA metrics such as *PSNR-DE1000*, *HDR-VDP-2* and *PSNR-Lx* can reliably detect visible difference between the reference and reconstructed frames. Similar evaluations were also conducted by Dehkrodi et al. [16], Rerabek et al. [17] and Narwaria et al. [18].

III. Background

In this section we provide an overview of some of the underlying concepts based on which the proposed algorithm has been designed.

A. Color spaces

HDR data is generally stored in linear RGB format which has a high correlation in-between the channels [19]. To minimize the effect of pixel manipulation on one channel affecting the others, RGB pixel values are typically converted to luma-chroma color spaces such as YC_bC_r or $Yu'v'$ where u' and v' represent uniform chromaticity scales. Also, for efficient compression purposes perceptual uniformity is desirable where the perceived difference in-between two colors is equal to the Euclidean distance between them [19]. Although the CIE-XYZ space can be used, it is not perceptually uniform and contains imaginary primaries with a large number of values which do not correspond to realizable colors leading to an inefficient use of available bit-depth [19]. Therefore, existing algorithms [5], [7], [9], [12] have used luma-chroma spaces as stated earlier. However, these color spaces are again not perfectly uniform. Thus, to address both the essential and desirable properties, the RGB data can be converted to device independent *hue, saturation and lightness* (HSL) color opponent spaces such as CIELAB/LUV [20]. However, further research [21] have confirmed issues with hue compressibility in CIELAB/LUV. Thus, the proposed algorithm uses the IPT color opponent space proposed by Ebner and Fairchild [22] which maintains the perceptual uniformity of CIELAB/LUV and mitigates the hue compressibility issues. Further details about the usage of IPT is discussed later in Section IV-B.

B. Perceptual Transfer Functions (PTFs)

In HDR video compression, a transfer function (TF) is ideally a reversible function which maps a range of input pixel values R_i to a range of output code values R_o such that R_o is suitable for video encoding. HDR pixel values are typically stored in floating point formats which are unsuitable for existing codecs limited to 14-bit integer representation [23]. Moreover, most commercial video codecs [3] are limited to a 10-bit representation only. Thus, a TF (say $f(\cdot)$) maps $R_i \in (10^{-5}, 10^{8}]$ cd/m^2 to a codec suitable n $-$ bit integer representation $R_o \in [0, 2^n - 1]$, such that $f : R_i \rightarrow R_o$. Such an operation however can cause banding artefacts which can be perceived by the human eye. Thus, to ensure that banding artefacts are not visible several TFs derived from previous psychophysical studies [24] have been proposed to date [25] taking into account the contrast sensitivity of the HVS at varying luminance. To simplify explanation, we focus on luminance Y to luma L mapping only in this section such that $f : Y \rightarrow L$.

Based on the psychophysical data, a function $f^{-1}(\cdot)$ i.e. $f^{-1} : L \rightarrow Y$ can be created such that an integer range L $\in [0, 2^{\text{n}} - 1]$ can be mapped to the physical luminance range Y $\in (10^{-5}, 10^{8}]$ where the difference between two successive code values ΔL $(L_i - L_{i-1})$ when mapped to ΔY $(Y_i - Y_{i-1})$ is below the human perception threshold as shown in Figure 2. Such a function is known as a *threshold vs intensity (t.v.i)* function. The *t.v.i* function along with the input and output boundary value conditions can be used to derive a TF, also known as a Perceptual Transfer Function (PTF). Such a PTF ($PTF : Y \rightarrow L$), used for compression related purposes

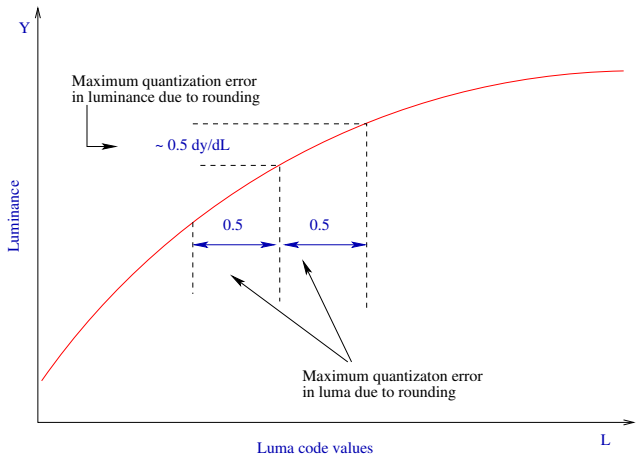


Fig. 2: Quantisation error in luma code values expressed in terms of luminance such that error $<$ 1 JND [26].

should ideally satisfy three conditions such that a) the output units L should be expressed in integers as that simplifies video compression b) a unit distance in L correlates with the Just Noticeable Difference (JND) which simplifies control of distortions for lossy compression algorithms and c) a half-unit distance in L should be ideally below 1 JND which ensures that the maximum quantization errors due to rounding operation are not noticeable [26].

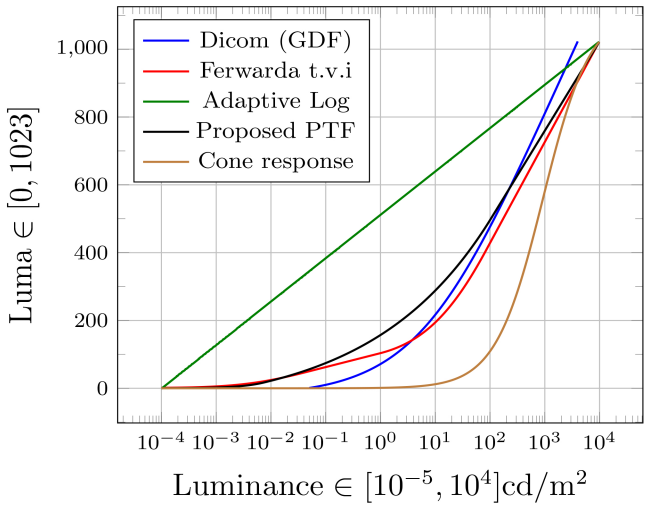


Fig. 3: A log-linear plot of five perceptual transfer functions (including a novel proposed PTF).

In this work, we have considered four widely used PTFs in conjunction with the proposed algorithm to non-linearly encode the scaled intensity channel values as described later in section IV-C. This includes the Adaptive LogLuv TF [8], the DICOM standard Grayscale Display Function [27], the *t.v.i* proposed by Ferwarda et al. [28] with Ward's modification [29] and the cone response model [30]. Here, we discuss the characteristics of each PTF and their usability in compression.

