



HDR image encoding using a companding-based nonlinear quantization approach without metadata

Asif Siddiq¹ · Jameel Ahmed¹ · Monagi H. Alkinani² · Ishtiaq Rasool Khan²

Received: 23 September 2021 / Revised: 20 January 2022 / Accepted: 21 January 2022 / Published online: 22 February 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

High dynamic range (HDR) images contain more details of dark and bright regions, which cannot be captured by the standard cameras. Raw HDR data is represented with floating-point precision, and the commonly used lossless/near-lossless HDR image encoding formats produce files that require large storage and are not suitable for transmission. Methods have been proposed for lossy encoding using single- and dual-layer structures, but the codecs are generally more complex and often require additional metadata for decoding. We propose a dual-layer codec in which companding, which is a closed-form transformation and hence can be reversed without any additional data, is used to generate the first layer. The residual data stored in the second layer is quantized linearly and do not need metadata either for decoding. The proposed codec is computationally light, has higher accuracy, and produces smaller size files compared to the existing codecs. This has been validated through an extensive evaluation using different metrics, and the results are reported in the paper.

Keywords HDR image reconstruction · Image compression · Lossy non-backward compatible coding · Companding · Image quality assessment

1 Introduction

The paradigm shift from a conventional low dynamic range (LDR) to high dynamic range (HDR) imaging for better viewer experience came with new challenges and has opened new avenues of research. These challenges include visualization of extended dynamic range, storage of high-precision data, streaming the content of large size over limited bandwidth, and gauging the performance of proposed solutions. The way raw HDR data is represented in the floating-point format, compared to the 8-bit LDR data, poses an entirely different realm of challenge in encoding. Different near-lossless formats with little or no compression have been proposed to encode raw HDR data [1–3]. The conventional LDR displays designed for 8 bits per channel cannot display the HDR content encoded in the raw formats. In other words, the raw formats are non-backward compatible (NBC), and the HDR

contents encoded in them require tone-mapping for viewing on the existing displays. Other formats that encode HDR content to render on LDR displays without additional processing are termed as backward compatible (BC) formats [4–6].

Initially, when HDR displays were few, the rationale behind developing BC formats was strongly justifiable. With HDR displays making their way in the consumer market and HDR contents becoming more prevalent, the requirement of backward compatibility started appearing as a bottleneck to achieving higher coding efficiency. Therefore, the trend of developing more efficient non-backward compatible (NBC) formats started gaining popularity [7, 8]. Both BC and NBC formats can use single- or dual-layer structures; however, traditionally, BC formats used two layers, and NBC formats used one layer to encode the whole dynamic range. In dual-layer structures, the first layer is the base layer (BL), and the second layer is generally termed as the enhancement layer (EL). For BC formats, BL contains a tone-mapped version of the HDR image, which can be directly displayed on an LDR display, while the EL contains the residual information necessary to reconstruct the HDR image. In general, some additional information is also needed for reconstruction and is stored as metadata besides the two layers. The single-layer BC and NBC formats only have the BL and the metadata.

✉ Asif Siddiq
3195@students.rii.edu.pk

¹ Department of Electrical Engineering, Riphah International University, Islamabad 44000, Pakistan

² College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia

Metadata hold a pivotal role in reconstructing HDR content regardless of BC or NBC format and can be embedded in the first layer or encoded separately. In any case, extracting and handling the metadata and using it for prediction add to the decoder's design complexity.

This paper proposes a dual-layer NBC format in which BL and EL reconstruct HDR images at the decoder end without requiring any metadata. The proposed algorithm is based on the companding technique that uses a continuous mapping function in closed form to reshape the content while encoding and decoding. This means that no metadata is required to reconstruct the HDR image, which distinguishes the proposed solution from other two-layer structures and makes the decoding process very efficient. The proposed format shows promising results with a size reduction of more than 50% compared to the images in the raw formats. Moreover, the proposed decoding algorithm can faithfully reconstruct the HDR images and require lesser time than the existing dual-layer decoders. We have carried out extensive evaluations using several metrics to validate the accuracy of the encoding and speed of decoding operations for the proposed format.

The rest of the paper is organized as follows: Section 2 describes the related work. Section 3 describes the structure of the proposed format. A brief description of the forward and inverse companding functions and a detailed explanation of different modules of the encoder and decoder design are given in this section. The results and detailed analysis of the experimental evaluations are presented in Sect. 4 followed by some conclusions in Sect. 5.

2 Related work

Raw floating-point HDR data cannot be stored in the traditional LDR image formats and require new encoding techniques. Different near-lossless formats with little or no compression have been proposed to encode raw HDR data. A 96-bit per-pixel TIFF [1] can encode uncompressed data with 32-bit precision for each channel. However, even though TIFF is very accurate and easy to read and write, the file size becomes very large. Industrial Light and Magic (ILM) introduced OpenEXR format [2] that encodes data as 16 bits per channel in "half" floating-point type reserving one sign bit, 10 bits for the mantissa, and 5 bits for the exponent. OpenEXR supports 72-bit and 96-bit formats too. Radiance picture format known as RGBE [3] encodes three floating-point channels of the HDR data into four 8-bit integer channels. The fourth channel is the common exponent of the rest three. These formats target encoding the entire dynamic range of the raw data in a nearly lossless way, leaving less room for compression.

