High Dynamic Range Video Coding

Konstantinos Konstantinides, Guan-Ming Su, and Neeraj Gadgil

Abstract Methods for the efficient coding of high-dynamic range (HDR) still-images and video sequences are reviewed. In dual-layer techniques, a base layer of standard-dynamic range data is enhanced by additional image data in an enhancement layer. The enhancement layer may be additive or multiplicative. If there is no requirement for backward compatibility, adaptive HDR-to-standard dynamic range (SDR) mapping schemes in the encoder allow for improved coding efficiency versus the backward-compatible schemes. In single-layer techniques, a base layer is complemented by metadata, such as supplementary enhancement information (SEI) data or color remapping information (CRI) data, which allow a decoder to apply special "reshaping" or inverse-mapping functions to the base layer to reconstruct an approximation of the original HDR signal. New standards for exchanging HDR signals, such as SMPTE 2084 and BT. 2100, define new mapping functions for translating linear scene light captured by a camera to video and are replacing the traditional "gamma" mapping. The effect of those transforms to existing coding standards, such as high efficiency video coding (HEVC) and beyond, are reviewed, and novel quantization and coding schemes that take these new mapping functions into consideration are also presented.

1 Introduction

In digital imaging, the term dynamic range refers to the ratio of the highest over the lowest luminance values in an image or a scene. For example, the human visual system (HVS) can perceive the brightness of an object from the darkest shadows

K. Konstantinides (✉) · G.-M. Su · N. Gadgil
Dolby Laboratories, San Francisco, CA, USA
e-mail: k.konstantinides@ieee.org; guanmingsu@ieee.org; njgadg@dolby.com

S. S. Bhattacharyya et al. (eds.), *Handbook of Signal Processing Systems*,
https://doi.org/10.1007/978-3-319-91734-4_5

or faint starlight (about 10^{-6} cd/m^2 or nits),[1] all the way to direct sunlight (about 10^8 cd/m^2). These intensities may be referred to as "scene-referred" intensities. Thus, the physical word represents approximately 14 orders of dynamic range, typically referred to as high dynamic range (HDR). In contrast, the dynamic range that humans may simultaneously perceive is approximately 5–6 orders of magnitude. While this dynamic range, sometimes referred to as visual dynamic range or extended dynamic range, is significantly lower than HDR, it may still be referred to as a high dynamic range.

In display technology, dynamic range (say, 5000:1) refers to the ratio of the brightest white (e.g., 500 nits) over the darkest black (e.g., 0.1 nits) a display can render. These intensities are referred to as "display-referred" intensities. Most consumer desktop displays and high definition televisions (HDTVs) currently support peak luminance of 200–500 nits. Such conventional displays thus typify a lower dynamic range (LDR), also referred to as a standard dynamic range (SDR). Modern, HDR-branded, displays support peak luminance of 800–1000 nits; however, studio HDR monitors are known to support peak luminance values exceeding 4000 nits.

Traditional, 24-bit, digital photography represents a low dynamic range of about 2–3 orders of dynamic range and has dominated both software and hardware architectures in both the consumer electronics and the broadcast industry. However, many display manufacturers are now entering the market with HDR displays and thus there is increased interest and need for the efficient representation, compression, and transmission of HDR content. This Chapter provides a brief overview of past efforts in backwards-compatible coding of HDR video and presents some recent developments using both dual-layer and single-layer architectures. More specifically, Sect. 2 describes early work for coding still HDR images. Before covering HDR video coding, in Sect. 3, we review non-linear transfer functions commonly being used to translate linear light to video signals. Backward-compatible video coding methods are reviewed in Sect. 4, and non-backward compatible methods are reviewed in Sect. 5. Both single-layer and multi-layer coding schemes are examined.

2 Early Work: HDR Coding for Still Images

The image processing community has shown interest in HDR images since the late 1980s, using a variety of logarithmic-based or floating-point encodings for intensity, such as the Radiance RGBE format, LogLuv TIFF, and OpenEXR [1]. None of these formats allowed for backward compatibility, typically a necessary (but not always sufficient) requirement for wider adoption. In 2004, encouraged by the wide adoption of the JPEG image compression standard, Ward and Simmons

[1] Candela per square meter (cd/m^2), also referred to as *nit*, is the international standard unit of luminance.

[1] introduced JPEG-HDR, a JPEG-backwards-compatible format that includes a baseline image representing a tone-mapped version of the input HDR image,[2] and a ratio image representing pixel by pixel the ratio of luminance values in the input HDR image over those in the tone-mapped image. The ratio image is log-encoded and compressed as an 8-bit grayscale image, and it is stored as part of the JPEG image using a dedicated JPEG application marker. The format allows legacy JPEG decoders to ignore the ratio image and simply decode the tone-mapped version of the JPEG image. JPEG-HDR-enabled decoders can reconstuct the HDR image by a simple multiplication with the ratio image. JPEG-HDR was later enhanced to include chroma residuals, and the new format was standardized as Profile A in JPEG XT, which includes also two alternative Profiles (B and C) for coding HDR still images [1–3]. Following the notation by Richter [3], the three profiles may be expressed as follows:

Profile A
In this profile, based on JPEG-HDR, the reconstructed HDR image is generated as

$$HDR = \mu(r)\,(\Phi(SDR) + \chi)\,, \tag{1}$$

where SDR is the base JPEG image (typically encoded in a gamma-corrected space), Φ denotes an inverse-gamma correction, χ denotes a function of chroma (e.g., CbCr) residuals, and $\mu(r) = \exp(r)$ denotes an exponential function of the ratio log-image r (that is, the logarithm of the ratio of luminance in the original HDR image over luminance in the tone-mapped image).

Profile B
Profile B is similar to Profile A, except that the ratio is expressed for each color channel ($i = 1, 2, 3$). Ratios are also coded using a logarithmic representation, thus

$$HDR_i = \sigma\ \exp\left(\log\left(\Phi\left(HDR_i\right)\right) - \log\left(\Psi\left(RES_i\right)\right)\right), \tag{2}$$

where Φ denotes an inverse-gamma correction, Ψ is typically selected to be a gamma correction derived by the encoder, and σ is a scaler.

Profile C
In profile C, the floating-point residuals between the HDR image and the SDR image, in each color component (RES_i, $i = 1, 2, 3$), are expressed as integers using a pseudo-log2 representation. Then, in the decoder, one may apply a pseudo-exponent to recover them. Thus, while the reconstructed image may be represented as

$$HDR_i = \Phi\left(SDR_i\right) RES_i\,,$$

[2] "Tone mapping" refers to the process of mapping luminance values in a high dynamic range to luminance values in a lower dynamic range.

in a decoder, under Profile C,

$$HDR_i = \psi \exp\left(\widehat{\Phi}\left(SDR_i\right) + RES_i - o\right), \tag{3}$$

where $\psi \exp$ represents a pseudo-exponential function, o is an offset that ensures the residual image is not negative, and $\widehat{\Phi}$ represents an inverse gamma followed by an inverse log approximation, typically optimized by the encoder and passed to the decoder.

Despite the adoption of HDR image capture in many cameras and smart-phones (typically, by combining three separate exposures), to the best of our knowledge, by mid-2017, none of the camera manufacturers had adopted any of the JPEG XT Profiles. All captured HDR images are simply represented and stored as tone-mapped JPEG images. Before we continue with HDR video coding schemes, it is worth revisiting another important topic: the role of cathode ray tube (CRT) technology in traditional SDR coding, especially as it is related to quantizing linear scene light (e.g., light as captured by a camera sensor) to a non-linear signal for efficient processing of captured images in a video pipeline.

3 Signal Quantization: Gamma, PQ,[3] and HLG[4]

Due to signal-to-noise constraints in analog and digital video signals and the characteristics of the traditional CRT display, scene light was never represented in a linear form in the video processing pipeline. Captured images are quantized using a non-linear opto-electrical transfer function (OETF) which converts linear scene light into the camera's (non-linear) video signal. Then, after encoding and decoding, on the receiver, the signal would be processed by an electro-optical transfer function (EOTF), which would translate the input video signal to output screen color values (e.g., screen luminance) produced by the display. Such non-linear functions include the traditional "gamma" curve documented in Recommendations ITU-R BT.709, BT.1886, and BT.2020 [4]. The combination of an OETF, the EOTF, and any artistic adjustments (either during content creation or content display) is referred to as the system opto-optical transfer function or OOTF [5].

Currently, most digital interfaces for video delivery, such as the serial digital interface (SDI) are limited to 12 bits per pixel per color component. Furthermore, most compression standards, such as H.264 (or AVC) and H.265 (or HEVC), are limited, at least in practical implementations, to 10-bits per pixel per component. Therefore, efficient encoding and/or quantization is required to support HDR content, with dynamic range from approximately 0.001–10,000 cd/m^2 (or nits), within existing infrastructures and compression standards.

[3] PQ stands for "Perceptual Quantizer" EOTF, as defined by Miller et al. [7].

[4] HLG stands for a non-linear transfer function known as "Hybrid Log-Gamma."