*1) **Adaptive Logarithmic TF**:* The adaptive logarithmic transform, proposed by Motra and Thoma [8], is a modification of a logarithmic function which adjusts and scales the output values based on the input and output boundary conditions. This enables the TF to encode the entire range of visible luminance

into $n-bit$ code values. A logarithmic PTF (see Figure 3) exhibits conservative quantization at lower luminance and coarser quantization at higher luminance pixel values. This can be attributed to the shape of the curve where a steeper curve results in a finer quantization [26]. However, previous psychophysical experiments have shown that the contrast detection thresholds of the HVS at scotopic and mesopic ranges are higher than at photopic luminance ranges. Therefore, the use of a logarithmic TF results in an inefficient usage of available bit depth [26].

*2) **Grayscale Display Function**:* The Digital Imaging and Communications in Medicine (DICOM) standard Grayscale Display Function (GDF) [27] is a polynomial fit derived from Barten's CSF experiments [11], which maps the input luminance $Y \in [0.05, 4000]$ $\mathrm{cd/m^2}$ to a 10-bit perceptually uniform JND space. Although the GDF is suitable for existing high-fidelity commercial HDR displays [31], it is limited to 4000 $\mathrm{cd/m^2}$ and future displays might exceed the encoding capabilities of this function. Also, the GDF exhibits exceedingly coarse quantization below 1000 $\mathrm{cd/m^2}$ and redundantly conservative quantization for higher luminance values which renders it unsuitable for accurate scotopic and mesopic luminance preservation.

*3) **Ferwarda's t.v.i** based PTF:* Ferwarda et al. [28] proposed another *t.v.i* function which takes into account the non-linear response of rods and cones separately. The proposed function models input luminance $Y \in [10^{-6}, 10^{9})$ $\mathrm{cd/m^2}$ to a JND space for rods and cones separately. The responses can be further approximated (by curve fitting) to create a single function as shown by Ward in [29]. However, this *t.v.i* function is based on data from 18 subjects, and the detection thresholds are higher for low luminance values and banding artefacts might be visible because the authors used a pulsating target on a constant background and perception thresholds are higher for transient stimuli compared to static stimuli [29]. Therefore, Ward [29] proposed a modification where the threshold luminances are divided by a factor of nine which brings the function in better agreement with the Barten model and yet preserves detail below $10^{-2}\mathrm{cd/m^2}$.

*4) **Global Cone Response Model**:* The final PTF in consideration was the Global Cone Response Model (GCRM) [30] primarily targeted to model the HVS response at photopic levels. The shape of GCRM (see Figure 3) indicates a conservative preservation of high luminance values at the cost of lower luminance values.

In addition to the mentioned functions, several other PTFs have also been proposed to date. A detailed overview and derivation of several PTFs along with their effect on the visibility of contouring artefacts are given in [25], [26], respectively.

As stated earlier, the algorithm proposed in this work uses the IPT color opponent space where the intensity channel is subsequently scaled and non-linearly mapped to a JND space by means of a PTF (see Section IV-C for details). To that end, we tested the encoding performance of the four mentioned PTFs in conjunction with the rest of the proposed algorithm and based on the results, a novel TF was proposed to incorporate the advantages of existing PTFs while mitigating some of their issues. The proposal of a novel TF (PATF), specifically designed for *intensity* channel encoding is one of the major contributions of this work.

IV. Overview of the Proposed Algorithm

The three major contributions of the proposed algorithm are; a) the usage of the IPT color opponent space, b) the proposal of a novel PTF with a straightforward analytical solution to perceptually encode the *intensity* channel information and c) proposal of a novel error minimization function (EMF) to accurately preserve the chroma information. Both the PTF and EMF, described in this section has been optimized for 10-bit encoding. In this section, we provide a detailed overview of the proposed algorithm and also outline the details of the novel PTF along with the details of the EMF.

A. Overall data-flow

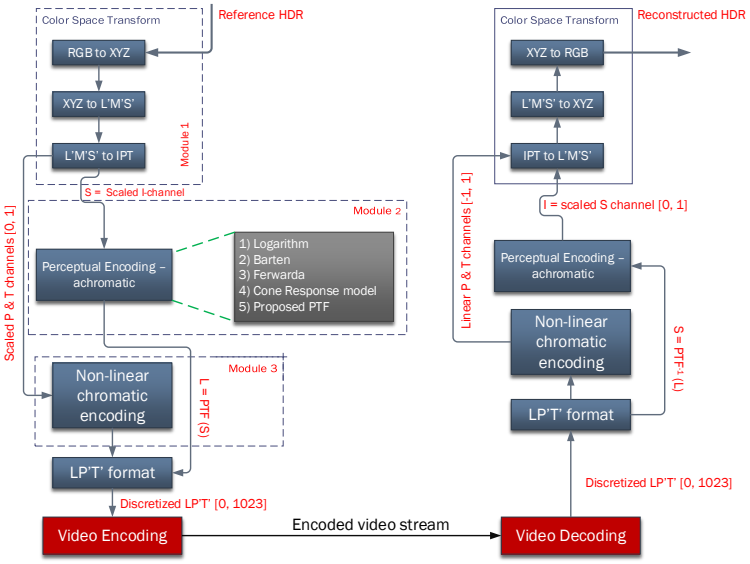


Fig. 4: Schematic diagram of the proposed algorithm and framework. **Note:** Modular breakdown is not shown in schematic clarity.

The proposed algorithm can be broadly classified into three modules. The first module normalizes and performs color space transform of input HDR frames (linear RGB) to perceptually uniform IPT color opponent space (see Figure 4 and Section IV-B). The second module extracts the *intensity* channel information from the resultant IPT frames and linearly scales the *intensity* information according to the requirements of a chosen PTF. Subsequently, the scaled *intensity* values are perceptually encoded to JND scaled luma code values using the chosen PTF (see Figure 4 and Section IV-C). The third module extracts the chroma components from IPT and applies the EMF (see Figure 4 and Section IV-D) to non-linearly encode the chroma components. Finally, the luma and chroma components are merged and passed to the video codec for encoding. On the decompression side, the encoded video stream is decoded and decompressed to reconstruct the HDR frames by reversing the data flow. The algorithm also uses meta-data information (see Section IV-E) to accurately reconstruct the HDR frames. The overall data-flow is visually described in Figure 4.

B. Module 1: Color space transform

The psychophysical experiments conducted by Ebner and Fairchild demonstrated that perceptually uniform color spaces such as CIELAB and CIELUV have hue compressibility issues [19], [21]. To mitigate the limitations of CIELAB/LUV, the algorithm converts the input HDR frame (in linear RGB) to the IPT color opponent space. This transforms input data into a perceptually uniform space for ease of image manipulation, facilitates de-correlation of the light and color information for compression purposes and finally exploits the advantages of CIELAB/LUV without affecting the hue changes that occur when compressing chroma along the lines of compressed hue [19]. A brief outline of the color space transformation is given in Algorithm 1, the details of which are described in [19].