The development of BC formats was led by the large predominance of LDR displays in the consumer market. A

dual-layer JPEG-XT format [9] standardized some backward compatible structures for HDR images by extending the JPEG format. The HDR image is preprocessed, generating two layers, the LDR layer and the extension layer, which are quantized using discrete cosine transform (DCT) and encoded using 8-bit JPEG. To improve the bit precision of the legacy 8-bit codec, JPEG-XT defines two alternate mechanisms: (i) refinement coding and (ii) residual coding, and both can also be combined. Kobayashi et al. [10] propose a two-layer near-lossless BC method using an extended histogram packing technique. The base layer is identical to JPEG-XT except for the refinement scan, which is not performed. In the extended/residual layer, the histogram of each color is quantized using zero-skip quantization. Mantiuk et al. [11] proposed a dual-layer BC design in which HDR movies are distributed over DVDs. The encoder takes two inputs, the HDR frames, and their tone-mapped LDR versions. LDR frames are encoded using standard MPEG in the BL, forming the LDR stream. LDR and HDR frames are transformed into a custom color space to minimize the difference between them. The reconstruction function predicts the HDR content, and a residue frame is generated by computing the difference between the original and the predicted HDR frames. After undergoing a filtering step to improve compression, the residue frames are encoded using the standard MPEG generating the residue stream. Quantization factors along with the reconstruction function are compressed losslessly and stored in the auxiliary stream. The authors reported a 30% storage overhead in encoding HDR data with respect to the LDR encoding. Wang et al. [12] proposed a dual-layer BC compression technique based on visual saliency. The saliency map was extracted from the tone-mapped image (base layer) of the HDR image. This map was used to adaptively tune the residual image in the extension layer. Some other dual-layer BC structures using slightly different algorithms for generating BL and EL can be found in [13, 14].

Single-layer BC formats consist of only the BL and the metadata. As in the dual-layer BC formats, the BL in single-layer BC formats is a tone-mapped version of the HDR image, which can be decoded and shown on LDR displays. BL, along with the metadata, can be used to recover the HDR data. Lasserre et al. [15] proposed a single-layer BC system in which the input HDR frame is converted from RGB to Y'CbCr 4:2:0, and an LDR frame is generated using a tone-mapping function. Two lookup tables, one for luminance and chrominance each, are generated as the dynamic metadata. LDR image along with the dynamic metadata is used to generate the HDR image at the decoder. Goris et al. [7] proposed a parameter-based single-layer BC system wherein parameters to reconstruct LDR and HDR are embedded in the bitstream. LDR frames are generated by assigning weights ' ω ' to HDR frames. At the receiver, LDR frames along with the metadata ' ω ' are used to reconstruct the HDR frames.

In BC formats mentioned earlier, the tone-mapping process in generating the BL can be seen essentially as a quantization process to convert the HDR values of floating-point precision to 8-bit integers. If the condition of viewing the BL on the LDR displays is dropped, better quantization methods can be designed to encode more information in the BL. Restriction of backward compatibility also increases the storage size, as was noted by Mantiuk et al. [11]. The flexibility of picking the best quantization method for producing the BL enables dual-layer NBC formats to store more information than their dual-layer BC counterparts. On the downside, these formats cannot be displayed on LDR screens without tone-mapping the decoded content. However, compared to the raw formats mentioned above [1–3], dual-layer NBC formats have two notable advantages. Firstly, they achieve a significant reduction in file sizes without any considerable information loss, as shown later in this paper. Secondly, they can be encoded utilizing the existing codecs like PNG or JPEG. In other words, they are not compatible with the existing displays, but they are still compatible with the existing codecs and applications that do not require visualization. Therefore, some dual-layer structures dropped the condition of compatibility with LDR displays. The dual-layer NBC formats provide more flexibility as the BL is not optimized for display; instead, the goal is changed to encode as much information as possible in the BL to leave less residual information for the EL to improve the overall performance of the codec.

In a dual-layer NBC system proposed by Su et al. [8], the input HDR image is converted to 4:2:0 YC_bC_r color space and quantized to 8 bits. The quantized data is encoded in a lossy format and forms the BL. A predictor is used to approximate the HDR luminance channel, which, along with the original luminance channel, is used for residual generation. The residual data's dynamic range can be large, so a nonlinear quantization (NLQ) algorithm is used to encode more information in the EL. Parameters used for prediction and NLQ are encoded and used at the decoder to reconstruct the HDR image. Dufaux et al. [16] proposed signal splitting and recombination at the coder and decoder sides, respectively. The input image is split into MSB and LSB, forming two layers for HDR transmission. A preprocessing step is used in [17, 18], where the input HDR image is split into two images, a monochromatic modulation picture and a residual picture using a nonlinear mapping function. The modulation picture consists of relatively low-frequency components of the input image, and a residual picture represents the remaining relatively high-frequency components. The residual picture can be used for backward compatibility or to improve compression. For the latter case representing an NBC scenario, the split images, after going through a number of color space and signal format changes, are encoded and transmitted to the decoder. At the decoder end, inverse transformation and

recombination of signals from the two layers reconstruct the HDR image.