Gamma encoding was satisfactory for delivery of SDR (e.g., 8–10 bits) content, but has been proven rather inefficient when coding HDR content. The human visual system responds to increasing light levels in a very non-linear way. A human's ability to see a stimulus is affected by the luminance of that stimulus, the size of the stimulus, the spatial frequencies making up the stimulus, and the luminance level that the eyes have adapted to at the particular moment one is viewing the stimulus [6, 7]. In 2013, Miller et al. [7] proposed an alternative EOTF to the gamma function, commonly referred to as "PQ" (for perceptual quantizer). PQ maps linear input gray levels to output gray levels that better match the contrast sensitivity thresholds in the human visual system. Compared to the traditional gamma curve, which represents the response curve of a physical CRT device and coincidently may have a very rough similarity to the way the human visual system responds, the PQ curve imitates the true visual response of the human visual system using a relatively simple functional model, shown in Eq. (4) [8].

$$V = EOTF^{-1}[L] = \left(\frac{c1 + c2 * L^{m1}}{1 + c3 * L^{m1}}\right)^{m2}, \tag{4}$$

where, V represents the result non-linear signal (say, (R′, G′, B′)) in a range [0,1], L represents luminance of a displayed linear components in cd/m^2 (assuming a peak luminance of 10,000 cd/m^2), and the constants are:

$$m1 = 2610/16384 = 0.1593017578125,$$

$$m2 = 2523/4096 \times 128 = 78.84375,$$

$$c1 = 3424/4096 = 0.8359375 = c3 - c2 + 1,$$

$$c2 = 2413/4096 \times 32 = 18.8515625,$$

$$c3 = 2392/4096 \times 32 = 18.6875.$$

For comparison, according to Rec. BT. 709 [9], the traditional gamma curve on an encoder is given by

$$V = \begin{cases} 1.099L^{0.45} - 0.099; & 1 \geq L \geq 0.018 \\ 4.500\ L; & 0.018 > L > 0 \end{cases}, \tag{5}$$

where L denotes input luminance in [0,1].

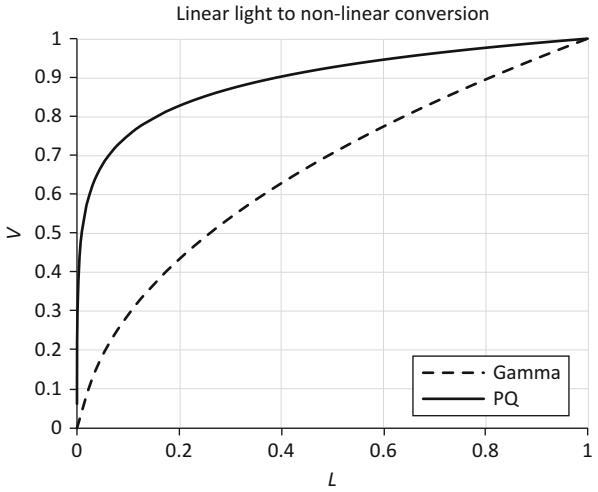

Fig. 1 BT. 709 (gamma) versus SMPTE 2084 (PQ) encoding for HDR signals

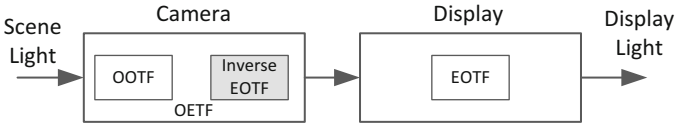

Fig. 2 PQ-oriented system, OOTF is in the camera

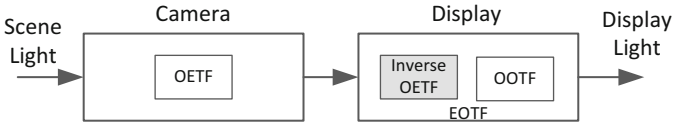

Fig. 3 HLG-oriented system, OOTF is in the display

The PQ mapping functions were adopted in SMPTE ST 2084 and Rec. ITU-R BT.2100 [8]. Figure 1 provides an example of the gamma and PQ coding for L in (0, 1].

The PQ EOTF was designed assuming a camera-centric OOTF that applies an inverse PQ EOTF to generate the signal to be processed (see Fig. 2). In Rec. BT. 2100, an alternative, display-centric nonlinear transfer function is also presented, commonly referred to as Hybrid Log-Gamma (HLG) (see Fig. 3). HLG was designed with backward-compatibility in mind, especially as related to ITU-R BT. 2020 color displays. PQ and HLG may co-exist, and Annex 2 of BT. 2100 provides examples of converting between PQ-coded signals and HLG-coded signals. Given a video stream, information about its EOTF is typically embedded in the bit stream as metadata.

4 Backward-Compatible HDR Coding

By the end of 2016, there was a variety of HDR content available for consumers with branded HDR displays; however, the majority of consumers around the world have only SDR displays. Backward compatibility; that is, support for both SDR and HDR content, is considered by many critical for the wider adoption of HDR displays. Depending on the operational cost, including content storage, network bandwidth, and processing time, content providers have two different approaches to support both formats.

For a system or distribution channel with abundant storage, but limited network bandwidth, most system providers prefer to use a unicast solution: namely, store both the SDR and HDR bit streams in the server side, and transmit either the SDR or the HDR bit stream, depending on the targeted display.

For a system with limited storage, or when the storage cost is high, the system providers prefer to provide a bitstream which can be viewed on a legacy SDR display, but also allows users with an HDR display to reconstruct the HDR content given some additional information. This design is referred to as a backward compatible codec. A legacy receiver can decode the base layer (BL), and the viewers can watch the SDR-grade video. With additional metadata and/or an enhancement layer (EL), the viewers can reconstruct and watch the HDR-grade video. The additional amount of information is often much smaller than the base layer. In this section, we will discuss dual-layer and single-layer backward-compatible HDR video codecs.

The backward compatible HDR codec can be categorized as either a dual-layer system or single-layer system. Each layer is encoded using a legacy video codec, such as AVC (H.264) or HEVC (H.265). Single-layer systems include a single coded bitstream and rely on metadata that define a parametric model to reconstruct the HDR signal. Dual-layer systems include a coded base layer, metadata for reconstructing an HDR estimate signal, and an additional layer of coded residual data which can be added to the HDR signal reconstructed by the base layer, thus, typically providing a more accurate representation of the HDR signal, but at the expense of more bandwidth.

4.1 Dual-Layer Coding

Motivated by the work of Ward and Simmons, in 2006, Mantiuk et al. [10] introduced an MPEG-backwards-compatible coding scheme for HDR video based on layered predictive coding. An example encoder of their scheme is shown in Fig. 4. As depicted in Fig. 4, the encoder receives an HDR video sequence. An HDR to SDR transformation block (e.g., a tone-mapper) applies any of the known HDR to SDR transformation techniques to generate an SDR version, representing the backward-compatible video to be rendered on an SDR display. The SDR version

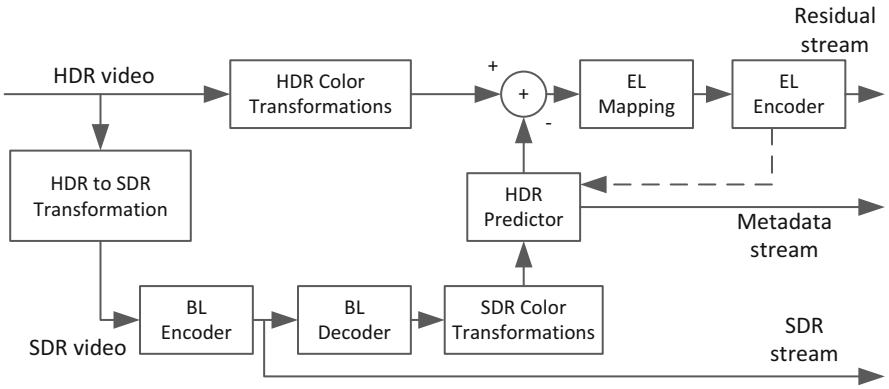

Fig. 4 Block diagram of a dual-layer HDR encoder

is coded with a legacy MPEG encoder. A decoded version, after optional color transformations to match the color format of the HDR video, is passed to a predictor, which estimates the input HDR signal based on the SDR input. The output of the predictor is subtracted from the original HDR input to generate a residual, which is further compressed by another encoder. The SDR stream, the residual stream, and information about the predictor (e.g., metadata) are all multiplexed to form an HDR-coded bitstream. The term "metadata," as used here, relates to any auxiliary information that is transmitted as part of a coded bitstream and assists a decoder to render a decoded image. Such metadata may include, but are not limited to, color space or gamut information, reference display parameters, and other auxiliary signal parameters that will be described later on in this chapter.

In a decoder (Fig. 5), a legacy decoder simply decodes the base layer to be displayed on an SDR display. An HDR decoder, applies the predictor function received from the encoder to the baseline SDR input to generate a predicted HDR signal, which is added to the decoded residual signal to reconstructs the output HDR signal to be displayed on an HDR display. Additional color transformations may be needed to match the color transformations in the encoder or the capabilities of the display pipeline.