Algorithm 1 ColorConvert(hdr)

1: $\nu \leftarrow max(hdr)$ //get normalization factor
2: $rgb_{norm} \leftarrow \frac{hdr}{\nu}$ //normalization
3: $XYZ \leftarrow [REC.709] \times rgb_{norm}$ //rgb to xyz
4: $LMS \leftarrow [HPE] \times XYZ$ //xyz to lms
5: $L'M'S' \leftarrow |LMS|^{0.43}$ //lms spectral sharpening
6: $IPT \leftarrow [IPTConv] \times L'M'S'$ //lms to ipt
7: $P \leftarrow IPT(x, y, 2)$ //extract the 2^{nd} channel
8: $T \leftarrow IPT(x, y, 3)$ //extract the 3^{rd} channel
9: $P_{scale} \leftarrow \frac{P-min(P)}{max(P)-min(P)}$ s.t $P \in (0, 1]$ //P scaling
10: $T_{scale} \leftarrow \frac{T-min(T)}{max(T)-min(T)}$ s.t $T \in (0, 1]$ //T scaling
11: $IPT_{out} \leftarrow I(x, y, 1), P_{scale}, T_{scale}$ //scaled IPT space

In Step 4, HPE refers to the Hunt-Pointer-Estevez fundamentals [32] and the metadata includes the normalization factor ν along with the minimum and maximum pixel values of the P and T channels prior to scaling.

C. Module 2: Perception based intensity encoding

Module 2 extracts the *intensity* channel from IPT_{out}. The *intensity* information when linearly scaled and discretized for 10-bit encoding exhibits visible contouring artefacts due to rounding errors. Thus, to minimize the quantization errors, the scaled I' channel can be perceptually encoded by any one of the four mentioned PTFs such that the resultant luma space satisfies the properties mentioned in Section III-B. Since $I \in (0, 1]$, it can be scaled to any range, suitable for a chosen PTF. The linear scaling operation is performed by a multiplying factor ψ followed by the application of the PTF as shown in equation 1.

For instance, if $I \in (0, 1]$, $f(\cdot)$ is the chosen PTF (say GDF) and L is the 10-bit JND quantized luma then the scaling and JND mapping operation is given as:

$$\begin{aligned} I' &= I \cdot \psi \text{ such that } I' \in [0.05, 4000] \\ \therefore L &= f(I') \text{ such that } L \in [0, 1023] \end{aligned} \tag{1}$$

*1) **Evaluation of existing PTFs**:* In order to evaluate and determine the *intensity* encoding efficiency of each PTF as well as to evaluate the reconstruction quality of the algorithm upon the application of the specific PTF, the scaled I' channel is encoded using each of the four existing PTFs (one at a time). The rest of the data flow remains unchanged (see Figure 4). Subsequently, the reconstruction quality of the proposed algorithm is determined using the evaluation methodology described later in Section V-B. The mean RD characteristics (averaged across 39 HDR video sequences) across a set of different quality levels determines the overall HDR reconstruction quality of the algorithm when using each of the four PTFs. This indirectly indicates the *intensity* channel encoding efficiency of each PTF.

The RD characteristics discussed later in Section VI show that amongst the existing PTFs, the algorithm demonstrates the best reconstruction quality using either GCRM or the PTF based on modified Ferwarda's *t.v.i*. However, both PTFs have certain issues as mentioned in Section III-B. To mitigate these issues, this paper also proposes a novel PTF which incorporates the advantages of both along with the added advantage of a straightforward analytical solution.

*2) **Design of the proposed PTF (PATF)**:* Following recommendation REC 1886 [33], the proposed PATF has been designed as a three-part analytical solution such that $f(\cdot) : I' \longrightarrow L$. The conditional equation 2 bears similarity to sRGB-non-linearity with linear and power function segments with an additional logarithmic segment to encode high *intensity* values.

$$L = \begin{cases} a \cdot I' & \text{if } I' < I'_s; \\ b \cdot I'^{(\frac{1}{c})} + d & \text{if } I' \in [I'_s, I'_p); \\ e \cdot log_{10}(I') + f & \text{if } I' \in [I'_p, I'_h]; \end{cases} \tag{2}$$

Similarly, $f^{-1}(\cdot)$ can be formulated as in equation 3.

$$I' = \begin{cases} \frac{L}{a} & \text{if } L < L_s; \\ (\frac{L-d}{b})^c & \text{if } L \in [L_s, L_p); \\ 10^{(\frac{L-f}{e})} & \text{if } L \in [L_p, L_h]; \end{cases} \tag{3}$$

The boundary value conditions I' was assumed to be similar to [9]. Therefore, $I \in (0, 1]$ is scaled by ψ such that $I' \in [10^{-5}, 10^4]$. Also, the JND quantized $L \in [0, 1023]$. The goal of the PATF is to facilitate a conservative quantization throughout the range of I' for low-, mid- and high-*intensity* values. Since the shape of GCRM shows bias towards preservation of high-*intensity* regions, it was taken out of consideration. Amongst the existing PTFs, the shape of Ferwarda's *t.v.i* based PTF was a close fit to the model proposed in Daly's VDP [26], [34] for the power segment and also a close fit to Barten's CSF based PTF for the logarithmic segment. Therefore, the analytical solution was initially fitted to Ferwarda's *t.v.i* (with Ward's modification) using non-linear regression techniques for initial calculation of the interval boundaries I'_s and I'_p. I'_h was always fixed to 10^4 as the upper bound of the *intensity*, considered in this work. This not only produced the initial interval boundaries and the co-factors of equation 2, it also ensured that the fitted solution adheres to the conditions shown in Figure 2. Finally, L_s and L_p are computed using the co-factors and interval boundaries. Similar to I'_h, L_h was again fixed to 1023 as the upper bound for 10-bit encoding.

Since the PATF is a piecewise-nonlinear model, it is important to enforce C^0 continuity at the intervals bounds I'_s and I'_p. Also, a luminance to luma plot (in semi-log scale) as shown in Figure 3 is unable to decipher contrast jumps and discontinuities which might result in visible contouring artefacts. Therefore, using a Contrast vs. Intensity (*c.v.i*) plot [29], the function was tested for contrast jumps and discontinuities and the parameters in equation 2 were adjusted to eliminate visible contouring artefacts especially at low-*intensity* regions. A *c.v.i* plot can also be used to indirectly measure the effectiveness of the bit-depth allocation in L. Re-plotting the *c.v.i* with the PATF's modified co-factors showed that the bit-depth allocation was not optimal. Thus, a second round of optimization was performed on both the boundary values and co-factors to ensure optimal bit-depth allocation. The final co-factors of the PATF are given in Table I.

$a = 2285.712$	$b = 224.174$	$c = 5.000$
$d = -67.100$	$e = 263.500$	$f = -31.000$
$I'_s = 0.007$	$I'_p = 100.000$	$I'_h = 10000.000$
$L_s = 16.000$	$L_p = 496.000$	$L_h = 1023.000$

TABLE I: Co-factors used for the proposed PATF.