In addition to the formats mentioned above that give main consideration to encoding accuracy, there are other formats that primarily focus on efficient end-to-end delivery and display of HDR content. HDR10 and Dolby Vision (DV) are two leading structures in this category. Both HDR10 and DV use the perceptual quantizer (PQ) to quantize the HDR content. HDR10 employs PQ10, which uses 10 bits per pixel for quantization and static metadata [19]. DV uses PQ10 or PQ12 (using 12 bits for quantization) and dynamic metadata [20]. The metadata in HDR10 and Dolby Vision is contained in the supplemental enhancement information (SEI) messages of the High-Efficiency Video Coding (HEVC) format [21]. Philips has designed its own HDR delivery system using a single layer along with metadata [7]. It utilizes the 10-bit HEVC codec, along with Philips HDR EOTF and display tuning, to optimize the image's peak luminance. Technicolor HDR supports HDR content production and delivery to HDR and legacy SDR displays [15]. The technologies and standards used in this workflow facilitate an open approach, including a single-layer LDR/HDR HEVC encoding, the MPEG standardized Color Remapping Information metadata (CRI) for HDR to LDR conversion, a Parameterized Electro-Optical Transfer Function (P-EOTF), and SHVC. Hybrid Log-Gamma (HLG) is another transfer function developed for camera capture (OETF) jointly developed by BBC and NHK Japan specified in ARIB STD-B67 [22]. The HLG curve maintains compatibility with the LDR displays using the standard gamma curve for the lower luminance values and the log curve for the higher luminance values.

Raw HDR formats discussed in this section require huge storage space and are not suitable for transmitting the content due to increased bandwidth requirement and cost. On the other hand, the existing BC and NBC formats are computationally complex and require metadata for reconstruction at the decoder end. The proposed format explained in the next section attempts to address these issues.

3 The proposed dual-layer codec

Quantization is inherently a lossy process and induces an error in the data known as the quantization error. It is easy to show that in uniform quantization methods where quantization intervals are equal in width, the percentage quantization errors for smaller values are more than the percentage quantization errors for larger values. A nonlinear quantization method can overcome this issue, as the quantization intervals can be optimized to reduce the loss. A dual-layer BC HDR encoding scheme proposed by Khan [23] utilizing nonlinear quantization in the EL outperformed the dual-layer format proposed by Mantiuk et al. [11], which uses uniform

quantization. However, the metadata generated for Khan [23] increased the complexity of the decoder. Companding, a non-linear transformation discussed below in detail, becomes a good candidate for use in the quantization process because of its simplicity. Another desirable feature of companding is that it does not require any metadata for inversion; thus, the complexity of both reshaping and reconstruction processes can be reduced.

Companding refers to a technique of compressing and then expanding a signal. In a typical streaming scenario, the compression function reduces the dynamic range of the transmitted signal by raising the amplitude of the weak signals, and the expansion function recovers the dynamic range at the decoder by lowering the amplitude of the raised signals. Companding has been standardized by ITU-T G.711 [24]. The compression and expansion are carried out with nonlinear log functions as given by Eqs. (1) and (2), respectively.

$$y = F(x) = \text{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)}, -1 \leq x \leq 1 \quad (1)$$

$$x = F^{-1}(y) = \text{sgn}(y) \frac{(1 + \mu)^{|y|} - 1}{\mu}, -1 \leq y \leq 1 \quad (2)$$

where μ is a constant that determines the slope of the transformation curves, \ln refers to the natural log, y is the compressed signal, and x is the expanded signal. It can be noted that Eqs. (1) and (2) require the signals x and y to be in the $[-1, 1]$ range. HDR images have very different ranges and therefore, the original companding algorithm cannot be utilized. We adapt the algorithm to cater to any dynamic ranges and validate its effectiveness in encoding HDR images. HDR images capture better details of shadows, which are lost due to the limitations of traditional imaging techniques and the quantization process. Our modified companding algorithm can successfully enhance those subtle details before quantization and restore them back while decoding. Figure 1 explains the process using a typical HDR image. The image on the left shows the original and the compressed pixel values, sorted in the ascending order. The y-axis is shown in log scale to highlight the difference. The compressed signal significantly amplifies the low amplitude pixels, which would lead to minimization of quantization loss. Enhancement of smaller values also leads to reduction in the dynamic range of the compressed data (one order of magnitude on average in our experiments), which allows a better precision during the quantization process. These two factors help the expansion process to faithfully reconstruct HDR values, as seen in Fig. 1.

Figure 2 shows the design of the proposed encoder used to produce the base and enhancement layers. The input image is read from a raw format (such as RGBE or OpenEXR) in RGB channels in block-1 and converted into $YCbCr$ color space in block-2 using Eq. (3) in accordance with ITU-R recommendation BT-709.

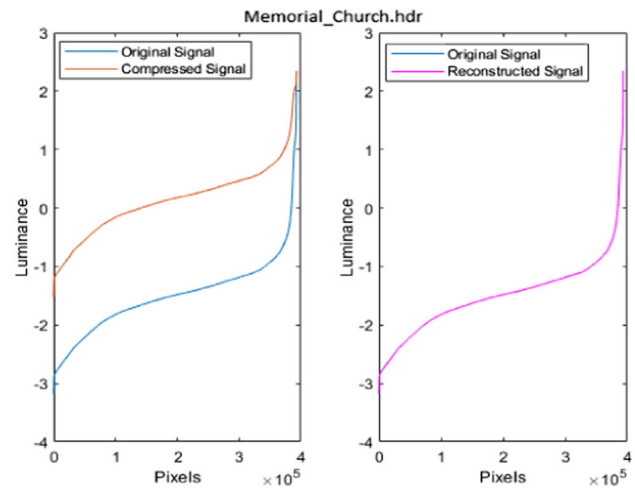


Fig. 1 Compression (left) and expanding (right) phases of companding for a typical image