To improve coding compression of the residual information, Mantiuk et al. proposed using custom color spaces for both the coded SDR and the HDR streams. For SDR, the input RGB data is converted to a luma-chroma space where, chroma is encoded using a format similar to logLuv encoding and luma is encoded by a special transformation using linear and power function segments [23]. For the HDR signal, after translating YCbCr or RGB data to XYZ, XYZ-float data are converted to a luma-chroma space where: chroma is encoded the same way as SDR chroma, and luma is encoded using a perceptual uniform luminance encoding.

For the prediction function, Mantiuk et al. used a method that combines "binning" with pixel averaging. For example, for SDR images with an 8-bit bit depth there will be 256 bins, where each bin represents all the HDR pixels

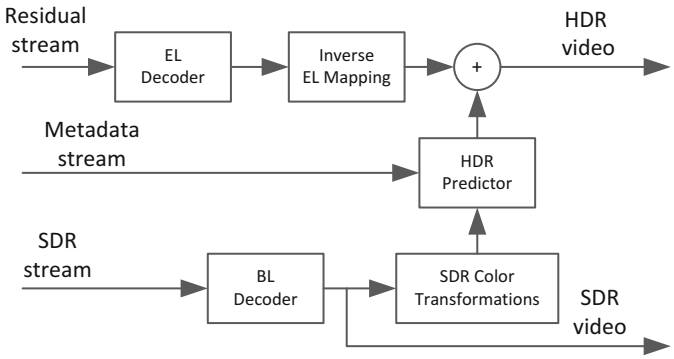

Fig. 5 Block diagram of dual-layer decoder

mapped to the specific (SDR) bin value. Thus, given N pixels in a frame, bin $\Omega_l = \{i = 1, \ldots, N : I_{sdr}(i) = l\}$, where $l = 0, 1, \ldots, 255$, and $I_{sdr}(i)$ denotes the tone-mapped SDR luma value for the i-th input (HDR) pixel ($I_{hdr}(i)$). Then, given an input SDR pixel value l, the corresponding predicted HDR pixel value $(\widehat{I}_{hdr}(l))$ is generated using the arithmetic mean of all HDR pixels that are mapped to the SDR value l in bin Ω_l, or

$$\widehat{I}_{hdr}(l) = \frac{1}{|\Omega_l|} \sum\nolimits_{i \in \Omega_l} I_{hdr}(i), \tag{6}$$

where $|\Omega_l|$ denotes the cardinality of the l-th bin.

Looking at the block diagram in Fig. 4, one can identify three key components that may affect dual-layered coding: (a) the proper color representation of the signals, (b) the choice of the HDR to SDR transformation, and (c) the choice of a prediction function. As discussed earlier, input HDR content may now be represented using both gamma- and PQ-quantization. Regarding the HDR to SDR transformation block, in many cases, this block may be omitted, since the SDR video may be provided separately. For example, the SDR video may represent the output of manual or semi-automatic color grading by a professional colorist who performed color grading according to a director's intent on a reference SDR display. Regarding the HDR predictor, there are multiple alternatives to Eq. (6) and will be discussed later.

In many applications, the residual signal may not be in a format suitable for encoding by a legacy encoder. For example, it may be in floating point, or it is possible that its dynamic range exceeds the dynamic range of the EL Encoder. Then, as shown in Fig. 4, the residual video signal may also be further processed by an enhancement layer mapping function (EL Mapping), such as a linear or non-linear quantizer. If that is the case, as shown in Fig. 5, a decoder needs to apply to the decoded residual an inverse mapping (Inverse EL Mapping).

In some encoder designs, information about the compressed EL bitstream can be fed back to the HDR prediction module as in-loop processing, or it can be accessed independently, as an open-loop process. In-loop designs can generate the reconstructed HDR signal by using inter-layer prediction, that is, by combining the inverse mapped SDR and the residual. This approach can enable temporal domain prediction to further reduce bit rates; however, the in-loop designs need to modify legacy encoders and makes the encoder design more complex. The open-loop approach is a much simpler design, but it loses compression efficiency owing to the lack of a temporal, inter-layer, prediction path.

Besides the temporal-prediction domain option, the EL Mapping block may also support re-scaling and/or sub-sampling in the spatial domain, so that the residual can be down-sampled to a lower resolution to reduce the bit rate. In this case, the EL inverse mapping block in the decoder should support performing the corresponding inverse re-scaling to up-sample the received EL signal to the same resolution as the BL signal. Rescaling may reduce the EL stream bit-rate requirements; however it may also introduce additional artifacts, such as blurred edges and texture. Applying spatial re-scaling is a design tradeoff that depends on the system's bandwidth and picture quality requirements.

One of the key components of any dual-layer system is the design of the HDR predictor. Some of the proposed designs are briefly reviewed next.

4.1.1 Piecewise Linear Model Representation

In [10], Mantiuk et al. proposed sending to the decoder a simple one-dimension look-up Table (1D-LUT) representing the one-to-one SDR to HDR mapping based on Eq. (6); however, the overhead to transmit an 1D-LUT for each frame or even a small collection of frames as metadata is big. Furthermore, Eq. (6) does not guarantee that the SDR to HDR mapping will satisfy certain important properties to reduce coding artifacts. One such property is that the inverse mapping function should be monotonically non-decreasing. This property is important for images with areas with smooth gradient. It has been observed that if the SDR-to-HDR mapping function has some ranges with decreasing slope at its first derivative, then it often creates "hole" artifacts for those smooth areas. To address these problems, in [11], the authors proposed communicating the prediction function to a decoder using a piecewise linear model.

Let $A^L()$ denote the HDR-to-SDR transformation for the luma component (e.g., a global tone-mapping function) and let $B^L()$ denote the inverse luma mapping (e.g., SDR-to-HDR), also referred to as the HDR prediction function. To construct the piecewise linear model for the HDR to SDR mapping, one needs to define: (a) the pivot or end points $\{s_l, s_0 < s_1 < \ldots < s_L\}$ separating/connecting two nearby pieces (or segments) in the piecewise function, and (b) the parameters of the linear model for each piece (e.g., for $y = m_l{*}x + b_l$, m_l and b_l). As the number of possible segments increases, the optimal solution problem becomes intractable very quickly. To simplify the solution, a fixed interval length in logarithm domain is

used to eliminate the need to compute the pivot points. The problem can be further simplified by only computing the slope $\{m_l\}$ in each piece. For the SDR to HDR inverse mapping, the pivot points can remain the same as $\{s_l\}$, and the slope will be the inverse value, as $\{1/m_l\}$, in each piece. The entire system can be formulated as a mean-square-error (MSE) optimization problem to reduce the end-to-end distortion between the original HDR picture and the reconstructed HDR picture:

$$\arg\min_{\{m_L\}} \left\| B^L\left(A^L\left(v_i^L\right)\right) - v_i^L \right\|^2, \tag{7}$$

where v_i^L denotes the intensity of the i-th HDR pixel (e.g., $I_{hdr}(i)$). Assuming a local linearity of the mapping function, the problem can be further simplified and solved with a closed form solution.

4.1.2 Multivariate Multiple Regression Predictor

In order to preserve the film-director's look or intent, content providers may use a colorist to manually color grade an HDR (or SDR) version of a movie based on an existing SDR (or HDR) version. In this case, the HDR to SDR mapping is often a complex function that cannot be represented by existing tone-mapping curves. Furthermore, this complex HDR to SDR process imposes great difficult for the inverse mapping. A single-channel (e.g., luma) HDR predictor cannot capture any chroma-related transformations. For example, chroma saturation and hues may be different in the SDR and HDR versions.

To address these issues, Su et al. [12] proposed separate schemes for predicting luma and chroma. For the luminance component, the HDR predictor uses a piecewise higher order polynomial which allows for more accurate mapping, but with fewer segments than the linear model. With the same overhead, a higher order polynomial can produce better accuracy and a smaller residual. The small residual often leads to smaller energy to encode in the enhancement layer, which yields a lower bit rate requirement for the coded residual stream.

Although a piecewise higher order polynomial can provide better inverse mapping accuracy, the required computation to find the optimal solution is very expensive. As in the piecewise linear model, the parameters in higher order polynomial include the pivot point selection and the polynomial coefficients. By stating the solution as an MSE minimization problem, one can take advantage of the matrix structures and pre-compute many parameters offline as look-up Tables. A faster algorithm (by a factor of 60) to achieve the same optimal solution as full search was proposed in [13].