The interval boundaries of the final configuration, I'_s, I'_p and I'_h (as used in equation 2) represent the scaled *intensity* channel values where the HVS exhibits linear, power and logarithmic response [25]. Correspondingly, the *c.v.i* plot ensured that the JND space L was divided into three blocks with optimal bit-depth allocation within intervals where $L \in (0, L_s)$, $L \in [L_s, L_p)$ and $L \in [L_p, L_h]$ such that each block can facilitate a conservative quantization of low-, mid- and high-*intensity* regions. Also, when the proposed PTF is plotted with a semilog plot (I' vs. L) and compared with the existing PTFs, the shape of the curve shows the following characteristics:

- In the low-*intensity* regions, the curve exhibits a more conservative quantization than exhibited by the PTF based on modified Ferwarda's *t.v.i* while not as conservative as a logarithmic PTF.
- In the mid-*intensity* regions, the curve exhibits a similar quantization to the modified Ferwarda's *t.v.i*.
- In the high-*intensity* regions, the curve exhibits a conservative quantization similar to Barten's CSF based PTF.

Furthermore, the bit-depth allocation effectiveness was tested against existing EOTFs and found to be a close fit with the PQ algorithm [9]. The *c.v.i* plot is given Figure 5. For a further confirmation, the PATF and its inverse were rigorously tested as a part of the proposed algorithm. Results obtained from the evaluation demonstrate that the performance of the proposed algorithm using the PATF is better than the existing PTFs (see Figure 7).

D. Module 3: Error minimization function (EMF)

Similar to Module 2, this module extracts the chroma information (P & T channels) from IPT_{out} and performs a non-linear encoding which minimizes quantization errors frequently encountered in video compression. Typically, non-linear encoding is performed by a power function, say $\lambda < 1.0$

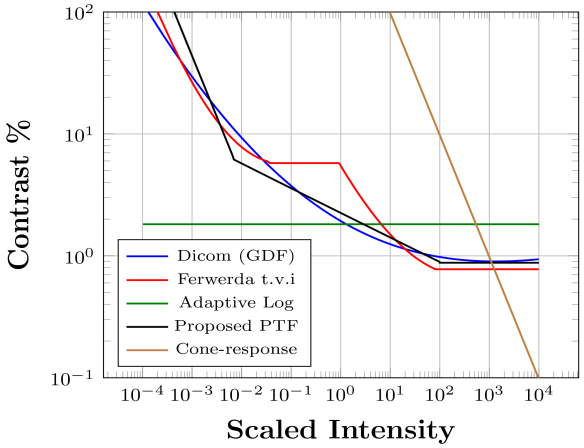


Fig. 5: Comparative Contrast vs. Intensity plot of the proposed PTF compared to existing PTFs and EOTFs used in other algorithms.

applied to the input values such that more bits are allocated to lower magnitudes where perceptual differences are more visible thus minimizing the quantization errors.

To encode chroma information, existing algorithms such as *hdrv* and *fraunhofer* encode the chroma channels using the procedure similar to LogLuv [6]. Although, the proposed algorithm uses the IPT color space which introduces a degree of non-linearity during conversion from the LMS cone excitation space to IPT, direct scaling and discretization of the chroma channels to 10-bit integer representation leads to rounding error based visible contouring artefacts. Therefore, we introduce a further non-linear encoding step to the P and T channels by deriving the most appropriate power value(s) which when applied to the chroma information minimizes the quantization errors during discretization.

The EMF is an optimization function which minimizes the difference between discretized floating point values such that $\mathrm{P}, \mathrm{T} \in [0, 1023]$ and their nearest integer calculated via a floor operation. The power value λ is derived as follows:

Let λ be the power value to be used for non-linear encoding, n be the targeted bit depth (10 in this case), $P_{inp} \in (0, 1]$ be the input channel and $P_{out} \in [0, 1023]$ be the output discretized channel. The application of the power function can be formulated as in equation 4.

$$P_{out} = \left\lfloor (P_{inp})^{\lambda} \cdot (2^n - 1) \right\rfloor \tag{4}$$

where the power function λ is computed via an optimization pass, where each step of the optimization replicates the quantization and de-quantization steps, evaluates different values of $\lambda \in (0, 1]$ such that the difference between 10-bit scaled floating point values and its nearest integer representation is minimal as shown in equation 5.

$$\underset{\lambda}{\operatorname{argmin}} \left(\frac{1}{MN} \sum_{j=1}^{N} \sum_{i=1}^{M} \left| \left(\frac{\lfloor (P_{inp})^{\lambda} \cdot (2^n - 1) \rfloor}{(2^n - 1)} \right)^{\frac{1}{\lambda}} - P_{inp} \right| \right) \tag{5}$$

where M and N represent the horizontal and vertical resolu-

tion, respectively. Upon application of the power values to the chroma channels, the λ values applied to each chroma channel is then stored as metadata and used later during reconstruction. In our implementation, the search for the optimal value of $\lambda \in [0, 1]$ was run in a straightforward uniform search with a step size of 0.02 which reduces the number of trials, albeit at the cost of minuscule precision loss. The optimization takes ≈ 1 second per frame using a single thread on an Intel Xeon 24 core workstation with a clock frequency of 2.6 GHz and 32 GB of RAM. Better optimization techniques, such as gradient descent, could be employed to speed up this process, and will be investigated in future work.

E. Module 4: Metadata information

As a result of frame normalization, intensity scaling, chroma scaling and non-linear encoding of chroma channels, the proposed algorithm produces metadata information containing the scaling information of the intensity channel, the minimum and maximum values of the chroma channels prior to scaling and finally the power values applied to each chroma channel. This data is then stored in the form of a look-up table (LUT) for each frame and the final LUT is stored as a secondary metadata stream. A visual description of the LUT along with additional information is given in the supplementary material.

V. Objective Evaluation of Compression Algorithms

The proposed algorithm was evaluated against four state-of-the-art compression algorithms (see Section II-A for details) using 39 HDR video sequences across a range of energy-difference, structural and perceptual QA metrics. Each sequence was encoded at 11 different quality levels (output bitrates) controlled by the quantization parameter (QP) of the codec. Here, we briefly discuss the evaluation methodology and the materials required to conduct the objective evaluation.

A. Materials

The materials used for this evaluation were the five compression algorithms including the four mentioned in Section II-A, the 39 HDR video sequences which represent a wide variety of scenes and overall dynamic range, seven QA metrics including the perceptual and structual QA metrics and the x265 [3] video codec.