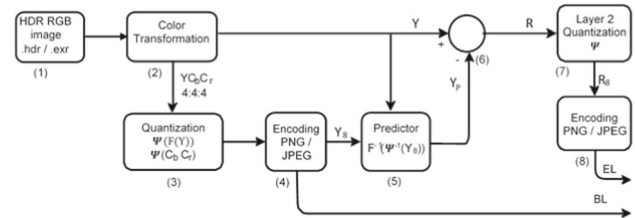


Fig. 2 Structural diagram of the proposed encoder

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = [A] \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3)$$

where

$$[A] = \begin{bmatrix} 0.2126 & 0.7152 & 0.0722 \\ -0.09991 & -0.33609 & 0.436 \\ 0.615 & -0.55861 & -0.05639 \end{bmatrix}$$

The Y channel from block-2 is quantized nonlinearly according to Eq. (1), while the chroma channels are quantized linearly. Since the BL is not meant for viewing, this combination of nonlinear–linear quantization can be utilized at the BL.

Mathematically the process of generating the BL can be expressed as:

$$Y_8 = \Psi(F(Y)) \quad (4)$$

$$C_8 = \Psi(C) \quad (5)$$

where F is defined in Eq. (1), Ψ represents 256-level linear quantization process, the subscript 8 refers to 8-bit data, and C represents the C_b and C_r channels. The quantized output

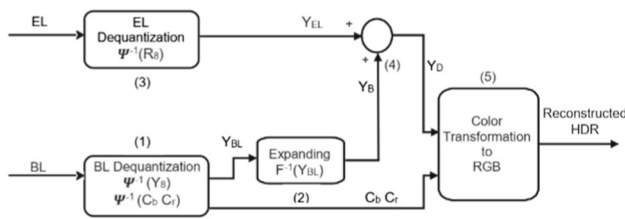


Fig. 3 Structural diagram of the decoder

contains integer values therefore any existing lossy or lossless coding format (such as JPEG or PNG) can be utilized in block-4. The output of block-4 forms the BL for transmission to the decoder.

Block-5 de-quantizes the 8-bit channel Y_8 by linearly mapping to the original range of HDR luminance. It then expands it using Eq. (2) predicting Y_P , which is subtracted from the original Y channel in block-6 to generate the residue luminance R. Note that the residue is only calculated for the Y channel. The residue consists of floating-point data, which is linearly quantized at block-7 and encoded to generate EL at block-8. The residue generation and quantization can be mathematically expressed as:

$$Y_P = F^{-1}(\Psi^{-1}(Y_8)) \tag{6}$$

$$R = Y - Y_P \tag{7}$$

$$R_8 = \Psi(R) \tag{8}$$

where F^{-1} is defined in Eq. (2).

Figure 3 shows the structure of the decoder. The BL from the coder is decoded linearly in block-1. Equations (9) and (10) describe the process:

$$Y_{BL} = \Psi^{-1}(Y_8) \tag{9}$$

$$C = \Psi^{-1}(C_8) \tag{10}$$

where Y_8 , C_8 , and C are defined in Eqs. (4) and (5). The Y_{BL} channel from block-1 is expanded using Eq. (2), generating Y_B signal. The decoded signal Y_{EL} from EL in block-3 is then added to Y_B in block-4 to generate the final luminance channel Y_D at the decoder. Equations (11) to (13) describe the process:

$$Y_{EL} = \Psi^{-1}(R_8) \tag{11}$$

$$Y_B = F^{-1}(Y_{BL}) \tag{12}$$

$$Y_D = Y_{EL} + Y_B \tag{13}$$

where R_8 is defined in Eq. (7). Y_D from block-4 and C_b and C_r from block-1 are transformed from $Y C_b C_r$ to RGB color

space in block-5 using Eq. (14) to generate the reconstructed HDR image as

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = [A]^{-1} \begin{bmatrix} Y_D \\ C_b \\ C_r \end{bmatrix} \tag{14}$$

where $[A]$ is defined in Eq. (3).

4 Results and discussion

4.1 Methods and the datasets

To evaluate the performance of the proposed codec, we carried out comparative studies using dual-layer codecs based on Ward et al. [25], Mantiuk et al. [11], Khan [23], and the proposed structure. We provide some comparisons with the recent single-layer HDR delivery format, HDR10, as well. Three datasets, Funt dataset of 105 images [26], Pfstools dataset of 8 images [27], and a mixed dataset of 22 images collected randomly from various public sources were used as the ground truth. Funt dataset and Pfstools datasets have less variance in the size of images, while the mixed dataset is a collection of images with more diversity in size.

In implementing the proposed method, we used some general best practices on chroma channels to reduce the overall error. We applied the same to the codes of other methods to make a fair comparison and observe the effect of the proposed companding-based quantization alone. For the same reason, in most of the evaluations, both layers were encoded losslessly for all the methods using PNG codec to exclude the impact of coding loss. Even the layers of Ward et al. [25] were encoded as PNG, although the original method (referred to as JPEG-HDR) and its JPEG-XT implementation in profile A use JPEG. However, in one set of experiments, we used JPEG encoding at different levels of compressions to plot rate-distortion (RD) curves to observe the performance at lower bitrates. Here it is worth mentioning that the original codes by the authors obtained lower scores than our implementations of these methods. For a fair comparison and to observe only the effect of the algorithm used to generate the BL, we encoded HDR images in $Y C_b C_r$ 4:4:4 format in the implementation of all four methods.