For the chroma components (say, *Cb* and *Cr* in YCbCr), a cross-color channel inverse mapping, $B^C()$, is adopted to cope with the color space differences, saturation and hue adjustments, and bit depth requirement [14]. The cross-color channel inverse mapping takes the three color channels ($s_i{}^L$, $s_i{}^{C0}$, $s_i{}^{C1}$) for each SDR pixel, and converts them to HDR chroma values $\left(\widehat{v}_i^{C0}, \widehat{v}_i^{C1}\right)$ with parameter set

$\left\{m_{\alpha\beta\gamma}^{C0}, m_{\alpha\beta\gamma}^{C1}\right\}$ using a multivariate multiple regression (MMR) prediction model, which in its most general form can be expressed as

$$\begin{aligned}\widehat{v}_i^{C0} &= \sum_{\alpha}\sum_{\beta}\sum_{\gamma} m_{\alpha\beta\gamma}^{C0}\cdot\left(S_i^L\right)^{\alpha}\cdot\left(S_i^{C0}\right)^{\beta}\cdot\left(S_i^{C1}\right)^{\gamma},\\ \widehat{v}_i^{C1} &= \sum_{\alpha}\sum_{\beta}\sum_{\gamma} m_{\alpha\beta\gamma}^{C1}\cdot\left(S_i^L\right)^{\alpha}\cdot\left(S_i^{C0}\right)^{\beta}\cdot\left(S_i^{C1}\right)^{\gamma}.\end{aligned} \tag{8}$$

For example, using a second-order MMR predictor, Eq. (8) for color component C0 may be expressed as

$$\begin{aligned}\widehat{v}_i^{C0} = {} & m_0 + m_1 s_i^L + m_2 s_i^{C0} + m_3 s_i^{C1}\\ & + m_4 s_i^L s_i^{C0} + m_5 s_i^L s_i^{C1} + m_6 s_i^{C0} s_i^{C1} + m_7 s_i^L s_i^{C0} s_i^{C1}\\ & + m_8\left(s_i^L\right)^2 + m_9\left(s_i^{C0}\right)^2 + m_{10}\left(s_i^{C1}\right)^2\\ & + m_{11}\left(s_i^L s_i^{C0}\right)^2 + m_{12}\left(s_i^L s_i^{C1}\right)^2 + m_{13}\left(s_i^{C0} s_i^{C1}\right)^2 + m_{14}\left(s_i^L s_i^{C0} s_i^{C1}\right)^2.\end{aligned} \tag{9}$$

The optimal solution for the MMR prediction coefficients $\left(\left\{m_{\alpha\beta\gamma}^{C0}, m_{\alpha\beta\gamma}^{C1}\right\}\right)$ can be obtained via multivariate multiple regression which minimizes the mean-squared-error between the original HDR chroma and the reconstructed HDR chroma pixel values [14]. MMR-based prediction is used in the Dolby Vision® HDR format from Dolby Laboratories.

4.1.3 MPEG Color Gamut Scalability

In contrast to using a prediction polynomial for luma and/or chroma inverse mapping, one could also deploy a 3D-LUT method, which is commonly used in color science and display management system as a quantization process. This 3D-LUT method is standardized in MPEG as color gamut scalability [15].

In general, the 3D-LUT solution will partition the 3D space into multiple cubes. For each cube (x, y, z), one can have eight corresponding vertices with values denoted as $N(x, y, z)$. Each cube covers a range of code words in three dimensions. For any given input code word, $(s_i{}^L, s_i{}^{C0}, s_i{}^{C1})$, for pixel i, one can find which cube, (x, y, z), contains this triplet. Then, one finds out the eight corners (vertices) of this cube (for both chroma values) and performs interpolation to get the HDR values. In tri-linear interpolation, illustrated in Fig. 6, the interpolated value for one particular color channel can be computed as

$$\widehat{v}_i^{C0} = \sum_{\alpha=0}^{1}\sum_{\beta=0}^{1}\sum_{\gamma=0}^{1} w_{\alpha\beta\gamma} N^{C0}(x+\alpha, y+\beta, z+\gamma), \tag{10}$$

where $w_{\alpha\beta\gamma}$ are weighting factors depended on the distances to the eight vertices.

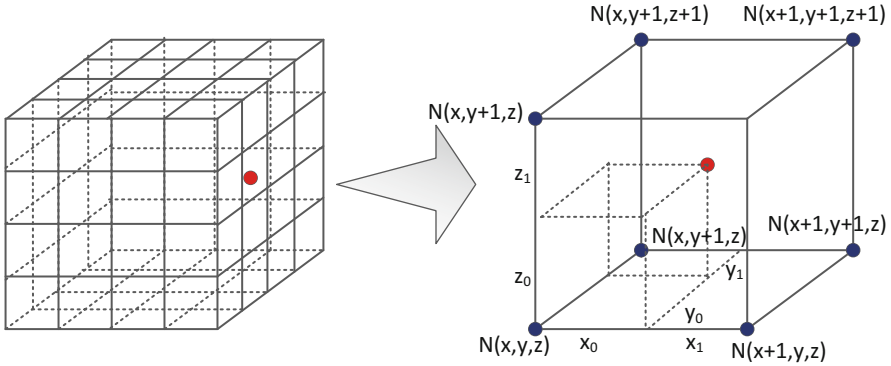

Fig. 6 3D LUT for Color-gamut scalability

The cube can be partitioned evenly in each dimension so one may have a regular cube size. However, the major problem for uniform partition requires a larger overhead to achieve required color precision. Consider a $17 \times 17 \times 17$ 3D LUT, then we need to store $17^3 = 4913$ node values for each color channel. Another approach is to adopt an octree-based structure where each parent cube contains eight smaller cubes via a non-uniform quantization process. The advantage of a tree based partition is to explore the color sensitivity diversity: give more precision for interpolation for more sensitive color regions, and assign less prevision for less sensitive color regions. The overhead for such a cube representation can be smaller. On the other hand, color artifacts along the cube boundaries should be carefully handled as they often represent a discrete value selection and might cause sudden color changes on a flat area.

4.1.4 System-Level Design Issues in Dual-Layer Systems

Dual-layer systems, in general, demand more system resources and higher data management flow for the entire pipeline. From a processing and computational point of view, at the encoder side, the encoding process will be longer, owing to two encoder instances. The overall processing time includes the BL encoding time, BL inverse mapping time, EL mapping time, and EL compression time. The decoder also needs to double its decoding time to handle both layers. It often requires more memory to handle EL information.

From the transmission's point of view, a dedicated multiplexing design is needed to transmit the BL stream, the EL stream, and metadata. To decode one HDR frame, the bitstreams need to be synchronized so the composing can be done based on the current frame. Thus, system level transport at the video elementary stream level is needed to ensure correct synchronization and decoding.

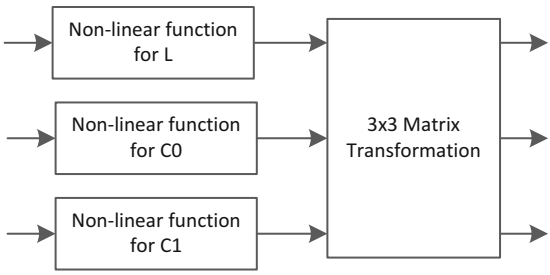

Fig. 7 Parametric model of HDR to SDR transformation

4.2 Single-Layer Methods

In a single-layer system there is only a base layer and metadata to help transform the base layer SDR signal into an HDR signal. The base layer bitstream can be directly decoded and shown on legacy SDR display. To construct the HDR signal, metadata is used to convert the SDR signal to the HDR signal. A single layer requires less processing and computation, and less bandwidth. On the other hand, to recreate a high quality HDR signal, the correlation between the SDR and the HDR signals should be able to be expressed using a well-defined parametric model. This approach, in general, limits the ability to match the director's intent in the reconstructed HDR signal.

As shown in Fig. 7, in a typical color grading process, it is observed that during the generation of the SDR color-graded version from the HDR original, the dynamic range is adjusted by applying: (a) a non-linear tone-mapping curve to each color channel (say, L, $C0$, and $C1$) to adjust the dynamic range; and (b) a 3×3 color rotation matrix to all three channels to adjust the hue and saturation. If one could express these operations in a simple parametric model, then, in a decoder, given the SDR signal, one could recreate the HDR content by doing the reverse, that is, by applying: (a) an inverse 3×3 transformation, followed by (b) inverse non-linear functions in each of the color channels.

To model this mapping, in [16] Su et al. proposed a non-linear matrix mapping model. The most common non-linear function in color grading is the slope, offset, and power (SOP) model [17] (also referred to as the lift, gain, and gamma (LGG) model), where each pixel is adjusted according to gain, offset, and power parameters, as

$$\widehat{v}_i = (S \cdot s_i + O)^P, \tag{11}$$

where s_i denotes an input SDR value, S denotes the slope (or gain), O denotes the offset (or lift), and P denotes the power (or gamma). The rotation matrix is a simple 3×3 matrix. The order of applying the non-linear function and the 3×3 matrix can be switched depending on the prediction accuracy. The SOP parameters and the elements of the 3×3 matrix may be recorded by the color-grading process or they could be estimated by the encoder. They can be passed to the decoder as metadata.