B. Objective evaluation method

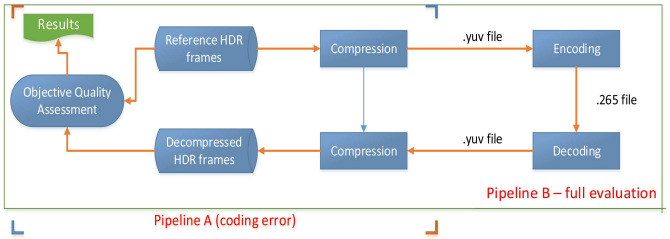


Fig. 6: Schematic diagram of the evaluation methodology. Pipeline 'A' is used to evaluate coding errors and Pipeline 'B' is used to evaluate rate-distortion characteristics.

The evaluation method can be classified into two parts. In *Pipeline A* (see Figure 6), the reference HDR frames from each of the 39 sequences were compressed using the five algorithms creating intermediate codec suitable files (labeled as HDRVs). Subsequently, the HDRVs are decompressed using the decompression part of the algorithms to reconstruct the HDR frames. The reference and reconstructed HDR frames are then evaluated using the objective QA metrics. In video compression, the results obtained by this procedure compute the coding errors produced by each algorithm which determines compression quality without the external influence of the codec.

Pipeline B, extends *Pipeline A* and introduces the x265 codec. The HDRVs are passed to the codec which encodes the frames into a raw video stream which is subsequently decoded and decompressed to reconstruct the HDR frames.

For a comprehensive evaluation at different quality levels, 150 frames from each of the 39 sequences were compressed using the five algorithms producing HDRVs which were subsequently 4:2:0 sub-sampled (typical for video compression) and then encoded at 11 different quality levels by controlling the QP of the codec. The encoding profile of the x265 video codec was *main444-10* and the QP values were set such that $QP \in [0, 5, 10, ..., 50]$, where $QP = 0$ represents lossless encoding and $QP = 50$ represents a highly lossy compression. The group of pictures (GOP) sequence was **I-B-B-B-P** with an intra-frame period of 30 and all sequences were encoded with a single-pass encoding scheme.

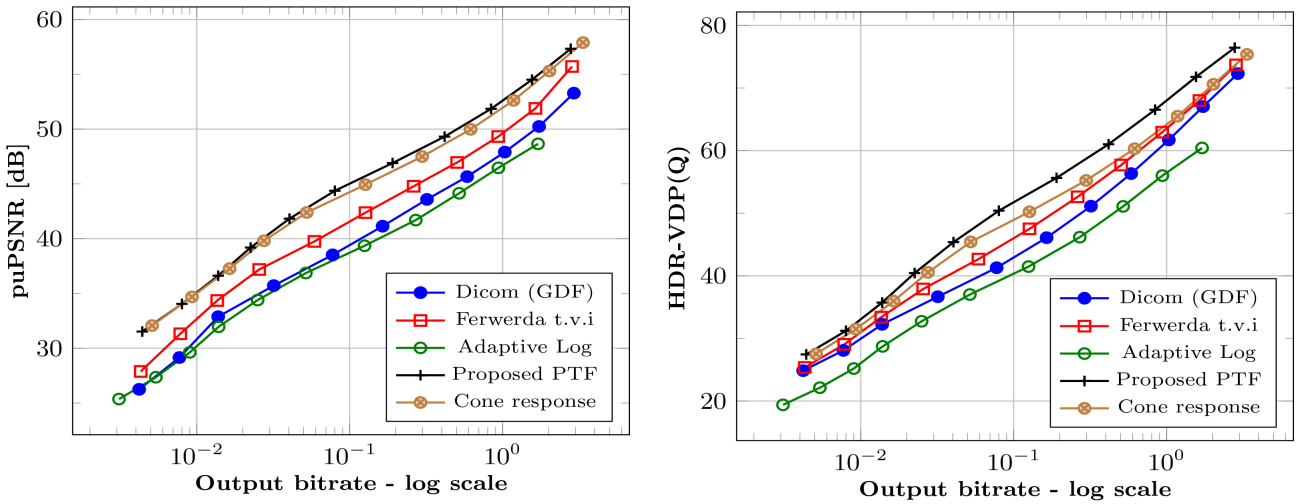
The reference and reconstructed HDR frames were evaluated against a set of QA metrics and results obtained were first averaged over the number of frames (per sequence) followed by a cumulative average over 39 sequences. The averaged results are then used to plot the mean Rate Distortion (RD) graphs which exhibit the overall performance of the algorithms. However, the mean RD graphs do not provide the complete picture since there is a significant variation in image reconstruction quality and bitrates required to encode the variety of scenes. Therefore, the obtained raw data was used to plot interpolated RD graphs of type a) variation in image quality (with 95% confidence interval) at fixed bitrates $\in [0.2, 2]$ bits/pixel (bpp) and b) variation in bitrates at fixed quality levels. A combination of mean and interpolated RD plots facilitates an in-depth understanding of the compression performance.

VI. Objective Evaluation Results

In this section, we present three sets of results obtained from the objective evaluation.

A. PTF evaluation results

The first set of results shown in Figure 7 demonstrate the overall HDR reconstruction quality of the proposed algorithm with the five PTFs (one at a time), evaluated against the set of 39 HDR sequences following the evaluation method mentioned in Section V-B - Pipeline B. The mean RD characteristics establish the overall superiority of the proposed PATF.



(a) Mean RD characteristics exhibited by each of the five PTFs: puPSNR - higher is better

(b) Mean RD characteristics exhibited by each of the five PTFs: HDR-VDP(Q) - higher is better

Fig. 7: Rate Distortion characteristics of proposed algorithm with five different PTFs - averaged over 39 sequences. This evaluation follows *Pipeline B* shown in Figure 6.

B. Overall objective evaluation results

With the performance of the PATF established, we now present the two sets of results which demonstrate the coding errors and the mean RD characteristics of the proposed algorithm when used in conjunction with the PATF and EMF. The coding error results presented in Table II and RD characteristics presented in Figure 8 demonstrate the comparative compression performance of the evaluated algorithms.

Metrics	Algorithms				
	hdrv	fraunhofer	pq	bbc-hlg	proposed
puPSNR [dB]	64.4542	70.5541	69.7689	58.3398	72.7814
puSSIM	0.9997	0.9999	0.9999	0.9982	1.0000
HDR-VDP(Q)	89.2398	91.9987	92.3235	82.0000	93.5157
HDR-VQM	0.9984	0.9964	0.9974	0.9957	0.9975

TABLE II: Coding error of five algorithms - averaged over 39 sequences. This evaluation follows *Pipeline A* mentioned in Figure 6.

Although several metrics were used, we present the most relevant results obtained from perceptual and structural metrics such as puPSNR [35], puSSIM [36], HDR-VDP [37] and HDR-VQM [38] as the predicted results have a high to very high correlation with subjective evaluation [1], [15], [39]. The remaining results including the interpolated RD characteristics for fixed quality (variation in bitrate) and fixed bitrate (variation in quality) are given in the supplementary materials.