4.2 Comparison using objective metrics

We use a large set of objective metrics, including traditional and some metrics designed specifically for HDR images. These metrics measure the difference between the ground-truth images and the HDR images reconstructed at the decoder. Figure 4 shows the average results of these metrics for the three datasets. The star on the graphs shows the

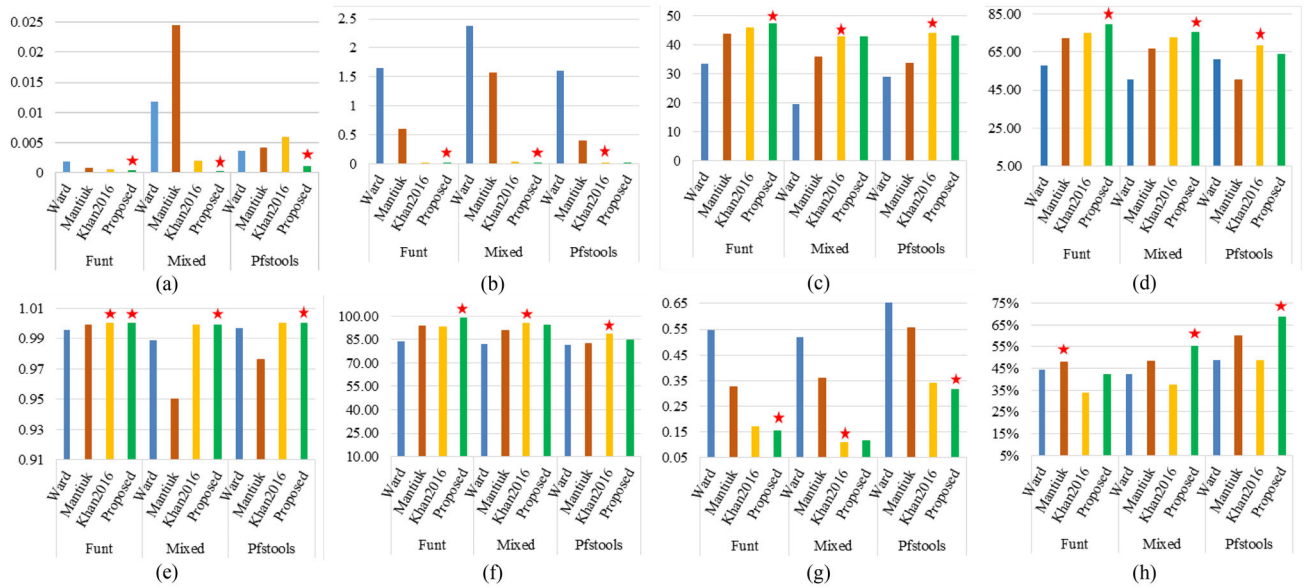


Fig. 4 Objective evaluations using three test datasets and metrics: **a** MSE, **b** MSE-PU, **c** mPSNR, **d** wPSNR, **e** PU_SSIM, **f** HDR-VDP2, **g** DE2000, and **h** reduction in size. Lower values of MSE, MSE-PU,

and DE2000 and higher values of mPSNR, wPSNR, SSIM_PU, and HDR-VDP2 mean better performance. The winner is indicated by a star in each experiment

best-performing codec. Details of each comparison and the results are explained below.

- i) Mean square error (MSE): Fig. 4a shows an 80% to 95% reduction in the MSE for the proposed codec when compared with the other three codecs.
- ii) Perceptually uniform mean square error (MSE-PU): MSE-PU is designed specifically for measuring the difference between HDR images. The proposed algorithm performed the best for this metric and showed an improvement of 69% in score on average for all the three datasets from the second-best Khan [23], as shown in Fig. 4b. The other two codecs have a very high MSE-PU.
- iii) Multi-Exposure Peak-Signal-to-Noise Ratio (mPSNR): mPSNR is another metric designed for HDR images, which takes multiple exposures to calculate the peak-signal-to-noise ratio [28]. Except for the Pfstools dataset, where the proposed codec stood second, it outperformed the other three codecs in all comparisons using mPSNR, as shown in Fig. 4c. An overall improvement of 38% and 15% for the proposed codec is observed compared to [25] and [11].
- iv) Weighted peak-signal-to-noise ratio (wPSNR): wPSNR [29] weighs spatial frequency in the error image using contrast sensitivity function to calculate the peak-signal-to-noise ratio. Comparing wPSNR as in Fig. 4d, the proposed codec outperformed all other codecs for Funt and Mixed datasets ranging from 3.6 to 32.7% improvement in performance. Only for the Pfstools dataset, the proposed codec stood at the

second position and lagged behind Khan [23] with a tiny margin.

- v) Structural similarity index measure for perceptually uniform space (SSIM_PU): SSIM_PU [30] is meant to be used for display-referred HDR images in perceptually uniform space. Figure 4e shows the results for SSIM-PU where all codecs performed well, but the proposed method remained the winner for all datasets, although the improvement over [23] was marginal. For the Funt dataset, it had 0.13% and 0.45% improvement over [11] and [25], respectively. The improvement over [11] and [25] was 4.9% and 1.03% for the mixed dataset, while it remained 2.37% and 0.30%, respectively, for the Pfstools dataset.
- vi) Mean opinion score of quality in HDR-VDP2 metric [31]: Instead of subjective evaluation of an image or video's visual aspects, the mean opinion score (mos) of HDR-VDP2 is used, which provides an alternative for reliable prediction of visibility and quality difference between the test and reference images. P. Hanhart et al. [32] analyzed 35 objective benchmarks on a dataset of 20 HDR images. They found HDR-VDP2 to be amongst the best and observed a strong correlation between HDR-VDP2 and subjective testing's mean opinion score. For calculating HDR-VDP2, we used 30 pixels/degree and surrounding light of 13 Lux, which are typical values for standard displays and viewing conditions. In terms of HDR-VDP2 scores in Fig. 4f, the proposed system outperforms the other codecs for the Funt dataset, with an improvement in 5% to 15% range. For the other two datasets, the proposed system stands