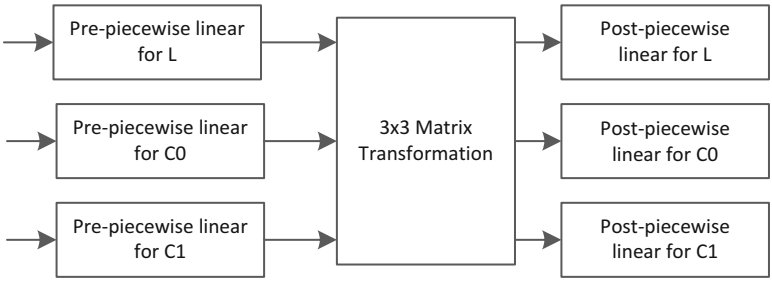

Fig. 8 HEVC model for color remapping information (CRI)

A recent version (v.4 or version 12/2016) of the HEVC (or H.265) standard [18] allows encoders to communicate more complex parametric models as color remapping information (CRI), by concatenating the encoder and decoder non-linear models. As shown in Fig. 8, the mapping process consists of three major stages: the first stage includes a piecewise-linear function (pre-piecewise linear) for each individual color channel to model the non-linear function; the second stage contains a 3×3 matrix; and the last stage includes another piecewise linear function (post-piecewise linear) to model the non-linear function.

The proposed system by Technicolor for single-layer HDR transmission uses similar principles and is discussed in detail in [19]. The design requires two LUTs (one for luma and one for chroma) and two dynamic scaling parameters (a and b). To accommodate multiple broadcasting scenarios, the authors propose that the two LUTs may be communicated either explicitly (table-based mode) or using a simplified set of parameters (parameter-based mode). A table-based mode may provide better quality, but at the expense of more bandwidth. The parameter-based mode assumes a fixed default LUT which can be adjusted by a piece-wise linear table of at most six points. In table-based mode, the two tables are explicitly coded using CRI data.

4.2.1 Philips HDR Codec

In 2015, in response to a call for evidence for HDR and wide-color-gamut coding, Philips [20] submitted a parameter-based, single layer, HDR plus SDR solution, where parameters related to the reconstruction of both the SDR and the HDR signals are embedded into the bitstream. Under the Philips model, each SDR RGB color plane (SDR_i, i = R, G, or B) is expressed as $SDR_i = \omega * HDR_i$, where the scaler ω is determined via a tone-mapping process of *MaxRGBY*; an input signal generated by a weighted combination of the luma (Y) and RGB values of the HDR input. The *MaxRGBY* signal is first translated into a perceptual-uniform signal via the inverse of a Philips-defined EOTF. The perceptual-uniform signal is mapped back to a linear signal via a multi-step process which includes (a) a black/white

level adaptation step, (b) a tone-mapping step using a non-linear curve expressed through parameters defining a "Shadow Gain control," a "Highlight Gain control," and a "Mid-Tones adjustment," and (c) a perceptual to linear SDR mapping step, which uses the Philips-defined EOTF. A receiver, given the received SDR signal and metadata defining ω, applies the inverse steps to generate an approximation of the HDR signal depending on the characteristics of the target display.

As will be discussed later, together with the Dolby Vision dual-layer HDR format, the Philips HDR format is one of the optional HDR formats in the Blu-Ray UHD specification. In 2016, Technicolor and Philips combined their formats into a single unified proposal.

5 Non-Backward-Compatible HDR Coding

Backward-compatible methods do not necessarily guarantee commercial success and they also face multiple challenges. First, a backward-compatible codec typically uses an advanced inter-layer prediction mechanism that demands a higher computational complexity. Second, the details in the higher- and lower-intensity regions are often asymmetric between the SDR and HDR versions, causing undesired clipping in those regions. Furthermore, they typically require a higher bit rate.

Recently, there has been an increasing interest in optimizing video quality for target HDR displays using the existing compression architectures, such as 8- or 10-bit legacy H.26x or VPx encoders, without the restriction of backward compatibility. In this case, the target application, say streaming using an over-the-top set-top box, is focused strictly on HDR playback. Codecs developed for such applications typically require less computational complexity, lower bit rate, and mitigate the clipping issues caused by the typical backward compatible codecs.

5.1 Multi-Layer Non-Backward-Compatible Systems

Given an HDR signal, multi-layer methods typically use multiple lower-bit depth encoders to provide an effective higher bit depth required to process HDR content. A base layer and one or more enhancement layers are used as a mechanism for coding and distributing the source video signals. A preprocessing step generally involves a signal splitting method to generate from the source signal multiple lower bit depth (8- or 10-bit) layers. These layers are then independently encoded to produce standard-compliant bitstreams so that they can be decoded by the existing decoding architectures which are in most cases hardware-based. In a decoder, the decoded layers are then combined to form the output HDR signal that is optimized for viewing on a HDR display. Note that the decoded SDR signal is never viewed.

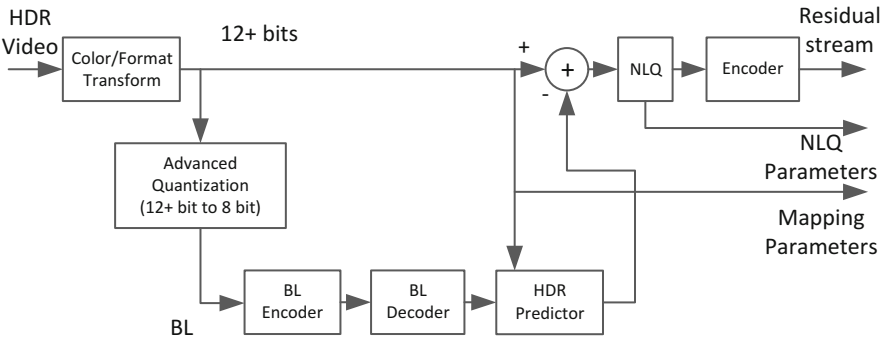

Fig. 9 Dual-layer, non-backward compatible, 8-bit Codec Architecture

5.1.1 Dolby Non-Backward-Compatible 8-Bit Dual-Layer Codec

In a technique developed by Su. et al. [21], an advanced quantization method is used to generate a non-backward-compatible BL and EL signals which may consist of residual values and quantization and mapping parameters that are obtained using a signal prediction mechanism. The proposed adaptive dynamic range adaptation techniques consider effects such as fade-in and fade-outs for an improved coding performance. Figure 9 shows a block diagram of the baseline profile architecture of this codec. The baseline profile restricts all video processing in the base and enhancement coding layers in the YCbCr 4:2:0 color space.

As shown in Fig. 9, an input 4:4:4 RGB (Rec. 709 or Rec.2020), 12+ bit, HDR sequence is first converted to 4:2:0 YCbCr color space. Then, advanced quantization is applied to generate an 8-bit BL sequence in the 4:2:0 YCbCr space. Unlike the backward-compatible case, the BL in not intended to be viewed on SDR displays. Rather the BL signal is optimized in such a way that it contains necessary information to minimize the overall bit requirement for HDR video data carried using multiple layers for the purpose of displaying it on HDR displays. Due to the absence of external color corrections in the BL signal, the clipping levels in the BL and EL are fully controlled by the codec itself. The Advanced quantization block supports many linear and non-linear mapping methods, such as linear quantization, linear stretching, curve-based or non-uniform quantization. Quantization can be done for each individual color channel or jointly, at the frame level or at the scene level.

As an example, a scene-adaptive linear stretching quantization method uses HDR values from each scene to generate corresponding BL values using a simple, invertible, linear mapping. Let v_k^i be the k-th pixel value of the i-th scene of an HDR sequence. Let $\left(v_{min}^i, v_{max}^i\right)$ denote the minimum and maximum pixel values in the i-th scene. Let $\left(s_{min}^i, s_{max}^i\right)$ be the min and max values of 8-bit YCbCr Rec.709 pixel values. Then, the scene adaptive linear stretching method generates base layer s_k^i as:

$$s_k^i = round\left(\frac{\left(s_{max}^i - s_{min}^i\right)}{\left(v_{max}^i - v_{min}^i\right)} \cdot \left(v_k^i - v_{min}^i\right) + s_{min}^i\right), \tag{12}$$

where $0 \leq v_{min}^i \leq v_{max}^i \leq (2^n - 1)$, for n-bit HDR signal.

The BL video sequence is encoded with a standard H.26x encoder using a compliant (such as 4:2:0 YUV) image container. The encoded BL is subsequently decoded to produce a 4:2:0 reconstructed BL to account for the approximations introduced by (possibly) the lossy BL encoder. As in the backward design (Fig. 4), an HDR predictor (e.g., using the inverse of Eq. (12)) may be used to approximate the input HDR signal and generate a residual which will be further coded as an enhancement layer. Since the dynamic range of the residuals may exceed the 8-bit dynamic range of the EL encoder, a non-linear quantizer (NLQ) is subsequently applied to generate 8-bit residuals. Mapping and NLQ parameters are transmitted as a part of metadata categories supported by the legacy video compression standard. For example, H.26x uses the SEI syntax to transmit metadata. For decoding, a similar decoder with the one shown in Fig. 5 may be used, except, that there is no need to display the decoded SDR signal.