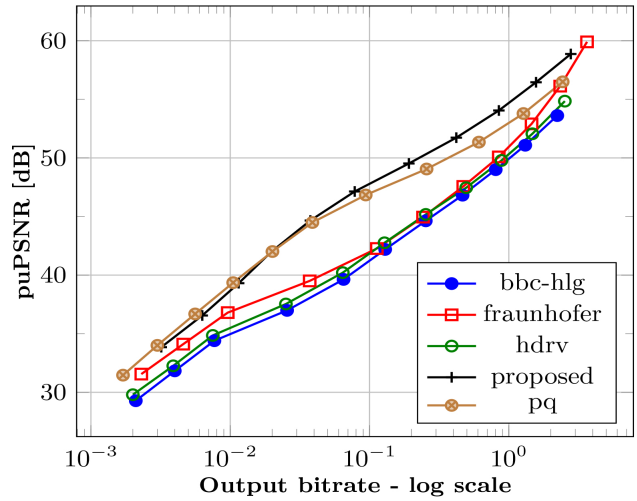
VII. Subjective Evaluation

In addition to the objective results, a subjective quality assessment experiment was conducted at two different quality levels. This section provides a brief outline of the subjective evaluation. **Design:** For the purpose of the subjective experiments, six sequences were short listed from the set of 39 sequences such that the sequences represent a variety of capturing techniques and a wide dynamic range. The selected sequences were compressed and decompressed using all five algorithms, including the proposed at the two designated quality levels. The design of the experiment involved participants seeing all six sequences for all algorithms. The ground truth was shown followed by a clip encoded and decoded by one of the candidate algorithms and the participants were asked to rate the clip based on its similarity with that of the ground truth. The first independent variable *algorithms* was composed of the five candidate algorithms: *hdrv*, *franhofer*, *pq*, *bbc-hlg* and the *proposed*. This was a within-participants independent variable as all participants had the opportunity to view all algorithms. The second independent variable *quality* was a between-participants variable and consisted of two conditions Q_1 and Q_2 corresponding to the quality setting of 0.09 bpp ($\approx$ 6 Mbps) and 0.24 bpp ($\approx$ 15 Mbps); the latter corresponds to the 10-bit settings recommended for online streaming and the former as lower quality than recommend 8-bit streaming chosen to stress test the algorithms. Six sequences used in the experiment were randomly displayed to the participants. The dependent variable *score* was a value from 1 to 10. Participants were asked to allocate *score* to each scenario computed for each algorithm for a total of 30 clips ranked per participant. The question asked was *"Rate the quality of decoded video sequences, on a scale of 1 to 10, based on their similarity to the uncompressed video sequence"*.

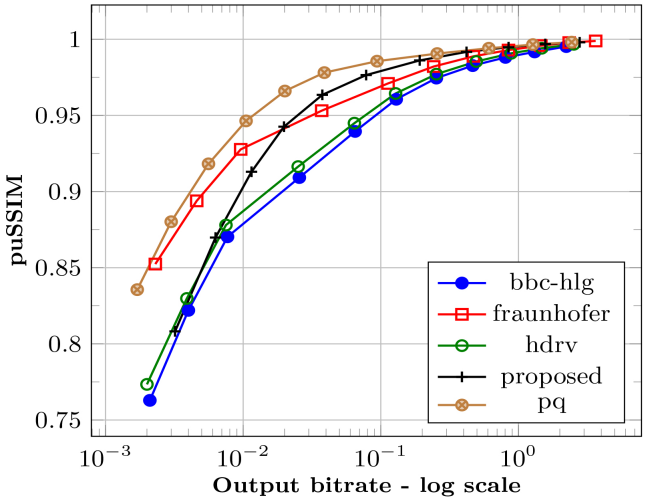
Materials: A SIM2 HDR display [31] with a peak luminance rating of 4,000 $\mathrm{cd/m^2}$ along with an LG 22" LED display with peak luminance rating of 300 $\mathrm{cd/m^2}$ and a computer with a solid state drive for quick loading of HDRVs was used for both experiments. All clips were prepared using a similar methodology to the objective experiments.

Participants: A total of 40 participants in two mutually exclusive groups of 20 participants (age $\in [22, 51]$), randomly allocated to one of the *quality* conditions volunteered for the experiments. All participants had normal or corrected to normal vision with no color blindness.

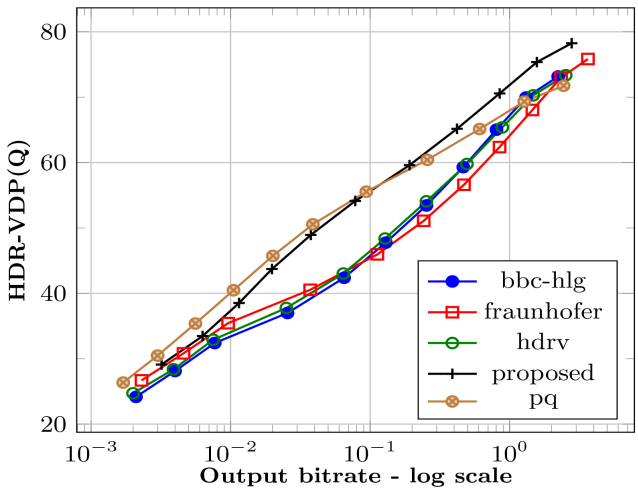
Analysis and Results: Inductive results were analyzed using a factorial 2 (*quality*) $\times$ 5 (*algorithms*) repeated measures ANOVA, with the scores averaged across all the scenarios for each algorithm. The main effect of *quality* was insignif-



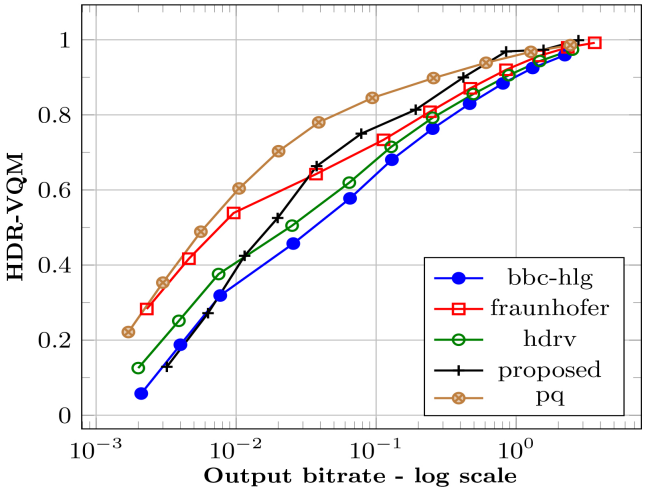
(a) Mean RD characteristics: puPSNR - higher is better