Table 1 Size comparison (in kb) for ten randomly chosen images from the three dataset

	Small Bottles	Memorial Church	Desk	Forest path	S0040	S0700	S0800	S1050	S1190	S1240	Overall reduction (%)
Ground truth	2451.07	1312	1880.99	10,420	9681.09	9742.98	9793.69	9448.89	9731.2	9774.89	
Ward	929.072	801.541	1026.93	6603.2	5515.7	4052.7	4747.4	3715.42	3984.7	5119.16	45.27
Mantiuk	759.163	599.743	830.192	5346.7	4334.29	3479.44	3852.48	3070.7	3225.93	4283.53	52.03
Khan	1063.39	714.306	1003.71	6760.6	5723.48	4741.86	5327.25	3931.32	4841.91	5737.64	39.92
Proposed	816.25	285.147	706.787	5536.4	3243.06	4039.68	3469.76	1934.87	2542.26	4736.16	55.27

The bold and the italics values signify the highest compression achieved

second to Khan [23] by a very small margin. However, in the overall average score for the three datasets, the proposed system outperformed other methods, including Khan [23], with an overall improvement of up to 10%.

- vii) CIE-DE2000: CIEDE2000 computes color difference between the sample between a reference with CIELab coordinate. Details can be found at [33] for the interested readers. Figure 4g compares the performance based on the color difference metric DE2000. The proposed codec outperformed all three codecs while being second to Khan [23] only in the mixed dataset. For the other two datasets, an improvement of 8% from the second-best [23] to 250% from [17] is observed.

Based on the results of the quantitative metrics above, the proposed codec stands a clear winner among the four codecs compared. In 15 out of the total 21 comparisons, it remained at the first position. For the remaining 6 cases, it lost to the winner by a tiny margin and remained at the second position.

4.3 Compression efficiency

Compression efficiency refers to the reduction in the size of the encoded data compared to the size of the raw data. It is evident from Fig. 4h that the proposed system has the best percentage reduction in size for the mixed and Pfstools datasets compared with the other codecs. The improvement ranges from 13 to 32%. Only for the Funt dataset, the proposed system slightly lags behind Ward et al. [25] and Mantiuk et al. [11] in percentage reduction. Note that the performance of these two codecs was very poor in all other quality metrics. Khan [23], which remained the closest competitor to our method for all the quality metrics above, had the worst compression for all datasets. To give some actual numbers of the obtained file sizes, we summarize the size comparison of 10 randomly chosen files from the three datasets in Table 1. All sizes are recorded in kilobytes. The ground truth represents the original size of the image being

encoded. In six out of ten images, the proposed codec outperformed the other codecs.

The file sizes of Khan [23] remained the largest and had the worst compression in eight cases. An overall average reduction in the file sizes of all the three datasets achieved by the codecs is mentioned in the last column, and here again, the proposed codec shows the best performance.

4.4 Visual comparisons using tone-mapped images

Tone-mapping refers to compressing the dynamic range of HDR images/video for visualization on standard displays. For further validation of the proposed codec's accuracy, Fig. 5 shows tone-mapped LDR images produced by applying the algorithm given by [34] on an HDR image reconstructed by the decoders of all four methods compared above. The test HDR image "bottles_small.hdr" taken from the mixed dataset is used for this comparison. The tone-mapped version of the original and the reconstructed HDR images is shown in Fig. 5a–e. Visible disparities are obvious in the tone-mapped version of the image reconstructed by Ward et al. [25] shown in Fig. 5b. This can also be ascertained from Fig. 3f where Ward et al. [25] showed poor performance on the visual metric. In the lower right corner of the output of Mantiuk et al. [11] shown in Fig. 5c, the grove in the table appears to be darker than the ground-truth image. The result of Khan [23] shown in Fig. 5d seems to be brighter than the ground truth image. The proposed codec's output image shown in Fig. 5e is almost identical to the ground truth. We calculated the wPSNR values of the images shown in Fig. 5b–e, taking the image in Fig. 5a as reference. The image produced by the proposed method showed an improvement of 36% over Ward et al. [25], 25% over Mantiuk et al. [11], and 16% over Khan [23].

4.5 Decoding time

The simplicity of decoding operation is a desirable feature for faster operations. We present the average decoding times of the four compared methods for the three datasets in Table 2.



Fig. 5 Tone-mapped images of reconstructed bottles_small image using different codecs: **a** Ground truth, **b** Ward, **c** Mantiuk, **d** Khan, **e** Proposed

Table 2 Decoding time of decoders for the datasets

	Ward	Mantiuk	Khan	Proposed
Mixed	1.15	2.96	3.00	1.33
Pfstools	1.93	4.67	4.51	1.97
Funt	2.00	4.75	4.75	2.18

The proposed codec has better decoding time when compared with decoders by Mantiuk et al. [11] and Khan [23]. There is a 56% improvement in the decoding time for the proposed decoder when compared to these two. The decoding time for the proposed decoder is slightly more yet comparable with Ward et al. [25]. Note that Ward et al. [25] does not reshape the data or use metadata, and therefore, it has the simplest decoding operation (but the worst accuracy for all quality metrics, as shown earlier). A decoding time close to this method indicates that the companding-based transformation in our method, despite its best accuracy, is extremely efficient and does not place much computational burden on the decoder.