5.1.2 HEVC Range Extension Proposals

In an effort to standardize extended range coding (e.g., more than 10 bits and additional color formats) in accordance with the existing HEVC standard, many proposals had been considered by the standardization bodies [22]. Range extensions support up to 16-bits per sample. Given the limitations of using 10-bit coding to encode HDR content, a number of alternative dual-layer systems have been proposed.

Some proposed methods use signal splitting at the encoder and recombination at the decoder [23]. The signal is split in additive layers that are encoded using HEVC 8/10-bit coders. In another method [24], the samples are split into the most significant bits (MSBs) and the least significant bits (LSBs) to obtain two layers. Some methods use overlapped bits in the layers [25–27]. The split signals are typically either packed in 4:4:4, low-bit-depth pictures, or two layers, BL and EL.

Figure 10 [23] shows a dual layer MSB/LSB splitting architecture proposed by Qualcomm. In this design, the input 16-bit picture is split into MSB and LSB layers by separating the most and the least significant bits of each sample to form two layers. In this way, the existing infrastructure in the distribution pipeline can be used for sending HDR signals.

In another proposal [28, 29], a preprocessing step is used to split the input HDR signal (P_{HDR}^i)into two limited dynamic range (LDR) signals using a non-linear mapping function (f), a modulation picture (P_{mod}^i),and a residual picture $\left(P_{LDR}^i\right)$. The modulation picture consists of a low frequency monochromatic version of the input signal, whereas the LDR residual picture represents the remaining relatively-high frequency portion of the signal. The function f is conceptually similar to an

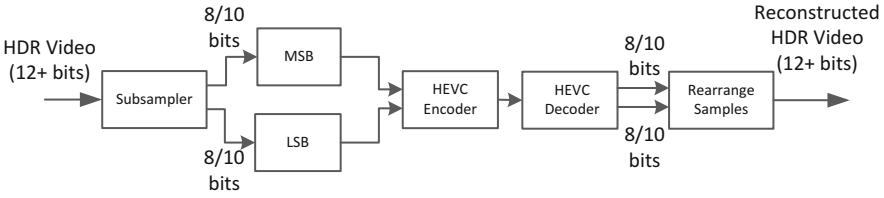

Fig. 10 Dual Layer MSB/LSB splitting architecture

OETF described earlier.

$$P^i_{HDR} = P^i_{mod} * f^{-1}\left(P^i_{LDR}\right). \tag{13}$$

The LDR picture is designed either to maintain backward compatibility or to improve the compression performance. In the latter (non-backward compatible) case, the LDR picture is built in the perceptual color space with the intention of optimizing the chroma quantization for a superior picture quality. Next, the split signals undergo a series of color space and signal format conversions before encoding them to be distributed to the receiver side. At the receiver, the decoder reconstructs the HDR signal by following the inverse operations. It involves inverse transforms and recombination of the layer signals. Arguably, this method has also many similarities with the proposed techniques in coding HDR still-pictures.

HEVC Range extensions (RExt) are now part of the second (or later) edition of the HEVC specification [18]. Various encoding profiles of HEVC RExt provide support for up to 12 or 16 bit signals. The high bit depth support is enabled using an extended precision mode that allows representing the transformed coefficients using 16-bit values [30] and by using a specific set of binarization [31] and entropy coding methods [32]. An overview of the HEVC RExt standard is described in [22].

5.2 Single-Layer Solutions Using Signal Reshaping

In the backward-compatible case, dual-layer systems typically demand more computational and management resources. Therefore, several single layer solutions have been proposed. Generally, the base layer signal is accompanied with metadata to help reconstruction of HDR signal at the decoder. The process of deriving this metadata based on the HDR signal and the base layer is called "reshaping." There have been a number of approaches to design a reshaper which can effectively provide the mapping function for base layer to HDR conversion. Single layer solutions require a re-quantization of the HDR signal to form the base (or the only) layer of the signal that can be encoded and distributed via a standard pipeline.

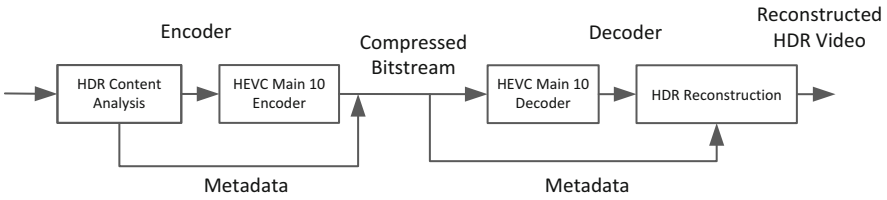

Fig. 11 System diagram of HEVC-based coding using non-backward compatible adaptive reshaping

5.2.1 MPEG Proposals for Reshaping Methods

There has been a considerable interest in the MPEG community to improve the HEVC Main 10 compression performance for HDR/WCG[5] video [33]. The HDR Exploratory Test Model (ETM) is a platform designed for coding Y′CbCr 10-bit 4:2:0 signals [34]. Figure 11 shows a typical system diagram of a non-backward compatible HDR codec using the HDR-10 framework [20, 35, 36].

One consideration for designing a good "reshaper" for a single layer codec is the usage of the PQ EOTF [8] and the practical limitations of HDR displays. PQ is designed to support luminance values of 0 to 10,000 nits, yet a majority of commercially available HDR displays do not exceed 1000 nits. Also, it is important to consider the director's (or the content producer's) creative intent (or "look"). Therefore, during coding, a better utilization of code words can be performed if the display range is known. In a recent study by Lu et al. [36], the impact of baseband quantization on the HDR coding efficiency was analyzed. The study states that matching code words to the target display range improves the coding efficiency. The term "baseband quantization" is defined as the range reduction step that uses a linear or non-linear mapping to convert a higher bit-depth signal to a lower (e.g., 8 or 10) bit signal that is encoded using the legacy encoder. The goal here is to quantify the mapping between the strength of the baseband quantizer and the coding efficiency measured in terms of peak-signal-to-noise-ratio (PSNR). The authors propose a method to estimate the error in reconstructed residues to be used later in a joint analysis of baseband and codec quantizers. The analysis shows that the coding efficiency is mainly lowered by the baseband quantization, and is less affected by the codec quantization [37].

There are many ways to design a reshaping function. Designing linear, piecewise linear, and non-linear, power functions are a few examples among those proposed by various study groups [20, 35, 36]. More advanced proposals apply adaptive codeword re-distribution and signal re-quantization of the three color components, which ultimately changes the bit-rate allocation. For example, in [38] the input

[5]WCG stands for wide color gamut, referring to any color gamut larger than the color gamut supported by the original analog television systems and CRTs. For example, Rec. BT. 2020 [51] defines a WCG container for ultra-high-definition TVs.

HDR signal is partitioned into various non-overlapping luma intensity bands. Each band is processed according to its perceptual significance (referred to as "band importance"), based on the HVS model. A reshaping curve that is non-linear in nature is constructed using this band importance to achieve a better compression efficiency.

Another major consideration in designing a reshaping function is that typically HDR/WCG signals are represented using a much larger color volume than SDR signals. A set of reshaping functions can be designed for each color space such as Y′CbCr, ICtCp [39], Y′CoCg, etc., in which each color channel can be reshaped on its own based on the input HDR signal. One of the main problems with the Y′CbCr color space is non-constant luminance (NCL), in which the mapping between the PQ luminance and Y' luma component from Y′CbCr is not linear. To counter this, a luma adjustment method is proposed in [40] by Ström et al. in Ericsson. This method uses a set of premises: (a) Y', the color-transformed luma value, can be changed independently at each pixel, and (b) Y' increases monotonically with the original (linear space) Y_0. Therefore, Y' is matched with the desired Y_0, the desired luma value. To speed up the implementation for practical purposes one may apply a 3D LUT that maps Cb, Cr and Y_0 values of a pixel to the desired Y' value. More details of the method, with a summary of experiments and performance comparison with the MPEG Call for Evidence (CfE) [33], are described in [40].

5.2.2 Encoder Optimization for PQ-Coded HDR Signals

Legacy encoders, such as HEVC, are highly tuned for compressing gamma-coded SDR signals. As described in Sect. 3, HDR contents are often encoded using the PQ EOTF. Encoding PQ-coded content by directly using gamma-based encoders, such as AVC or HEVC, may result into many undesirable artifacts. To address these issues, several authors have proposed alternative methods to control quantization within the HEVC codec.

A recent study by Lu et al. [41] shows the variance of the input signal to be an important statistic for encoder optimization, since it determines the bit rate allocation and quantization level for each pixel-block in a typical legacy encoder. For example, in a typical H.26x encoder rate control, a variance-based block quantization parameter (QP) model is applied in which, for lower block variance, QP is set to a smaller value. For gamma signal, typically darker areas have less codewords, leading to less variance and smaller QP values, whereas, for brighter areas, there are more codewords, hence higher QP values. However, for a PQ signal, the brighter areas are typically assigned less codewords than the darker areas, which demonstrates opposite behavior from the gamma domain signal. Therefore, applying the gamma-based model directly to PQ-domain signal leads to assigning higher QP values for highlight areas leading to artifacts, while the darker regions are assigned with smaller QP values, thus wasting bits.