(b) Mean RD characteristics: puSSIM - higher is better



(c) Mean RD characteristics: HDR-VDP(Q) - higher is better



(d) Mean RD characteristics: HDR-VQM (RGB)

Fig. 8: Mean RD characteristics of the five algorithms averaged across 39 sequences. This evaluation follows *Pipeline B* mentioned in Figure 6.

icant F(1,38) = 0.005, p = 0.945, while the main effect of *algorithm* was significant, F(2.008, 76.318) = 108.394 , $p < 0.01$ (with Greenhouse Geiser corrections applied as Mauchly's test of Sphericity was violated $p < 0.05$). Since *quality* was insignificant, results were collapsed across *quality* for further pairwise comparisons. Table III demonstrates the results including means. The groupings in the results show groups of *algorithms* for which the scores were not considered significantly different, in particular the *proposed* and *pq* and similarly *pq* and *fraunhofer*. There was statistically significant difference between the *proposed* and *fraunhofer*. The last two algorithms e.g. *hdrv* and *bcc* form a group of their own indicating statistically significant difference. Kendall's co-efficient of Concordance W which provides a value agreement amongst the participants choices with 0 being complete disagreement and 1 complete agreement was relatively high at 0.68 and significant at $p < 0.01$.

In addition to the mean rankings given in Table III, the objective results of the six short-listed sequences at the specified bitrates were correlated with the subjective results using Pearson's correlation; see Table IV. As can be seen the correlation is quite strong and is significant at $p < 0.05$ for all results and demonstrates a strong relationship between the objective and subjective results. Further correlation results are given in the supplementary materials.

Averaged results over 6 sequences across 2 quality levels	Algorithms					Kendall's coefficient of concordance (W)
	proposed	pq	fraunhofer	hdrv	bbc-hlg	
	7.396	7.263	7.071	5.422	5.033	0.68

TABLE III: Subjective mean ranks with Kendall W, averaged across the two experiments.

	puPSNR	puSSIM	HDR-VDP(Q)	HDR-VQM	Subjective
puPSNR	-	$0.977^{\ddagger}$	$0.974^{\ddagger}$	$0.989^{\ddagger}$	$0.954^{\dagger}$
puSSIM	$0.977^{\ddagger}$	-	$0.919^{\dagger}$	$0.990^{\ddagger}$	$0.990^{\ddagger}$
HDR-VDP(Q)	$0.974^{\ddagger}$	$0.919^{\dagger}$	-	$0.960^{\ddagger}$	$0.910^{\dagger}$
HDR-VQM	$0.989^{\ddagger}$	$0.990^{\ddagger}$	$0.960^{\ddagger}$	-	$0.985^{\ddagger}$
Subjective	$0.954^{\dagger}$	$0.990^{\ddagger}$	$0.910^{\dagger}$	$0.985^{\ddagger}$	-

TABLE IV: Pearson's correlation between objective and subjective results. † denotes significance at $p < 0.05$ level and ‡ denotes significance at $p < 0.01$ level. Further non-parametric correlation results using Kendall's Tau and Spearman's correlation coefficient are given in the supplementary materials.

VIII. Discussion

In this section, we combine the objective and subjective results and analyze the overall performance of the proposed algorithm against the state-of-the-art solutions.

The mean results in Figure 7 suggest that the PATF facilitates superior image reconstruction than existing PTFs. Amongst the established PTFs, GCRM and Ferwara's *t.v.i* based PTF produces similar encoding results while the logarithmic TF [8] exhibits the least desired performance. The superior performance of the PATF can be attributed to the balanced quantization of scaled *intensity* values where the PATF facilitates a finer quantization in the darker regions as well as maintaining the conservative quantization for high-*intensity* regions.

The overall performance of the proposed algorithm as shown in Table II demonstrate that the proposed algorithm (with the PATF) exhibits the smallest coding error amongst the five algorithms as evaluated by structural and perceptual QA metrics. While *bbc-hlg* demonstrates the maximum coding error, the coding error performance of *fraunhofer* and *pq* is quite similar to the proposed algorithm.

The mean RD characteristics of puPSNR and HDR-VDP in Figure 8 indicate that overall the proposed algorithm outperforms existing solutions at $\text{bitrates} \geq 0.4$ bpp. However, it can also be observed from Figure 8 that the proposed algorithm sometimes fails to outperform existing algorithms such as *pq* and *fraunhofer* at low bitrates i.e. $\text{bitrates} < 0.4$ bpp. This is primarily because the algorithm relies on an analytical solution for luminance compression as opposed to a smooth non-linear PTF used by *pq* and *fraunhofer*.

Also, Figure 8c shows that overall both *pq* and *bbc-hlg* perform marginally better at $\text{bitrates} < 0.4$ bpp. The puSSIM results in Figure 8b demonstrate that the structural reconstruction of both *pq* and the proposed algorithm are very similar to each other. Overall, the objective results indicate that the performance of the proposed algorithm is at par or better than existing solutions.

Finally, the subjective evaluation mean ranks given in Table III indicate that the proposed algorithm performs slightly better than existing solutions. Moreover, the correlation results given in Table IV indicate a high to very high correlation between the perceptual/structural QA metrics and subjective results. The correlation within the objective metrics is also quite strong. Furthermore, the correlation results in Table IV are analogous to the objective results presented in Figure 8.

IX. Limitations

A limitation of the proposed algorithm is that the normalization factor mentioned in Algorithm 1 is the maximum pixel value of the HDR frame; this may cause issues with a single aberrant high luminance pixel. The tested scenes did not feature any such cases. However, since the algorithm is intended for generic usage, an alternative normalization factor can be computed by considering the top select percentage, for example top 5%, of the frame values or the median value of the frame luminance.

Another limitation of this study is that all five algorithms have been modified/optimized for 10-bit encoding, sufficient for a luminance range of $\mathrm{L} \in [10^{-4}, 10^{4}]$ $\mathrm{cd/m^2}$. Although this does not significantly degrade the output quality, a 12-bit optimization can accommodate a much larger range of $\mathrm{L} \in [10^{-4}, 10^{8}]$ $\mathrm{cd/m^2}$. The only modification required to the proposed algorithm, in this case, is the re-computation of the PATF cofactors by changing the luminance and luma boundaries, respectively.

X. Conclusion and Future Work

In this paper, we have proposed a novel HDR video compression algorithm optimized for 10-bit encoding. The algorithm uses the IPT uniform color opponent space, a novel PATF providing better image reconstruction than existing PTFs and a novel EMF for accurate chroma preservation. Objective evaluation results demonstrate that the proposed algorithm performs at par or better than the four published and/or patented state-of-the-art algorithms. The proposed algorithm can also use existing PTFs albeit at the cost of reconstruction quality.