4.6 R-D curves

In all the comparisons above, we saved both base and enhancement layers of all methods using lossless PNG, to observe the effect of the quantization methods alone. It is, however, a common practice to plot the bitrate vs. distortion, i.e., the R-D curves, to observe the performance of the codec when data are saved in a lossy format at different bitrates. In Fig. 6, we draw R-D curves of all methods compared above for randomly chosen images from Funt dataset. It can be seen that the proposed method performs better than other codecs at lower bitrates as well.

4.7 Comparison with HDR10

HDR10 has arguably become a de facto standard for end-to-end delivery of HDR content. HDR10 principally relies on the PQ exploiting the response of the human visual system to brightness levels. Since HDR10 does not target the high accuracy of encoded data, it cannot compete with the codecs mentioned above in traditional metrics like MSE and PSNR.

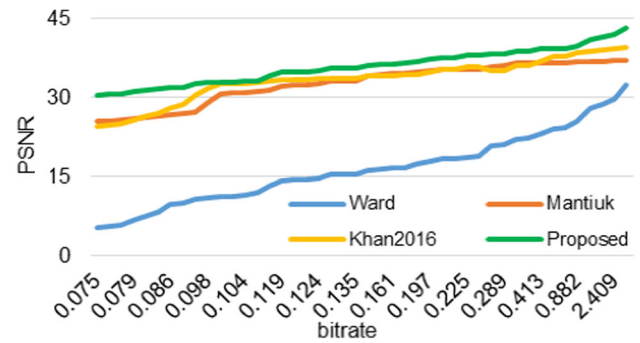


Fig. 6 Rate-distortion curve for Funt dataset

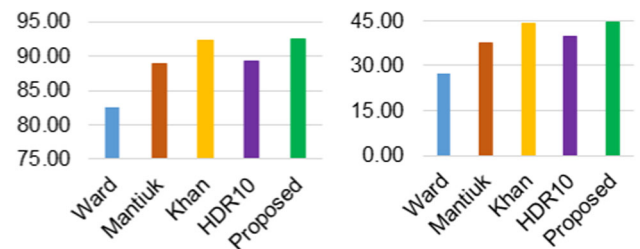


Fig. 7 Mean HDR-VDP2 (left) and mPSNR (right) results for all datasets

However, it would be reasonable to compare them in visual metrics like HDR-VDP2 and mPSNR. Figure 7 shows results for HDR-VDP2' quality (Q) metric. The reason of HDR10's lower score can be explained by the fact that some PQ10 codes in the lower luminance levels lie above the Barten's ramp [35], a commonly used measure to observe banding artifacts in gradients. The proposed codec has the best score with an improvement of 3.4% over HDR10 which remained third best. Figure 7 also shows the results of mPSNR. Here too, the proposed codec outperforms all the codec with an improvement of 10.75% over HDR10.

5 Conclusion

A new two-layer format for encoding HDR images has been proposed. The existing codecs generate metadata in general, which requires the receiver to read and process it in addition to the main image content, thereby adding to the computational complexity of the decoding process. In the proposed format, the need for metadata is eliminated, which reduced

the computational complexity and makes the proposed codec efficient in terms of decoding time. Metadata is generally used for higher encoding accuracy; however, the proposed method can reconstruct the HDR contents very accurately, even in the absence of metadata. It is shown that the tone-mapped versions of the reconstructed and the original images are visually identical, and up to 10% improvement on average is achieved in HDR-VDP2 scores compared to the existing state-of-the-art methods. The accuracy is due to a reduction in the quantization error achieved using the nonlinear companding process for quantization. A reduction of more than 80% in MSE and an increase of 18% and 27% in mPSNR and wPSNR on average, respectively, is observed in our experiments with a large set of images. In addition, more than 50% reduction in the file sizes of the reconstructed images is observed compared to the raw data formats, which is better than other encoders evaluated in this work. Smaller file size leads to smaller bandwidth requirements and faster delivery of content. Based on the presented results, the proposed codec comes as a good choice among dual-layer codecs when the base layer is not to be used for viewing.

Acknowledgements This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-21-ICI-17). The authors, therefore, acknowledge with thanks the University of Jeddah technical and financial support.