A group within the ITU/ISO Joint Collaborative Team on Video Coding (JCT-VC) [42] has been working on the study and standardization of the use of

AVC/HEVC coders for HDR/WCG contents with PQ transfer characteristics. This effort primarily provides a set of recommended guidelines on processing of consumer distribution HDR/WCG video. In [42], two different models: the simple reference model and the enhanced reference model, are described for the pre-encoding and encoding processes. The simple reference model corresponds to the reference configuration used in the MPEG CfE on HDR and WCG [33], while the enhanced reference model corresponds to a new reference configuration that was developed in MPEG following the CfE.

To improve the coder performance, Liu et al. [41] proposed a method to adaptively select block QP for PQ-coded HDR signal based on luminance levels. To further improve the rate-distortion (RD) curve, block signal properties such as edge, texture, contour map or model fitting can be used along with luminance and region of interest (ROI). Another method for assigning QP values for a pixel block based on the luminance range is jointly proposed by Ericsson, Sharp Laboratories and Apple in [43]. It makes use of a fixed LUT for luma channel and a negative QP offset for chroma channels to make the encoder adaptive to the PQ content. In [44], as a pre-processing step to the above compression methods, Norkin proposed a fast down-sampling method to speed up the conversion of 4:4:4 RGB HDR video to the Y′CbCr 4:2:0 non-constant luminance format.

5.2.3 Perceptual Quality-Based Quantization Models

Encoding HDR signals often requires understanding of the perceived HDR image quality. To assess the quality of HDR images with respect to their quantization levels, many model-based approaches are studied. In [45], HDR-VDP-2.2, a calibrated method for objective quality prediction of high-dynamic range and standard images is proposed. HDR-VQM [46] is an alternative proposal for measuring an objective quality of HDR video.

A recent visual study suggests that there exists a relatively strong correlation between the required bit-depth for representing an HDR signal and the standard deviation in code values [47]. Based on this study, a dynamic signal quantization approach called “content-aware quantization” (CAQ) was developed to exploit signal noise and texture in images for reducing the effective bit-depth of the HDR signal. CAQ predicts the required quantization per intensity bin for an HDR image based on a noise/texture estimation model.

As shown in Fig. 12, this model uses a Gaussian high-pass filter to selectively enhance high-frequency content (such as noise/texture) followed by a spatial blurring filter. This model is consistent with HVS models such as the ones described in [48, 49]. To calculate minimum allowed quantization level per pixel, a calibration LUT is applied. A comparative image-based experimental study of CAQ and HDR-VDP is reported in [47].

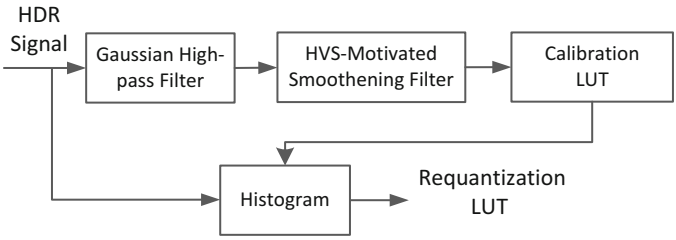

Fig. 12 Content-Aware Quantization (CAQ) for Single Layer Reshaping

5.3 The Ultra HD Blu-Ray Disc Format

As described in [50], Ultra HD Blu-ray™ can support video resolutions at up to 3840 × 2160 pixels (4K/UHD), at up to 60 frames per second. It can also support three different HDR formats: a mandatory Blu-Ray HDR format (also referred to as HDR10 or BDVM[6] HDR), and two optional formats, the Dolby Vision dual-layer format, and the Philips format.

The mandatory BDVM HDR format is a single-layer format that uses 10-bit, 4:2:0, YCbCr signals encoded using the SMPTE 2084 EOTF (PQ), in a BT. 2020 container [51]. The video is encoded using the Main10 Profile of the H.265/HEVC coding standard.

The Dolby Vision stream is composed of a BDVM HDR base layer and a Dolby Vision enhancement layer, with embedded Dolby metadata, which allows the reconstruction of 12-bit HDR video data. Reconstruction of a Dolby Vision HDR stream using the base layer and the enhancement layer follows the decoding scheme discussed in Fig. 5.

As described earlier, the Philips HDR format [20] includes a BDVM HDR video stream and Philips HDR SEI messaging which allows the conversion of the BDVM video to a format suitable for the target display.

6 Conclusions

The field of image and video coding for HDR signals has seen a tremendous growth in the last few years. HDR displays are now commercially available and HDR content is available for streaming. Video coding standards, like HEVC, do support video images with more than 8 bits per pixel, per color component; however, these standards were optimized for 8-bit, gamma-coded, YCbCr video signals, which do not represent the majority of new content, typically coded using either the PQ or

[6]BDVM stands for Blu-Ray Disc Movie.

HLG EOTFs. For most of the schemes we discussed in this chapter, the core SDR video encoder is treated as a "black box" that can be replaced by any video codec of choice (say, 8-bit AVC, 10-bit HEVC, and the like). For HDR coding, the codec's functionality may be enhanced using either preprocessing (say, PQ-coding and/or reshaping), or additional layers of information. We do not expect this to last for long. The "gamma" curve is already challenged, and new color formats, like ICtCp are challenging YCbCr domination as well. As the members of the MPEG committee start work on the next generation codec (to be completed sometime in 2020), we expect efficient HDR coding to be an integral part of any of the proposed coding tools, and not a simple after-thought. The future of HDR video coding is bright indeed.

Appendix: List of Abbreviations

AVC	Advanced Video coding
BDVM HDR	Blu-ray Disc Movie HDR
BL	Base Layer
CAQ	Content Adaptive Quantization
CfE	Call for Evidence
CRI	Color Remapping Information
CRT	Cathode Ray Tube
EL	Enhancement Layer
EOTF	Electro-Optical Transfer Function
HDR	High Dynamic Range
HDR ETM HDR	Exploratory Test Model
HDTV	High Definition Television
HEVC	High Efficiency Video Coding
HLG	Hybrid Log-Gamma
HVS	Human Visual System
ITU	International Telecommunication Union
JPEG	Joint Photographic Experts Group
LDR	Lower Dynamic Range
LSB	Least Significant Bit
LUT	Look-up Table
MMR	Multivariate Multiple Regression
MPEG	Moving Picture Experts Group
MSB	Most Significant Bit
MSE	Mean-Squared Error
NLQ	Non-Linear Quantizer
OETF	Opto-Electrical Transfer Function
OOTF	Opto-Optical Transfer Function
PQ	Perceptual Quantizer
PSNR	Peak Signal-to-Noise Ratio

QP	Quantization Parameter
RD	Rate-Distortion
ROI	Region of Interest
SDI	Serial Digital Interface
SDR	Standard Dynamic Range
SEI	Supplementary Enhancement Information
SMPTE	Society of Motion Picture and Television Engineers
TIFF	Tagged Image File Format
UHD	Ultra-high-definition
VDP	Visual Difference Predictor
VQM	Video Quality Measure
WCG	Wide Color Gamut

References

1. G. Ward and M. Simmons, "JPEG-HDR: A Backwards-Compatible, High Dynamic Range Extension to JPEG," ACM SIGGRAPH 2006.
2. A. Artusi et al. "JPEG XT: A compression standard for HDR and WCG images," *IEEE Signal Processing Magazine,* pp. 118-124, March 2016.
3. T. Richter, T. Bruylants, P. Schelkens, and T. Ebrahimi, "The JPEG XT Suite of standards: Status and Future Plans," SPIE Optical Engineering+ Applications, International Society for Optics and Photonics, Sept. 2015.
4. Report ITU-R BT. 2390-0, "High dynamic range television for production and international programme exchange," ITU, 2016.
5. W. Gish and S. Miller, "Unambiguous video pipeline description motivated by HDR." In Proc. IEEE Intern. Conf. on Image Processing (ICIP 2016), pp. 909-912. IEEE, 2016.
6. P.G.J. Barten, "Contrast sensitivity of the human eye and its effects on image quality," SPIE Optical Engineering Press: Bellingham, WA, 1999.
7. S. Miller et al., "Perceptual Signal Coding for More Efficient Usage of Bit Codes," *SMPTE Motion Imaging Journal,* vol. 122:(4), pp. 52-59, May-June 2013.
8. Rec. ITU-R BT. 2100, "Image parameter values for high dynamic range television for use in production and international programme exchange," ITU, July 2016.
9. Rec. ITU-R BT. 1866, "Reference electro-optical transfer function for flat panel displays used in HDTV studio production," ITU, 03/2011.
10. R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, "Backward Compatible High Dynamic Range MPEG Video Compression," ACM Trans. on Graphics 25(3):713-723, July 2006.
11. Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. K. Ward and W. Heidrich, "Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image/Video Compression," IEEE Trans. on Image Processing, Vol. 20, No. 6, pp. 1558 – 1571, June 2011.
12. G-M. Su, R. Atkins, and Q. Chen, "Backward-Compatible Coding for Ultra High Definition Video Signals with Enhanced Dynamic Range," US 9,549,207, January 17, 2017.
13. Q. Chen, G-M. Su, and P. Yin, "Near Constant-Time Optimal Piecewise LDR to HDR Inverse Tone Mapping," IS&T/SPIE Electronic Imaging, 2015.
14. G-M. Su, S. Qu, H. Koepfer, Y. Yuan, and S. Hulyalkar, "Multiple Color Channel Multiple Regression Predictor," US 8,811,490 B2, 2014.
15. P. Bordes, P. Andrivon, X. Li, Y. Ye, and Y. He, "Overview of Color Gamut Scalability," IEEE Trans. on Circuits and Systems for Video Technology, March 2016.