Future work also includes further refinement of the PATF and EMF to further reduce quantization errors. Moreover, the use of other color spaces such as $L\alpha\beta$ for the color space transform phase can be also be explored as well as adapting curves per scenario [40]. Furthermore, the algorithm can also be further evaluated across other HDR video sequences especially those processed with the BT.2020 primaries [41] at higher bit-depths such as 12- and 14-bits/pixel/channel when optimized codec support becomes available. Finally, the objective and subjective evaluations can also be extended using two recently developed quality assessment metrics using eye-tracking data such as the no-reference metric proposed in [42] and spatiotemporal correlation data using eye tracking as proposed in [43]. However, their suitability for HDR video content also needs to be verified before they can be used to evaluate the HDR video compression algorithms.

XI. Acknowledgement

The authors would like to thank Technicolor SA for the "Seine" footage, University of Tübingen and Stuttgart Media University for providing with some the HDR video sequences [44] used in this work. This work is partially funded by *EPSRC EP/K014056/1* with Jaguar Land Rover: "PSi Theme 7: Visualization and Virtual Experience". Debattista was partially supported by a Royal Society Industrial Fellow (IF130053), for whose support we are most grateful.

References

[1] R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, M. Bessa, B. Waterfield, and A. Chalmers, "Objective and subjective evaluation of high dynamic range video compression," *Signal Processing: Image Communication*, vol. 47, pp. 426–437, 2016.

[2] F. Ebner and M. D. Fairchild, "Development and testing of a color space (ipt) with improved hue uniformity," in *Color and Imaging Conference*, vol. 1998, pp. 8–13, Society for Imaging Science and Technology, 1998.

[3] V. Organization", "The x265 video codec." https://www.videolan.org/developers/x265.html.

[4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[5] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Perception-motivated high dynamic range video encoding," in *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, (New York, NY, USA), pp. 733–741, ACM, 2004.

[6] G. W. Larson, "Logluv encoding for full-gamut, high-dynamic range images," *Journal of Graphics Tools*, vol. 3, no. 1, pp. 15–31, 1998.

[7] J.-U. Garbas and 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*, pp. 829–832, May 2011.

[8] A. Motra and H. Thoma, "An adaptive logluv transform for high dynamic range video compression," in *2010 IEEE International Conference on Image Processing*, pp. 2061–2064, Sept 2010.

[9] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," *SMPTE Motion Imaging Journal*, vol. 122, pp. 52–59, May 2013.

[10] F. Dufaux, P. Le Callet, R. Mantiuk, and M. Mrak, *High Dynamic Range Video: From Acquisition, to Display and Applications*. Academic Press, 2016.

[11] P. G. Barten, "Physical model for the contrast sensitivity of the human eye," in *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pp. 57–72, International Society for Optics and Photonics, 1992.

[12] T. Borer and A. Cotton, "A "display independent" high dynamic range television system," September 2015.

[13] S. S. Stevens, "On the psychophysical law.," *Psychological review*, vol. 64, no. 3, p. 153, 1957.

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

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

[16] A. Banitalebi-Dehkordi, M. Azimi, M. T. Pourazad, and 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*, pp. 8–12, Aug 2014.

[17] 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)*, no. EPFL-CONF-203874, 2015.

[18] M. Narwaria, M. Perreira Da Silva, and P. Le Callet, "Study of high dynamic range video quality assessment," vol. 9599, pp. 95990V–95990V–13, 2015.

[19] E. Reinhard, E. A. Khan, A. O. Akyz, and G. M. Johnson, *Color imaging: fundamentals and applications*. AK Peters, Ltd., 2008.

[20] M. D. Fairchild, *Color appearance models*. John Wiley & Sons, 2013.

[21] F. Ebner and M. D. Fairchild, "Finding constant hue surfaces in color space," in *Photonics West'98 Electronic Imaging*, pp. 107–117, International Society for Optics and Photonics, 1998.

[22] F. Ebner and M. D. Fairchild, "Development and testing of a color space (ipt) with improved hue uniformity," in *Color and Imaging Conference*, vol. 1998, pp. 8–13, Society for Imaging Science and Technology, 1998.

[23] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560–576, 2003.

[24] O. M. Blackwell and H. R. Blackwell, "Ieri: Visual performance data for 156 normal observers of various ages," *Journal of the Illuminating Engineering Society*, vol. 1, no. 1, pp. 3–13, 1971.

[25] M. I. Sezan, K.-L. Yip, and S. J. Daly, "Uniform perceptual quantization: Applications to digital radiography," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 17, no. 4, pp. 622–634, 1987.

[26] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Lossy compression of high dynamic range images and video," in *Electronic Imaging 2006*, pp. 60570V–60570V, International Society for Optics and Photonics, 2006.

[27] M. Mustra, K. Delac, and M. Grgic, "Overview of the DICOM standard," in *ELMAR, 2008. 50th International Symposium*, vol. 1, pp. 39–44, Sept 2008.

[28] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A model of visual adaptation for realistic image synthesis," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 249–258, ACM, 1996.

[29] G. Ward, "Defining dynamic range," in *ACM SIGGRAPH 2008 Classes*, SIGGRAPH '08, (New York, NY, USA), pp. 30:1–30:3, ACM, 2008.

[30] R. A. Normann, B. S. Baxter, H. Ravindra, and P. J. Anderton, "Photoreceptor contributions to contrast sensitivity: applications in radiological diagnosis," *IEEE transactions on systems, man, and cybernetics*, no. 5, pp. 944–953, 1983.

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

[32] Y. Nayatani, K. Hashimoto, K. Takahama, and H. Sobagaki, "A nonlinear color-appearance model using estévez-hunt-pointer primaries," *Color Research & Application*, vol. 12, no. 5, pp. 231–242, 1987.

[33] B. Series", *Reference electro-optical transfer function for flat panel displays used in HDTV studio production*. Radiocommunication sector of ITU, 03 2011.

[34] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pp. 2–15, International Society for Optics and Photonics, 1992.

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

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.

[37] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and 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*, vol. 24, no. 1, p. 010501, 2015.

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

[39] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. G. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for hdr image quality assessment," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 39, 2015.

[40] K. Debattista, "Application-specific tone mapping via genetic programming," in *Computer Graphics Forum*, vol. 37, pp. 439–450, Wiley Online Library, 2018.

[41] I.-R. Recommendation, "BT-2020 "parameter values for ultra-high definition television systems for production and international programme exchange"," *International Telecommunication Union, Geneva*, 2012.

[42] P. K. Podder, M. Paul, and M. Murshed, "Qmet: A new quality assessment metric for no-reference video coding by using human eye traversal," in *2016 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1–6, Nov 2016.

[43] P. K. Podder, M. Paul, and M. Murshed, "A novel quality metric using spatiotemporal correlational data of human eye maneuver," in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8, Nov 2017.

[44] 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," vol. 9023, pp. 9023 – 9023 – 10, 2014.