References

1. A. D. Association: TIFF revision 6.0. Adobe Systems Incorporated, Mountain View (1992)
2. Bogart, R., Kainz, F., Hess, D.: OpenEXR image file format. SIGGRAPH, Sketches & Applications (2003)
3. Ward, G.: Real pixels (Graphics Gems II), pp. 80–83. Morgan Kaufmann, San Francisco (1991)
4. Khan, I.R., Rahardja, S., Khan, M.M., Movania, M.M., Abed, F.: A tone-mapping technique based on histogram using a sensitivity model of the human visual system. *IEEE Trans. Industr. Electron.* **65**(4), 3469–3479 (2017)
5. Khan, I.R., Aziz, W., Shim, S.-O.: Tone-mapping using perceptual-quantizer and image histogram. *IEEE Access* **8**, 31350–31358 (2020)
6. Rana, A., Singh, P., Valenzise, G., Dufaux, F., Komodakis, N., Smolic, A.: Deep tone mapping operator for high dynamic range images. *IEEE Trans. Image Process.* **29**, 1285–1298 (2019)
7. Goris, R., Brondijk, R., van der Vleuten, R.: Philips response to CfE for HDR and WCG. In Presented at the M-36266, ISO/IEC JTC1/SC29/WG11, 112th MPEG meeting, Warsaw, Poland, September (2015)
8. Su, G.-M., Qu, S., Hulyalkar, S.N., Chen, T., Gish, W.C., Koepfer, H.: Layer decomposition in hierarchical VDR coding (2016)
9. 10918-1. ITU-T Rec. T. 81: Digital compression and coding of continuous-tone still images: requirements and guidelines (1993)
10. Kobayashi, H., Watanabe, O., Hitoshi, K.: Two-layer near-lossless HDR coding with backward compatibility to JPEG. In: 2019 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 3547–3551 (2019)
11. Mantiuk, R., Efremov, A., Myszkowski, K., Seidel, H.-P.: Backward compatible high dynamic range MPEG video compression. Presented at the ACM SIGGRAPH, July (2006)
12. Wang, J., Li, S., Zhu, Q.: High dynamic range image compression based on visual saliency. *APSIPA Trans. Signal Inf. Process.* **9**, e16, Art. no. e16 (2020)
13. Khan, I.R.: Effect of smooth inverse tone-mapping functions on performance of two-layer high dynamic range encoding schemes. *J. Electron. Imaging* **24**(1), 013024 (2015)
14. Khan, I.R.: HDR image encoding using reconstruction functions based on piecewise linear approximations. *Multimedia Syst.* **21**(6), 615–624 (2015)
15. Lasserre, S., François, E., Le Léannec, F., Touzé, D.: Single-layer HDR video coding with SDR backward compatibility. In Applications of Digital Image Processing XXXIX; 997108, San Diego, California, United States, vol. 9971: International Society for Optics and Photonics (2016)
16. Dufaux, F., Le Callet, P., Mantiuk, R., Mrak, M.: High Dynamic Range Video: From Acquisition, to Display and Applications. Academic Press, London (2016)
17. Lasserre, S., Le Léannec, F., Lopez, P., Olivier, Y., Touze, D., François, E.: High dynamic range video coding. In Joint Collaborative Team on Video Coding (JCT-VC), 16th Meeting, San Jose, CA, pp. 1–8 (2014)
18. Le Léannec, F. et al.: Modulation channel information SEI message. In Document JCTVC-R0139 (m33776), 18th JCT-VC Meeting, Sapporo, Japan (2014)
19. [MaxCLL]: CTA 861-G A DTV Profile for Uncompressed High Speed Digital Interfaces. http://www.techstreet.com/standards/cta-861-g?product_id=1934129 (2002)
20. SMPTE: [SMPTE ST 2086] Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images. The Society of Motion Picture Television Engineers Journal (2014)
21. ISO/IEC: Doc. ISO/IEC 23008-2:2015 Information Technology—High Efficiency Coding and Media Delivery in Heterogeneous Environments—Part 2: High Efficiency Video Coding (2013)
22. Richter, T., Bruylants, T., Schelkens, P., Ebrahimi, T.: The JPEG XT suite of standards: status and future plans. In: Applications of Digital Image Processing XXXVIII, vol. 9599: International Society for Optics and Photonics (2015)
23. Khan, I.R.: A nonlinear quantization scheme for two-layer hdr image encoding. *Signal Image Video Process.* **10**(5), 921–926 (2016)
24. CCITT: Rec G. 711, Pulse Code Modulation of Voice Frequencies. CCITT Blue Book, vol. 3, pp. 175–184 (1984)
25. Ward, G., Simmons, M.: JPEG-HDR: A backwards-compatible, high dynamic range extension to JPEG. In ACM SIGGRAPH 2006 Courses, pp. 3–11 (2006)
26. Funt: HDR Dataset, Computational Vision Lab. https://www2.cs.sfu.ca/~colour/data/funt_hdr/ (2020)
27. SorceForge: Pfstools Dataset for HDR images. <https://sourceforge.net/projects/pfstools/files/pfstools/2.0.1/>
28. Banterle, F.: Hdr toolbox for matlab. https://github.com/banterle/HDR_Toolbox
29. Erfurt, J., Helmrich, C.R., Bosse, S., Schwarz, H., Marpe, D., Wiegand, T.: A study of the perceptually weighted peak signal-to-noise ratio (WPSNR) for image compression. In: 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan. IEEE, pp. 2339–2343 (2019)
30. Aydın, T., Mantiuk, R., Seidel, H.-P.: Extending quality metrics to full luminance range images (Electronic Imaging). SPIE (2008)
31. Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W.: HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.* **30**(4), 1–14 (2011)

32. Hanhart, P., Bernardo, M.V., Pereira, M., Pinheiro, A.M., Ebrahimi, T.: Benchmarking of objective quality metrics for HDR image quality assessment. *EURASIP J. Image Video Process.* **1**, 1–18 (2015)
33. Sharma, G., Wu, W., Dalal, E.N.: The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Res. Appl.* **30**(1), 21–30 (2005)
34. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic tone reproduction for digital images. In: *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, vol. 21, no. 3. *ACM Transaction on Graphics*, pp. 267–276 (2002)
35. Peter, G.J.B.: Formula for the contrast sensitivity of the human eye. In: *Proceedings of SPIE*, vol. 5294 (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.