16. G-M. Su, S. Qu, W. Gish, H. Koepfer, Y. Yuan, and S. Hulyalkar, “Image Prediction based on Primary Color Grading Model,” US 8,731,287 B2, 2014.
17. “ASC Color Decision List (ASC CDL) Transfer Functions and Interchange Syntax,” ASC Technology Committee Digital Intermediate Subcommittee, 2008
18. ITU Rec. H.265, “High efficiency video coding,“ Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – Coding of Moving Video, ITU, Dec 2016.
19. S. Lasserre, E. François, F. Le Léannec, and D. Touzé, “Single-layer HDR video coding with SDR backward compatibility,” SPIE Optical Engineering+ Applications (pp. 997108-997108), September, 2016.
20. R. Goris, R. Brondijk, R. van der Vleuten, “Philips response to CfE for HDR and WCG,” m36266, ISO/IEC JTC1/SC29/WG11, Warsaw, Poland, July 2015.
21. G-M. Su, S. Qu, S. Hulyalkar, T. Chen, W. Gish, and H. Koepfer, “Layered Decomposition in Hierarchical VDR Coding,” US 9,497,456 B2, November 15, 2016.
22. D. Flynn, D. Marpe, M. Naccari, T. Nguyen, C. Rosewarne, K. Sharman, J., and J. Xu “Overview of the range extensions for the HEVC standard: Tools, profiles, and performance,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 26, no. 1, pp 4-19, January, 2016.
23. F. Dufaux, P. Le Callet, R. Mantiuk, and M. Mrak, eds. “High Dynamic Range Video: From Acquisition, to Display and Applications, “Academic Press, 2016.
24. E. Francois, C. Gisquet, G. Laroche, P. Onno, “AHG18: On 16-bits Support for Range Extensions, Document,” JCTVC-N0142, 14th JCT-VC Meeting Vienna, Austria, Jul-Aug. 2013.
25. W. S. Kim, W. Pu, J. Chen, Y. K. Wang, J. Sole, M. Karczewicz, “AHG 5 and 18: High Bit-Depth Coding Using Auxiliary Picture, Document,” JCTVC-O0090, 15th JCT-VC Meeting, Geneva, Switzerland, Oct.-Nov. 2013.
26. A. Aminlou, K. Ugar, “On 16 Bit coding,” Document JCTVC-P0162, 16th JCT-VC Meeting, San Jose, CA, Jan. 2014.
27. C. Auyeung, J. Xu, “AHG 5 and 18, Coding of High Bit-Depth Source with Lower Bit-Depth Encoders and a Continuity Mapping,” Document JCTVC-P0173, 16th JCT-VC Meeting, San Jose, CA, Jan. 2014.
28. S. Lasserre, F. Le Leannec, P. Lopez, Y. Olivier, D. Touze, E. Francois, “High Dynamic Range Video Coding,” JCTVC-P0159 (m32076), 16th JCT-VC Meeting, San Jose, CA, Jan. 2014.
29. F. Le Leannec, S. Lasserre, E. Francois, D. Touze, P. Andrivon, P. Bordes, Y. Olivier, “Modulation Channel Information SEI Message,” Document JCTVC-R0139 (m33776), 18th JCT-VC Meeting, Sapporo, Japan, Jun.-Jul. 2014.
30. K. Sharman, N. Saunders, and J. Gamei, “AHG5 and 18:Internal Precision for High Bit Depths,” document JCTVC-N0188, 14th Meeting, JCT-VC, Vienna, Austria, Jul. 2013.
31. M. Karczewicz and R. Joshi, “AHG18: Limiting the Worst-Case Length for Coeff_Abs_Level_Remaining Syntax Element to 32 Bits,” document JCTVC-Q0131, 17th Meeting, JCT-VC, Valencia, Spain, Apr. 2014.
32. K. Sharman, N. Saunders, and J. Gamei, “AHG5 and AHG18: Entropy Coding Throughput for High Bit Depths,” document JCTVC-O0046, 15th Meeting, JCT-VC, Geneva, Switzerland, Oct. 2013.
33. A. Luthra, E. Francois, W. Husak, “Call for Evidence (CfE) for HDR and WCG Video Coding”, MPEG2014/N15083, 110th MPEG Meeting, Geneva, 2015.
34. K. Minoo, T. Lu, P. Yin, L. Kerofsky, D. Rusanovskyy, E. Francois, “Description of the Exploratory Test Model (ETM) for HDR/WCG extension of HEVC”, JCT-VC Doc. W0092, San Diego, CA, Feb. 2016.
35. L. Kerofsky, Y. Ye, and Y. He. “Recent developments from MPEG in HDR video compression,” IEEE Intern. Conf. on Image Processing (ICIP), pp. 879-883. IEEE, 2016.
36. T. Lu, F. Pu, P. Yin, Y. He, L. Kerofsky, Y. Ye, Z. Gu, D. Baylon, “Compression Efficiency Improvement over HEVC Main 10 Profile for HDR and WCG Content,” Proc. of the IEEE Data Compression Conference (DCC), Snowbird, March 2016.

37. C. Wong, G-M. Su, M. Wu, "Joint Baseband Signal Quantization and Transform Coding for High Dynamic Range Video," IEEE Signal Processing Letters, 2016.
38. T. Lu, F. Pu, P. Yin, J. Pytlarz, T. Chen, and W. Husak. "Adaptive reshaper for high dynamic range and wide color gamut video compression," SPIE Optical Engineering+ Applications, pp. 99710B-99710B, International Society for Optics and Photonics, 2016.
39. T. Lu, F. Pu, P. Yin, T. Chen, W. Husak, J. Pytlarz, R. Atkins, J. Fröhlich, G-M. Su, "ITP Colour Space and its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution," ZTE Communications, Feb. 2016.
40. J. Ström, J. Samuelsson, and K. Dovstam, "Luma Adjustment for High Dynamic Range Video," Proc. of the IEEE Data Compression Conference (DCC), Snowbird, March 2016.
41. T. Lu, P. Yin, T. Chen, and G-M. Su, "Rate Control Adaptation for High-Dynamic Range Images," U.S. Patent Application Publication US 2016/0134870, 2016.
42. J. Samuelsson et al., "Conversion and coding practices for HDR/WCG YCbCr 4:2:0 video with PQ transfer characteristics," Draft new Supplement 15 to the H-Series of Recommendations, JCTVC-Z1017, 26-th meeting, Geneva, CH, Jan. 2017.
43. J. Ström, K. Andersson, M. Pettersson, P. Hermansson, J. Samuelsson, A. Segall, J. Zhao, S-H. Kim, K. Misra, A. M. Tourapis, Y. Su, and D. Singer, "High Quality HDR Video Compression using HEVC Main 10 Profile," in Proc. of the IEEE Picture Coding Symposium (PCS), Nuremberg, 2016.
44. A. Norkin, "Fast algorithm for HDR video pre-processing," in Proc. of the IEEE Picture Coding Symposium (PCS), Nuremberg, 2016.
45. R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," ACM Trans. on Graphics (TOG), vol. 30, no. 4, p. 40. ACM, 2011.
46. M. Narwaria, M. P. Da Silva, and P. Le Callet. "HDR-VQM: An objective quality measure for high dynamic range video," Signal Processing: Image Communication, Vol. 35, pp. 46-60, 2015.
47. J. Froehlich, G-M. Su, S. Daly, A. Schilling, and B. Eberhardt. "Content aware quantization: Requantization of high dynamic range baseband signals based on visual masking by noise and texture," IEEE International Conf. on Image Processing (ICIP), pp. 884-888. IEEE, 2016.
48. S. Daly, "A visual model for optimizing the design of image processing algorithms," Proc. Intern. Conf. on Image Processing, (ICIP-94), vol. 2, pp. 16-20, 1994.
49. A. Lukin, "Improved visible differences predictor using a complex cortex transform," International Conf. on Computer Graphics and Vision, 2009.
50. Blu-Ray Disc Read-only Format, "Audio Visual Application Format Specifications for BD-ROM Version 3.1," White Paper, August 2016, Blu-Ray Disc Association.
51. Rec. ITU-R BT. 2020-1, "Parameter values for ultra-high definition television systems for production and international programme exchange," ITU, June 2014.